Leuphana Universität Lüneburg Wintersemester 2023/2024

Modul „Methodenorientierte Zugänge zu den Naturwissenschaften „

Seminar: „ Applied Algorithms in Estimation and in Control of Technical, Economical, and Biological Dynamical Systems “

Paolo Mercorelli

Estimation of the Gross Fixed Kapital using linear Regression

Timo Heyn	4. Semester
Matr-Nr. 3047219	timo.f.heyn@stud.leuphana.de
Juluis Weber	4.Semester
Matr-Nr. 3047688	julius.weber@stud.leuphana.de

Table of Contents

Abstract

Investment is a determinant of an economies long run economic growth and therefore the long run economic circumstances of the respective market participants as well. Therefor it is a parameter of interest for science and policy. The goal of this paper is an estimation model for the Investment rate of a country. To do that we use multiple linear regression models, containing the regressands income and real interest rate with different orders. To decide which model is best suited for this problem we use the adjusted coefficient of determination, which is aimed to be minimized in order to avoid overfitting.

Introduction

General

The determinants of long run economic growth, that is output Growth per capita, is one of the big topics in macroeconomics. The basic theory is that relative long run output growth per capita is determined by technological progress (Blanchard, 266 et seq.). But the absolute output growth that results from the relative output growth through technological advancement, is determined by the level of capital per capita. Which in turn depends positively on investment (Blanchard 2020, 238.et seq) “Investment refers to all economic activity which involves the use of resources to produce goods and services.” (Muhamed S.Anwer and R.K. Sampath 1999, S. 3)

Long term economic goals of governments and institutions like central banks, as well as the standard of living of the people living in the respective countries, are influenced by the investment level. That makes the determinants of investment which are partly influenced by governments and central banks relevant for them.

Determinants of investment

Basic theory suggests that investment depends positively on income, and negatively on the real interest rate plus a risk premium (Blanchard 2020, 137 et seq.)Income has a positive relation with investment as a higher income leads to more available resources for investment. A higher real interest rate or risk premium leads to higher explicit or opportunity costs of investments, which lowers investment overall.

All of these can be directly or indirectly influenced by governments or central banks through different policies like monetary policy or investment guarantees, and the key interest rate or money supply.

In the Following we will cut out the factor of risk, using the Formula Bader and Malawi (Malawi und Bader 2010, 202 et seq) used.

$$\mathrm{GFCFt} = \beta 0 + \beta 1\,\mathrm{Rt} + \beta 2\,\mathrm{GDPt} + \mathrm{Ut}$$

Which is equal to:

$$Gross\ fixed\ Kapital = GDP * a1 + interest\ Rate * a2 + a0$$

Problem and Method

As explained above, the investment level is a significant parameter for decisions of politics and central banks. Because of that we want to build a model which estimates the investment level (proxied by gross capital formation) of a country given its income level (proxied by GDP per capita) and its real interest rate that in the following.

To do that we decided to use a multiple least squared regression model like Bader and Malawi did.

A least squared regression model aims to predict the size of the regressand (Y), which is the dependant variable, for some given values of the regressors (X), which are the in the regression model included independent variables and determinants of the regressand. In order to achieve a prediction as precise as possible the method estimates a coefficient β for each independent variable which gives the change in the regressand for a unit increase in the regressor c.p.. And an additional constant coefficient β0 which is independent of the regressors values.(Stock/ Watson, 218 et seq.)

$$Y_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_{ki} + u_i$$

It does that, using a given data sample drawn from the population of interest containing unit values of the regressors and the regressand and determines the coefficients β such that the sum of squared errors of prediction is minimized for the sample values. (Stock/ Watson, 221 et seq.)

$$min_{b0,b1,\dots,bk} \sum (\hat{u}_i)^2$$

Where:

$$\hat{u}_i = Y_i - \hat{Y}_i$$

$$\hat{Y}_i = \hat{}\beta_0 + \hat{}\beta_1 x_1 + \cdots + \hat{}\beta_{ki} x_{ki}$$

In the case of our problem, the investment level is then the regressand and the income and real interest level are the regressors.

