Courant Research Centre
'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 94

The Africa-Dummy in Growth Regressions

Max Köhler, Stefan Sperlich, Julian Vortmeyer

September 2011

Wilhelm-Weber-Str. 2 · 37073 Goettingen · Germany
Phone: +49-(0)551-3914066 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

The Africa-Dummy in Growth Regressions

Max Köhler *, Stefan Sperlich†, Julian Vortmeyer ‡

September 5, 2011

Abstract

The Africa-Dummy has been identified and different explanations for its appearance have been published. In this paper, the issue of the empirical identification of the Africa-Dummy is addressed. We introduce a fixed effects regression model to identify the Africa-Dummy in one regression step so that its correlations to other coefficients can be estimated. A semiparametric extension of this model checks whether the Africa-Dummy is a result of misspecification of the functional structure. Furthermore, we show that sub-Saharan African countries have a positive return to the population growth and when adding interaction effects, the Africa-Dummy is even positive. Moreover, we show that the Africa-Dummy changes dramatically over time and the punishment for sub-Saharan African countries decreases incrementally since the mid-nineties. According to the Augmented Solow Growth model, it was even insignificant since the end-nineties.

*Corresponding author: Max Köhler, Max.Koehler@wiwi.uni-goettingen.de, University of Goettingen, CRC Poverty, Equity and Growth, Göttingen, Germany

†Université de Genéve, Département d'économétrie, Genéve, Switzerland

‡University of Göttingen, Faculty of Economics, Göttingen, Germany

1 Introduction

This paper focuses on the Africa-Dummy. There is no doubt that groups of countries possess certain characteristics that are hard to measure and to incorporate in growth models, but represent systematic drivers for growth. These characteristics could for example consist of international capital markets (see Barro, Mankiw and Sala-i-Martin (1995)), the differences of the country's production function (see Islam (1995)), political factors (see Collier and Gunning (1999)), diseases especially AIDS (see Were and Nafula (2003)), geographical factors and trade openness (see Sachs and Warner (1997)), ethnic diversity (see Easterly and Levine (1997)) or historical reasons such as the colonial heritage (see Price (2003)), to mention a few. One example is the group of sub-Saharan African countries, meaning that economic growth models are not able to explain the growth in sub-Saharan Africa, because their economic fundamentals incorporated in the growth model are not as bad as their actual performance. The result is that, if an additional variable is added, that only indicates the membership to sub-Saharan Africa, namely the Africa Dummy, it has a significant coefficient with a negative sign (see for example Barro (1991)). There is a lot of literature that adds explanatory variables to growth regressions in order to explain the significance of the Africa-Dummy (see for example Sachs and Warner (1997)). This is critical, as explanatory variables that are added in growth regressions do not necessarily identify drivers for growth because their unique output identifies certain groups of countries and therefore act like a dummy variable. However, there is only one paper that addresses the problems of statistical estimation of growth regressions and the Africa-Dummy (see Hoeffler (2002)). In this paper, we concentrate on discussing the estimation of the Africa-Dummy and deriving statistical facts about it.

Section (2) is divided into four subsections. Subsection (2.1) describes how the data are collected and subsection (2.2) describes how business cycles are removed. Many authors conduct growth regression with numerous variables and understand growth as a theory of everything. In this paper growth regressions are all justified by the augmented Solow model. It is briefly described in subsection (2.3). Subsection (2.4) explains what the Africa-Dummy is and gives a literature review.

Section (3) deals with statistical methods to identify the Africa-Dummy. It is divided into five subsections. Subsection (3.1) is about the underlying statistical model and contains some notes about running the growth regressions. Subsection (3.2) deals with the System GMM estimator and comments on its disadvantages. Subsection (3.3) discusses estimators based on error components models and subsection (3.4) concentrates on estimating with fixed effects. The Two-Groups Least-Square Dummy-Variable estimator is introduced and identified as the best estimator to estimate country-specific dummy variables. Finally, subsection (3.5) gives results on identifying the Africa-Dummy and estimating the correlations of the Africa-Dummy and other coefficients.

Section (4) uses extensions of the Two-Groups Least-Square Dummy-Variable estimator to derive facts about the Africa-Dummy. First of all, subsection (4.1) relaxes the functional structure of the regression equation and checks if the Africa-Dummy is a result of a misspecification of the functional structure. Second subsection (4.2) estimates the interaction effects of the Africa-Dummy. Third, in subsection (4.3), a model is introduced that estimates one Africa-Dummy for each year in the observation period.

Section (5) finally concludes.

2 Growth Regression and the Africa-Dummy

2.1 Data Collection

The objective is to collect long time-series for as many countries as possible for which we can guarantee good data quality. The information sources for the empirical investigation are the Penn World Table 6.3 (PWT), World Bank's World Development indicators and Barro and Lee (2010). Except of population growth and human capital, all data come from the PWT. It collects a broad range of macroeconomic time-series for almost all countries published by Heston, Summers and Aten (2009). The beginning of a widespread availability is 1960. Most variables are published until 2007, so that observations are obtained for 48 periods. The sample could have been increased significantly but the quality of the data for some countries is insufficient. Heston, Summers and Aten (2009) introduce a country rating system based on the number of participations in worldwide benchmark surveys, the variation of the accessible data and the quality of the statistical methods applied. This results in a grading scheme from A to D with descending order. A rating of D is regarded as too weak to be included in the sample. Therefore, only countries with a grading from A to C are incorporated in the sample. Furthermore, we only incorporate complete time-series for the relevant variables from 1960 to 2007. This also excludes countries that where separated in a sub-period, for example Germany and the countries of the Soviet Union. We excluded these countries because their incorporation would have made it necessary to unify several countries to one country or to split one country in a given period in several countries. The loss of data quality when doing this is unclear. We ended up with 81 complete time-series, one for each of the 81 countries. The time-series are 48 years long. The total sample size is therefore 3888.

The selection process of the data can cause a problem. Figure 1 shows that sub-Saharan

African countries are much more often affected by D grading than other regions. In general, poor countries have weaker databases and are more likely to be excluded. The question is, whether the exclusion of the D graded data causes a significant violation of the information that the whole sample would inhabit. Since we cannot reliably compare the excluded and the included data, we cannot fully answer this question.

The preparation of the variables mainly follows Hoeffler (2002) and Caselli (2005). Because economic growth is a consequence of changes in the production function, the output of the economy is measured as the real per worker gross domestic product. This is a more precise measure of the country's potential than the per capita GDP because it answers the question how much each productive factor contributes on average to the growth in its country. Per capita figures give information about the available income for the average individual but since the participation rate in the workforce differs a lot, the per capita GDP would be a distorted indicator of the production volume of the total workforce. We denote the logarithm of the per worker GDP of country i at time t by y_{it}.

The population growth refers to the working age population, i.e. to all individuals from 15 to 64 years. We use the data for the total population and multiply them with the share of adults in working age. We denote the growth rate of the working age population of country

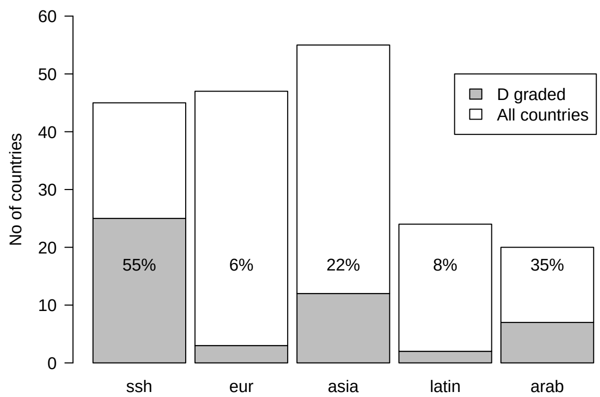

Figure 1: D grading in the PWT

i at time t by n_{it}. Data for depreciation rates are not available. In the literature there is accordance, as explained by Mankiw, Romer and Weil (1992), to expect the capital to wear out by 3% per year. Similarly, the advance in productivity is 2% per year for all countries. Therefore, the term $\ln(\delta + g + n_{it})$ is approximated by $\ln(0.05 + n_{it})$. Three values are replaced by $ln(0.05) = -2.995732$, because the reduction in population exceeds 0.05. We denote the logarithm of the depreciation rate of country i at year t by lnn_{it}.

The saving rate of the economy is approximated by the relative investment share of the real GDP. These data should correctly measure the savings in the case of a closed economy. We denote the logarithm of the share of country i at year t by $lnsk_{it}$.

While the PWT contains yearly data, the information from Barro and Lee (2010) about schooling are given in five years frequencies. The beginning of the observation period is 1950 and the end is 2010. In order to transfer this variable into a yearly frequency, the missing values are extrapolated by interpolation splines. When doing this, we have to be careful that we do not add an artificial parametric structure to the data. Figure (2) shows a graph of the educational attainments when applying natural splines. The points show the data obtained from Barro and Lee (2010) and the lines represent the natural spline functions. Figure (2) is representative for all countries. They all have monotonically and linearly increasing shape. Since the points do no not fluctuate a lot, we assume that the approximation error is sufficiently small. We denote the logarithm of the educational attainment data by Barro and Lee (2010) of country i and year t by $lnattain_{it}$.

2.2 Smoothing

We collected four time-series, namely lny_{it}, lnn_{it}, $lnsk_{it}$ and $lnattain_{it}$ for each country i. These time-series have a short term cyclical component and a trend component. The Solow model addresses long run growth but not the cyclical fluctuations. Therefore, we smooth the data. As the series have different magnitudes of short term fluctuations they have to be treated in different ways. The series lnn_{it} and $lnattain_{it}$ have only negligible short term fluctuations and are therefore not to be smoothed. The series $lnsk_{it}$ and lny_{it} have severe cyclical components.

First of all, we consider the GDP per worker time-series. The easiest approach is linear smoothing. It suggests taking the arithmetical averages over several years of the GDP's per worker so that for this sub-period only the mean enters the dataset. The most common choice is the average over five years. Figure (3) shows the resulting growth rates when applying five years averages. It shows the time-series of four countries that serve as examples. The grey points are the unsmoothed data. The horizontal lines demonstrate the choice of the time periods. Their heights show the reduction of these five points to one value. The black bullets are the middle time points of each period. The sample lasts from 1960 to 2007 so that data have to be excluded in order to obtain time periods of the same length, namely five years. We excluded the values of the years 1960, 2006 and 2007. These points are labeled with a black star in figure (3) and their information is fully lost. Especially in case of the Philippines where the last two observations represent unusual jumps it seems not adequate to exclude this information from the sample. This is the first disadvantage of

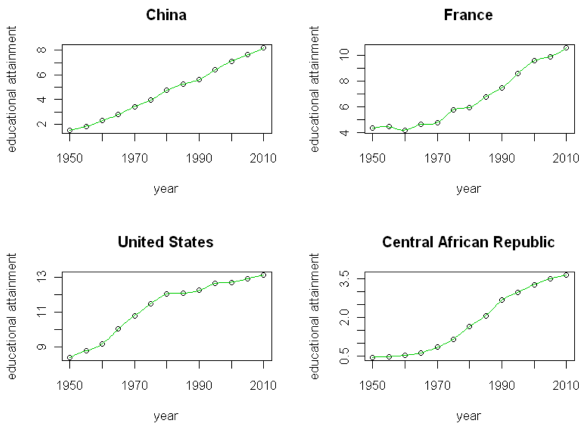

Figure 2: Interpolation of schooling

linear smoothing. Moreover, the long run growth variation within the time periods is fully lost. Another problem is the simultaneous smoothing of different time-series that interact. For example the series $lnsk_{it}$ has a different cyclical component than lny_{it}. Taking five year averages smoothes these series in the same naive manner so that the interactions of the long term components of the series are distorted. This problem is especially severe when combining linear smoothed series with unsmoothed series. It is not clear which values of the unsmoothed series should represent each time period. The average however leads to over-smoothing and taking the starting values of each time period would mean to make lagged variables enter the regression.

