Time-Series Modelling

ECON20222 - Lecture 9

Ralf Becker and Martyn Andrews

March 2019

Aim for today

- Understand the basic features of time-series data
- Understand autocorrelation
- Understand the difference between stationary and nonstationary data
- Understand how to build dynamic models that can be used for forecasting

Purpose of time-series modelling (TS modelling)

There are typically three things econometricians want to achieve with time-series modelling

1. **Establishing causal relationships** between time-series.
   This is very difficult and in general causal relationships are more difficult to establis with TS modelling. It is not impossible but getting convincing exogenous variation is difficult.
2. Understanding the **dynamics in relationships between variables**.
   Questions like, "If the Central Bank changes the base rate, how long will it take for this to carry through to mortgage rates?" This is perfectly possible as long as we don't make strong causal statements (the CB may change base rates because mortgage rates are very low!!!)
3. **Forecasting one or several time series**.
   This is possibly the most common purpose of TS modelling. We will focus on this.

Import some data into R

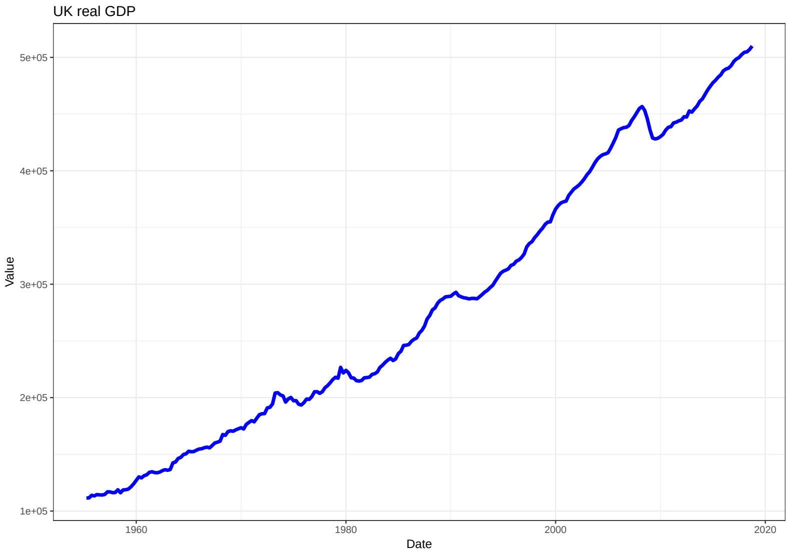
```
rGDP <- pdfetch_ONS("ABMI","UKEA")
periodicity(rGDP)   # check data availability
names(rGDP) <- "real GDP" # give a sensible name

# keep all the data including 2018-Q3
# this was the last observation available at the time this was
# remove this line if you want to use updated data
rGDP <- rGDP["/2018-9"]
```

`pdfetch` functions allow you to directly tap a umber of large data depositories:

- Bundesbank
- Office for National Statistics (ONS)
- Eurostats
- FRED, etc

An example


UK real GDP
Value
5e+05
4e+05
3e+05
2e+05
1e+05
1960
1980
2000
2020
Date

The autocorrelation function (ACF)

Recall the correlation coefficient between two random variables z_i and p_i for two cross-sectional variables (index i)

$$Corr(z_i, p_i) = \frac{Cov(z_i, p_i)}{sd(z_i)sd(p_i)}$$

- a measure that expresses the strength of relationship between z_i and p_i
- it takes values in the interval $[-1, 1]$

The autocorrelation function (ACF)

Consider the $y_t = rGDP_t$ time series

- we now use the subscript t
- the subscript goes from $t = 1, ..., T$ where T indicates how many observations we have
- here observations are quarterly (but other series can have other frequencies: e.g. annual, monthly, weekly, daily, hourly, etc.)
- here observations are from Q1 1955 to Q3 2018 (255 observations)

The ACF expresses how observations are correlated to observations 1, 2, 3 or k observations prior.

How can you calculate a correlation of a series with itself?

The autocorrelation function (ACF)

Let's consider the time series y_t and the series one period prior, y_{t-1}. We also call y_{t-1} a one period lag of y_t.