In order to account for the possibility of a nonlinear reationship between the regressors and the regressand we run polynomial regression models additional to the linear regression model depicted above. The working of the estimation of the least squared regression model stays the same. The difference is that a polynomial model includes higher order versions of the independent variables additional to the linear versions of the independent variables. So while the coefficients β stay linear, the regressors themselves are nonlinear, which leads to a nonlinear marginal effect on the regressand. (Stock/ Watson 21, 286 et seq.)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \beta_3 X_{2i} + \beta_4 X_{2i}^2 + u_i$$

To then decide which of the order least squared regression model is best suited to our purposes there needs to be an indicator which acts as a decision rule.

For that we use the adjusted coefficient of determination or R^2 adjusted. It is in essence the same as the regular R^2 which gives “the fraction of the sample variance of Y explained by (or predicted by) the regressors” (Stock/ Watson, 223). The difference is that the regular R^2 increases when additional variables are added to the model which impacts the comparability between R^2 values of different models. The R^2 adjusted corrects for property of R^2 by correcting the value of R^2 downwards with an increasing number of independent variables. (Stock/ Watson, 223 et seq.)

The decision rule is now to pick the order for the least squared regression model with the lowest R^2 adjusted. So, to minimize R^2 adjusted with respect to the model’s order, with the same sample and influence variables, to avoid overfitting. Overfitting describes a situation in which a model fits the sample data too precisely, capturing noise or random variations instead of the intended general relationships which leads to poor results when applying the model to units not included in the sample.

We apply this method to a data sample of the regressors and the regressand. Our data sources are the real interest rate, the gross fixed capital formation, and the GDP. The sources include data for all 195 countries and some summaries of multiple values, for example, the European Union. The year of the data ranges from 1960 to 2022. (The World Bank Data 2022a, 2022b, 2022c) To ensure that all countries are weighted equally, the summaries of countries were removed from all three data sources, so the value of all countries is only counted once. Then one new data sheet is created with all three values for all countries for one year. This data sheet is later used to calculate the linear regression. For our data, the year 2016 was chosen because it is still very recent data, but it has fewer countries with empty values than more recent years. After creating the data sheet, all country data with empty values is removed because, without values, the linear regression can't calculate the coefficients correctly. This leaves 106 countries for the linear regression.

Estimation and Interpretation

To estimate the gross fixed capital, we use this formula:

$$Gross\ fixed\ Kapital\ =\ GDP\ * a1 + interest\ Rate * a2\ + a0$$

To calculate the coefficients a0, a1, and a2, we create a matrix with all GDP and interest rate values and an equal amount, so 106. This matrix is called X1. Using the lecture-learned formula, we can use X1 to calculate the coefficient and save it in the variable theta (1, 2, 3).

$$Theta = inv(X1' * X1) * X1' * GFC;$$

To decide which order of the estimation model is best suited, we calculate R2 and adjust R2. We are adjusting R2 with the factors: the number of measured points is

106, N is equal to the order of the polynomial, and the number of coefficients is equal to (N*2)+1. So for the first order, the formulas for R2 and R2 adjusted are:

$$R\_squared = 1 - sum((GFC - (Theta(1) * GDP + Theta(2) * IR + Theta(3))).\hat{}2)/(sum((GFC - mean(GFC)).\hat{}2))$$

And:

$$R\_squared\_adj = 1 - ((m-1)/(m-p) * sum((GFC - (Theta(1) * GDP + Theta(2) * IR + Theta(3))).\hat{}2)/(sum((GFC - mean(GFC)).\hat{}2)))$$

To visualize the results, we chose a country at random, Japan, and estimated the gross fixed capital using the data and the calculated coefficients. Then we plotted the estimated value from the years 2007–2016 (blue line) and the actual measured value in the same years (red dots).

Then we repeated the process with one order higher until we found an R^2 adjusted with a lower value than the previous R^2 adjusted while still being positive. To calculate the coefficients and R^2 adjusted for a higher order, we add GDP^n and interest rate ^{n,} with n being the highest exponent of the order before +1, to the matrix. Then we multiply the newly calculated coefficients with the new variables when calculating R^2 adjusted. For example, the calculation of R^2 adjusted for the second order:

$$R\_squared\_adj = 1 - (m-1)/(m-p) * sum((GFC - (Theta(1) * GDP.\hat{}2 + Theta(2) * GDP + Theta(3) * IR.\hat{}2 + Theta(4) * IR + Theta(5))).\hat{}2)/(sum((GFC - mean(GFC)).\hat{}2))$$