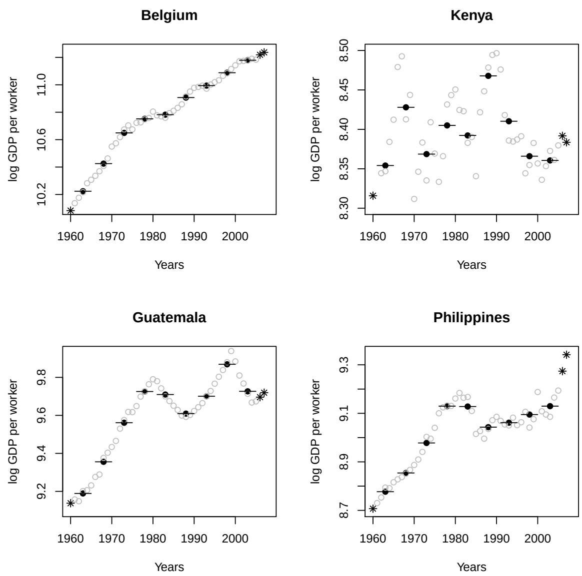

Figure 3: Five years averages

The disadvantages of linear smoothing give rise to look for another technique to remove business cycles. A prominent example is the Hodrick and Prescott filter. The HP filter decomposes a macroeconomic time-series $\tilde{\tau}_t$ in a structural trend component τ_t, which accounts for sustainable long-run growth and a cyclical component c_t. In Hodrick and Prescott

(1997) it is shown how these elements can be separated. The series τ_t is obtained due to

$$\min_{\tau_t} \sum_{t=1}^{T} (\tilde{\tau}_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} ((\tau_{t+1} - \tau_t)(\tau_t - \tau_{t-1}))^2 .$$

The first term can be interpreted as measuring the goodness-of-fit of the trend component with respect to the original series. The second part punishes for a high variation in the transformed series $\tilde{\tau}_t$. Note that minimizing the variation and maximizing the goodness of fit at the same time is a trade-off problem which is quantified by λ. The higher λ, the more variation is removed from the data. For the choice of this parameter, there are rather weak causal rationales. Hodrick and Prescott (1997) argue that $\lambda = 1600$ is a reasonable choice for quarterly data which intuitively corresponds to a value of 400 for yearly data. On the other hand, Baxter and King (1999) argue that λ should be chosen as the fourth power of a change in the frequency. In our case this corresponds to 6.25. After observing the different outputs of the smoothing with different smoothing parameters, we decided to chose $\lambda = 100$. Figure (4) shows the smoothed series of the yearly growth rates of the four countries Belgium, Kenya, Guatemala and Philippines. The grey points are the unsmoothed data. The smoothed data are connected with lines. It can clearly be seen that the disadvantages of linear smoothing are not shared by the HP filter.

When smoothing the series of $lnsk_{it}$ it is hard to derive the adequate smoothing parameter of one series from that of the other series. On the one hand, the series $lnsk_{it}$ have more variation than lny_{it}. On the other hand the former series are of much smaller magnitude than the latter. Smoothing the two series simultaneously means that one series should not appear to be over-smoothed compared to the other. Having this in mind, we choose the smoothing parameter of $lnsk_{it}$ by visual judgment. After observing the outputs of smoothed series for different smoothing parameters we decided that $\lambda = 25$ is the appropriate parameter. The result is given in figure (5). The HP filter performs satisfying and is therefore selected to smooth the data.

2.3 The Augmented Solow Model

The neoclassical growth theory is based on the work by Solow (1956) and Swan (1956). The theory of human capital accumulation tries to account for enhancements in technology

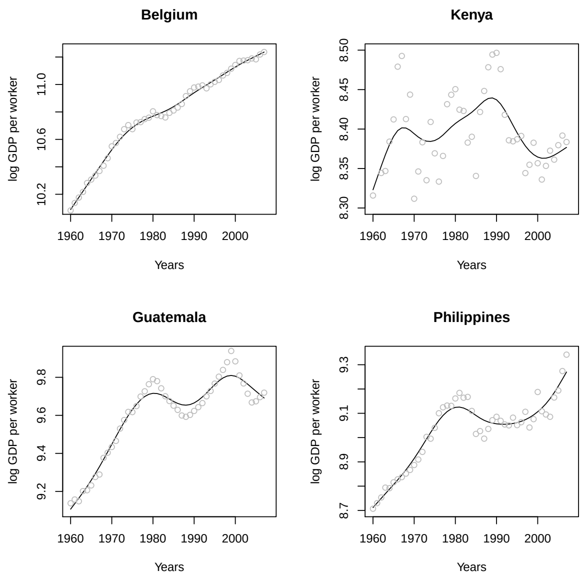

Figure 4: HP Smoothing of lny_{it}

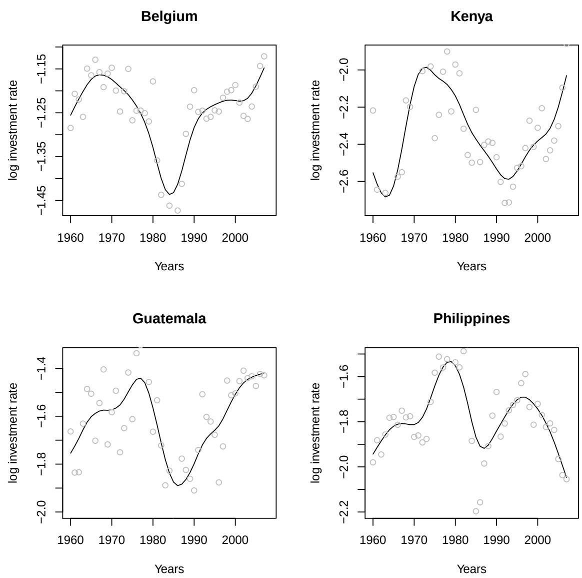

Figure 5: HP Smoothing of $lnsk_{it}$

by replacing homogeneous work with education-based improvements of workers that are regarded as investments in quality. Mankiw, Romer and Weil (1992) extend the Solow model by human capital, test this 'augmented' version and observe significant improvements in explanatory power. This model is the basis for growth regressions. We briefly describe it in what follows.

The economic agents are households and firms. Furthermore, there are three commodities and for each commodity there is a market. The commodities are output, capital and labor. When considering the corresponding markets we assume that all individuals behave rational and further information restrictions are not present. In the market for capital we think of households owning the capital $K(t)$ and lease it to the firms. The firms demand the capital $K^D(t)$. The price is $r(t)$ (real rental rate). In the market for labor the supply $\tilde{L}(t)$ comes from the households and the demand $\tilde{L}^D(t)$ comes from the firms. The price in the labor market is $w(t)$ (real wage rate). $\tilde{L}(t)$ is not a measure of headcount. It can be decomposed in a measure of working quality and a measure of the homogeneous supply per person. We decompose

$$\tilde{L}^D(t) = L(t)^{\frac{1-\alpha-\beta}{1-\alpha}} H(t)^{\frac{\beta}{1-\alpha}},$$

where $H(t)$ is the amount of human capital. In the market for output the supply consists of the total output of firms $Y(t)$ and the demand $Y^D(t)$ consists of what the households save ($S(t) = sY^D(t)$) and what they consume ($C(t) = (1-s)Y^D(t)$). We assume that investments equal savings ($I(t) = S(t)$). The households decide how to distribute total savings between gross investment and human capital. We assume that the summarized result of the households' decisions are that the fraction $s_K Y^D(t) = I_K(t)$ is invested in physical capital and the fraction $s_H Y^D(t) = I_H(t)$ is invested in human capital. Clearly $s_K + s_H = s$ and $I(t) = I_H(t) + I_K(t)$. The supply of output follows a production function with the input factors capital $K^D(t)$ and labor $L^D(t)$. The generated output is also influenced by the productivity $A(t)$ that characterizes the country's transformation capabilities. The improvement may consist of either of level of technology or of efficiency gains, meaning the ability to combine the input factors in an optimal way. The aggregated production function is

$$\begin{aligned} Y(t) &= K^D(t)^\alpha H(t)^\beta (L(t)A(t))^{1-\alpha-\beta} \\ &= A(t)^{1-\alpha-\beta} K^D(t)^\alpha \tilde{L}^D(t)^{1-\alpha} \end{aligned} \tag{1}$$

The price of the output market is normalized to one, so that other prices are measured in units of output. Note that the fundamental difference between the input factors of the production function is that capital and labor are rival goods while the applied technology can spillover to any entrepreneur in the economy what means that it is a public good.

All three markets are assumed to be perfectly competitive so that the economic agents take the prices as given and in each market the appropriate price adjusts such that $L^D(t) = L(t)$ and $K^D(t) = K(t)$ and $Y^D(t) = Y(t)$. Necessarily, the two inputs capital and labor are paid at their marginal products. Therefore, it holds

$$\frac{\partial Y(t)}{\partial K(t)} = r(t) \tag{2}$$

and

$$\frac{\partial Y(t)}{\partial L(t)} = w(t). \tag{3}$$

In this setting α is the capital intensity and β is the labor intensity in the production process. The labor force $L(t)$ and the productivity level $A(t)$ are assumed to grow exogenously at rates n and g respectively. Therefore, it holds

$$L(t) = L(0)\exp(nt) \tag{4}$$

and

$$A(t) = A(0)\exp(gt). \tag{5}$$

In any period, the investment of the prior period will be transformed into new capital minus the depreciation δ of the old capital stock. We express the production process in terms of effective worker units. ($k(t) = K(t)/A(t)L(t)$ and $y(t) = Y(t)/A(t)L(t)$). The growth of per effective worker capital over time is

$$\dot{k}(t) = s_K y(t) - (n + g + \delta)k(t). \tag{6}$$

Human capital behaves like its physical counterpart. The evolution of the per effective worker unit of human capital ($h(t) = H(t)/A(t)L(t)$) is

$$\dot{h}(t) = s_H y(t) - (\delta + g + n)k(t). \tag{7}$$

The model is (1), (2), (3), (4), (5), (6) and (7). The parameters of the model are α, β, s_K, s_H, δ, n and g. Given the initial values $A(0)$, $K(0)$, $H(0)$ and $L(0)$, the model will determine the dynamic evolution of the economy. Moreover, when assuming diminishing returns to capital input ($\alpha + \beta < 1$) the model converges. The situation of convergence is called steady state. It is identified by $\dot{k} = \dot{h} = 0$. In the steady state it holds that

$$\begin{aligned} k(t) \equiv k^* &= \left(\frac{s_k^{1-\beta} s_H^{\beta}}{\delta + g + n} \right)^{1/(1-\alpha-\beta)} \\ h(t) \equiv h^* &= \left(\frac{s_k^{\alpha} s_H^{1-\alpha}}{n + g + \delta} \right)^{1/(1-\alpha-\beta)} . \end{aligned} \tag{8}$$

Growth outside the steady state is determined by the evolution of $k(t)$ given in (6) and of $h(t)$ given in (7). The capital stock of human and physical capital increases if the economy fosters investments (s_H or s_K increases), or if the effective depreciation ($n+g+\delta$) decreases. The model makes quantitative predictions about the speed of convergence to steady state. Approximating around the steady state, the speed of convergence at a given time point t outside the steady state is given by

$$\frac{\partial \ln(y(t))}{\partial t} = \lambda(\ln(y^*) - \ln(y(t))), \tag{9}$$

with $\lambda = (n + g + \delta)(1 - \alpha - \beta)$. Equation (9) implies that

$$\ln(y(t)) - \ln(y(0)) = (1 - \exp(\lambda t)) \ln(y^*) - (1 - \exp(\lambda t)) \ln(y(0)), \tag{10}$$

where $y(0)$ is the income per effective worker at some initial date. This implies that if the economy moves from the initial state 0 to the time point t halfway to steady state, it holds

$$\frac{1}{2} = \frac{\ln(y(t)) - \ln(y(0))}{\ln(y^*) - \ln(y(0))} = 1 - \exp(\lambda t)$$

which is equivalent to $t = \ln(2)/\lambda$. If for example $\lambda = 0.02$ the economy moves halfway to steady state in 34.65736 years. In general the larger λ the less time it takes the economy to move halfway to steady state.

Equation (10) implies that outside the steady state it holds

$$\begin{aligned}
&\ln(\frac{Y(t)}{L(t)}) = (1 - \exp(-\lambda t)) \ln(A(0)) + gt + \exp(-\lambda t) \ln(\frac{Y(0)}{L(0)}) \\
&(1 - \exp(-\lambda t)) \frac{\alpha}{1 - \alpha - \beta} \ln(s_K) + (1 - \exp(-\lambda t)) \frac{\beta}{1 - \alpha - \beta} \ln(s_H) \\
&- (1 - \exp(-\lambda t)) \frac{\alpha + \beta}{1 - \alpha - \beta} \ln(n + g + \delta).
\end{aligned}$$

This justifies the following regression equation

$$lny_{it} = \rho * lny_{i(t-1)} + \beta_1 * lnn_{it} + \beta_2 * lnsk_{it} + \beta_3 * lnn_{it} + \eta_i + \nu_{it}, \tag{11}$$

where ν_{it} is an error with expectation zero.

