University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

PART 3

8. Time Series Modelling and Forecasting

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ljubljana, October 2022

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

I am indebted to Chris Brooks and Lisa Schopohl from the University of Reading and Roberto Martínez–Espiñeira from the Memorial University for permission to employ their lecture materials.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

8.1 Stationarity

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Basic definitions

A series (process) y_t is called **(weakly) stationary** if it has time-invariant first and second moments:

first and second moments are linear

first moment: mean

$$E(y_t) = \mu$$

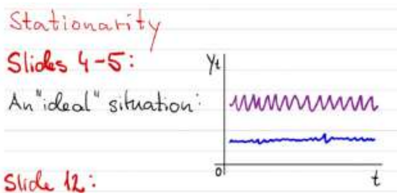
second moment: variance

$$\mathrm{var}(y_t) = \sigma^2$$

$$\mathrm{cov}(y_t, y_{t+s}) = \mathrm{cov}(y_t, y_{t-s}) = \gamma_s$$

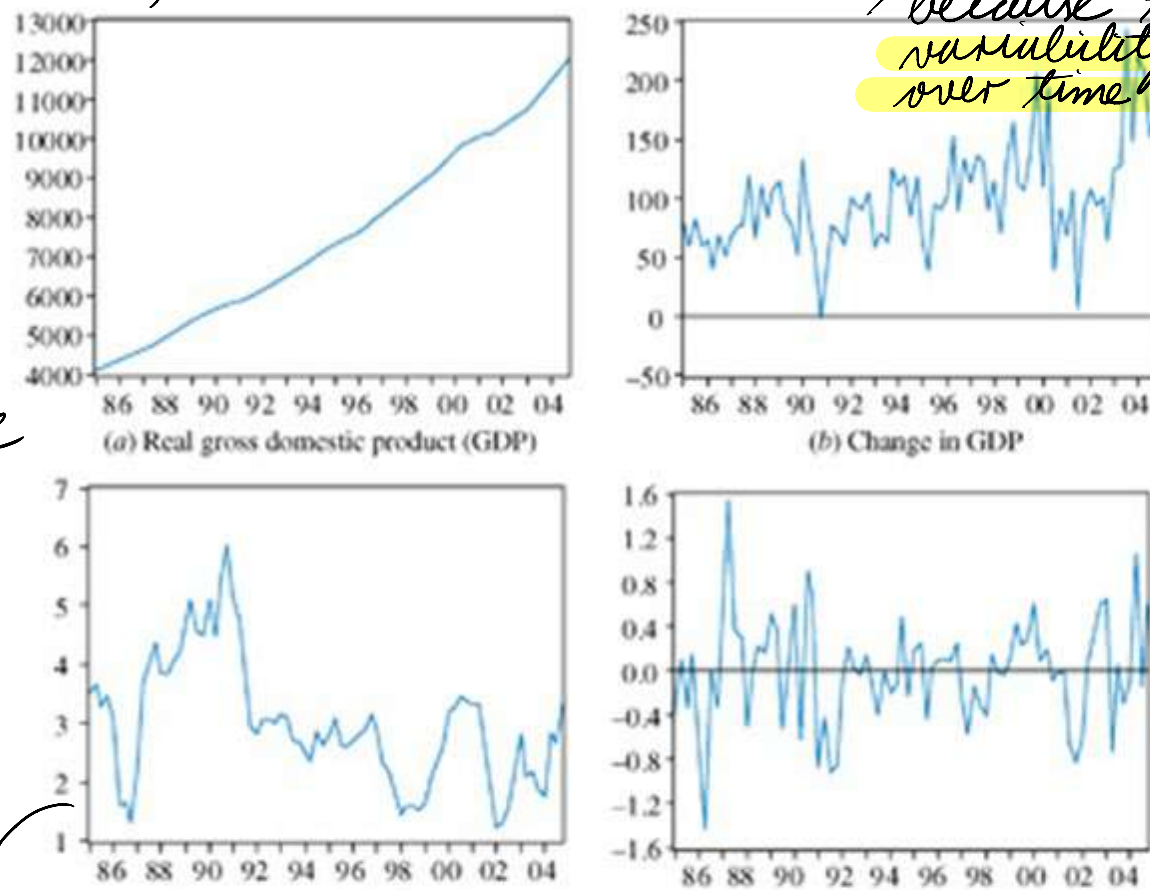
There also exists strict stationarity, which requires all joint distributions of y_t, …, y_{t-s} to be time invariant for every s.

Basic definitions



it is not stationary
because the first moment is linear

it is not stationary
because the variability changes over time



(*a*) Real gross domestic product (GDP)

(*b*) Change in GDP

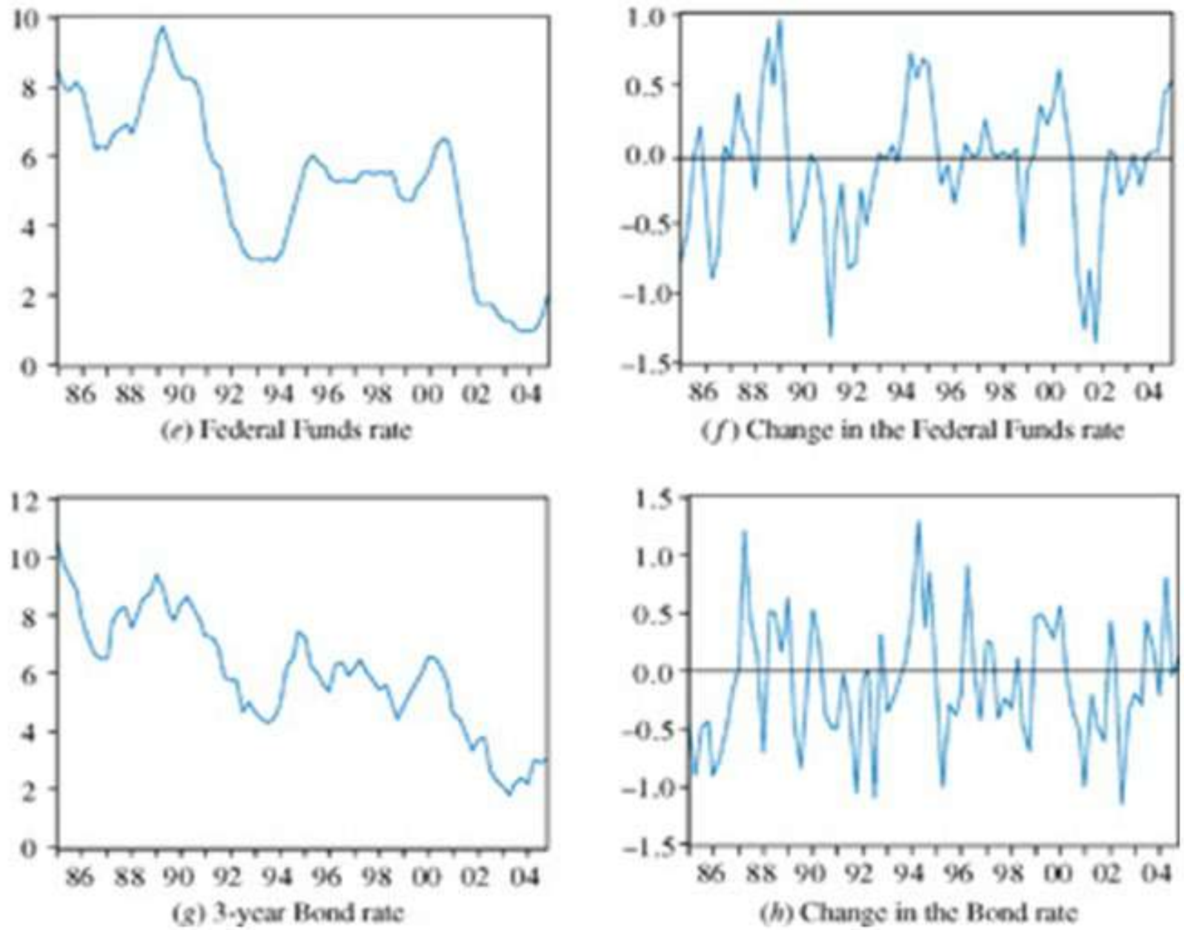
(*c*) Inflation rate

(*d*) Change in the inflation rate

the variability doesn't change over time

it is not stationary because we have a downwards trend

Basic definitions



Spurious regression

- If (weak) stationary is not satisfied for the variables included in a time-series regression model, we are likely to obtain the so-called spurious regression and invalid statistical inference (test statistics do *not* follow their respective statistical distributions).
- **Spurious regression** can be shown by generating two random walk (non-stationary) processes and regressing them on one another:

$$
\begin{aligned}
rw_1 &: \; y_t = y_{t-1} + v_{1t} \\
rw_2 &: \; x_t = x_{t-1} + v_{2t}
\end{aligned}
$$

random walk ⇒ value of the previous walk value and something stohastic

$v_{1t} + e$

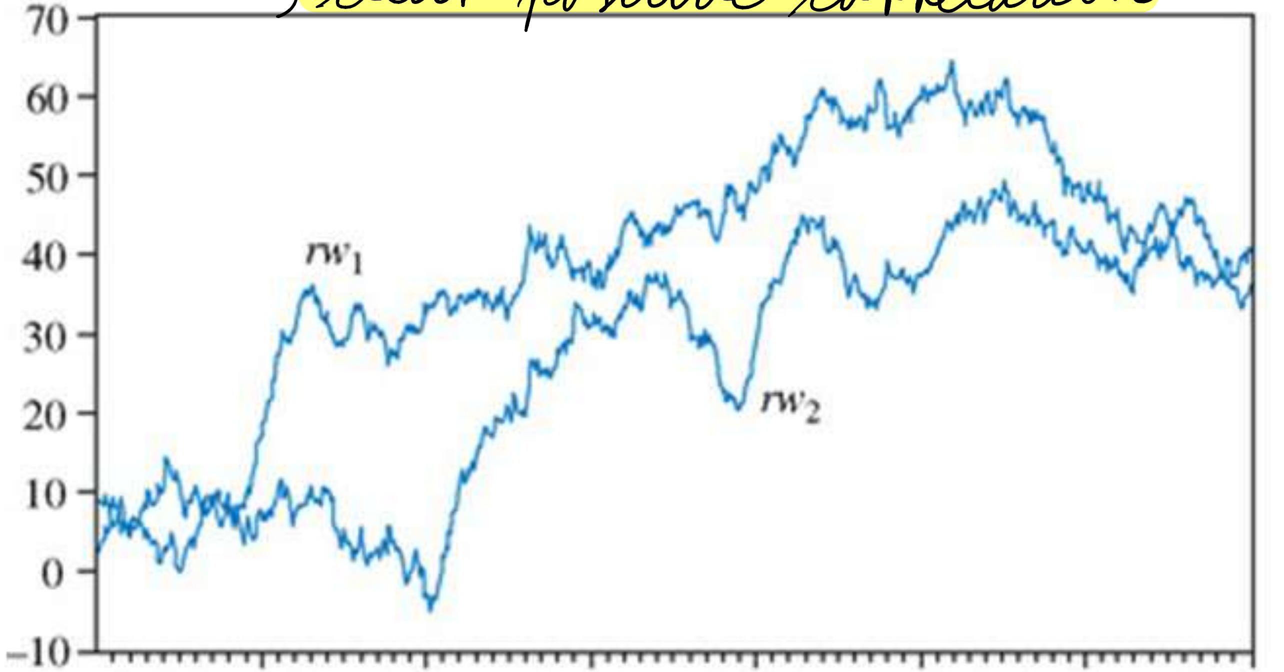
- Both processes are independent and were artificially generated, however, the relationship is strong and statistically significant:

$$
\widehat{rw_{1t}} = 17.818 + 0.842\, rw_{2t}, \quad R^2 = 0.70
$$
$$
(t) \qquad\qquad (40.837)
$$

Spurious regression

Time series of two random walk variables

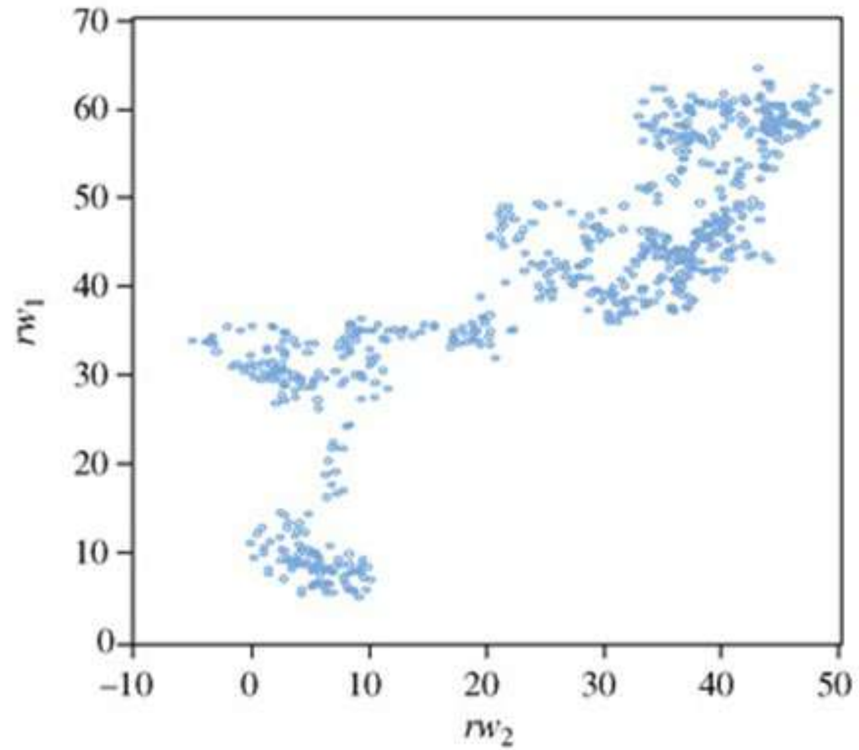
↳ clear positive correlation



(a) Time series

Spurious regression

Scatter plot of two random walk variables



(b) Scatter plot

If the variables are not stationary, then the results are not useful. Stationarity is a must in time series models.

Stationarity and integration

- If a time series is not stationary, we need to use suitable **transformations** to make it stationary.
- If a time series has a (stochastic) trend, we can apply (first) differences to remove it from the series.
- A data generating process (DGP) is called integrated of order *d*, denoted *I*(*d*), if first differences have to be applied *d*–times to make the process stationary, *I*(0):

$$y_t \sim I(d) \text{ if and only if } \Delta^d y_t \sim I(0),$$

$$\text{whereas } \Delta^{d-1} y_t \text{ is still non-stationary.}$$

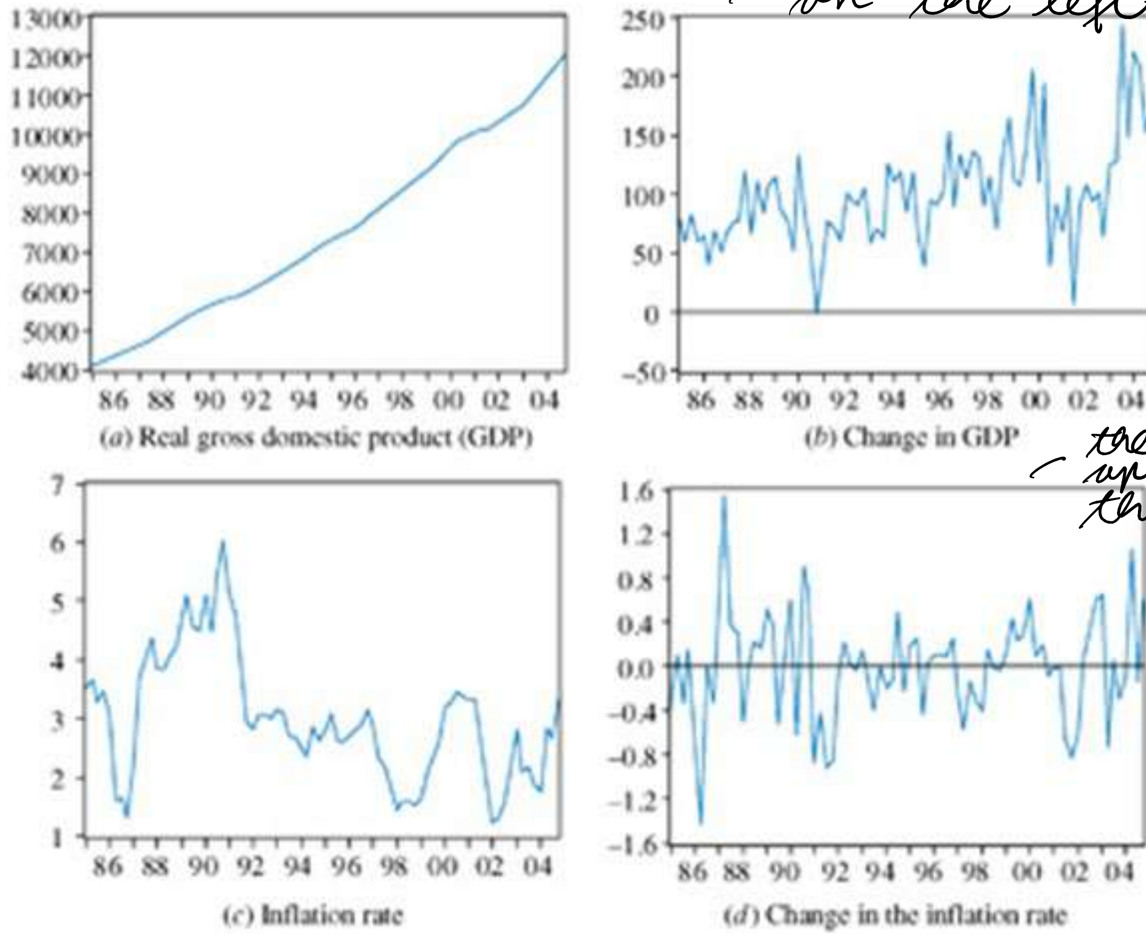
- We say that an *I*(1) non-stationary series (process) has a unit root.
- We can test for stationarity of (initial or transformed, first-differenced) time series with statistical tests. The most standard is the **Augmented Dickey–Fuller (ADF) test**.