Observation	y_t	y_{t-1}
0	y_{1955Q1}	NA
1	y_{1955Q2}	y_{1955Q1}
2	y_{1955Q3}	y_{1955Q2}
3	y_{1955Q4}	y_{1955Q3}
$\vdots$	$\vdots$	$\vdots$
253	y_{2018Q2}	y_{2018Q1}
254	y_{2018Q3}	y_{2018Q2}

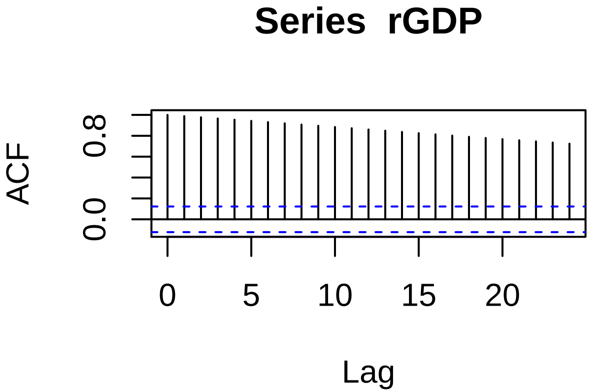
Now we have "two" series for which we can calculate a correlation coefficient. We call this the first order autocorrelation coefficient ρ_1.

An ACF is a collection of autocorrelation coefficients calculated for longer lags k, ρ_k.

The autocorrelation function (ACF)

In R this ACF is easily calculated using the `acf` function.

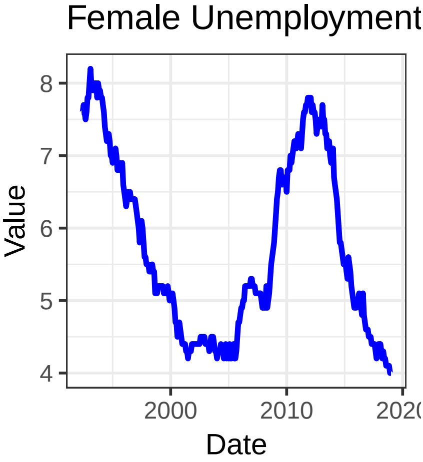
```
temp_acf <- acf(rGDP)
```

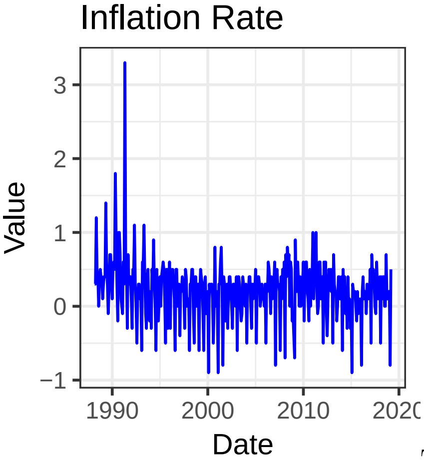


The autocorrelation function (ACF)

- The GDP series is strongly upward trending
- this is comon for many macroeconomic series
- this results in an ACF which has large ρ_k for fairly large values of k ($\rho_8 = 0.908$)
- we call this a persistant series