2.4 The Africa-Dummy

The model by Mankiw, Romer and Weil (1992) contains simplifications to keep the model simple. For example Barro, Mankiw and Sala-i-Martin (1995) mention that international capital markets have a significant impact on growth rates, especially on the convergence of the poor countries. Another unrealistic simplification is criticized by Islam (1995). He argues that countries have fundamentally differing production functions so that comparisons between their economies are difficult. Furthermore, the endowment with resources, that is modeled in A, can be infinitely substituted by capital. For example Georgescu-Roegen (1975) criticizes that this point of view is too optimistic with respect to the limitations of technological progress. Among others, these problems result in empirical weaknesses. For example Barossi-Filho, Goncalves Silva and Martins Diniz (2005) summarize that among most regressions the estimated capital share exceeds the value obtained from the national accounts and that the estimated convergence rate is usually too low. The Africa-Dummy combines all empirical weaknesses of sub-Saharan African countries compared to all other countries in a punishment term. It is a dummy variable that is one if and only if the country belongs to sub-Saharan Africa and zero else. When incorporating it in growth regressions many authors identify its significant negative coefficient. Barro (1991) for example run a cross-sectional regression. This means that he holds an initial and a final time point fixed and calculate the growth rates in this time horizon for each country before regressing them on several explanatory variables. We run a cross-sectional regression using the variables given by the model by Mankiw, Romer and Weil (1992). Table (1) shows the results. The time horizons are 1960 to 1985 (as chosen by Barro (1991)), 1960 to 2007 and 1986 to 2007. The explanatory variables are the initial values for lny, $lnsk$ and $lnattain$, the average of lnn according to the time horizon and the dummy variable that indicates if the country belongs to sub-Saharan Africa. The standard errors are given in parentheses. The first observation is that all coefficients follow their predicted influence on growth. Furthermore, for the full sample we obtain a significant Africa-Dummy. This simple regression shows that the Africa-Dummy is present in the data.

As African countries started with a lower level of income, they should converge to the income observed in regions that have similar characteristics. The presence of the Africa-Dummy

Table 1: Cross-sectional regression results

	60-85	60-07	86-07
	Estimate	Estimate	Estimate
	(S.E.)	(S.E.)	(S.E.)
$Intercept$	2.0692*	1.8895	1.0590
	(0.8869)	(1.3862)	(0.9256)
lny	-0.3214***	-0.5750***	-0.2292***
	(0.0625)	(0.0898)	(0.0665)
lnn	-0.6188*	-1.6244***	-0.6237*
	(0.3007)	(0.4721)	(0.3110)
$lnsk$	0.1723*	0.1786	0.3195**
	(0.0737)	(0.1067)	(0.1150)
$lnattain$	0.1337	0.2467*	0.1666
	(0.0717)	(0.1050)	(0.1069)
ssh	-0.2005	-0.4693**	-0.0498
	(0.1170)	(0.1690)	(0.1267)
* $p : \leq 0.05$	** ≤ 0.01	*** ≤ 0.001	

shows that this is not the case. There is a lot of literature addressing this issue. For example Collier and Gunning (1999) mention that in 1975 60% of all Africans lived in regimes that were not legally elected and democratic structures are often not achievable in the medium-term. Additionally, governments tend to implement lax monetary policies, not considering the inflationary long-run effects. They also report of high corruption, bureaucracy and a lack of public security. Another example is Were and Nafula (2003) who show how diseases and especially AIDS affect economic indicators.

In order to eliminate the Africa-Dummy, authors add variables to the growth regression. Sachs and Warner (1997) focus on the effects of trade openness and landlocked status. They conclude that a lack of liberalization and too restrictive foreign policies impair economic growth in sub-Saharan Africa. Additionally, countries without access to the sea suffer from comparative disadvantages. After controlling for these factors, the Africa-Dummy is no longer significant. Easterly and Levine (1997) point out that ethnic diversity, measured in units of spoken languages in a country, could influence the economic development in a country. They argue that a strong mixture of different racial groups causes discord about the public resources. Furthermore, diversified societies tend to civil war and lower democratization. The authors are able to explain a large share of the cross-country variation using this measure. Easterly and Levine (1997) link their result to the historical background of sub-Saharan Africa. Like Arcand, Guillaumont and Jeanneney (2000) express, the underlying problem of the continent stems from the 'carve-up' among its occupants during the 19th century. From the authors' viewpoint, this colonial heritage still causes economic drawbacks. Acemoglu, Johnson and Robinson (2001) bring up a historical explanation that is based on the origins of the colonization. Price (2003) also address the problem of determining the effects of colonial heritage on economic growth in sub-Saharan Africa.

Adding variables to the growth regression in order to explain the Africa-Dummy is critical. The extra variables identify unique characteristics of sub-Saharan Africa and therefore act like the Africa-Dummy. For example Levine and Renelt (1992) test the causality of different explanatory variables in growth regressions. They summarize that most of the included variables are not robust and dependent on the model. Many explanatory variables that are added in growth regressions do not necessarily identify drivers for growth. Moreover, Collier and Gunning (1999) note that the addition of explanatory variables transfers the puzzle elsewhere.

3 Identifying the Africa-Dummy

3.1 Growth Regressions

Sampling Process: We denote the information of the dependent variable from some initial time point 1 up to t by $y_i^t = (y_{i1}, \dots, y_{it})$ and the information of the exogenous variables from some initial time point 2 up to t by $x_i^t = (x'_{i2}, \dots, x'_{it})$. We assume that $\{(y_i^T, x_i^T)\ ,\ i = 1, \dots, n\}$ is a number of independent observations from the same probability distribution, with finite first and second order moments.

Regression Equation: We are aiming for estimating (11) with the Africa-Dummy. (11) is of the form

$$y_{it} = \rho y_{i(t-1)} + x'_{it}\beta + \eta_i + \nu_{it}. \tag{12}$$

The Africa-Dummy is a part of the country-specific effects

$$\eta_i = \eta_g + SSH * 1_{SSH,i} + \tilde{\eta}_i, \tag{13}$$

where $E(\tilde{\eta}_i) = 0$, $1_{SSH,i}$ equals 1 if country i belongs to the group of sub-Saharan African countries and 0 else and η_g is the common intercept. When plugging (13) in (12) we have

$$y_{it} = \eta_g + \rho y_{i(t-1)} + x'_{it}\beta + SSH * 1_{SSH,i} + \tilde{\eta}_i + \nu_{it}. \tag{14}$$

We aim to estimate the parameters ρ, β, η_g, SSH and each $\tilde{\eta}_i$.

Exogeneity: We assume

$$E(\nu_{it}|1_{SSH,i}, y_i^{t-1}, x_i^T, \tilde{\eta}_i) = 0. \tag{15}$$

An implication of the assumption is that the errors ν_{it} are conditionally serially uncorrelated. Namely for $j > 0$ it holds

$$E(\nu_{it}\nu_{i(t-j)}|1_{SSH,i}, y_i^{t-1}, x_i^T, \tilde{\eta}_i) = 0.$$

By the law of iterated expectations it also holds that

$$E(\nu_{it}\nu_{i(t-j)}) = 0.$$

Hoeffler (2002) made the feedback-assumption

$$E(\nu_{it}|1_{SSH,i}, y_i^{t-1}, x_i^t, \tilde{\eta}_i) = 0. \tag{16}$$

After smoothing there are no shocks to which the time-series could react. We have very slowly varying time-series and an error that fluctuates around zero. The major part of correlation between present regressors and past errors obviously comes from the lagged dependent variable. We will show in this section that even the endogeneity bias resulting from the lagged dependent variable is negligibly small when estimating the Africa-Dummy. We do not think that this stricter assumption is necessary.

Second Moments of the Errors: We assume

$$E(\nu_{it}\nu_{js}) = \begin{cases} \sigma_\nu^2, & \text{if } i = j \text{ and } s = t \\ 0, & \text{else.} \end{cases} \tag{17}$$

Furthermore, we assume that the second moments of the country-specific errors exist.

The Country-Specific Effects: We observe $\{(y_i^T, x_i^T)\ ,\ i = 1, \dots, n\}$ but we do not observe the country-specific intercepts. The model by Mankiw, Romer and Weil (1992) indicates that the total country-specific effect η_i is determined by the growth rate of technological change g, the convergence rate λ and the initial level of technology $A(0)$. g and λ are assumed not to change between countries and over time. The initial endowment with production technology cannot be expected to be constant in all countries. Mankiw, Romer and Weil (1992) mention several influences on $A(0)$ like resources, climate or institutions. They decompose $A(0)$ in a common component that reflects the general productivity and a component that reflects all country-specific characteristics. The assumption that $\tilde{\eta}_i$ and x_{it} are uncorrelated seems to be too strong. For example developed institutions can increase the level of human capital in the population. We assume that $\tilde{\eta}_i$ is in general correlated to the error and to every $y_{i(s-1)}$ and x_{is} for all i and s.

Vector-Matrix-Notation: First of all, we stack the time-series data of (12):

$$\iota = (1, \dots, 1)' \in \mathbb{R}^{T-1}$$

$$y_i = (y_{i2}, \dots, y_{iT})' \in \mathbb{R}^{T-1}$$

$$y_{i(-1)} = (y_{i1}, \dots, y_{i(T-1)})' \in \mathbb{R}^{T-1}$$

$$X_i = (x_{i2}, \dots, x_{iT}) \in \mathbb{R}^{K \times (T-1)}$$

$$\nu_i = (\nu_{i2}, \dots, \nu_{iT})' \in \mathbb{R}^{T-1}.$$

Equation (12) is

$$y_i = \rho y_{i(-1)} + X_i'\beta + \eta_i \iota + \nu_i \in \mathbb{R}^{T-1}.$$

Furthermore, we stack cross-sectional data:

$$
\begin{aligned}
y &= (y_1', \ldots, y_n')' \in \mathbb{R}^{n(T-1)} \\
y_{-1} &= (y_{1(-1)}', \ldots, y_{n(-1)}')' \in \mathbb{R}^{n(T-1)} \\
X &= (X_1, \ldots, X_n)' \in \mathbb{R}^{n(T-1)\times K} \\
C &= I_n \otimes \iota \in \mathbb{R}^{n(T-1)\times n} \\
\eta &= (\eta_1, \ldots, \eta_n)' \in \mathbb{R}^{n} \\
\nu &= (\nu_1', \ldots, \nu_n') \in \mathbb{R}^{n(T-1)}.
\end{aligned}
$$

Equation (12) is

$$
y = \rho y_{-1} + X\beta + C\eta + \nu \in \mathbb{R}^{n(T-1)}. \tag{18}
$$

(14) is stacked in the same way. We assume without loss of generality that the data are available in the form that exactly the first s rows belong to the group of sub-Saharan African countries. Denote

$$
\begin{aligned}
\tilde{\eta} &= (\tilde{\eta}_1, \ldots, \tilde{\eta}_n)' \in \mathbb{R}^{n}, \\
\iota_{n(T-1)} &= (1, \ldots, 1)' \in \mathbb{R}^{n(T-1)} \text{ and} \\
\iota_{n(T-1),SSH} &= (\underbrace{1, \ldots, 1}_{\in \mathbb{R}^{s(T-1)}}, \underbrace{0, \ldots, 0}_{\in \mathbb{R}^{(n-s)(T-1)}}) \in \mathbb{R}^{n(T-1)}.
\end{aligned}
$$

(14) is in stacked form

$$
y = \iota_{n(T-1)}\eta_g + \rho y_{-1} + X\beta + \iota_{n(T-1),SSH} * SSH + C\tilde{\eta} + \nu \in \mathbb{R}^{n(T-1)}. \tag{19}
$$

The Bias of the Within Group Estimator: Regression equations (12) and (14) have a lagged variable. This will cause an endogeneity bias when running OLS regressions. Orcutt and

Irwin (1948) and Kendall (1954) have shown the existence of a finite sample autoregressive bias in time-series models. Nickell (1981) has shown that this bias persists asymptotically in large panels when $n \to \infty$. In consequence, bias reduction procedures have been proposed, for example Kiviet (1995), Hahn and Kuersteiner (2002) or Phillips and Sul (2007). In the estimation methods that will be presented, the bias only occurs in the regression step where the β's are estimated. Except of the Random Effects estimator, this regression step is the same as applying the Within Group estimator. Therefore, we estimate the bias of Within Group estimator using the precise formulas as $n \to \infty$ given by Phillips and Sul (2007). Using

$$\bar{\eta}_j = y_{j\bullet} - \hat{\rho}_{WG} y_{-1j\bullet} - x'_{i\bullet} \hat{\beta}_{WG}$$

we can then see how mistakes in the Within Group estimation step affect the estimation of the fixed effects. Afterwards we can estimate the bias of SSH using

$$S\hat{S}H = \bar{\eta}_A - \bar{\eta}_{NA}.$$

Since the true ρ is not known, we calculate biases for different ρ's. The Within Group estimator of the coefficient of the lagged variable is biased downwards and therefore we use it as the smallest ρ to plug in. We calculate these biases for $\hat{\beta}_{WG}$ since fluctuations result in negligible small differences. The results are given in table (2). The biases of the fixed effects listed in this table are the maximum of all absolute values of the biases of each fixed effect. Table (2) shows that all biases, apart from that of the coefficient of the lagged variable, are negligible small. Calculating biases when adding more exogenous variables is not necessary since Phillips and Sul (2007) argue that the addition of exogenous variables result in smaller biases. We therefore assume in regressions using the Within Group estimator that the bias that results from the lagged variable (apart from that of the coefficient of the lagged variable itself) is negligible small.

