

Session 9: Data Analysis and Time Series*

Doing Economics with the Computer

1 The Topic

The objective of this section is to get an introduction into the possibilities of data analysis and to develop a sample script-file to do descriptive analysis of time series. Most data in macroeconomics come in the form of *time series*, which is a set of repeated observations of the same variable over time. Formally,

$$\{x_1, x_2, x_3, \dots, x_T\}, \quad \text{or} \quad \{x_t\}, t = 1, 2, \dots, T.$$

We treat x_t as a *random variable*. If the series y_t is generated by the process

$$y_t = \alpha + x_t\beta + \epsilon_t, \quad \text{with} \quad \mathbb{E}_0[\epsilon_t|x_t] = 0, \quad (1)$$

then we can apply basic OLS from econometrics to provide a consistent estimate of β . What makes a time series different from micro data over a cross section and therefore interesting for our analysis, is that time series are typically not *iid*, meaning we do not have $x_t = \epsilon_t$ with $\epsilon_t \sim i.i.d.\mathcal{N}(\mu, \sigma^2)$. Typically, we observe in time series that today's observation of per example GDP is related to last year's observation of GDP; and tomorrow's might as

well be related to today's value. Such a behavior can be observed in the following figure for the growth rate of real GDP in the USA.

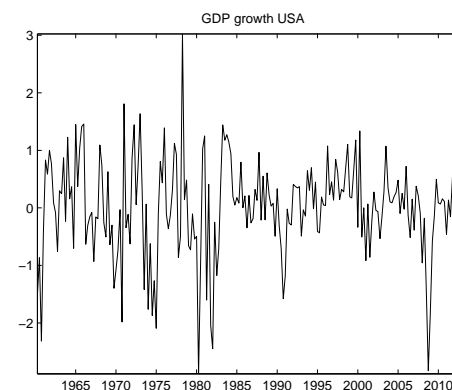


Figure 1: Real GDP growth in USA

It is not the aim of this session to give a methodological introduction to time series analysis, it is rather the aim to give a first insight of what can be done and how to get some insights into the time series dimension of macroeconomic data.

*This session was written by Philipp Wegmueller

2 The consumption function

In this first exercise we are interested in estimating the consumption function of the traditional IS-LM model we have already encountered in sessions 1 and 6. Remember that in this model we have that current private consumption expenditures, C , depend upon the available private income, Y_v . We can express a simple version of the consumption function as

$$C = c_0 + c_1 Y_v, \quad (2)$$

in which c_0 measures how much would be consumed if available income was zero. Note that it is reasonable to assume that $c_0 > 0$ also when $Y_v = 0$, since households also have to eat if income is zero. The coefficient c_1 measures the marginal propensity to consume, that is $mpc = dC/dY_v$. MPC is the proportion of additional income that an individual consumes. For example, if a household earns one extra dollar of disposable income, and the marginal propensity to consume is 0.65, then of that dollar, the household will spend 65 cents and save 35 cents. According to John Maynard Keynes, marginal propensity to consume is less than one. Let's have a look if this is the case in the data!

- (a) Open a new script file, and import the data into the Matlab workspace.

In the Excel file *session8_usdata.xls* you will find a selection of time series for the United States. For the present exercise, we only need a selection of series, namely DATE, GDP (nominal GDP in billions), GDPDEF (deflator), PCESV (consumption of services), PCND (consumption of non-durables), POP160V (working age population (16-64)). Total private consumption is given by the sum of services and non-durables.

