Edwin Paapa hMensa
Econ 312
Final Project
05/08/14

The Convergence Hypothesis

Introduction

Macroeconomic theory suggests that all things being equal, a country that saves more and has more productive technologies will have higher output per capita than a country that saves less. This is known as the Solow Growth Model. It suggests that countries with similar savings rates and comparable technological sophistication should have the same output per capita in the long run.

The model also suggests further that given similar underlying conditions, output per capita should converge in level and growth rate across countries. These suggestions are known as the convergence hypothesis.

This project examines the conditional convergence hypothesis in the Solow growth model. Conditional convergence across countries will be tested by estimating the equation below:

$$Y_{i,\,average} = \beta_1 + \beta_2 ln(Y_{i,\,1}) + \beta_3 ln(S_i) + \beta_4 ln(LE_i) + \beta_5 T_i + e_i\,; \text{ for country } i$$

where is average annual growth rate per capita, is the initial real GDP per capita, is the savings rate, is the life expectancy and is openness to trade. It is hypothesized that a country's average growth rate per capita for a period is linearly related to the logs of the real GDP per capita at the beginning of that period, the national savings rate, life expectancy and its openness to trade.

Assuming that is normal, and that all the dependent variables are fixed and not perfectly collinear, I will determine the quantitative impact of national savings on growth and the speed of convergence, and also determine the marginal effect of initial real GDP per capita If and significant, I will then conclude whether or not the data set exhibits conditional beta convergence and we would reject the null hypothesis .This will imply that countries which are further away from steady state (poorer countries) will grow at a faster rate than countries which are nearer to it (richer countries).

Data

This project uses the dataset used in Jeffrey D. Sachs and Andrew M. Warner's paper on "Sources of Slow Growth in African Economies" published in Journal of African Economies in December 1997. The cross-country data was downloaded from http://www.cid.harvard.edu/ciddata/ciddata.html. The dataset contains multiple economic variables for over 100 countries, majority of which are African countries.

The following table shows the variable names and their descriptions. In general the variables are those that several authors, including Sachs & Warner, have previously found to be related to cross-country growth.

Variable name	**Description**
gea7090	Average annual growth in real GDP per person between 1970 and 1990.
LGDPEA70	The log of real GDP per head of the economically active population in 1970.
OPEN6590	The fraction of years during the period 1965-1990 in which the country is rated as an open economy according to the criteria in Sachs and Warner
LIFEE	Life Expectancy in years, circa 1970.
CGB7090	Average value of central government savings over the period 1970-1990. Savings is defined as current revenues minus current expenditure, and is measured in percent of GDP.

GDP data are from the Penn World Tables, Mark 5.6, and are adjusted for differences in the purchasing power across countries (see Summers and Heston 1981). The economically active population is defined as the population between the ages 15-64. The population data is from the World Data CD-ROM, 1995, World Bank.

From this list of variables, I have determined based on theory that *ldgpea70, cgb7090, lifee1l* and *open6590* will have the most statistically and economically significant effects on *gea7090*. Though *cgb7090* is not a measurement of national savings' rates, for lack of actual values of countries' savings rates, it will serve as a proxy because I expect central government saving to highly correlated with national savings' rate, though it misses saving that comes from local governments, public enterprises, and private savings.

Moreover, I expect *sxp, infl6590, inv7089, ns7089* and *geap-gpop* to be highly endogenous to this model, and *tropics, icrge80, ssafrica and ethling* to be statistically insignificant. The summary statistics of the selected variables are directly below:

ariable	Obs	Mean	Std. Dev.	Min	Max
gea7090	105	1.161703	1.910552	-3.635249	5.770552
gdpea70	127	8.215905	.9169763	6.324667	9.948785
cgb7090	105	2.696704	5.462104	-5.243911	37.00037
lifee1l	128	3.959382	.2284858	3.465736	4.298645
pen6590	144	.2711325	.4103255	0	1

From the summary, it appears that there is a lot of missing data, as the different variables have different number of observations. This could introduce a bias into regression estimates and diminish internal validity if not curbed at this point. I now drop data items with missing values, and below are the summary statistics of my revised dataset.

riable	Obs	Mean	Std. Dev.	Min	Max
ea7090	88	1.212709	1.872502	-3.635249	5.770552
dpea70	88	8.336835	.9018107	6.425351	9.948785
gb7090	88	2.091408	3.963145	-5.243911	16.21274
ifee1l	88	3.98034	.2302656	3.465736	4.298645
en6590	88	.4017133	.4451852	0	1

Given that all the assumptions made in the introduction still hold for these variables, I now proceed to run a regression to conclude whether or not these countries exhibit beta convergence, and to estimate the quantitative effect of national savings on growth and the speed of convergence.