Two-Step Regressions: There are to ways to estimate the Africa-Dummy. The two-step

Table 2: Biases

True Value	Bias		
ρ	ρ (10^{-2})	lnn (10^{-4})	lnsk (10^{-4})
0.98971	-1.4298	7.1027	5.1380
0.99314	-1.3815	6.8523	3.7093
0.99657	-1.3302	6.5506	2.2294
1.00000	-1.2760	6.1961	0.7095
ρ	lnattain (10^{-4})	FE (10^{-17})	SSH (10^{-17})
0.98971	-0.6022	2.9554	4.0494
0.99314	0.5347	2.8576	3.9813
0.99657	1.7158	2.7540	3.9082
1.00000	2.9316	2.6441	3.8292

method first estimates (12) together with the country-specific effects which contain the Africa-Dummy according to decomposition (13). In the second step the estimated country-specific effects are used to estimate equation (13) and to obtain an estimator for the Africa-Dummy. This method has the disadvantage that it does not use all the available information from the correlations between the different variables of (14). The other estimation method estimates (14) directly and does not share this disadvantage.

Lags: Running the regressions using exactly (14) has three drawbacks. First, the one year growth time-series shows little variation so that the coefficient of the lagged dependent variable is almost one and all other coefficients are very small. Second, we only checked that the endogeneity bias caused by the lagged dependent variable is small. Since the economy can choose its growth driving parameters as reaction of a shock, the regression is suspected to suffer from an endogeneity bias. It is natural to assume that the bias caused by the explanatory variables is much smaller than that caused by the lagged dependent variable itself, which is already negligibly small. Nevertheless, we do not know the exact correlation of explanatory variables and the error and cannot give precise formulas for the bias as done by Phillips and Sul (2007). Third, we aim for comparison of our results with that of other authors, who refer their regressions to five year time horizons taking either averaged or initial explanatory variables to represent the time horizons (see Hoeffler (2002)).

Taking lagged variables has two drawbacks. First, we move away from the situation described by Mankiw, Romer and Weil (1992) and loose theoretic justification. Second, the model by Mankiw, Romer and Weil (1992) deals with the evolution of the differences of the logarithms of the subsequent GDP's. These can only be interpreted as growth rates if the subsequent GDP's are close to each other, since in this case a Taylor-Expansion shows that

$$\ln(GDP_t) - \ln(GDP_{t-1}) \approx \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}.$$

Time horizons from $t-5$ to t generate larger differences between the two growth rates than time horizons from $t-1$ to t.

Therefore, we always run the regression with a one year lagged dependent variable and contemporary explanatory variables and recheck the results with a regression with a five year lagged dependent variable and five year lagged explanatory variables.

3.2 Why we do not use System GMM

Caselli, Esquivel, Lefort (1996) applied the Difference GMM to growth regression using linear smoothed data with five year time horizons between 1960 and 1985. Bond, Hoeffler and Temple (2001) note that the Difference GMM uses weak instruments because the series of the logarithms of GDP's per capita is highly persistent and recommend the System GMM. Afterwards, many papers appear using System GMM. Roodman (2006) gives access to System GMM by implementing it in Stata. Hoeffler (2002) addresses the problem of estimating the Africa-Dummy in growth regressions and comes to the conclusion that System GMM is the preferred method. We have the impression that the System GMM is the leading method in growth regressions. As most authors use linear smoothing instead of applying the HP filter their time-series are shorter which leads to less instruments. The number of instruments when having time-series data with $T = 48$ is very large. This causes problems. Furthermore, Hoeffler (2002) applies a two step method for estimating the Africa-Dummy which leads to efficiency problems. Before discussing these problems we give an account of the System GMM.

First of all, we stack the time-series data of model (12) and write it as

$$y_i = W_i\alpha + \eta_i\iota + \nu_i$$

with $W_i = (w_{i2}, \ldots, w_{iT})' \in \mathbb{R}^{(T-1)\times(K+1)}$, $w_{it} = (y_{i(t-1)}, x'_{it})' \in \mathbb{R}^{K+1}$, $\nu_i = (\nu_{i2}, \ldots, \nu_{iT})' \in \mathbb{R}^{T-1}$ and $\alpha = (\rho, \beta')' \in \mathbb{R}^{K+1}$. We assume the feedback assumption

$$E(\nu_{it}|x_{i2}, \ldots, x_{it}, y_{i1}, \ldots, y_{i(t-1)}, \eta_i) = 0. \tag{20}$$

It follows from (20) that for $t = 3, \dots, T$

$$E((w'_{i2}, \dots, w'_{i(t-1)})'(\nu_{it} - \nu_{i(t-1)})) = 0 \in \mathbb{R}^{(K+1)(t-2)} \tag{21}$$

holds. These are $r_{Diff} = (K+1)(T-2)(T-1)/2$ moment conditions. Note that

$$\Delta\nu_i = D\nu_i = (\nu_{i3} - \nu_{i2}, \dots, \nu_{iT} - \nu_{i(T-1)})' \in \mathbb{R}^{T-2},$$

with

$$D = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(T-2)\times(T-1)}.$$

Stacking these moment conditions gives

$$E(Z'_i D\nu_i) = 0 \in \mathbb{R}^{r_{Diff}}, \tag{22}$$

where

$$Z_i = \begin{pmatrix} w'_{i2} & & & & & & \\ & w'_{i3} & w'_{i2} & & & & \\ & & & \ddots & & & \\ & & & & w'_{i(T-1)} & \dots & w'_{i2} \end{pmatrix} \in \mathbb{R}^{(T-2)\times r_{Diff}}.$$

One can derive the Difference GMM Estimator by applying the usual GMM procedure using (22). It was first proposed by Arellano and Bond (1991). Arellano and Bond (1998) show that the instruments of the Difference GMM estimator are weak when the autoregressive coefficient is close to one. The System GMM estimator augments the set of moments of the Difference GMM estimator by additionally assuming moment conditions for the level equation. When doing this, Arellano and Bond (1998) observe a dramatic efficiency gain when the autoregressive coefficient is close to one. Arellano and Bond (1995) introduce the use of lagged differences as possible instruments for the equation in levels. The equation in levels is (12) in the form

$$y_{it} = \eta_g + w_{it}'\alpha + \tilde{\eta}_i + \nu_{it} = \tilde{w}_{it}'\tilde{\alpha} + u_{it}$$

with $\tilde{w}_{it} = (1, w_{it}')' \in \mathbb{R}^{K+2}$, $\tilde{\alpha} = (\eta_g, \alpha')' \in \mathbb{R}^{K+2}$ and $u_{it} = \tilde{\eta}_i + \nu_{it}$. Stacking time-series data gives

$$y_i = \tilde{W}_i\tilde{\alpha} + u_i \in \mathbb{R}^{T-1},$$

with $\tilde{W}_i = (\tilde{w}_{i2}, \ldots, \tilde{w}_{iT})' \in \mathbb{R}^{(T-1)\times(K+2)}$, $u_i = (u_{i2}, \ldots, u_{iT}) \in \mathbb{R}^{T-1}$. We assume that

$$E(\Delta x_{it}\eta_i) = 0 \tag{23}$$

and the initial condition

$$E(\Delta y_{i2}\eta_i) = 0 \tag{24}$$

hold. In this case condition (24) also holds for all subsequent y_{it}. It follows that

$$E(\eta_i(y_{i(t-1)} - y_{i(t-2)}, x_{it}' - x_{i(t-1)}')') = E(\eta_i\Delta w_{it}) \text{ for } t = 3, \ldots, T.$$

(23) and (24) imply the $(T-2)(K+1)$ moment conditions

$$E(\Delta w_{it}u_{it}) = 0 \text{ for } t = 3, \ldots, T.$$

Furthermore, it clearly holds that $E(u_{it}) = 0$. We summarize all these $r_{Lev} = (T-2)(K+2)$ moment conditions by

$$E(Z_{li}'u_i) = 0 \in \mathbb{R}^{r_{Lev}} \tag{25}$$

with

$$Z_{li} = \begin{pmatrix} 0 & \cdots & 0 \\ (1, \Delta w_{i3}') & & \\ & \ddots & \\ & & (1, \Delta w_{iT}') \end{pmatrix} \in \mathbb{R}^{(T-1)\times r_{Lev}}.$$

The equation in levels is $y_i = \tilde{W}_i\tilde{\alpha} + u_i \in \mathbb{R}^{T-1}$ with moment conditions $E(Z'_{li}u_i) = 0$. The equation in differences is $Dy_i = DW_i\alpha + D\nu_i = D\tilde{W}_i\tilde{\alpha} + Du_i \in \mathbb{R}^{T-2}$ with moment conditions $E(Z'_iD\nu_i) = E(Z'_iDu_i) = 0$. Stacking the levels equation on the differenced equation yields

$$y_i^\dagger = W_i^\dagger\tilde{\alpha} + u_i^\dagger \in \mathbb{R}^{2T-3},$$

with

$$y_i^\dagger = \begin{pmatrix} y_i \\ Dy_i \end{pmatrix} \in \mathbb{R}^{2T-3},\ W_i^\dagger = \begin{pmatrix} \tilde{W}_i \\ D\tilde{W}_i \end{pmatrix} \in \mathbb{R}^{(2T-3)\times(K+2)} \text{ and } u_i^\dagger = \begin{pmatrix} u_i \\ Du_i \end{pmatrix} \in \mathbb{R}^{2T-3}.$$

Summarizing all $r = r_{Lev} + r_{Diff} = (T-2)(K+2) + (K+1)(T-2)(T-1)/2$ moment conditions yields

$$E((Z_i^\dagger)'u_i^\dagger) = 0 \in \mathbb{R}^r, \tag{26}$$

with

$$Z_i^\dagger = \begin{pmatrix} Z_{li} & 0 \\ 0 & Z_i \end{pmatrix} \in \mathbb{R}^{(2T-3)\times r}.$$

The System GMM estimator is

$$\hat{\tilde{\alpha}}_{SysGMM} = \left[(\sum_{i=1}^n (W_i^\dagger)'Z_i^\dagger)A_n(\sum_{i=1}^n (Z_i^\dagger)'W_i^\dagger) \right]^{-1} (\sum_{i=1}^n (W_i^\dagger)'Z_i^\dagger)A_n(\sum_{i=1}^n (Z_i^\dagger)'y_i^\dagger). \tag{27}$$

The optimal choice of the weighting matrix A_n is the inverse of $Var((Z_i^\dagger)'u_i^\dagger)$.