- (b) Generate a vector that contains the time, one for the GDP, and others for the deflator, the services, non-durables and the population.
- (c) Bring the population (millions) into the same unit as GDP (billions) and divide the deflator by 100. Finally, real GDP per capita is defined by the nominal GDP divided by the working age population and the deflator, $y_t = Y_t / (Pop_t * P_t)$. Denote y_t with **rinc**.
- (d) Generate a new time series of real private consumption per capita, **rcon**, which is the sum of deflated consumption of services per capita and deflated consumption of non-durables per capita.
- (e) Plot a graph for both **rcon** and **rinc**.
- (f) Construct a time variable, T, containing all quarters. Generate a new graph with T on the x-axis and **rcon** and **rinc** on the y-axis.
- (g) Plot **rcon** against **rinc**.
- (h) Now, calculate some basic descriptive statistics (like the mean of both series, the standard deviation of both series, the correlation between the two series) for **rcon** and **rinc**.
- (i) Now, you are going to estimate c_0 and c_1 in the consumption function as in equation (2), which here reads: **rcon** = $c_0 + c_1 \cdot \text{rinc}$. In matrix notation the linear regression model can be written as: $y = X \cdot b + e$ where y denotes a vector where each element is an observation of the dependent variable and X is the set of independent variables. The dependent variable is **rcon**. The independent variables are a vector containing ones, and the variable **rinc**. Each element is an observation.

b is the parameter vector, containing two elements. In a first step, generate a constant, C , as a ones-vector of the same length as `rcon` and `rinc`. Now, generate y and X .

- (h) Calculate the ordinary least squares estimator of b by $b = (X'X)^{-1}X'y$.
- (i) Calculate the OLS-residuals, e , which are given by $e = (y - Xb)$.
- (j) Insert the regression line in the plot of problem (e).
- (k)** For those of you who seek a challenge and have spare time, try the following: Write a function file that calculates the ordinary least squares (OLS) estimator with White robust standard errors. Use the following equation as basis:

$$Y_i = X_i * \beta + u_i, \quad \text{where } u_i \sim \mathcal{N}(0, \sigma^2). \quad (3)$$

The function shall return the estimator, $\hat{\beta}$, corresponding to the k regressors, the standard errors of the estimator, the White robust standard error of the estimator, the estimated variance of the residuals and plot the residual series of the regression.

3 In the search of the Kaldor Facts

In this exercise we want to go a little bit deeper into the time series properties and calculate some key characteristics of the data.

- (a) Make a new section in your current script file (use `%%`).
- (b) Generate new vectors for durable goods (`PCDG`), change in inventories (`DINVENT`), fixed private investment (`FPI`). Then deflate them by the

GDP deflator and express them in per capita terms (call these vectors `RCD`, `RFI` and `RDI`). Define finally another vector for total investment (`RI=RCD+RFI+RDI`).

- (c) Generate also a new vector for real per capita government spending, `RG=GCE./(POP160V.*GDPDEF)`.
- (d) Once we have defined all the relevant series, let's calculate and display some long-run quantities (also called *great ratios*). A great ratio is the mean share of some variable in total GDP, e.g. the share of government spending in GDP, G/Y , or the share of the different components of investment in GDP and so on. Try it out.¹
- (e) Next we want to have a look at the growth rates of some of these variables. Take the log of each series and use the command `diff()` to generate new variables `gy`, `gc`, `gi` and `gg` for the growth rate of output, consumption, investment and government spending in percent.
- (f) Display the mean of these growth rates and create a figure that contains 4 subplots, one for each series. Next, create a figure which contains a scatterplots against output (e.g. for `gy` and `gc`, etc.).
- (g) Calculate the standard deviation of each growth rate and compare them.

¹In 1957, Nicholas Kaldor drew attention to some of the “*stylized facts*” of economic growth. Importantly, Kaldor pointed out that any satisfactory model of economic growth must account for the constancy of several key macroeconomic ratios that have remained stable even in the presence of enduring and volatile growth paths. These stylized facts are often referred to as the “*Kaldor facts*” and the ratios are sometimes known as the “*Great Ratios*.” For a survey read “Resuscitating real business cycles” by Robert King and Sergio Rebelo, chapter 14 in the Handbook of Macroeconomics, Volume 1b

- (h) Find a way to create a matrix that contains the correlation coefficients
- (i)** If you have still some time and energy, then take the log of the series **RY**, **RC**, **RI** and **RG** and detrend them linearly, by using either a function or performing a regression to extract the cyclical component of each series. Make a figure that contains a subplot with the cyclical components of each series. Find the mean, standard deviation and correlation and make a table that contains the stylized facts.