Regression estimates

I assume endogeneity of the selected independent variables, and now run a regression of *gea7090* on *lgdpea70, cgb7090, lifee1l* and *open6590*. The results are on the following page:

Source	SS	df	MS
Model	186.18468	4	46.54617
Residual	118.860267	83	1.43205141
Total	305.044947	87	3.50626376

Number of obs =	88
F(4, 83) =	32.50
Prob > F =	0.0000
R-squared =	0.6104
Adj R-squared =	0.5916
Root MSE =	1.1967

gea7090	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lgdpea70	-1.843632	.2770625	-6.65	0.000	-2.394698	-1.292566
cgb7090	.1410498	.0340311	4.14	0.000	.0733634	.2087362
lifee1l	6.019395	1.189848	5.06	0.000	3.652836	8.385954
open6590	2.89339	.3965723	7.30	0.000	2.104624	3.682157
_cons	-8.833782	3.290923	-2.68	0.009	-15.3793	-2.288268

$$gea7090 = \beta_1 + \beta_2 lgdpea70 + \beta_3 cgb7090 + \beta_4 lifee1l + \beta_5 open6590 + e$$

where $\beta_2 = -1.843632$, $\beta_3 = 0.1410498$, $\beta_4 = 6.019395$, $\beta_5 = 2.89339$ and $\beta_1 + e = -8.833782$

Based on the table, I can now interpret the coefficients in terms of the effect of the log of real GDP per capita 1970 (*lgdpea70*) and (average value of central government savings over the period 1970-1990) *cgb7090* on the average growth rate in real GDP per capita between 1970-1990 (*gea7090*):

lgdpea70: A country with 1% greater real GDP per capita in 1970 will have a 1.844% less average GDP per capita growth rate between 1970-1990. The negative sign and small magnitude of the coefficient are plausible because I would expect a country with a richer country (higher real GDP) in 1970 to experience growth at a slower rate than a poorer country (lower real GDP) in 1970.

cgb7090: A country with 1% greater average value of central government savings over the period 1970-1990 will have a 0.141% greater average GDP per capita growth rate between 1970-1990. The negative sign and small magnitude of the coefficient are plausible because I would expect a country with higher savings to experience growth at a faster rate than a country with less savings.

The regression relationship is quite strong, as inferred from the . The other regressors *lifee1l* and *open6590* are both economically and statistically significant at the 5% level.

However, RMSE (root-mean-square error) = 1.1967 is very high, which suggests that my regressors are not excellent predictors of a country's average growth rate between 1970 and 1990. The following is a summary of predicted values *ghat* (based on my model) compared with the actual values *gea7090*. gives us the following summary:

Variable	Obs	Mean	Std. Dev.	Min	Max
ghat	88	1.212709	1.462892	-2.695695	5.052984
gea7090	88	1.212709	1.872502	-3.635249	5.770552

My predicted values have exactly the same mean as the actual values, but have smaller variance and range. This is a significant problem, showing that my model appears to not be a very good predictor given my regressors.

Bearing these problems in mind, the estimates of my regressors are all statistically and economically significant as expected, and their signs and magnitudes satisfy a priori expectations. β_2 is found to be statistically significant and negative (-1.844), which suggests that This will imply that countries which are further away from steady state (poorer countries) will grow at a faster rate than countries which are nearer to it (richer countries).