usually this is enough

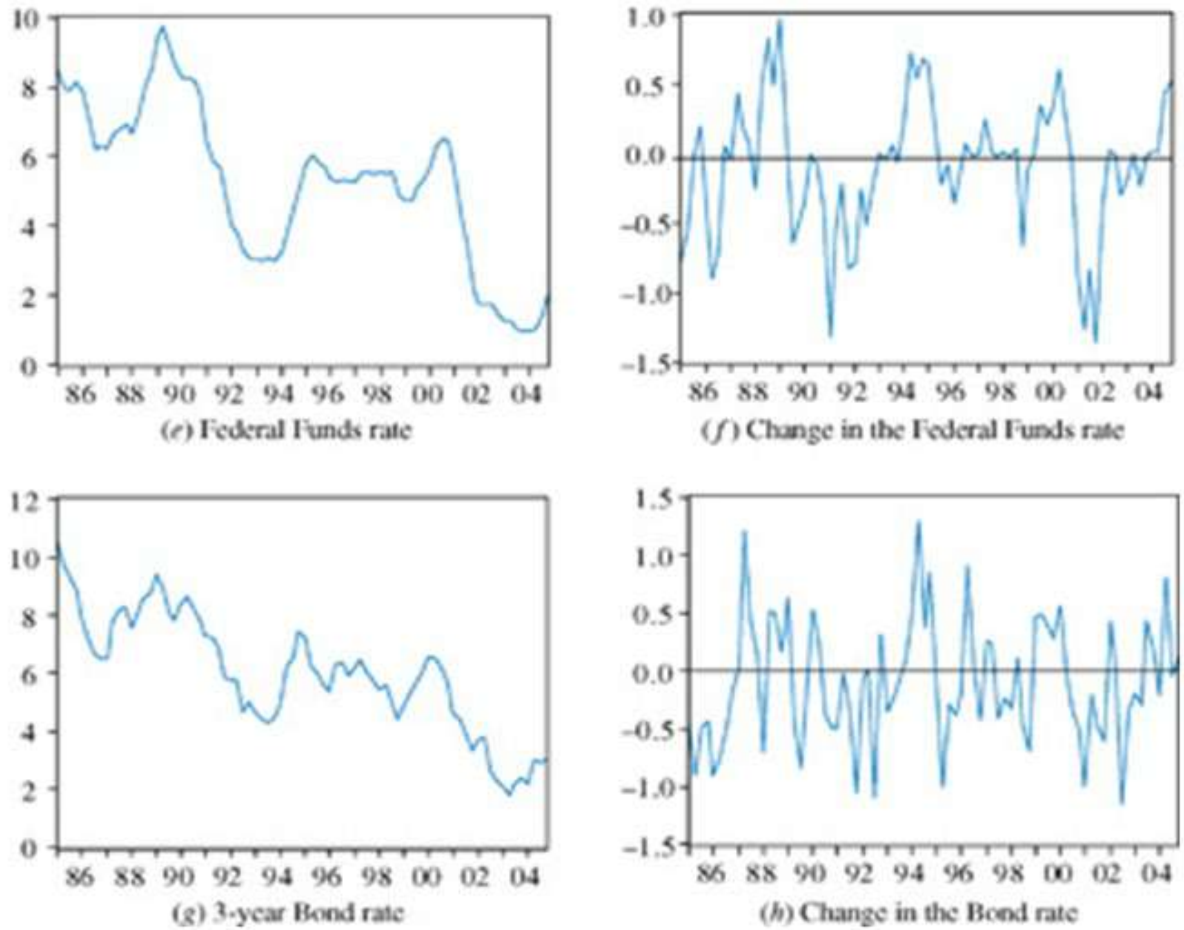
Examples of taking first differences

looks better than on the left

the same applies to this



Examples of taking first differences



Dickey–Fuller test

Version 1 (no constant and no trend):

$$\text{AR(1):}\ \ y_t = \rho y_{t-1} + v_t$$

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + v_t$$

$$\Delta y_t = (\rho - 1) y_{t-1} + v_t$$ – substract y_t on both sides

$$\Delta y_t = \gamma y_{t-1} + v_t$$ (Dickey–Fuller equation)

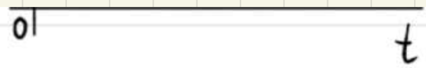
we have non stationarity

$H_0 : \rho = 1 \quad \leftrightarrow \quad H_0 : \gamma = 0 \quad$ y_t has a unit root

$H_1 : \rho < 1 \quad \leftrightarrow \quad H_1 : \gamma < 0 \quad$ y_t is stationary, $I(0)$

we have stationarity

We want to reject H_0 and conclude that we have stationarity.



Slide 12:

When $\rho = 1$ or $\gamma = 0$, we get from the Dickey-Fuller (DF) equation:

$$\Delta y_t = 0 \cdot y_{t-1} + v_t$$
$$\Delta y_t = v_t$$
$$y_t - y_{t-1} = v_t$$
$$y_t = y_{t-1} + v_t$$

random walk
(inherently non-stationary)

Dickey–Fuller test

Version 2 (with constant, but no trend):

$$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$$

Version 3 (with constant and with trend):

$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$$

We're testing these values.

Dickey–Fuller test

First, plot the time series of the variable:

- If the series appears to be wandering or fluctuating around a sample average of zero, use Version 1;
- If the series appears to be wandering or fluctuating around a sample average that is non-zero, use Version 2;
- If the series appears to be wandering or fluctuating around a linear trend, use Version 3.

Dickey–Fuller test

- The test statistic is based on the t–statistic from the least-squares estimation of the reparameterized model (coefficient γ).

we are unable to use tables in the handbook !!

- However, this test statistic follows a non-standard limiting distribution (standard critical values of the t–distribution cannot be used for statistical inference).
- Thus, either: a) critical values are used that are tabulated specifically for this test or b) specific p–values are simulated by statistical software (command `dfuller` in Stata or `adf.test` in R).

Augmented Dickey–Fuller test

An important extension of the Dickey-Fuller test allows for the possibility that the disturbance term is autocorrelated:

add lagged differences to correct autocorrelation

$$\Delta y_t = \gamma y_{t-1} + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$$

where: $\Delta y_{t-1} = (y_{t-1} - y_{t-2}),\ \Delta y_{t-2} = (y_{t-2} - y_{t-3}),\ \ldots$

Augmented Dickey–Fuller test

- The unit root tests based on the above three variants (with or without intercept and trend) are referred to as Augmented Dickey–Fuller tests.
- They are more general than the Dickey–Fuller tests.
- Enough lags of Δy_t should be included so that the disturbance term is serially uncorrelated, but not too many as this decreases the power of the test. Nonethelss, including too many lags is better than too few.
- The choice of the number of lags can be facilitated by using the *sequential-t method* (Ng–Perron), the *Schwarz criterion* (SC), the *modified AIC criterion* (MAIC) or the *Newey-West method*. !!
- If the sample size is sufficient, the SC criterion is the preferred choice. Otherwise, the MAIC criterion is most often used.