Hoeffler (2002) addresses the problem of estimating the Africa-Dummy in growth regressions. She applies a two-step regression, estimating (12) with an intercept first and then regressing the residuals on the Africa-Dummy. This method has efficiency problems that result from the variation induced by estimating the residuals in the first step and from the GMM method in general. It is not surprising that Hoeffler (2002) observes that the negative

Africa-Dummy becomes insignificant. Beside this, the System GMM has more problems. One general problem of GMM is a bias that occurs when too many instruments are used (see Tauchen (1986) or Ziliak (1997)). Windmeijer (2005) observes a decreasing bias when applying the Difference GMM if the instrument count is reduced. Arellano (2003) gives analytical evidence for the bias when the number of observations and the length of the time-series go to infinity.

Furthermore, problems occur when estimating the optimal weighting matrix A_n. The number of elements to be estimated is quadratic in the number of instruments and therefore quartic in T. Moreover, the elements of the optimal matrix are fourth moments of the underlying distributions because they are second moments of the result of differenced variables times variables. Roodman (2009) notes that a common symptom for estimations of the weighting matrix is that they are singular. Therefore, the generalized inverse rather than the inverse is calculated. This can give results that are far away from the theoretical ideal. The breakdown tends to occur as the number of instruments approaches n. Therefore, n can be seen as a general benchmark for the number of instruments. We have 4554 instruments when estimating with the System GMM and n equals 81.

The Hansen J-Test (see Hansen (1982)) usually checks the validity of instruments, but as for example Bowsher (2002) observes in simulation studies, a too large number of instruments weakens the test dramatically. Roodman (2009) notes that in case of too many instruments the weights of those moments that are least well satisfied are too small. We conclude that we do not have a reliable test available that tells us how many and which instruments to choose. This problem is especially severe as the initial condition (24) is least likely to be fulfilled in case of highly persistent time-series as in our case. To understand this we follow the arguments of Roodman (2009). If there exists a long-run mean, it holds that

$$E(y_{it}|\eta_g, \tilde{\eta}_i, x_i^T) = E(y_{i(t+1)}|\eta_g, \tilde{\eta}_i, x_i^T),$$

which is equivalent to

$$y_{it} = \frac{x'_{i(t+1)}\beta}{1-\rho} + \frac{\eta_g}{1-\rho} + \frac{\tilde{\eta}_i}{1-\rho} \;\forall t.$$

Assuming that there exists such a long-run mean, we define the correlation of the deviations from it to $\tilde{\eta}_i$ by

$$m_{it} = E((y_{it} - (\frac{x'_{i(t+1)}\beta}{1-\rho} + \frac{\eta_g}{1-\rho} + \frac{\tilde{\eta}_i}{1-\rho}))\tilde{\eta}_i).$$

m_{it} has got interesting properties. First of all, if the initial condition (24) holds for example in t and therefore $E(\Delta y_{i(t-1)}u_{it}) = 0$, then this is equivalent to $m_{i(t-2)} = 0$. This is because

$$0 = E(\Delta y_{i(t-1)}u_{it}) = E(((\rho-1)y_{i(t-2)} + x'_{i(t-1)}\beta + \eta_g + \tilde{\eta}_i)\tilde{\eta}_i)$$

is equivalent to

$$0 = E((y_{i(t-2)} - (\frac{x'_{i(t-1)}\beta}{1-\rho} + \frac{\eta_g}{1-\rho} + \frac{\tilde{\eta}_i}{1-\rho}))\tilde{\eta}_i) = m_{i(t-2)}.$$

Furthermore, if assumption (23) holds, it follows that $m_{it} = \rho m_{i(t-1)}$. This is because

$$\begin{aligned}
m_{it} &= E((y_{it} - (\frac{x'_{i(t+1)}\beta}{1-\rho} + \frac{\eta_g}{1-\rho} + \frac{\tilde{\eta}_i}{1-\rho}))\tilde{\eta}_i) \\
&= E((y_{it} - (\frac{x'_{it}\beta}{1-\rho} + \frac{\eta_g}{1-\rho} + \frac{\tilde{\eta}_i}{1-\rho}))\tilde{\eta}_i) \\
&= E((\rho y_{i(t-1)} - \frac{\rho}{1-\rho}x'_{it}\beta - \frac{\rho}{1-\rho}\eta_g - \frac{\rho}{1-\rho}\tilde{\eta}_i)\tilde{\eta}_i) \\
&= \rho E((y_{i(t-1)} - (\frac{x'_{it}\beta}{1-\rho} + \frac{\eta_g}{1-\rho} - \frac{\tilde{\eta}_i}{1-\rho}))\tilde{\eta}_i) \\
&= \rho m_{i(t-1)}.
\end{aligned}$$

This means that if the system has been generating numbers, such that (24) holds once, it also holds for all subsequent y_{it}. The initial condition for an individual is for example fulfilled if it has already achieved its long run steady state and is only fluctuating around it with respect to ν_{it}. If the country is in its transition phase to its steady state, then the difference to its long-run steady state can be uncorrelated to the individual error but this is not necessarily the case. However, if $\rho < 1$ the correlations of the differences to the steady state to the individual errors decrease with speed determined by ρ. The System GMM offers the most help if ρ is close one in which case the system is least likely to have achieved the initial condition when the observation time begins. Therefore, when the System GMM becomes especially necessary it is least likely to fulfill the underlying assumptions that allow to apply it. The Hansen J-Test does not offer help to test the validity of the moment conditions because of the large number of instruments.

Roodman (2009) provides methods to reduce the instrument count. Limiting the lag-depth to one gives an instrument count which is still far too large. Another method to reduce the instrument count is collapsing. Suppose we do not assume that

$$E(Z_i' Du_{it}) = E((w_{i2}'\Delta u_{i3}, w_{i3}'\Delta u_{i4} w_{i2}'\Delta u_{i4}, \dots, w_{i(T-1)}'\Delta u_{iT}, \dots, w_{i2}'\Delta u_{iT})') = 0 \in \mathbb{R}^{r_{Diff}},$$

but only assume that

$$E(Z_i' Du_{it}) = E((\sum_{t=3}^{T} w_{i(t-1)}'\Delta u_{it}, \sum_{t=4}^{T} w_{i(t-2)}'\Delta u_{it}, \dots, \sum_{t=T}^{T} w_{i(t-(T-2))}'\Delta u_{it})) = 0 \in \mathbb{R}^{(T-2)(K+1)}.$$

In the same way we can collapse the additional instruments for the System GMM estimator. Instead of assuming that $E((\Delta w_{i3}u_{i3}, \Delta w_{i4}u_{i4}, \dots, \Delta w_{iT}u_{iT})) = 0$, we assume that $E(\sum_{t=3}^{T} \Delta w_{it}u_{it}) = 0$. The instrument count is still far too large. The only way is to collapse and to reduce the lag-depth. If we reduce the lag depth to two and collapse, we have 13 instruments. Note that reducing the number of instruments makes it possible to apply the System GMM but has large drawbacks in terms of efficiency.

We conclude that there is a dramatic loss of efficiency due to reducing the instrument count. Additionally there is a dramatic loss of efficiency that results from the two-step method. This can give results that are far away from the theoretical ideal. Moreover, the significance of the Africa-Dummy is hard to determine. On the other hand, when estimating with Least Squares an endogeneity bias has to be accepted. Therefore, the decision of which estimation method to choose is a bias-variance trade-off. As shown in subsection (3.1) the bias is negligible small. Therefore, we do not use System GMM.

3.3 The Hausman-Taylor Estimator

We estimate the coefficients of equation (14)

$$\begin{aligned} y &= \iota_{n(T-1)}\eta_g + \rho y_{-1} + X\beta + \iota_{n(T-1),SSH} * SSH + C\tilde{\eta} + \nu \\ &= W(\eta_g, SSH, \rho, \beta')' + u, \end{aligned}$$

where $W = (\iota_{n(T-1)}, \iota_{n(T-1),SSH}, y_{-1}, X) \in \mathbb{R}^{n(T-1)\times(K+3)}$ and $u = C\tilde{\eta} + \nu \in \mathbb{R}^{T-1}$. We assume for the second moments of the country-specific errors, that they are independent and that their common variance is

$$Var(\tilde{\eta}_i) = \sigma_\eta^2. \tag{28}$$

The Random Effects model disregards the correlation of η_i to the exogenous regressors. The first approach to estimate this model is to pool all data and then apply OLS. The pooled estimator provides consistent estimates. As the errors u_{it} are correlated, a robust choice to estimate the coefficients yields more efficient estimates. The covariance matrix of the vector $u_i = (u_{i2}, \ldots, u_{iT})$ is

$$\Sigma_{u_i} = \begin{pmatrix} \sigma_\eta^2 + \sigma_\nu^2 & \sigma_\eta^2 & \cdots & & \sigma_\eta^2 \\ \vdots & & \ddots & & \vdots \\ \sigma_\eta^2 & & \cdots & \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\nu^2 \end{pmatrix} \in \mathbb{R}^{(T-1)\times(T-1)}.$$

Therefore, the covariance matrix of the vector $u = (u_1', \ldots, u_n')' \in \mathbb{R}^{n(T-1)}$ is

$$\Sigma = \begin{pmatrix} \Sigma_{u_1} & & \\ & \ddots & \\ & & \Sigma_{u_n} \end{pmatrix} \in \mathbb{R}^{n(T-1)\times n(T-1)}.$$

Applying GLS yields an unfeasible estimator of $(\eta_g, SSH, \rho, \beta')'$, namely

$$(W'\Sigma^{-1}W)^{-1}W'\Sigma^{-1}y.$$

The solution to this is the same as regressing the quasi-demeaned y on the quasi-demeaned columns of W. If vector z is

$$z = (z_{12}, \ldots, z_{1T}, z_{22}, \ldots, z_{2T}, \ldots, z_{n2}, \ldots, z_{nT})' \in \mathbb{R}^{n(T-1)},$$

then the quasi-demeaned z is

$$\tilde{z}_{QD} = (z_{12}-\theta\bar{z}_{1\bullet}, \ldots, z_{1T}-\theta\bar{z}_{1\bullet}, z_{22}-\theta\bar{z}_{2\bullet}, \ldots, z_{2T}-\theta\bar{z}_{2\bullet}, \ldots, z_{n2}-\theta\bar{z}_{n\bullet}, \ldots, z_{nT}-\theta\bar{z}_{n\bullet})' \in \mathbb{R}^{n(T-1)},$$

where

$$\bar{z}_{i\bullet} = \frac{1}{T-1}\sum_{t=2}^{T} z_{it}$$

and

$$\theta = 1 - \sqrt{\sigma_\nu^2/((T-1)\sigma_\eta^2 + \sigma_\nu^2)}.$$

To obtain a feasible version, we estimate the variances of the error components. The pooled OLS estimator gives consistent estimates for the residuals u_{it} which we denote by $\hat{\hat{u}}_{it}$ and a consistent estimator of its variance which we denote by $\hat{\sigma}_u^2$. Consistent estimators for the variances of the error components and θ are given by

$$\begin{aligned}
\hat{\sigma}_{\tilde{\eta}}^2 &= \frac{1}{n(T-1)(T-2)/2 - (K+3)} \sum_{i=1}^{n}\sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1} \hat{\hat{u}}_{it}\hat{\hat{u}}_{is}, \\
\hat{\sigma}_\nu^2 &= \hat{\sigma}_u^2 - \hat{\sigma}_{\tilde{\eta}}^2, \\
\hat{\theta} &= 1 - \sqrt{\hat{\sigma}_\nu^2/((T-1)\hat{\sigma}_\eta^2 + \hat{\sigma}_\nu^2)}.
\end{aligned} \tag{29}$$

We use $\hat{\theta}$ to obtain the quasi-demeaned W and y, namely $\tilde{W}_{QD}$ and $\tilde{y}_{QD}$. The Random Effects estimator is

$$(\hat{\eta_g}_{RE}, S\hat{S}H_{RE}, \hat{\rho}_{RE}, \hat{\beta}'_{RE})' = (\tilde{W}'_{QD}\tilde{W}_{QD})^{-1}\tilde{W}'_{QD}\tilde{y}_{QD}. \tag{30}$$

As the individual effects of (14) are correlated to the regressors, the Random Effects estimator suffers from an endogeneity bias. Hausman and Taylor (1981) present an Instrumental Variable estimator for estimating the coefficients of (14). Since the Africa-Dummy already rules out systematic differences of the group of the sub-Saharan African countries, we assume

$$E(1_{SSH,i} * \tilde{\eta}_i) = 0. \tag{31}$$

When demeaning the regression equation, all individual variables disappear. Furthermore, we disregard the endogeneity bias induced by the lagged variable (see subsection (3.1)) and use (31). Then

$$Z = (\iota_{n(T-1)}, \iota_{n(T-1),SSH}, (I_{n(T-1)} - I_n \otimes \iota\iota')y_{-1}, (I_{n(T-1)} - I_n \otimes \iota\iota')X) \in \mathbb{R}^{n(T-1)\times(K+3)}$$

is a matrix whose columns provide instruments. Hausman and Taylor (1981) propose to use some of the explanatory variables as additional instruments. In our case we do not find a reason for that one of the explanatory variables is uncorrelated to the individual effects. Applying the Instrumental Variable estimation method yields an estimator for $(\eta_g, SSH, \rho, \beta')'$, namely $(Z'W)^{-1}Z'y$. The solution to this is that β and ρ are estimated by the Within Group estimator and η_g and SSH are estimated by $\bar{\eta}_{NA}$ and $\bar{\eta}_A - \bar{\eta}_{NA}$ respectively, where we denote the average residual of country j by $\bar{\eta}_j = y_{j\bullet} - \hat{\rho}_{WG}y_{-1j\bullet} - x'_{i\bullet}\hat{\beta}_{WG}$, the average residual of all non sub-Saharan African countries by $\bar{\eta}_{NA} = \frac{1}{n-s}\sum_{j=s+1}^{n} \bar{\eta}_j$ and the average residual of all sub-Saharan African countries by $\bar{\eta}_A = \frac{1}{s}\sum_{j=1}^{s} \bar{\eta}_j$. Since we have error components, we apply 2SLS using (29). This is the Hausman-Taylor estimator.