Postestimation diagnostics

I suspect that heteroskedasticity may be present in my model, and so I now conduct a Breusch-Pagan test for heteroskedasticity:

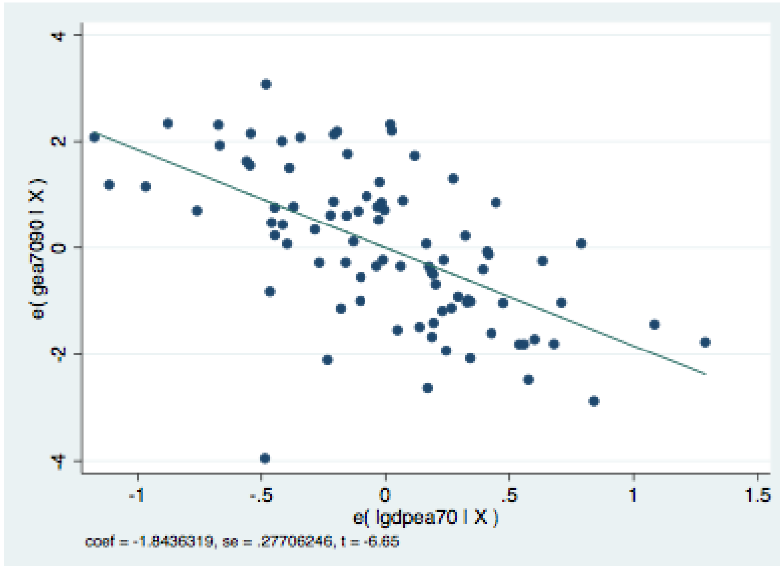
```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of gea7090

         chi2(1)      =     1.53
         Prob > chi2  =   0.2159
```

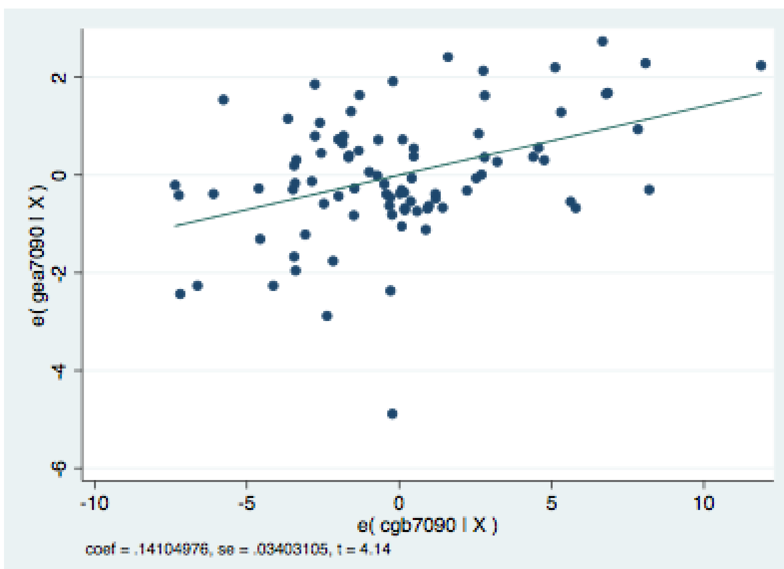
The p-value found is greater than my significance level of 5%, therefore I do not reject the null hypothesis that my model is homoskedastic. There is not enough evidence of heteroskedasticity in my model.

I also suspect that there may be several high impact outlier data points within my model. In line with the focus of my project, I will only investigate *lgdpea70* and *cgb7090*.

Below is a plot of residuals of regression of *gea7090* on *cgb7090*, *lifee1l* and *open6590* against residuals of regression of *lgdpea70* on *cgb7090*, *lifee1l* and *open6590*. This is to locate outliers.

