MOOC Econometrics

Case project

Erasmus School of Economics

Notes:

- See website for how to submit your answers and how feedback is organized.

- This exercise uses the datafile `Case_GDP` and requires a computer.

- The dataset `Case_GDP` is available on the website.

Goals and skills being used:

- Get hands-on experience with applying econometric methods.

- Apply techniques and interpret results related to discrete choice models.

- Apply techniques and interpret results using time series models.

Background

A good understanding of the macroeconomic cycle with alternating recession and expansion periods (also known as the business cycle) is important for various decision makers. Macroeconomic policy is often based on predictions of this cycle, and such predictions can influence investment decisions of large companies. Central banks and other institutions often publish so-called leading indicators that are helpful to predict the state of the economy. These indicators are based on macroeconomic series like job formation, interest rates, credit, demand, and supply.

In this case project you will predict GDP growth by using quarterly data on a hypothetical economy from 1950 quarter 1 to 2015 quarter 4. The data set contains the GDP of the economy and two leading indicators `li1` and `li2`. In order to evaluate the predictive performance of econometric models, you need to split the data in two parts. As estimation sample you take the period from 1951 to 2010 (240 observations), and as evaluation sample you take the period from 2011 to 2015 (20 observations). The first year of data (1950) is used only to create lags of variables.

The project consists of two parts. In the first part (a-c) you use logit models to predict whether the economic situation improves or declines, and in the second part (d-g) you use time series models to predict the size of the growth rate of the economy.

Data

The data file `Case_GDP` contains the following variables:

- `DATE`: Date of the observation;

- `GDP`: Gross Domestic Product of the economy;

- `GDPIMPR`: dummy variable indicating whether the GDP has increased (1) or decreased (0);

- `LOGGDP`: Log of Gross Domestic Product;

- `GrowthRate`: Relative growth of the economy: $\text{GrowthRate}_t = \log(\text{GDP}_t) - \log(\text{GDP}_{t-1})$;

- `li1`: First leading indicator;

- `li2`: Second leading indicator;

- `T`: Linear trend (where the first observation, for 1950 quarter 1, is defined as 0).

(a) The table below summarizes the outcomes of four logit models to explain the direction of economic development (GDPIMPR) for the period 1951 to 2010. Perform three Likelihood Ratio tests to prove both the individual and the joint significance of the 1-quarter lags of li1 and li2, where the alternative hypothesis is always the model with both indicators included.

Dependent variable: GDPIMPR Sample size: 240				
Variable	Coeff	Coeff	Coeff	Coeff
Constant	0.693	0.812	0.636	0.729
li1(-1)	x	-0.340	x	-0.372
li2(-1)	x	x	-0.087	-0.120
Log likelihood	-152.763	-139.747	-149.521	-134.178

(b) It could be that the leading indicators lead the economy by more than 1 quarter. The table below summarizes outcomes of four logit models that differ in the lags of the indicators. For what reason can we use McFadden R^2 to select the best lag structure among these four models? Compute the four values of McFadden R^2 (with four decimals) and conclude which model is optimal according to this criterion.

Dependent variable: GDPIMPR Sample size: 240				
	1	2	3	4
Variable	Coeff	Coeff	Coeff	Coeff
Constant	0.729	0.731	0.746	0.749
li1(-1)	-0.372	-0.366	x	x
li1(-2)	x	x	-0.429	-0.421
li2(-1)	-0.120	x	-0.131	x
li2(-2)	x	-0.121	x	-0.129
Log likelihood	-134.178	-134.126	-130.346	-130.461

(c) Use the logit model 3 of part (b) (with li1(-2) and li2(-1)) to calculate the predicted probability of economic growth for each of the 20 quarters of the evaluation sample. Assess the predictive performance by means of the prediction-realization table and the hit rate, using a cut-off value of 0.5. Evaluate the outcomes.

(d) Perform the Augmented Dickey-Fuller test on LOGGDP to confirm that this variable is not stationary. Use only the data in the estimation sample and include constant, trend, and a single lag in the test equation ($L = 1$, see Lecture 6.4). Present the coefficients of the test regression and the relevant test statistic, and state your conclusion.

(e) Consider the following model: $\text{GrowthRate}_t = \alpha + \rho \text{GrowthRate}_{t-1} + \beta_1 \text{li1}_{t-k_1} + \beta_2 \text{li2}_{t-k_2} + \varepsilon_t$. Here the numbers k_1 and k_2 denote the lag orders of the leading indicators. Estimate four versions of this model on the estimation sample from 1951 to 2010, by setting k_1 and k_2 equal to either 1 or 2. Show that the model with $k_1 = k_2 = 1$ gives the largest value for R^2, and present the four coefficients of this model in six decimals.

(f) Perform the Breusch-Godfrey test for first-order residual serial correlation for the model in part (e) with $k_1 = k_2 = 1$. Does the test outcome signal misspecification of the model?

(g) Use the model in part (e) with $k_1 = k_2 = 1$ to generate a set of twenty one-step-ahead predictions for the growth rates in each quarter of the period 2011 to 2015. Note that the required values of the lagged leading indicators are available for each of these forecasts. Calculate the root mean squared error of these forecasts and present a time series graph of the predictions and the actual growth rates.