If conditional on the regressors, individual effects can be viewed as random draws from a common population, we estimate with error components. One motivation for doing this could be that the common population characteristics are of interest. In growth regression, it is very unlikely that there is a common population. The effects of different countries are highly heterogeneous. Furthermore, the performance of individual countries is of interest. The disadvantage of Random Effects estimators is that it does not take this heterogeneity into account and it is not possible to examine the performance of individual countries.

3.4 The Two-Groups Least-Square Dummy-Variable Estimator

The Least-Square Dummy-Variable estimator is the OLS estimator of ρ, β and of each η_i in equation (12)

$$\hat{\rho}_{LSDV} = \hat{\rho}_{WG},\ \hat{\beta}_{LSDV} = \hat{\beta}_{WG} \text{ and } \hat{\eta}_{LSDV,i} = \bar{\eta}_i \text{ for } i = 1, \ldots, n. \tag{32}$$

Since $(C'C)^{-1} = \frac{1}{T-1} I_{n(T-1)}$ the model can be identified. Equation (14) has $n+2$ country-specific regressors (an intercept, n country-specific errors and an Africa-Dummy). When stacking this equation and considering the country-specific regressor matrix, it has $n+2$ columns and $n(T-1)$ rows from which only n rows are different to each other. Therefore, the country-specific regressor matrix has rank n at the highest and the model cannot be identified. Therefore, applying the Least-Square Dummy-Variable estimator yields in applying a two-step regression, which has efficiency problems.

To be able to estimate (14) directly, we assume that the errors of the sub-Saharan African countries sum up to zero and that the errors of the non-sub-Saharan African countries sum up to zero separately

$$\sum_{i=1}^{s} \tilde{\eta}_i = 0 \text{ and } \sum_{i=s+1}^{n} \tilde{\eta}_i = 0. \tag{33}$$

This assumption specifies two errors precisely

$$\tilde{\eta}_s = -\tilde{\eta}_1 - \tilde{\eta}_2 - \ldots - \tilde{\eta}_{s-1} \text{ and } \tilde{\eta}_n = -\tilde{\eta}_{s+1} - \tilde{\eta}_{s+2} - \ldots - \tilde{\eta}_{n-1}.$$

Plugging (33) into (14) yields

$$y = \rho y_{-1} + X\beta + C_{SSH}\eta_{SSH} + \nu \in \mathbb{R}^{n(T-1)}, \tag{34}$$

with

$$\eta_{SSH} = (\eta_g, SSH, \tilde{\eta}_1, \ldots, \tilde{\eta}_{s-1}, \tilde{\eta}_{s+1}, \ldots, \tilde{\eta}_{n-1})' \in \mathbb{R}^n$$

and

$$C_{SSH} = \left(\begin{array}{cc|ccc|ccc}
\iota & \iota & \iota & & & & & \\
\vdots & \vdots & & \ddots & & & & \\
\iota & \iota & & & \iota & & & \\
\iota & \iota & -\iota & \cdots & -\iota & & & \\
\hline
\iota & & & & & \iota & & \\
\vdots & & & & & & \ddots & \\
\iota & & & & & & & \iota \\
\iota & & & & & -\iota & \cdots & -\iota
\end{array}\right) \in \mathbb{R}^{n(T-1)\times n},$$

where the lower right box refers to the non-sub-Saharan African countries and has $n-s-1$ columns and $(n-s)(T-1)$ rows and the upper middle box refers to the sub-Saharan African countries and has $s-1$ columns and $s(T-1)$ rows. It is easy to check that

$$C'_{SSH}C_{SSH} = (T-1)\begin{pmatrix} Z_1 & & \\ & Z_2 & \\ & & Z_3 \end{pmatrix} \in \mathbb{R}^{n\times n},$$

with

$$Z_1 = \begin{pmatrix} n & s \\ s & s \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

$$Z_2 = \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{pmatrix} \in \mathbb{R}^{(s-1)\times(s-1)},$$

and

$$Z_3 = \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{pmatrix} \in \mathbb{R}^{(n-s-1)\times(n-s-1)}.$$

The inverses of Z_1, Z_2 and Z_3 exist and are given by

$$Z_1^{-1} = \frac{1}{n-s} \begin{pmatrix} 1 & -1 \\ -1 & n/s \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

$$Z_2^{-1} = \frac{1}{s} \begin{pmatrix} (s-1) & -1 & \dots & -1 \\ -1 & (s-1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \dots & -1 & (s-1) \end{pmatrix} \in \mathbb{R}^{(s-1)\times(s-1)},$$

and

$$Z_3^{-1} = \frac{1}{n-s} \begin{pmatrix} (n-s-1) & -1 & \dots & -1 \\ -1 & (n-s-1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \dots & -1 & (n-s-1) \end{pmatrix} \in \mathbb{R}^{(n-s-1)\times(n-s-1)}.$$

Therefore,

$$(C'_{SSH}C_{SSH})^{-1} = \frac{1}{T-1} \begin{pmatrix} Z_1^{-1} & & \\ & Z_2^{-1} & \\ & & Z_3^{-1} \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Note that the existence of $(C'_{SSH}C_{SSH})^{-1}$ is equivalent to that the columns of C_{SSH} are linear independent, meaning that the model can be identified. It is now easy to check that

$$M_{C_{SSH}} = I_{n(T-1)} - C_{SSH}(C'_{SSH}C_{SSH})^{-1}C'_{SSH} = I_{n(T-1)} - I_n \otimes \iota\iota' \in \mathbb{R}^{n(T-1)\times n(T-1)}.$$

Therefore, ρ and β are estimated by the Within Group estimator. Furthermore,

$$\hat{\eta}_{SSH} = (C'_{SSH}C_{SSH})^{-1}C'_{SSH}(y - \hat{\rho}_{WG}y_{-1} - X\hat{\beta}_{WG}).$$

Solving this gives the Two-Groups Least-Square Dummy-Variable estimator

$$\begin{aligned} &\hat{\rho} = \hat{\rho}_{WG},\ \hat{\beta} = \hat{\beta}_{WG},\ \hat{\eta}_g = \bar{\eta}_{NA},\ S\hat{S}H = \bar{\eta}_A - \bar{\eta}_{NA}, \\ &\hat{\tilde{\eta}}_j = \bar{\eta}_j - \bar{\eta}_A \text{ for } j \in \{1, \dots, s-1\} \text{ and } \hat{\tilde{\eta}}_j = \bar{\eta}_j - \bar{\eta}_{NA} \text{ for } j \in \{s+1, \dots, n-1\}. \end{aligned} \tag{35}$$

With (35) and $-\tilde{\eta}_1 - \ldots - \tilde{\eta}_{s-1} = \tilde{\eta}_s$ we have $\hat{\tilde{\eta}}_s = \bar{\eta}_s - \bar{\eta}_A$ and in the same manner $\hat{\tilde{\eta}}_n = \bar{\eta}_n - \bar{\eta}_{NA}$. The total country-specific effect of a sub-Saharan African country with index $j \in \{1, \ldots, s\}$ is $\hat{\eta}_g + S\hat{S}H + \hat{\tilde{\eta}}_j = \bar{\eta}_j$ and that of a non-sub-Saharan African country with index $j \in \{s+1, \ldots, n\}$ is $\hat{\eta}_g + \hat{\tilde{\eta}}_j = \bar{\eta}_j$. Note that these are the country-specific effects of the Least-Square Dummy-Variable estimator.

The advantage of the Two-Groups Least-Square Dummy-Variable estimator compared with the Hausman-Taylor estimator is that it does not need the assumption of a common population. Therefore, the effects of different countries are heterogeneous. Furthermore, it allows to examine the performance of individual countries. The formulas of the Hausman-Taylor estimator for estimating the intercept, the Africa-Dummy, ρ and β are exactly the same as those of the Two-Groups Least-Square Dummy-Variable estimator but the estimators for second moments are not. The Two-Groups Least-Square Dummy-Variable estimator allows to reliably estimate the correlations of the Africa-Dummy to other regressors. Furthermore, as it does not use the inefficient Instrumental Variable method, its confidence bands are smaller.

3.5 Results

Tables (3) and (4) show the estimated coefficients and the standard errors. The interpretation of the five year lagged model is similar to that of the one year lagged model for all estimation methods. The coefficient of lnn is almost zero in the one year lagged model and at least becomes negative significant on ten percent level in the five year lagged model. It is surprising to see that the coefficient of $lnattain$ is clearly negative. Figure (6) shows this negative correlation when multiplying the dependent variable and $lnattain$ by the projection matrix that projects each vector on the orthogonal column space of that spanned by all other explanatory variables. It can clearly be seen that the negative coefficient is not a result of a misspecification of the functional structure or of influential observations. The negative coefficient was for example also identified by Islam (1995). He argues that the observed effect of human capital is either a measurement problem or relates to a misspecification of this variable by the Augmented Solow model. The indicator by Barro and Lee (2010) does not take the quality of schooling into account. It can be observed that the school attainment according to Barro and Lee (2010) incrementally increases for almost all countries but the growth rate does not. The result is a negative coefficient.

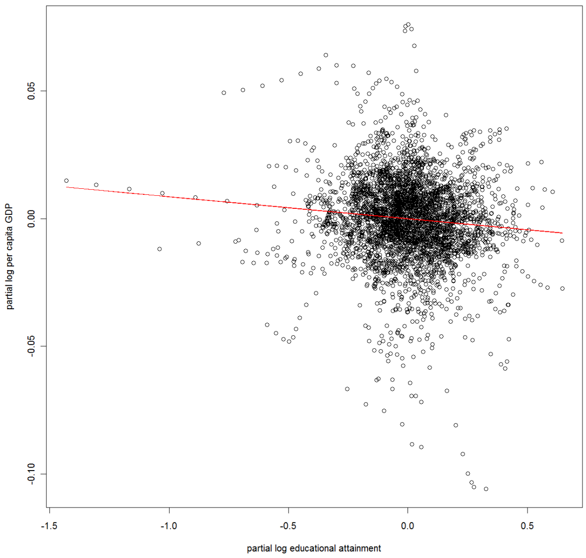

Figure 6: The negative coefficient of *lnattain* in the growth regression.

Table (3) shows the estimated coefficients of the error components models. Random Effects suffers from an endogeneity bias and its results are slightly different than Hausman-Taylor. Table (3) shows the estimated coefficients of the fixed effects models. Least-Square Dummy-Variable and Two-Groups Least-Square Dummy-Variable give similar results for the time- and country-varying coefficients but Least-Square Dummy-Variable estimates a larger intercept with smaller standard errors and an equal Africa-Dummy with much larger standard errors. Hausman-Taylor has larger standard errors than Two-Groups Least-Square Dummy-Variable. The advantage of Two-Groups Least-Square Dummy-Variable can also be seen when considering correlations. Table (5) shows the correlations of the estimated coefficients and the estimated Africa-Dummy. Least-Square Dummy-Variable does not estimate correlations at all. Random Effects and Hausman-Taylor give similar results because they are both based on the idea of error components. Two-Groups Least-Square Dummy-Variable gives very different results because it is based on the idea of including fixed effects as regressors. It does not need the rather strict group-wise homogeneity assumption which is why we identify it as the best estimator to calculate the correlations.