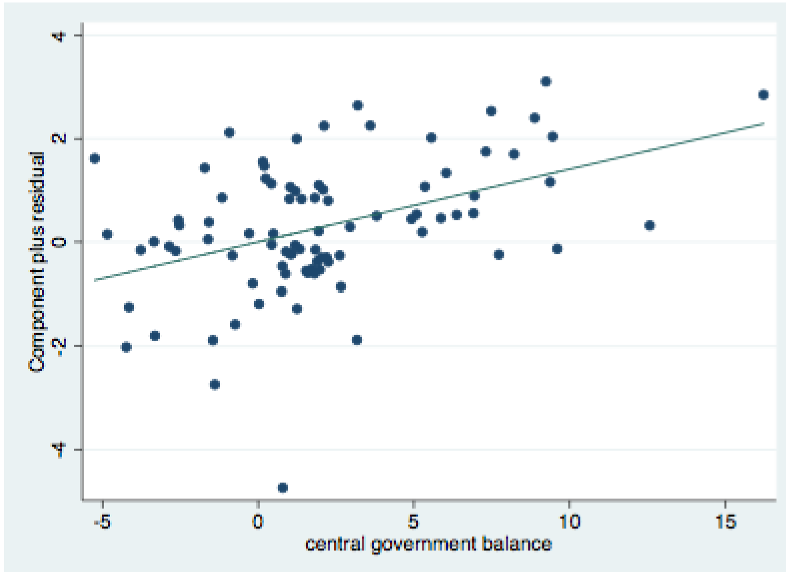


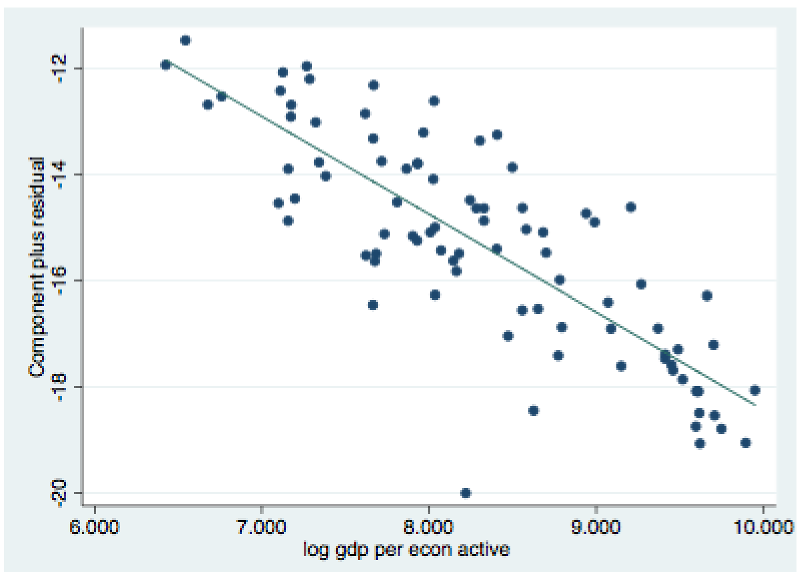
Outliers are observed for *lgdpea70*. A similar plot is output below for *cgb7090*.



Outliers are also observed for *cgb7090*. Implications of these outliers will be discussed in the validity assessment section of this report.

I now wish to assess the linearity of the relationship between *gea7090, cgb7090* and *lgdpea70*. Below are partial residual plots for my two main regressors.





Both *cgb7090* and *lgdpea70* appear to have fairly linear relationship with *gea7090*. *cgb7090* is more scattered away from the best-fit line, suggesting that it is not a purely linear relationship as my specification suggests. *lgdpea70* on the other hand is showing a more linear relationship and consistent variance about the best-fit line, suggesting that it indeed shares a linear relationship.

Conclusion and Assessment of Validity

This project examined the conditional convergence hypothesis in the Solow growth model using dataset used in a paper by . I determined that national savings has a statistically and economically significant effect of 0.141% on average growth rate, and initial real GDP per capita has also a statistically and economically significant effect of -1.844% on average growth rate. The countries in the data set exhibit conditional beta convergence, which implies that countries which are further away from steady state (poorer countries) will grow at a faster rate than countries which are nearer to it (richer countries).

With regards to internal validity, this project suffered from several problems. First was missing data, which I resolved by dropping countries with missing values. Second was the lack of actual values of countries' savings rates, which therefore had to be proxied with data on central government saving (expected to be highly correlated with national savings' rate) which misses saving that comes from local governments, public enterprises, and private savings. Postestimation diagnostics also showed that my model has several outlier data points which suggests that my regression estimates are biased. However, diagnostics showed no evidence of heteroskedasticity in my model.

Additionally, my selected variables were already in log form within the dataset, therefore limiting my specification to a linear-log functional form. I also cannot ascertain the accuracy of the measurements and data collection procedures. There is a high chance that my dataset suffers from sample selection bias also. Given all these problems of internal invalidity, it will be dubious to externalize these conclusions and extrapolate my results for all other regions in the world. More work should be done in the future to resolve some, if not all, of these issues.