



Dr. Marcel Bluhm
Assistant Professor of Economics

The Wang Yanan Institute for Studies in Economics
Xiamen University

**Special Topics in Applied Econometrics: Cross-Country Studies
Summer Term 2012**

Take Home Assignment:

Estimating a Growth Regression and a Simulation-Based Study

This take home exam consists of 2 parts: one empirical part and a simulation based part. Both parts need to be implemented in Matlab. Please comment for each single line in the codes what you are doing. Missing comments have negative repercussions on the final grade. Your solution package to be sent to the e-mail address bluhm.wise.xmu@googlemail.com should consist of:

Part 1: the initial .txt file of the data downloaded from the Penn World tables, the prepared dataset in Excel, two Matlab .m-files: the program to prepare the dataset and the program to solve the questions (codes will be checked), a Word document in which you give answers to the questions/summarize your results.

Part 2: two Matlab .m-files: the program to run the code to answer the questions, an .m function which contains the likelihood and is used for maximization, and a Word document in which you summarize your results.

The solutions need to be sent until 21 March at noon. Receipt of the solutions will be confirmed via mail. The noon deadline is **strictly** binding. Solutions sent after the deadline will not be accepted and the student will fail the course.

PART 1: Estimating a Growth Equation

1. Download the following variables (in constant prices) from the Penn World Tables over the time period from 1970 to 2009: population (pop), real gdp per capita (rgdpl), government share of real gdp (kg), investment share of real gdp (ki), openness (openk). Import the data into the first sheet of an Excel document. Add a documentation on the second sheet of the dataset where you describe the source of your data and what the abbreviations stand for.

2. Set up main dataset: Transform all data into quinquennial data (that is, take the mean over 5 years), generate the growth rate for real gdp per capita (grgdpl) and take log of real gdp per capita and population. Assign each country a panel id and a year. Since a balanced panel dataset shall be used for regression, remove all countries which lack some observations (you can use the function „isnan(,)“ in Matlab to detect missing data.).

The dataset used for regression has to look as follows:

Column order of variables: year; panel id; log population, log rgdpl, kg, ki, openness, growth of real gdp per capita. Save these data on the third sheet of your main dataset. The dataset has to be submitted and counts as part of the final grade.

Additional question to be answered for Part 2.2: What is the advantage of transforming the yearly data into quinquennial data?

3. Estimate an augmented static growth regression in a pooled panel framework outlined in chapter one of the course and summarize all results from the substeps, including the economic interpretation of statistically significant variables:

$$\text{grgdpl}_{it} = c + b_1 \log(\text{pop}_{it}) + b_2 \log(\text{init per capita rgdpl}_{it-1}) + b_3 \text{kg}_{it} + b_4 \text{ki}_{it} + b_5 \text{openness}_{it} + \varepsilon_{it}$$

where c is a constant, ε_{it} is a white noise error term, and the other variables are outlined above
Additional question to be answered for Part 3.2: Regressing the growth rate of real GDP per capita on a single constant and the initial GDP per capita only is referred to as „unconditional convergence assumption“. Why is the regression framework above referred to as „conditional convergence assumption“ and why is this framework more realistic?

- calculate coefficients and t-statistics based on non-robust standard errors. Here and in the following statistical significance is taken at the 5% significance threshold, that is, all critical values have to be chosen at the 5% significance level.

- test for serial correlation; interpret results

- test for heteroscedasticity, including all explanatory variables as well as their square as explanatory variables (omit cross products in the h-function), interpret results

- if serial correlation or heteroscedasticity is present, calculate a robust variance covariance matrix and robust t-statistics for the coefficient estimates

4. Estimate an augmented static growth regression in a random effects panel framework seen in chapter one of the lecture and summarize all results from the substeps, including the economic interpretation of statistically significant variables:

$rgdpl_{it} = c_i + b_1 \log(\text{pop}_{it}) + b_2 \log(\text{init per capita } rgdpl_{it-1}) + b_3 kg_{it} + b_4 ki_{it} + b_5 \text{openness}_{it} + \varepsilon_{it}$
where c_i is assumed to be randomly distributed as outlined on the slides of chapter 1.

- calculate coefficients and t-statistics

- test for the presence of an unobserved effect and based on your finding reject either the pooled model or the random effects model

5. Estimate an augmented static growth regression in a fixed effects panel framework seen in chapter 1 of the lecture and summarize all results from the substeps, including the economic interpretation of statistically significant variables:

$rgdpl_{it} = c_i + b_1 \log(\text{pop}_{it}) + b_2 \log(\text{init per capita } rgdpl_{it-1}) + b_3 kg_{it} + b_4 ki_{it} + b_5 \text{openness}_{it} + \varepsilon_{it}$
where c_i can be correlated with the explanatory variables as outlined on the slides of chapter 1.

- calculate coefficients and t-statistics based on non-robust standard errors

- test for serial correlation

- if serial correlation is present, calculate a robust varcov matrix and robust t-statistics for the coefficient estimates

- make a Hausman test whether the model should contain a random or a fixed effect

6. Critically outline (at maximum on one page, font times new roman, single spacing, 12 pt; more text will have negative repercussions on the points given for this sub question) **the caveats** which one needs to be aware of when estimating and interpreting growth equations such as the ones in the above framework.

PART 2: Simulation of a Pooled Tobit Model

1. Simulate the model given in Equation (2.14) on slide 20 of chapter 2 with

- x containing a constant and two further variables distributed as $N(0,5)$

- $\theta = [2, 5, -4]'$, where 2 is the constant

- u_{it} is distributed as $N(0,5)$

- $N=100$

- $T=5$

2. Obtain starting values for the maximum likelihood (ML) maximization via running a POLS regression of the model given in Equation (2.14).

3. Estimate the coefficient vector θ in Equation (2.14) via ML (the log likelihood is given in Equation 2.16, summed up over the cross section and time series dimension).

4. Calculate the asymptotic variance of the estimated parameter vector.

5. Describe what happens to the t-values of the estimated parameters if (i) you change the number of observations, N or T , or, (ii) increase the variance of the error term.

6. To find out whether the ML estimation yields unbiased results run a Monte Carlo Simulation with 1000 replications and calculate the mean of each estimated parameter.