Additional considerations

The same as Frisch-Waugh-Lovell theorem

you get rid of the trend

1. regress on the linear trend
2. take the residuals

- Instead of applying (first) differences to a non-stationary time series, the following approaches are also used in practice:
  - ✓ Detrending a non-stationary time series;
  - ✓ Transforming a non-stationary time series to growth rates (usually by taking differences of logarithms) or
  - ✓ Applying a time series filter to a non-stationary time series (e.g. Hodrick–Prescott filter that extracts the cyclical component).
- As before, we need to test for stationarity of the transformed variable(s).
- **Stationarity needs to be ensured for every (dependent and explanatory) variable of a time-series regression model.**

Slide 18:

Approximation of growth rate of y_t:

$$\Delta \ln y_t = \ln y_t - \ln y_{t-1}$$

Approximation of percentage changes in y_t:

$$100 \cdot \Delta \ln y_t = 100 \cdot (\ln y_t - \ln y_{t-1})$$

In Finance this is used as a variable for growth rate

Example 1: We will analyse monthly UK house price series for the period from January 1991 to May 2013, already prepared in Stata data file `ukhp.dta`. The programming code is given in Stata Do file `ukhp-commands.do`.

First, we will inspect, test and ensure stationarity of our monthly UK house price series.

a) Inspect the UK house price (time series *HP*) visually. Does it look stationary? Perform the Augmented Dickey-Fuller (ADF) test on the time series. How many lags should you use? Should you include a trend into the DF regression? What do you find?
b) Also perform the Dickey-Fuller GLS test (DFGLS), which has in general more statistical power than the ADF test. What do you find? You can use information on the number of lags from the DFGLS test in the ADF test. Has the conclusion changed?
c) How can you make a time series stationary? Apply first differences on the time series *HP* and inspect the new time series visually. Does it look stationary? Repeat the ADF test on the new time series. What do you find?
d) Another way of making a time series stationary, often used in finance, is to generate continuously compounded percentage returns (percentage changes). Do this and inspect the new time series visually. Repeat the ADF test on this time series. What do you find?

Now, we will build an ARMA model for the house price changes. Recall that there are three stages involved: identification, estimation and diagnostic checking.

e) Identification is carried out by looking at the autocorrelation and partial autocorrelation coefficients to identify any structure in the data. Generate a table of autocorrelations, partial correlations and related test statistics. Draw the autocorrelation function (ACF) and the partial autocorrelation function (PACF). Does the time series appear to be an autoregressive (AR) process, a moving average (MA) process or a mixed ARMA process?
f) Use information criteria to decide on model orders p and q. Consider at least the first four lags. Which information criterion will you focus on? What do you find?
g) Estimate the model, proposed by the information criteria analysis, by employing the maximum likelihood estimator. Interpret the Stata output, but keep in mind that because the construction of these models is not based on economic or financial theory, it is often best not to interpret the individual parameter estimates, but rather to examine the plausibility of the model as a whole, and to determine whether it describes the data well and produces accurate forecasts. Also obtain the impulse responses (IRs).
h) Check for stability of the chosen model by examining inverse AR and MA roots of the characteristic equation. What do you find?
i) Now, suppose that the chosen time series model was estimated using observations from the period February 1991 – December 2010, leaving 29 remaining observations for constructing forecasts (for the period January 2011 – May 2013). Generate static and dynamic forecasts and present them graphically, together with the actual data, which are in fact available for the forecasting period. What do you find?

Static forecasts imply a sequence of one-step-ahead forecasts, rolling the sample forwards one observation after each forecast. Dynamic forecast calculates multi-step forecasts starting from the first period in the forecast sample. The dynamic forecasts quickly converge upon the long-term unconditional mean value as the horizon increases. This does not occur with the series of one-step-ahead forecasts, which seem to more closely resemble the actual series.

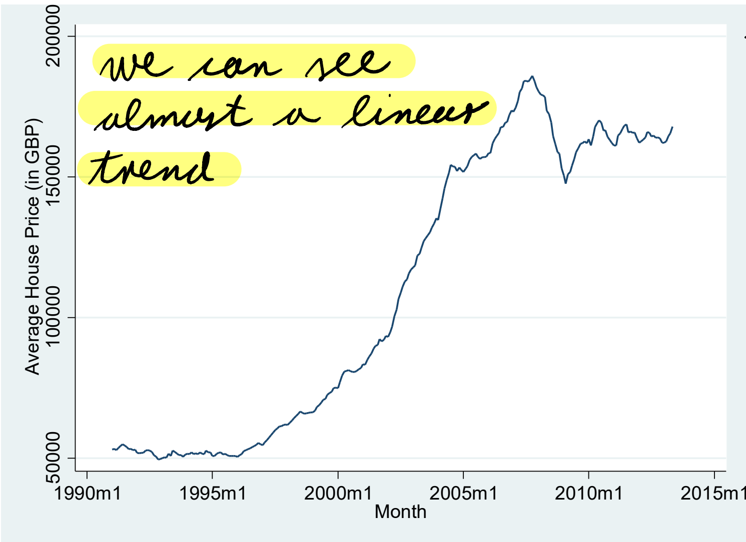
Computer printout of the results in Stata:

Data exploration

```
. tsset Month
        time variable:  Month, 1991m1 to 2013m5
                delta:  1 month
```

this tells stata that month is a time series variable

```
. twoway line HP Month
```



it isn't stationary

Example 1

Ad a) & b) PDF, pp. 2-4.

When $|\tau| > |\tau_c|$ or $p < \alpha$, we reject the unit root (non-stationarity).

Performing the Dickey-Fuller tests

```
. dfuller HP, lags(12)

Augmented Dickey-Fuller test for unit root         Number of obs   =       256

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.960            -3.460            -2.880            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7676
```

We can't reject the null.

The null is non stationarity

p-value

this could be deleted on the exam

```
. dfuller HP, regress lags(12)

Augmented Dickey-Fuller test for unit root         Number of obs   =       256

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.960            -3.460            -2.880            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.7676

------------------------------------------------------------------------------
        D.HP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          HP |
         L1. |  -.0014392   .0014994    -0.96   0.338    -.0043926    .0015143
         LD. |   .2859919   .0633351     4.52   0.000     .1612334    .4107504
```

```
        L2D. |   .3226638   .0656779     4.91   0.000     .1932906    .4520371
        L3D. |   .0217977   .0687253     0.32   0.751    -.1135784    .1571738
        L4D. |   .0249136   .0685106     0.36   0.716    -.1100396    .1598668
        L5D. |  -.0288369   .0684909    -0.42   0.674    -.1637514    .1060775
        L6D. |   .0128432   .0683735     0.19   0.851    -.1218399    .1475263
        L7D. |  -.1132738   .0683326    -1.66   0.099    -.2478763    .0213288
        L8D. |   .0238635   .0687485     0.35   0.729    -.1115583    .1592852
        L9D. |   .0884816   .0686746     1.29   0.199    -.0467948     .223758
       L10D. |  -.1151423   .0689386    -1.67   0.096    -.2509386     .020654
       L11D. |   .0911545   .0662825     1.38   0.170    -.0394098    .2217187
       L12D. |   .1653165   .0640052     2.58   0.010      .039238    .2913949
             |
       _cons |   271.1407   183.6549     1.48   0.141    -90.62553    632.9069
------------------------------------------------------------------------------
```

. dfuller HP, trend lags(12) - include trend

```
Augmented Dickey-Fuller test for unit root         Number of obs   =       256

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.364            -3.990            -3.430            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3987
```