Results

First Order:

R_squared = 0.0044

R_squared_adj = -0.0149

Plot:

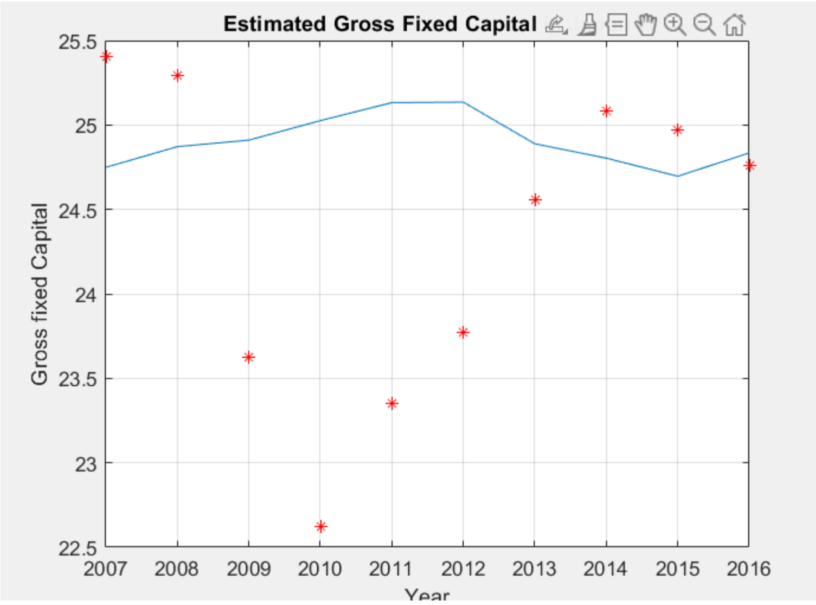

Figure 1 First Order: Estimation of Gross fixed Kapital (self-Created Picture)

Second Order:

R_squared = 0.0368

R_squared_adj = -0.0013

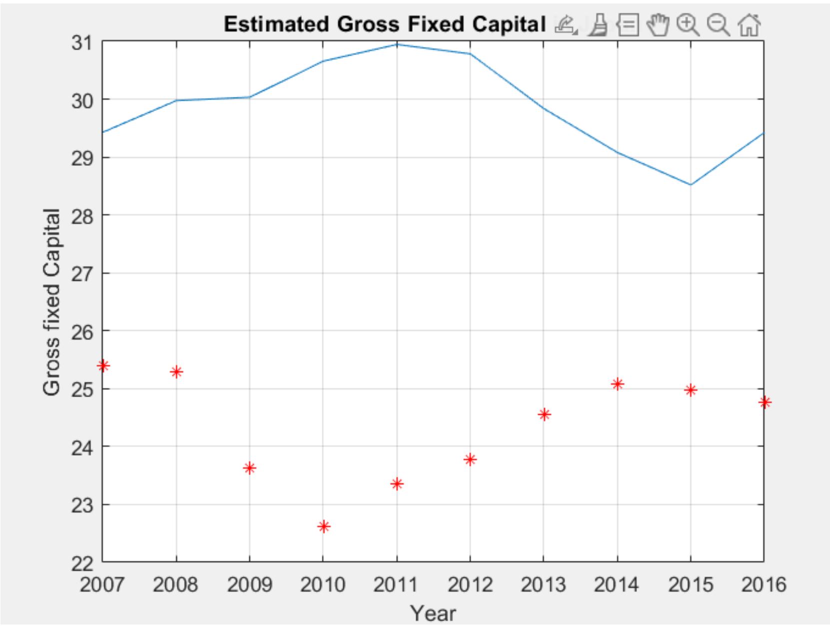

Figure 2 Second Order: Estimation of Gross fixed Kapital (self-Created Picture)

Third order:

R_squared = 0.0773

R_squared_adj =0.0407

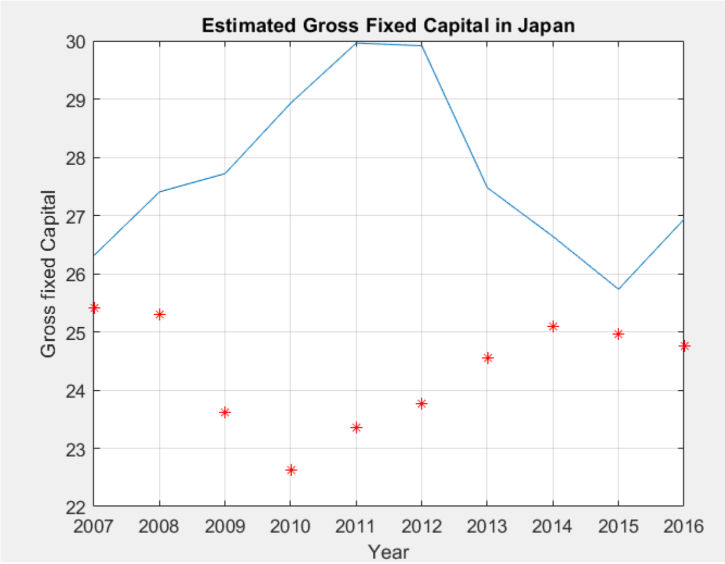

Figure 3 Second Order: Estimation of Gross fixed Kapital (self-Created Picture)

Forth order:

R_squared = 0.0895

R_squared_adj = 0.0534

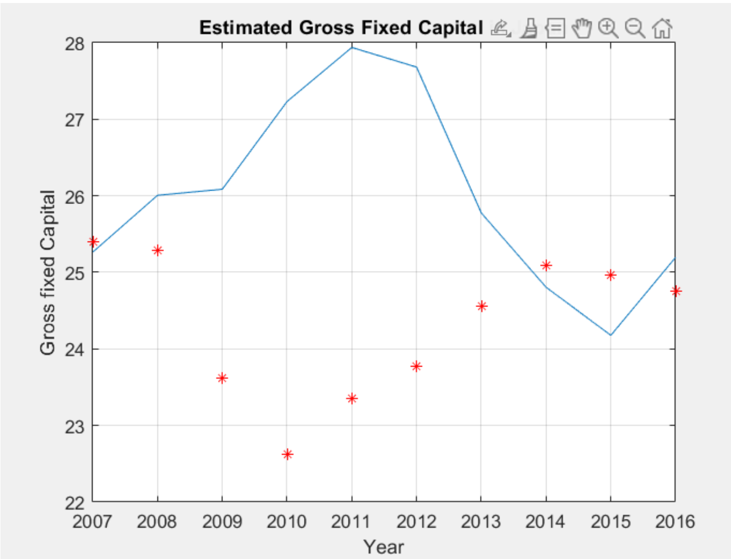

Figure 4 Second Order: Estimation of Gross fixed Kapital (self-Created Picture)

The estimation of the first two orders has negative R^2 adjusted. The value of the third order is lower than the value of the fourth order, so the third order is the best estimation according to our factors.

Conclusion

The correlation between Gross fixed Kapital GDP and interest Rate is strongest with a linear Regression of the third order. However, the measured Value to describe the accuracy of the correlation R^2 adjusted is very low, and when comparing the estimated value with the actual Value of an Example the model was not able to estimate ups and downs which is an important part of estimation in Economics. This Result contradicts the Formular found by Bader and Malawi.

Lists

List of Contents

Blanchard, Olivier (2020): Macroeconomics (international edition). 8e edition: Pearson.

Malawi, Ahmad; Bader, Majed (2010): The Impact of Interest Rate on Investment in Jordan: A Cointegration Analysis. In: *eco* 24 (1), S. 199–209. DOI: 10.4197/Eco.24-1.6.

Muhamed S.Anwer and R.K. Sampath (1999): INVESTMENT and ECONOMIC GROWTH. In: *Presented at Western Argricultural Economics Assosiation ANual Meting*.

Stock/ Watson (21): Introduction to Econometrics, eBook, Global Edition: Pearson.

The World Bank Data (2022a): GDP (current US$). Online accessible at:13.04.2024.https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.last check on:

The World Bank Data (2022b): Gross fixed capital formation (% of GDP). Online accessible at https://data.worldbank.org/indicator/NE.GDI.FTOT.ZS, last check on: 13.04.2024.

The World Bank Data (2022c): Real interest rate (%). Online accessible at https://data.worldbank.org/indicator/FR.INR.RINR?end=2022&start=1986, last check on: 13.04.2024.