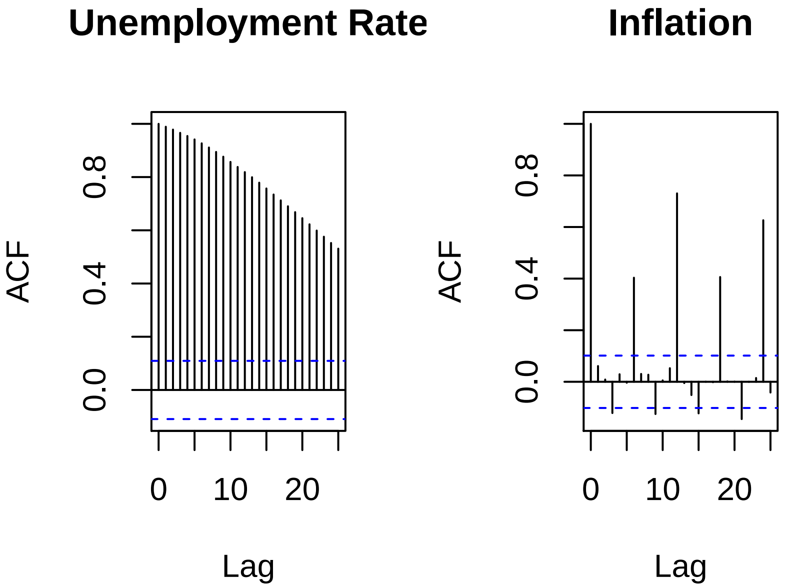
Two further examples





These are data at a monthly frequency

The ACF



The ACF

- the ACF shows that the unemployment rate is also very persistant and the ACF only slowly converges to 0. $\Rightarrow$ a series can be oersitant without time trend
- The inflation rate is not persistant and the autocorrelation quickly drops towards 0.
- But there are peaks of autocorrelation at frequencies 6 and 12 indicating some seasonal variation.

The ACF tells us something about how informative today's observation is for that in 1, 2, 3 or k periods ahead.

- If the ACF decays quickly to 0 then we will not be able to forecast a long period ahead.
- If the ACF decays slowly then todays info is valuable for future observations

Stationary and Nonstationary Series

The ACF expresses how persistant a series is.

- A series that is extremely persistant is called a **nonstationary** series.
- A series that is not very persistant is called a **stationary** series.

Here: rGDP and unemployment rate are nonstationary. Inflation rate is stationary.

- In general series with a time-trend are nonstationary
- **BUT** there is a huge grey area inbetween.

Formal statistical tests exist (e.g. Augmented Dickey-Fuller test) to decide (but they can be contradictory) and are not dealt with here. Here we eye-ball the series and look at how slowly the ACF converges to 0.

Transformations

An important time-series transformation we consider is that of differencing a series.

Observation	y_t	y_{t-1}	Δy_t
1	y_{1955Q2}	y_{1955Q1}	$y_{1955Q2} - y_{1955Q1} = \Delta y_{1955Q2}$
2	y_{1955Q3}	y_{1955Q2}	$y_{1955Q3} - y_{1955Q2} = \Delta y_{1955Q3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Often we are actually much more interested in the difference of a series rather than the level. GDP is a case in point, the growth rate is what we are really interested in!

The GDP growth rate can be approximated for small growth rates by (assuming that y_t is the GDP series)

$$gGDP_t = \frac{y_t - y_{t-1}}{y_{t-1}} \texttt{ or} \tag{1}$$

$$gGDP_t = ln(y_t) - ln(y_{t-1}) \tag{2}$$

ACF of differenced series

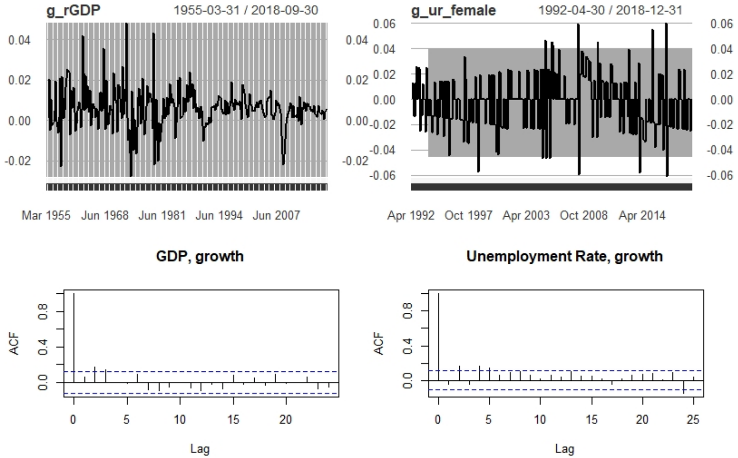


Figure 1: TS and ACF plots for growth in GDP and unemployment rate, ONS

Differencing can turn nonstationary series into a stationary series.

A simple regression

Cross-Section Data

$$y_t = \alpha + \beta x_t + u_t \tag{3}$$

$$\begin{aligned} E(u) &= 0 && (4) \\ E(u|x) &= 0 && (5) \end{aligned}$$

this implied that x_t was exogenous.