According to Two-Groups Least-Square Dummy-Variable the coefficient of the Africa-Dummy is larger, the smaller the coefficient of lnn and $lnattain$ and the larger the coefficient of $lnsk$. Nevertheless, its correlations to the coefficient of $lnattain$ and lnn are small. In other words, if the return to investment in physical capital increases, the punishment of belonging to sub-Saharan Africa decreases. Furthermore, if the return to the depreciation rate or the school attainment increases, the punishment of belonging to sub-Saharan Africa increases slightly.

We analyze the fixed effects estimated by the Two-Groups Least-Square Dummy-Variable estimator. The total fixed effects are $\bar{\eta}_i$. The Two-Groups Least-Square Dummy-Variable estimator is able to estimate the decomposition $\tilde{\eta}_i + \eta_g + SSH * 1_{SSH;i}$. We denote $\tilde{\eta}_i + \eta_g + SSH * 1_{SSH;i}$ by fixed effects and $\tilde{\eta}_i + \eta_g$ by corrected fixed effects. The corrected fixed effects are larger than the fixed effects in case of a sub-Saharan African country and equal for all other countries. Figure (7) shows boxplots of the fixed effects in the one year lagged case. We observe that the distribution of the fixed effects is slightly skewed to the left. In the one year lagged model it can be seen that adding the Africa-Dummy as a regressor results in a more symmetric distribution of the remaining parts of the fixed effects. The two outliers of the one year lagged model correspond to the sub-Saharan African country Niger and the Latin American country Nicaragua. Even though Niger is affected by the correction, it remains being an outlier when considering the corrected country-specific errors. Figure (8) shows that the skewness is only partially evaporated when taking the corrected the five year lagged model into account. The tails of the corrected fixed effect support a symmetric distribution but as the median is closer to the first quartile than to the third quartile, the distribution is slightly skewed to the left. Nevertheless, the corrected fixed effects of the five year lagged model are slightly skewed to the left.

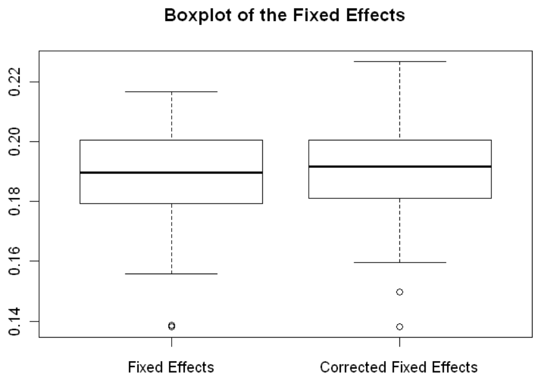

Figure 7: Boxplot of the fixed effects for the one year lagged model.

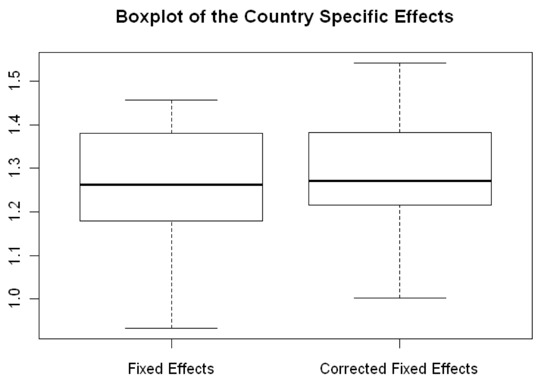

Figure 8: Boxplot of the fixed effects for the five year lagged model.

When taking the residuals into account, we observe a similar behavior for the one year lagged model and the five year lagged model. Its distribution is extremely heavy tailed and slightly skewed to the left. This indicates that more regressors than those given by Mankiw,

Table 3: Random Effects Estimators

	RE	RE (5)	HT	HT (5)
Intercept	0.1771***	1.2100***	0.1905***	1.2894***
	(0.0112)	(0.0633)	(0.0118)	(0.0649)
lag lny	0.9898***	0.9000***	0.9897***	0.8926***
	(0.0011)	(0.0059)	(0.0011)	(0.0061)
lnn	-0.0002	-0.0282*	0.0008	-0.0240
	(0.0025)	(0.0126)	(0.0025)	(0.0127)
lnsk	0.0277***	0.0837***	0.0275***	0.0813***
	(0.0012)	(0.0062)	(0.0012)	(0.0063)
lnattain	-0.0148***	-0.0496***	-0.0150***	-0.0493***
	(0.0010)	(0.0052)	(0.0010)	(0.0053)
SSH	-0.0090	-0.1428***	-0.0109*	-0.1551***
	(0.0049)	(0.0353)	(0.0046)	(0.0301)

* $p : \leq 0.05$ **≤ 0.01 ***≤ 0.001

Romer and Weil (1992) contribute to explaining growth. However, when adding a regressor to the growth model it is not clear whether it drives growth or is only somehow correlated to what cannot be explained by the model without that regressor.

4 More about the Africa-Dummy

4.1 Semiparametric Modeling

The growth model by Mankiw, Romer and Weil (1992) suggests the regression equation (14) which has a linear functional structure. We investigate if a misspecification of this functional structure is responsible for that the Africa-Dummy is negative and significant.

Table 4: Fixed Effects Estimators

	LSDV	LSDV (5)	2G LSDV	2G LSDV (5)
Intercept	0.1905***	1.2894***	0.1795***	1.1343***
	(0.0020)	(0.0134)	(0.0117)	(0.0635)
lag lny	0.9897***	0.8926***	0.9897***	0.8926***
	(0.0011)	(0.0061)	(0.0011)	(0.0061)
lnn	0.0008	-0.0240	0.0008	-0.0240
	(0.0025)	(0.0127)	(0.0025)	(0.0127)
lnsk	0.0275***	0.0813***	0.0275***	0.0813***
	(0.0012)	(0.0063)	(0.0012)	(0.0063)
lnattain	-0.0150***	-0.0493***	-0.0150***	-0.0493***
	(0.0010)	(0.0053)	(0.0010)	(0.0053)
SSH	-0.0109*	-0.1551***	-0.0109***	-0.1551***
	(0.0044)	(0.0293)	(0.0017)	(0.0090)

* $p : \leq 0.05$ **≤ 0.01 ***≤ 0.001

Table 5:

	Corr lnn	Corr lnsk	Corr lnattain	Method
RE	-0.5587	-0.0129	-0.0757	Direct
RE(5)	-0.5588	-0.0487	-0.0266	Direct
HT	-0.5605	-0.0112	-0.0753	Direct
HT(5)	-0.5589	-0.0481	-0.0240	Direct
LSDV	.	.	.	Two Step
LSDV(5)	.	.	.	Two Step
2G LSDV	-0.1170	0.5641	-0.0938	Direct
2G LSDV(5)	-0.1279	0.5252	-0.0537	Direct

We use B-Splines of degree three with equidistant knots to relax the functional structure of the variables lnn, $lnsk$ and $lnattain$. The number of knots have to be chosen in a reasonable way that takes the sample size as well as the number of regressors into account. The punishment term of Akaike's Information Criterion does not vary when the sample size is large meaning that models with a lot of parameters seem advantageous. The Bayesian Information Criterion punishes harder for choosing a lot of explanatory variables. Therefore, we chose the number of knots with respect to that it minimizes the Bayesian Information Criterion. More precisely, we vary the number of knots between three and ten and choose the combination that minimizes the Bayesian Information Criterion. The result for the one year lagged model is zero knots for the variables lnn and $lnattain$ and one knot for the variable $lnsk$. The result for the five year lagged model is one knot for all variables. When running these regressions we observe that the coefficient of the lagged dependent variable increases from 0.9897 to 0.9920 in the one year lagged model and decreases from 0.8926 to 0.8911 in the five year lagged model. Therefore, less variation is explained by the remaining variables. The intercept decreases from 0.1905 to 0.0322 in the one year lagged model and from 1.2894 to 0.8834 in the five year lagged model. The magnitude of the Africa-Dummy increases slightly from -0.0109 to -0.0113 in the one year lagged model and from -0.1551 to -0.1582 in the five year lagged model. However, in the one year lagged and five year lagged case we observe a highly significant Africa-Dummy. We conclude that the significance of the Africa-Dummy cannot be explained by a misspecification of the functional structure.

4.2 Interaction Effects

In this subsection we discuss how the beta coefficients of (14) differ for sub-Saharan African countries. We consider model (14) with interaction effects. Interaction effects also allow for time varying punishments of sub-Saharan African countries. The results are given in table

(6).

First of all, we observe a positive significant interaction effect of the coefficient of lnn. This means that it needs to be corrected for sub-Saharan African counties such that the resulting coefficient is positive. For the one year lagged model the total coefficient of lnn is $-0.0129 + 0.0357 = 0.0228$ and for the five year lagged model $-0.0760 + 0.1535 = 0.0775$. This is counterintuitive. Figure (9) shows boxplots for the time-series of lnn_{it} for the sub-Saharan African countries and other countries. We observe that sub-Saharan African countries have a larger depreciation rate because of the larger rate of population growth. Furthermore, the Inter Quartile Range is smaller with more outliers. The positive coefficient of lnn of sub-Saharan African countries shows that the larger population growth is advantageous for the growth of sub-Saharan African countries. The difference of the total coefficient of lnn for sub-Saharan African countries and the coefficient for all other countries overemphasizes sub-Saharan Africa's punishment. The Africa-Dummy is positive. A low population growth rate means that there is a low birth rate or people die. For example conflicts or diseases cause high death rates but both reduce the GDP as for example war costs money or diseases cause people not to work.

Furthermore, it can be seen from the interaction effect of the estimated coefficient of the five year lagged model that the time-series of GDP per worker entails less autocorrelation than that of the other countries. This also means that less variation is explained by the GDP per worker time-series itself and indicates that other explanatory variables, such as for example those given by the model of Mankiw, Romer and Weil (1992), contribute to growth.

Moreover, in the one year lagged model, the interaction effect of $lnattain$ is small and positive but significant. However, the resulting coefficient is still negative and of large magnitude.

4.3 The Development of the Africa-Dummy

In this subsection we investigate how the Africa-Dummy evolves over time. Consider the model

$$y_{it} = \eta_g + \rho y_{i(t-1)} + x'_{it}\beta + \sum_{s=2}^{T} SSH_s * d_{SSH,t}(i,s) + \tilde{\eta}_i + \nu_{it}, \tag{36}$$

with $t = 2, \ldots, T$ and $i = 1, \ldots, n$, where $d_{SSH,t}(i,s) = 1$ if country i belongs to sub-Saharan Africa and $s = t$ and $d_{SSH,t}(i,s) = 0$ else. We assume that this model has the same statistical properties concerning the error structure and the fixed effects as (14). This includes $\sum_{i=1}^{s} \tilde{\eta}_i = 0$ and $\sum_{i=s+1}^{n} \tilde{\eta}_i = 0$ to be able to identify the model. Stacking first time-series and then cross-sectional data yields

$$y = \rho y_{-1} + X\beta + (\iota_{SSH} \otimes I_{T-1})SSH + C\eta + \nu \in \mathbb{R}^{n(T-1)},$$