- we can't reject the null

. dfuller HP, trend regress lags(12)

```
Augmented Dickey-Fuller test for unit root         Number of obs   =       256

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.364            -3.990            -3.430            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.3987

------------------------------------------------------------------------------
D.HP         |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          HP |
         L1. |  -.0109715   .0046402    -2.36   0.019    -.0201121   -.0018309
         LD. |   .2790635   .0629371     4.43   0.000     .1550865    .4030405
        L2D. |   .3231062   .0651813     4.96   0.000     .1947084    .4515039
        L3D. |   .0278044   .0682616     0.41   0.684    -.1066611    .1622698
        L4D. |   .0331733   .0680988     0.49   0.627    -.1009716    .1673182
        L5D. |  -.0203513   .0680853    -0.30   0.765    -.1544694    .1137669
        L6D. |   .0162966   .0678748     0.24   0.810    -.1174071    .1500003
        L7D. |  -.1099157   .0678333    -1.62   0.106    -.2435375    .0237061
        L8D. |   .0258148   .0682343     0.38   0.706    -.1085969    .1602265
        L9D. |   .0919464   .0681738     1.35   0.179    -.0423462     .226239
       L10D. |  -.1083164   .0684894    -1.58   0.115    -.2432307    .0265979
       L11D. |   .0991286   .0658837     1.50   0.134    -.0306528      .22891
       L12D. |   .1798951   .0638756     2.82   0.005     .0540692    .3057209
      _trend |   6.705113   3.091602     2.17   0.031     .6151001    12.79512
       _cons |   369.3101   187.8018     1.97   0.050    -.6325458    739.2527
------------------------------------------------------------------------------
```

. sum HP

```
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
          HP |        269    109363.5    50086.37   49601.66   186043.6
```

```
. scalar nwl=floor(4*(r(N)/100)^(2/9))
. display nwl
4
```
– number of lags should be 4

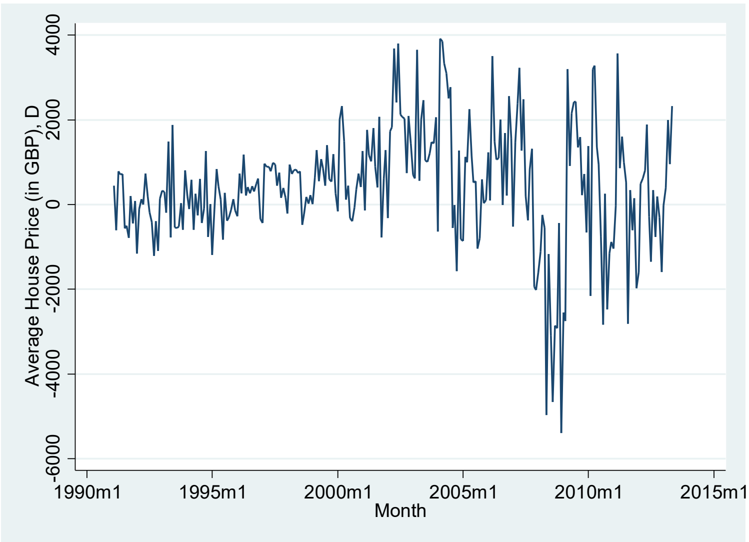
```
. dfuller HP, trend lags(4)
```
– include trend

```
Augmented Dickey-Fuller test for unit root         Number of obs   =       264

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -1.831            -3.989            -3.429            -3.130
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.6898
```
– we still can't reject the null.

```
. dfgls HP, maxlag(12)

DF-GLS for HP                                            Number of obs =   256

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    12           -2.019           -3.480            -2.836            -2.554
    11           -1.706           -3.480            -2.844            -2.561
    10           -1.478           -3.480            -2.851            -2.568
    9            -1.510           -3.480            -2.859            -2.575
    8            -1.317           -3.480            -2.866            -2.582
    7            -1.223           -3.480            -2.873            -2.589
    6            -1.296           -3.480            -2.880            -2.595
    5            -1.269           -3.480            -2.887            -2.601
    4            -1.344           -3.480            -2.893            -2.606
    3            -1.349           -3.480            -2.899            -2.612
    2            -1.310           -3.480            -2.905            -2.617
    1            -0.897           -3.480            -2.910            -2.622

Opt Lag (Ng-Perron seq t) = 12 with RMSE  1139.738
Min SC    =  14.22714 at lag  2 with RMSE  1189.249
Min MAIC =  14.19142 at lag  2 with RMSE  1189.249
```

Applying first differences and repeating the ADF test

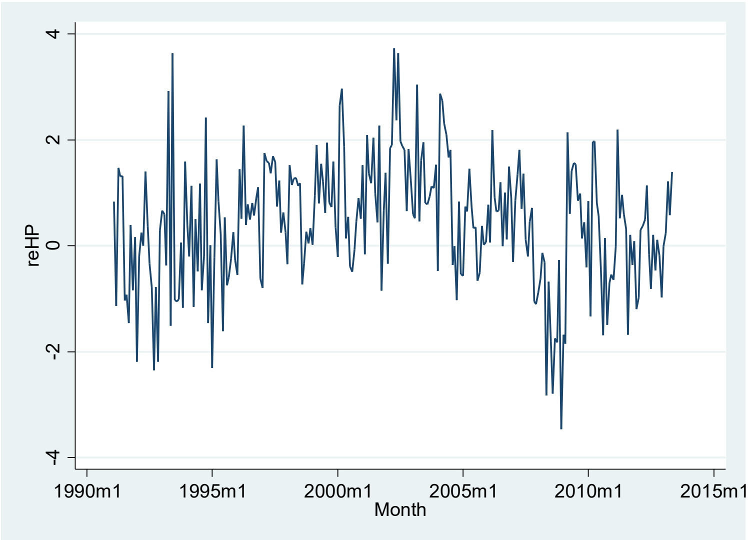
```
. twoway line d.HP Month
```



```
. dfuller d.HP, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =       263

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.044            -3.459            -2.879            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Generating continuously compounded percentage returns (percentage changes) and repeating the ADF test page 41 in chapter 8

```
. generate reHP=100*(ln(HP/L.HP))
(1 missing value generated)

. twoway line reHP Month
```




```
. dfuller reHP, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =       263

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.152            -3.459            -2.879            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Constructing an ARMA model

```
. corrgram reHP, lags(12)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.3566   0.3575   34.468  0.0000          |--                |--
2        0.4329   0.3504   85.444  0.0000          |---               |--
3        0.2407   0.0179   101.27  0.0000          |-                 |
```

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

8.2 ARMA (Box-Jenkins) Models

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Basic definitions

- In this section, we will deal with **univariate time series models**.
- We will focus on models, where the dependent variable is explained using information contained only in its own past values and current and past values of a disturbance term (stochastic variable).
- Such models are usually atheoretical, as they they are not based upon any underlying theoretical model of the behaviour of a variable.
- They are used when structural models are not feasible (due to data limitations, lack of theory, or forecasting underperformance).
- A multivariate extension to such models will be briefly mentioned, where exogenous explanatory variables can be added to the model.
- First, we need to introduce some additional concepts.

Basic definitions

- If a process is weakly stationary, all the variances are the same and all the covariances depend on the difference between t_1 and t_2. The moments:

$$E(y_t - E(y_t))(y_{t+s} - E(y_{t+s})) = \gamma_s, \quad s = 0, 1, 2, \ldots$$

are known as the autocovariance function.

- The covariances, γ_s, are known as autocovariances.
- However, these values depend on the units of measurement of y_t.
- It is thus more convenient to use the autocorrelations, which are the autocovariances normalised by dividing by the variance:

$$\boxed{\tau_s = \frac{\gamma_s}{\gamma_0}, \qquad s = 0, 1, 2, \ldots}$$

- If we plot τ_s against $s = 0,1,2, \ldots$, then we obtain the autocorrelation function (ACF) or correlogram.

Basic definitions

- A white noise process is one with (virtually) no discernible structure. A definition of a white noise process is:

$$
\begin{aligned}
&E(y_t) = 0 \\
&Var(y_t) = \sigma^2 \\
&\gamma_{t-r} = \begin{cases} \sigma^2 & if \quad t = r \\ 0 & \quad otherwise \end{cases}
\end{aligned}
$$

Autoregressive (AR) Processes

- An autoregressive model of order p, AR(p), can be expressed as

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + u_t$$

- By using the lag operator notation, $Ly_t = y_{t-1}$ and $L^i y_t = y_{t-i}$, we get alternative formulations:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$$

$$y_t = \mu + \sum_{i=1}^{p} \phi_i L^i y_t + u_t$$

$$\phi(L) y_t = \mu + u_t \text{ , where } \phi(L) = 1 - (\phi_1 L + \phi_2 L^2 + \ldots \phi_p L^p) \text{ .}$$

ARMA (Box-Jenkins) models

Slide 23:

We distinguish among 3 types of processes:

1) AR processes;
2) MA processes;
3) ARMA processes.