Time-Series Data

$$ur_t = \alpha + \beta \; rGDP_t + u_t \tag{6}$$

A simple regression

let's look at the end of the data table

```
# we multiply by 100 to express in percentage points, i.e. 0.5 is 0.5% or 0.005
reg_data$d_lgdp <- 100*diff(log(reg_data$real.GDP))
reg_data$d_lur <- 100*diff(log(reg_data$ur_female.Close))
tail(reg_data,10)
```

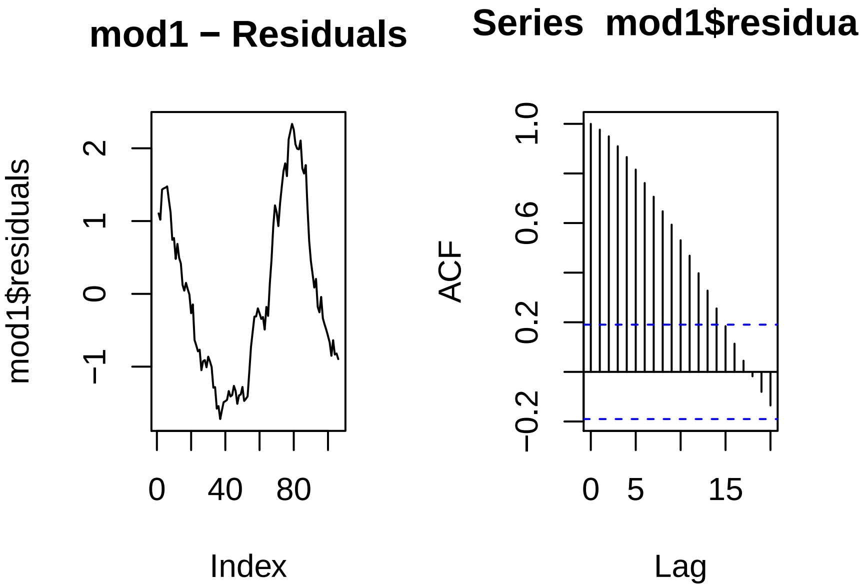
```
##            real.GDP ur_female.Close     d_lgdp     d_lur
## 2016-09-30   492816             4.7 0.47065228 -2.105341
## 2016-12-31   496470             4.6 0.73871795 -2.150621
## 2017-03-31   498582             4.5 0.42450107 -2.197891
## 2017-06-30   499885             4.4 0.26100026 -2.247286
## 2017-09-30   502473             4.2 0.51638352 -4.652002
## 2017-12-31   504487             4.4 0.40001642  4.652002
## 2018-03-31   504829             4.2 0.06776867 -4.652002
## 2018-06-30   506928             4.2 0.41492236  0.000000
## 2018-09-30   510013             4.1 0.60672339 -2.409755
## 2018-12-31       NA             4.0         NA -2.469261
```

A simple regression

```
mod1 <- lm(ur_female.Close~real.GDP,data = reg_data)
stargazer_HC(mod1)
```

```
##
## ==========================================================
##                               Dependent variable:
##                      -------------------------------------
##                                 ur_female.Close
## ----------------------------------------------------------
## real.GDP                          -0.00001***
##                                    (0.00000)
##
## Constant                            8.653***
##                                     (0.730)
##
## ----------------------------------------------------------
## Observations                          106
## R2                                   0.136
## Adjusted R2                          0.128
## Residual Std. Error             1.161 (df = 104)
## F Statistic                16.368*** (df = 1; 104)
## ==========================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
##                      Robust standard errors in parenthesis
```

A simple regression - residual autocorrelation



A simple regression - testing for residual autocorrelation

We can again apply a hypothesis test the Breusch-Godfrey test (`bgtest`). The null hypothesis is that there is no autocorrelation.

```
bgtest(mod1,order=4)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to
##
## data:  mod1
## LM test = 102.23, df = 4, p-value < 2.2e-16
```

A simple regression - HAC standard errors

When we estimate a regression which has autocorrelated error terms we need to apply a different formula to calculate standard errors for coefficients in a regression model (autoregressive heteroscedasticity consistent - HAC).

- They are called Newey-West standard errors.
- They are implemented in `stargazer_HAC.r`.
- This will not change the coefficient estimates.
- Will change the standard errors to the coefficients and hence inference (which will be incorrect if you don't use them).
- If you have time-series data, in doubt, use Newey-West standard errors
- But crucial problems remain (see next slides)

In this example the standard errors only change marginally and hence are not shown here.