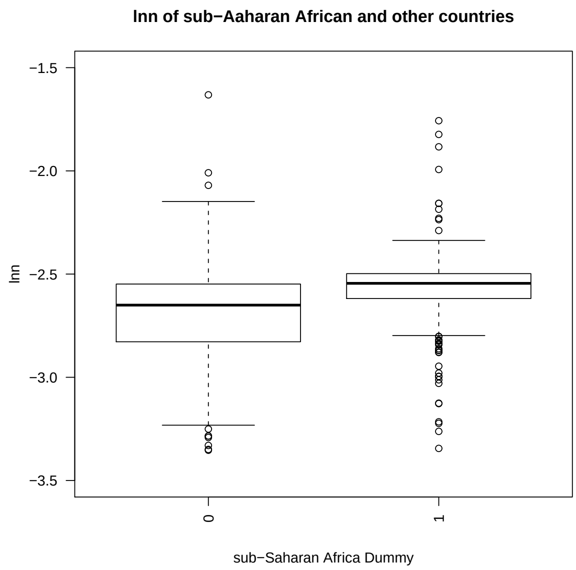

Figure 9: lnn_{it} stratified by sub-Saharan African and other countries

Table 6: Estimating the coefficients of the growth regression with interaction effects

	one year	five year
	Estimate	Estimate
	(S.E.)	(S.E.)
Intercept	0.1588***	1.0938***
	(0.0134)	(0.0724)
SSH	0.0646*	0.6151***
	(0.0266)	(0.1451)
lag lny	0.9895***	0.8976***
	(0.0013)	(0.0070)
Int. lag lny	0.0020	-0.0397**
	(0.0027)	(0.0147)
lnn	-0.0129***	-0.0760***
	(0.0031)	(0.0159)
Int. lnn	0.0357***	0.1535***
	(0.0052)	(0.0265)
lnsk	0.0268***	0.0752***
	(0.0016)	(0.0081)
Int. lnsk	0.0028	0.0145
	(0.0025)	(0.0129)
lnattain	-0.0175***	-0.0498***
	(0.0013)	(0.0070)
Int. lnattain	0.0047*	0.0017
	(0.0020)	(0.0108)
*$p : \leq 0.05$	**≤ 0.01	***≤ 0.001

where $SSH = (SSH_2, \dots, SSH_T)' \in \mathbb{R}^{T-1}$, $\eta = (\eta_g, \tilde{\eta}_1, \dots, \tilde{\eta}_{s-1}, \tilde{\eta}_{s+1}, \dots, \tilde{\eta}_{n-1})' \in \mathbb{R}^{n-1}$ and

$$C = \left(\begin{array}{c|ccc|ccc}
\iota & \iota & & & & & \\
\vdots & & \ddots & & & & \\
\iota & & & \iota & & & \\
\iota & -\iota & \cdots & -\iota & & & \\
\hline
\iota & & & & \iota & & \\
\vdots & & & & & \ddots & \\
\iota & & & & & & \iota \\
\iota & & & & -\iota & \cdots & -\iota
\end{array}\right) \in \mathbb{R}^{n(T-1)\times(n-1)}.$$

The lower right box refers to the non sub-Saharan African countries and has $n-s-1$ columns and $(n-s)(T-1)$ rows, the upper middle box refers to the sub-Saharan African countries and has $s-1$ columns and $s(T-1)$ rows and the first column refers to the intercept. The dummy matrix $(\iota_{SSH} \otimes I_{T-1}, C) \in \mathbb{R}^{n(T-1)\times(n+(T-1))}$ has full column rank. In the same way we formulate the five year lagged model

$$y_{it} = \eta_g + \rho y_{i(t-5)} + x'_{i(t-5)}\beta + \sum_{s=6}^{T} SSH_s * d_{SSH,t}(i,s) + \tilde{\eta}_i + \nu_{it}.$$

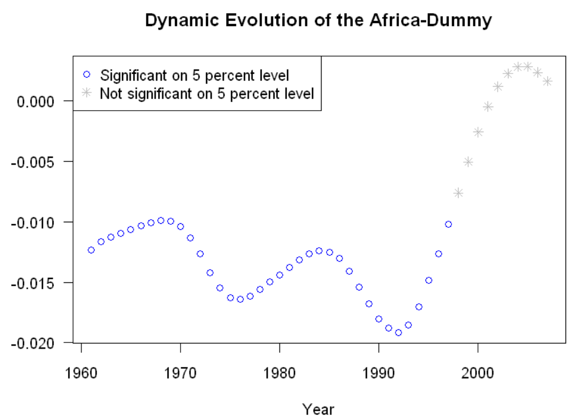

Figure 10: The Evolution of the Africa-Dummy in the one year lagged model

Table 7: Coefficients with a time-varying Africa-Dummy

	one year Estimate (S.E.)	five year Estimate (S.E.)
Intercept	0.1832***	1.2654***
	(0.0117)	(0.0636)
lag lny	0.9911***	0.8964***
	(0.0011)	(0.0062)
lnn	0.0012	-0.0214
	(0.0025)	(0.0128)
lnsk	0.0277***	0.0834***
	(0.0013)	(0.0065)
lnattain	-0.0175***	-0.0510***
	(0.0012)	(0.0065)
* $p : \leq 0.05$	**≤ 0.01	***≤ 0.001

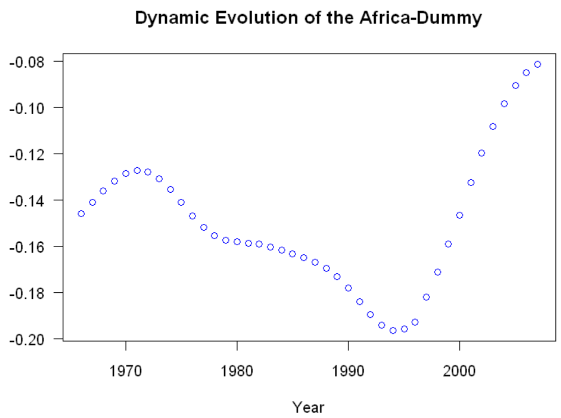

Figure 11: The Evolution of the Africa-Dummy in the five year lagged model

The results for the estimators of the coefficients are given in table (7). We observe that the estimators of the coefficients of (36) are similar to those of (14). Figures (10) and (11) show that the Africa-Dummy varies a lot over time. Apart from small bumps it incrementally decreases until the beginning to mid-nineties and then increases rapidly in the recent years. When considering the one year lagged model it even becomes insignificant. Furthermore, in the one year lagged model, the two very recent Africa-Dummies are smaller than the ones before. It is not clear if this is related to a small bump or a dramatic increase of Africa's punishment. However, in the most recent years, Africa's punishment was of much smaller magnitude than before.

5 Conclusion

By smoothing with the Hodrick-Prescott filter, we obtain yearly time-series that represent the connection of one time-series of an economy to another. When doing this, the length of the time-series is sufficiently large, so that the endogeneity bias that results from the lagged dependent variable in growth regressions is negligibly small. Estimating the coefficients of the growth regression with the Two-Groups Least-Square Dummy-Variable estimator identifies a negative significant Africa-Dummy. This clear punishment for sub-Saharan African economies increases if the return to investment in physical capital decreases, if the return the depreciation rate increases, or if the return to school attainment increases.

The Two-Groups Least-Square Dummy-Variable estimator is also used to relax the functional structure of the growth regression equation. We observe that the significance of the Africa-Dummy does not disappear when applying a semiparametric model so that it cannot be explained by a misspecification of the functional structure.

We observe that sub-Saharan African countries have clearly positive returns to the depreciation rate. When adding interaction effects, the Africa-Dummy is even positive and significant.

Finally, an extension of the Two-Groups Least-Square Dummy-Variable estimator estimates the evolution of the Africa-Dummy within the period we observe data. It can clearly be seen that Africa-Dummy changes over time. Apart from small bumps it incrementally decreases until the beginning to mid-nineties and then increases rapidly in the recent years. When estimating exactly the regression equation that is motivated by the Augmented Solow Model, we even observe that is becomes insignificant in the recent years.

References

ACEMOGLU, D. AND JOHNSON, S. AND ROBINSON, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *The American Economic Review* **91**(5) 1369–1401.

ARCAND, J. L. AND GUILLAUMONT, P. AND JEANNENEY, S. G. (2000). How to make a Tragedy: On the Alleged Effect of Ethnicity on Growth. *Journal of International Development* **12**(7) 925–938.

ARELLANO, M. AND BOND, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58**(2) 277–97.

ARELLANO, M. AND BOVER, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* **69** 29–51.

ARELLANO, M.(2003) Modeling Optimal Instrumental Variables for Dynamic Panel Data Models. CEMFI, Madrid.

BAROSSI-FILHO, M. AND GONCALVES SILVA, R. AND MARTINS DINIZ, E. (2005). The Empirics of the Solow Growth Model: Long-Term Evidence. *Journal of Applied Economics* **8**(1) 31–51.

BARRO, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics* **106**(2) 407–443.

BARRO, R. J. AND MANKIW, N. G. AND SALA-I-MARTIN, X. (1995). Capital Mobility in Neoclassical Models of Growth. *American Economic Review* **85**(1) 103–15.

BAXTER, M. AND KING, R. G. (1999). Measuring Business Cycles: Approximate Band-

Pass Filters for Economic time-series. *The Review of Economics and Statistics* **81**(4) 575–593.

Blundell, R. and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* **87**(1) 115–143.

Bond, S. and Hoeffler, A. and Temple, J.(2001). GMM Estimation of Empirical Growth Models. Discussion Paper No 01/525

Bowsher, C. G. (2002). On testing overidentifying restrictions in dynamic panel data models. *Economics Letters* **77** 211–220.

Barro, R. J. and Lee, J. W. (2010). A New Data Set pf Educational Attainment in the World, 19502010. National Bureau of Economic Research, Working Paper 15902.

Caselli, F. and Esquivel, G. and Lefort, F. (1996). Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics. *Journal of Economic Growth* **1**(3) 363–89.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. *Handbook of Economic Growth*, Chapter 9, 679-741.

Collier, P. and Gunning, J. W. (1999). Explaining African Economic Performance. *Journal of Economic Literature* **37**(1) 64–111.

Easterly, W. and Levine, R. (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *The Quarterly Journal of Economics* **112**(4) 1203–50.

Georgescu-Roegen, N. (1975). Dynamic Models and Economic Growth. *World Development* **11-12**(3) 765–783.

Hahn, J. and Kuersteiner, G. (2002). Asymptotically Unbiased Inference for a Dynamic

Panel Model with Fixed Effects When Both n and T Are Large. *Econometrica* **70**(4) 1639–1657.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* **50**(4) 1029–1054.

Hausman, J. A. and Taylor, W. E. (1981). Panel Data and Unobservable Individual Effects. *Econometrica* **49**(6) 1377–1398.

Heston, A. and Summers, R. and Aten, B. (2009). Penn World Table Version 6.3. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.*

Hodrick, R. J. and Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking* **29**(1) 1–16.

Hoeffler, A.E. (2002). The augmented Solow model and the African growth debate. *Oxford Bulletin of Economics and Statistics* **64**(2) 135–58.

Islam, N. (1995). Growth Empirics: A Panel Data Approach. *The Quarterly Journal of Economics* **110**(4) 1127–70.

Kendall, M. G. (1954). Note on the Bias in the Estimation of Autocorrelations. *Biometrika* **41** 403–404.

Kiviet, J. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* **68** 53–78.

Levine, R. and Renelt, D. (1992). A Sensitivity Analysis of Cross-Country Growth Regressions. *The American Economic Review* **82**(4) 942–63.

Mankiw, N.G. and Romer, D. and Weil, D.N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics* **107**(2) 407–37.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* **49**(6) 1417–26.

Orcutt, G. H. and Irwin, J. O. (1948). A Study of the Autoregressive Nature of the time series Used for Tinbergen's Model of the Economic System of the United States. *Journal of Royal Statistical Society* **10**(1) 1–53.

Phillips, P. C. B. and Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* **137** 162–188.

Price, G. N. (2003). Economic Growth in a Cross-section of Nonindustrial Countries: Does Colonial Heritage Matter for Africa? *Review of Development Economics* **7**(3) 478–495.

Roodman, D.(1996). How to Do xtabond2: An Introduction to Difference and System GMM in Stata. Center for Global Development, Working Paper Number 103.

Roodman, D. (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics* **71**(1) 135–158.

Sachs, J. D. and Warner, A. M. (1997). Sources of Slow Growth in African Economies. *Journal of African Economies* **6**(3) 335–76.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics* **70**(1) 65–94.

Swan, T. W. (1956). Economic Growth and Capital Accumulation. *Economic Record* **32**(1) 334–361.

Tauchen, G. (1986). Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data. *Journal of Business and Economic Statistics* **4**(4) 397–416.

Were, M. and Nafula, N. N. (2003). An Assessment of the Impact of HIV/AIDS on Economic Growth: The Case of Kenya. CESifo Working Paper Series 1034.

Windmeijer, F. (2005). A Finite Sample Correction For The Variance of Linear Efficient Two-Step GMM Estimators. *Journal of Econometrics* **126** 25–51.

Ziliak, J. P. (1997). Efficient Estimation with Panel Data when Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators. *Journal of Business and Economic Statistics* **15**(4) 419–431.