AR(p) process

Slide 23

$$y_t = \mu + \sum \phi_i y_{t-i} + u_t$$

$$y_t = \mu + \sum \phi_i L^i y_t + u_t$$

$$y_t - \sum \phi_i L^i y_t = \mu + u_t$$

$$y_t - \phi_1 L y_t - \phi_2 L^2 y_t - \ldots - \phi_p L^p y_t = \mu + u_t$$

$$y_t \underbrace{(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)}_{\phi(L)} = \mu + u_t$$

$$\phi(L) y_t = \mu + u_t$$

Autoregressive (AR) Processes

- The **condition for stationarity** of a general AR(p) model is that the roots of the characteristic equation:

$$\boxed{1-\phi_1 z-\phi_2 z^2-\ldots-\phi_p z^p=0}$$

  all lie outside the unit circle.

- In addition to the standard stationarity benefits, a stationary AR(p) model will have a MA(∞) representation.
- Example 1: Is $y_t = y_{t-1} + u_t$ stationary?

  The characteristic root is 1, so it is a unit root process (non-stationary).
- Example 2: Is $y_t = 3y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + u_t$ stationary?

  The characteristic roots are 1, 2/3, and 2. Since only one of these lies outside the unit circle, the process is non-stationary.

EXAMPLE 1:

Slide 24

$$y_t = y_{t-1} + u_t$$ (random walk process)

$$y_t - y_{t-1} = u_t$$

$$y_t - Ly_t = u_t$$

$$y_t(\underline{1 - L}) = u_t$$

Characteristic equation:

$$1 - z = 0$$

$$z = \underline{1} \times$$

Non-stationary
(for stationarity, z should be above 1).

EXAMPLE 2:

$$y_t = 3y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + u_t$$

$$y_t - 3y_{t-1} + 2.75y_{t-2} - 0.75y_{t-3} = u_t$$

$$y_t - 3Ly_t + 2.75L^2 y_t - 0.75L^3 y_t = u_t$$

$$y_t(\underline{1 - 3L + 2.75L^2 - 0.75L^3}) = u_t$$

Characteristic equation:

$$1 - 3z + 2.75z^2 - 0.75z^3 = 0$$

$$(1-z)(1-1.5z)(1-0.5z) = 0$$

$$z_1 = \underline{1} \times$$

$$z_2 = \underline{\tfrac{2}{3}} \times$$

$$z_3 = \underline{2} \checkmark$$

Non-stationary
(for stationarity, all z_j should be above 1).

Autoregressive (AR) Processes

- **Wold's decomposition theorem** states that any stationary series can be decomposed into the sum of two unrelated processes, a purely deterministic part and a purely stochastic part, which will be an MA(∞).

- For the AR(p) model, $\phi(L)y_t = u_t$, ignoring the intercept for simplicity, the Wold decomposition is:

$$y_t = \psi(L)u_t, \quad \text{where} \quad \psi(L) = (1-\phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1}.$$

- If the AR model is stationary, the autocorrelation function (ACF) will decay exponentially to zero.

Moving Average (MA) Processes

- Let u_t (t = 1, 2, 3, ...) be a sequence of independently and identically distributed (IID) random variables with $E(u_t) = 0$ and $Var(u_t) = \sigma^2$. Then:

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q}$$

is a moving average model of order q, MA(q).

- Its properties are: $$E(y_t) = \mu$$

$$Var(y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma^2$$

Covariances: $$\gamma_s = \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \ldots + \theta_q\theta_{q-s})\sigma^2 & for \quad s = 1,2,\ldots,q \\ 0 & for \quad s > q \end{cases}$$

- The **invertibility condition:** similarly to the stationarity condition, we typically require the MA(q) model to have roots of $\theta(z) = 0$ greater than one in absolute value.

MA(q) process

Slide 26

$$y_t = \mu + u_t + \Theta_1 u_{t-1} + \Theta_2 u_{t-2} + \dots + \Theta_q u_{t-q}$$

$$y_t = \mu + u_t + \Theta_1 L u_t + \Theta_2 L^2 u_t + \dots + \Theta_q L^q u_t$$

$$y_t = \mu + u_t \underbrace{(1 + \Theta_1 L + \Theta_2 L^2 + \dots + \Theta_q L^q)}_{\Theta(L)}$$

$$y_t = \mu + \Theta(L) u_t$$

ARMA Processes

- By combining both the AR(p) and the MA(q) model, we can obtain an ARMA(p,q) model:

$$\phi(L)y_t = \mu + \theta(L)u_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$.

- Thus the current value of y depends linearly on its own past values and a combination of current and past values of a white noise disturbance term.

- Alternatively, we can write:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q} + u_t$$

- The properties of an ARMA model:

$$E(u_t) = 0;\ E(u_t^2) = \sigma^2;\ E(u_t u_s) = 0,\ t \neq s\ .$$

ARMA Processes

- **Partial autocorrelation function (PACF), denoted τ_{kk}**, measures the correlation between an "observation" k periods ago and the current "observation", after controlling for "observations" at intermediate lags (i.e. all lags $< k$).
- In other words, τ_{kk} measures the correlation between y_t and y_{t-k} after removing the effects of $y_{t-k+1}, y_{t-k+2}, \ldots, y_{t-1}$.
- At lag 1, the ACF is always equal to the PACF.
- At lag 2, $\tau_{22} = (\tau_2 - \tau_1^2) / (1 - \tau_1^2)$.
- For higher lags, the formulae get complex.

ARMA Processes

Summary of the behaviour of the ACF and PACF for autoregressive and moving average processes

An autoregressive (AR) process has:

- a geometrically decaying ACF and
- number of spikes in the PACF equal to the AR order.

ravno

obrutno

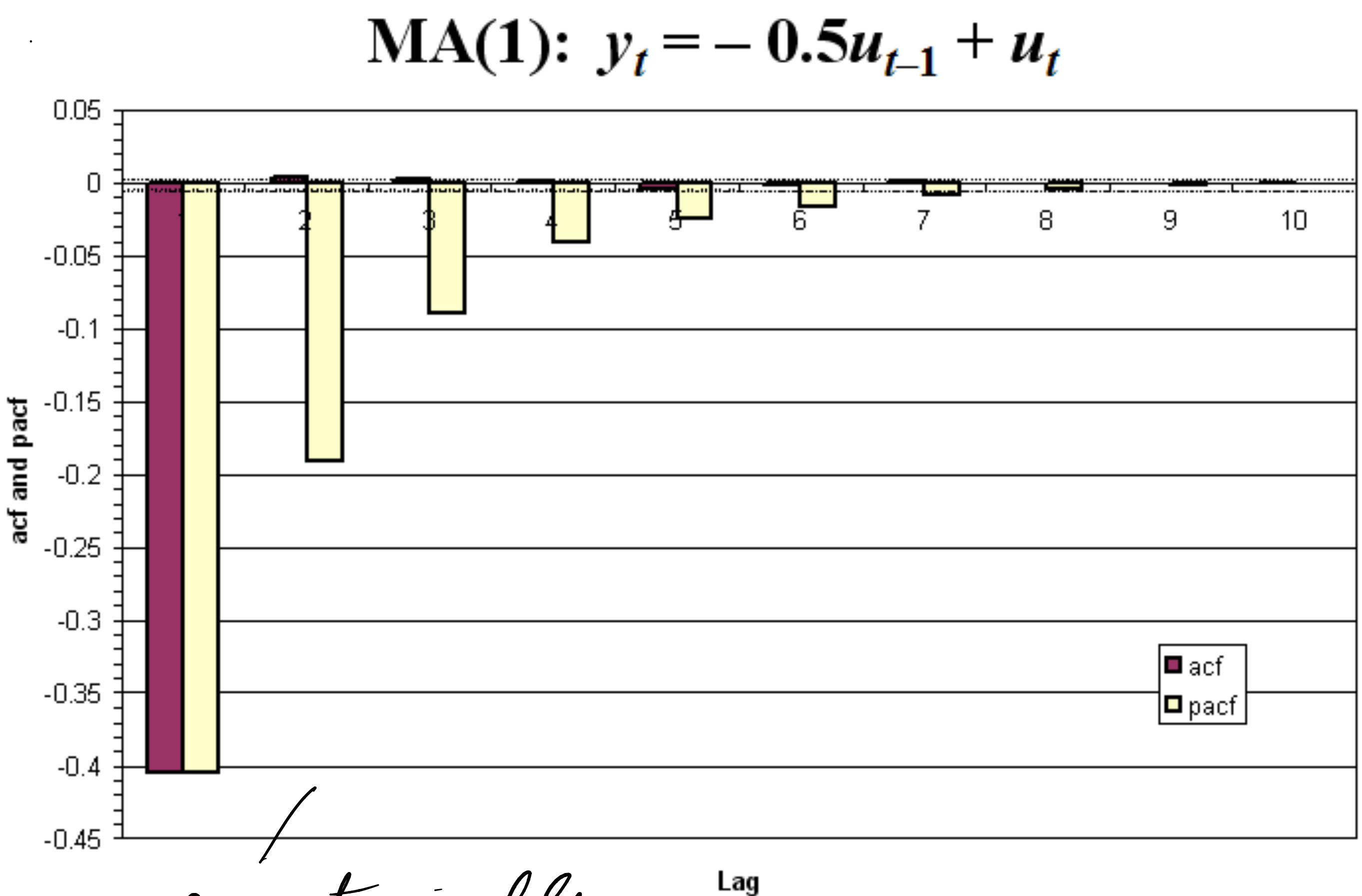
A moving average (MA) process has:

- number of spikes in the ACF equal to the MA order and
- a geometrically decaying PACF.

An autoregressive moving average (ARMA) process has:

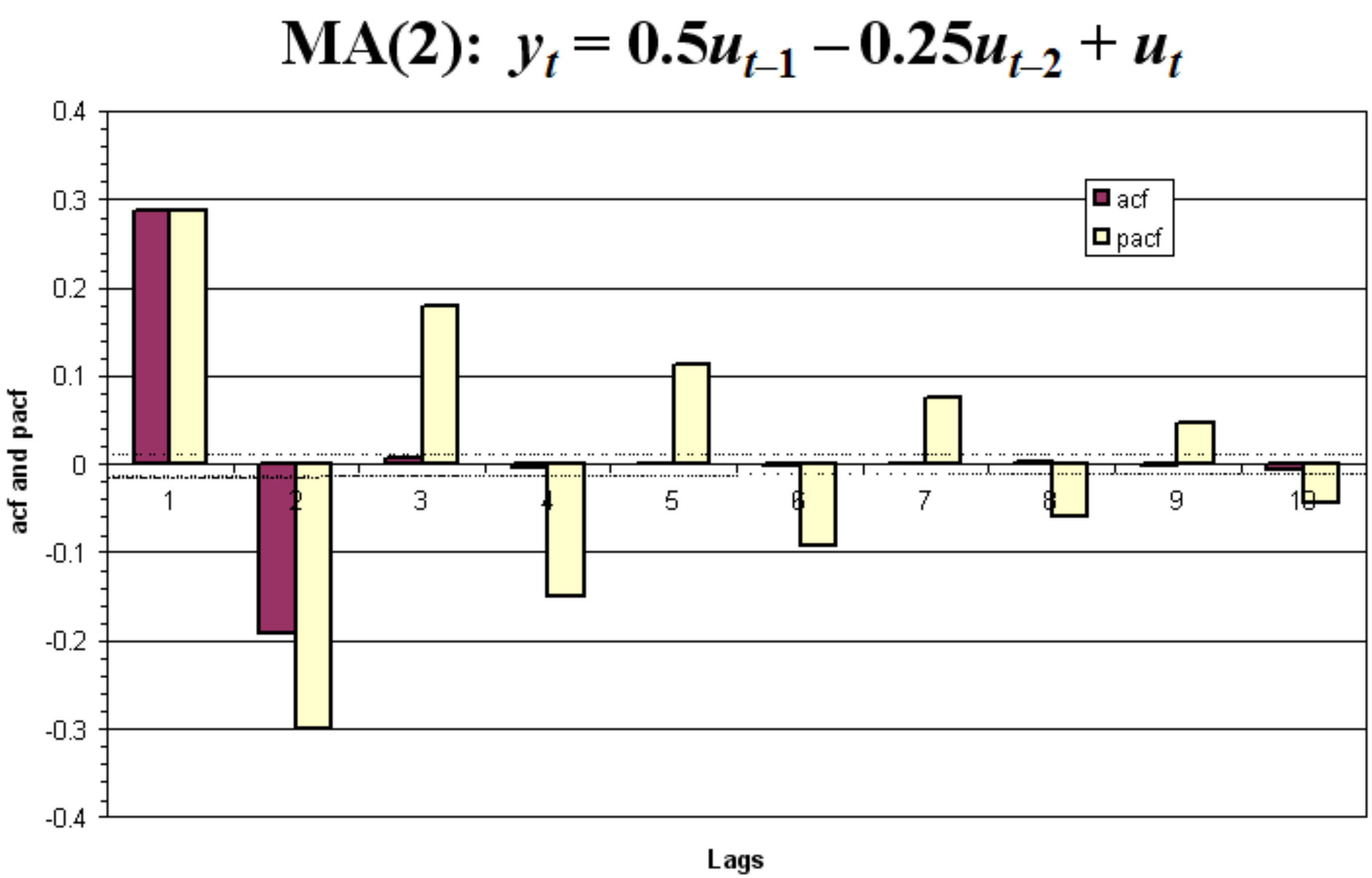
- a geometrically decaying ACF and
- a geometrically decaying PACF.

ACF and PACF plots for standard processes



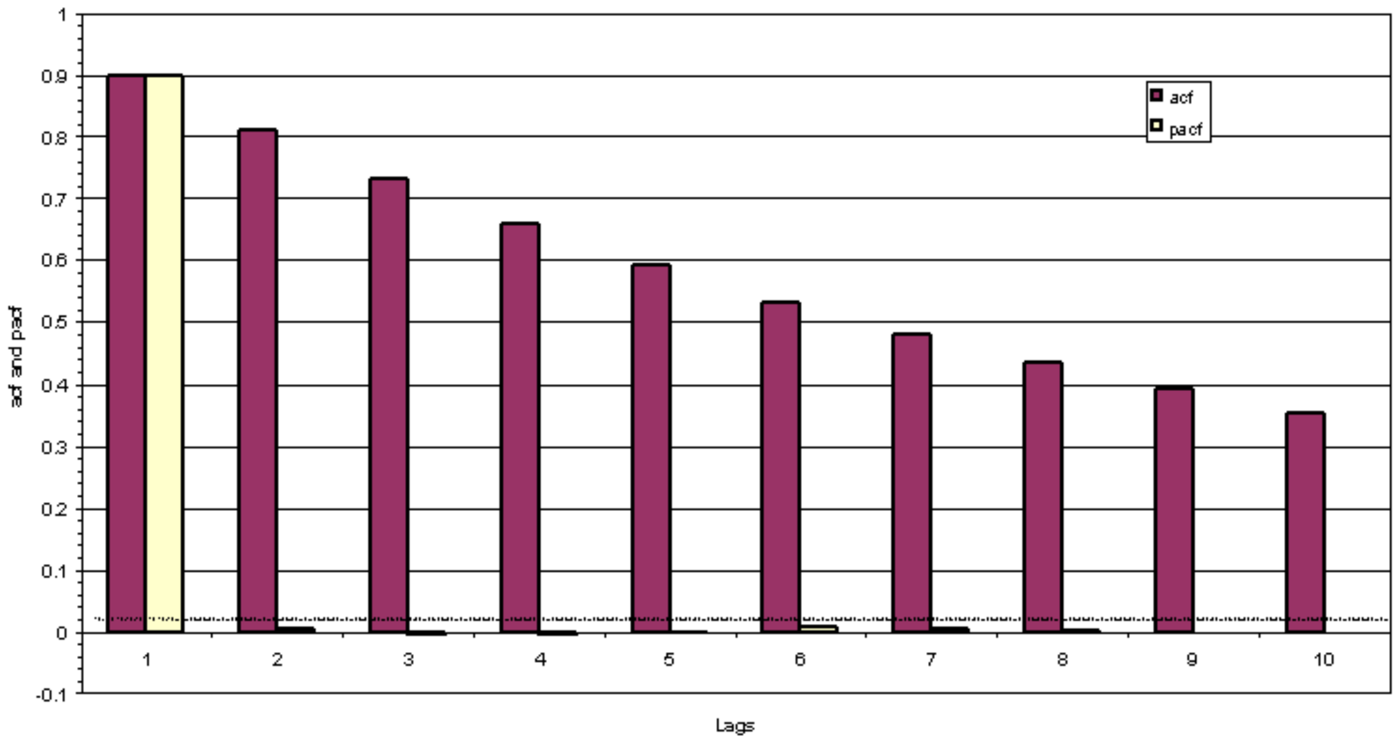
geometrically decaying regression coefficient

ACF and PACF plots for standard processes

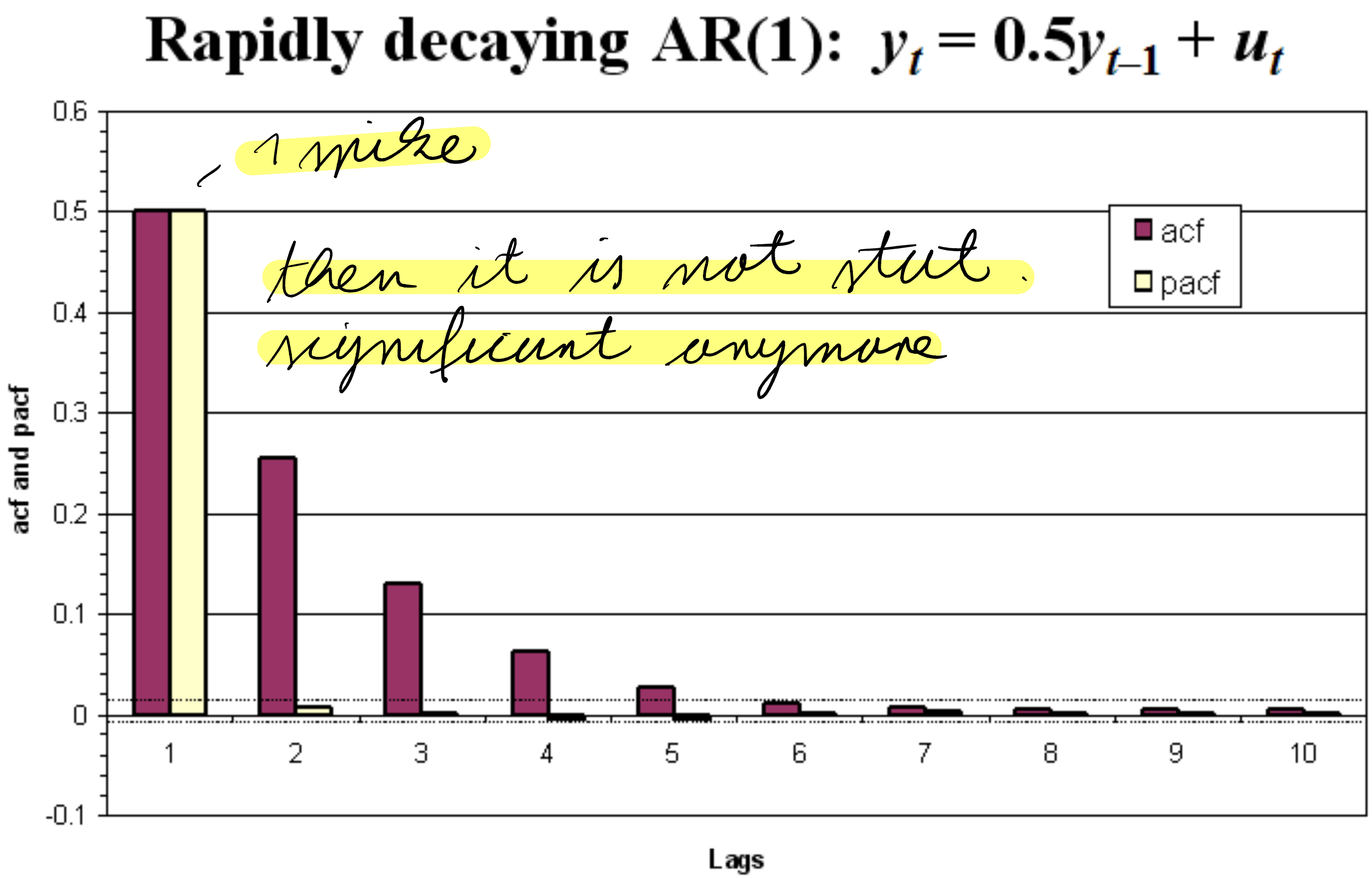


ACF and PACF plots for standard processes

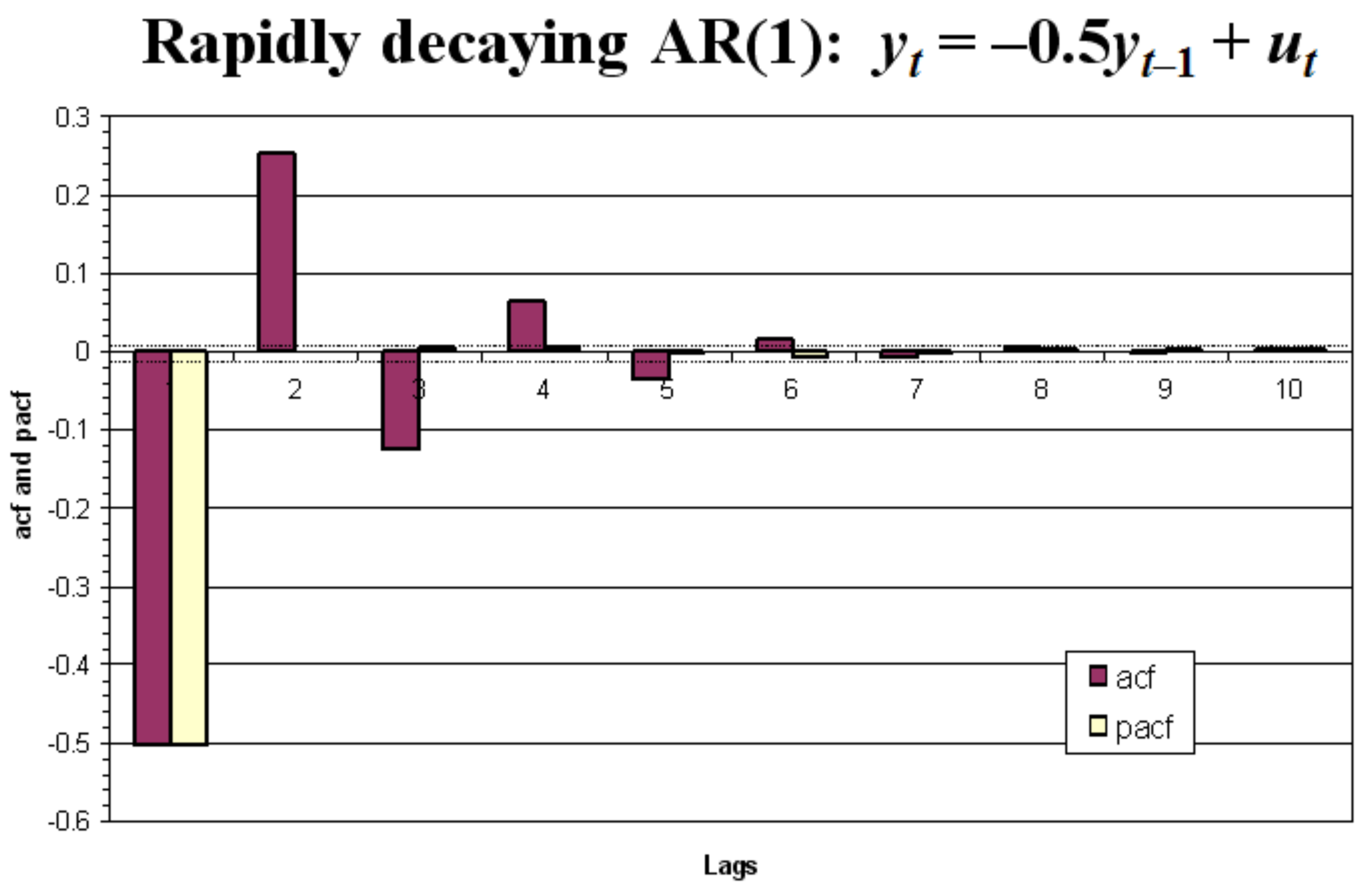
Slowly decaying AR(1): $y_t = 0.9y_{t-1} + u_t$



ACF and PACF plots for standard processes



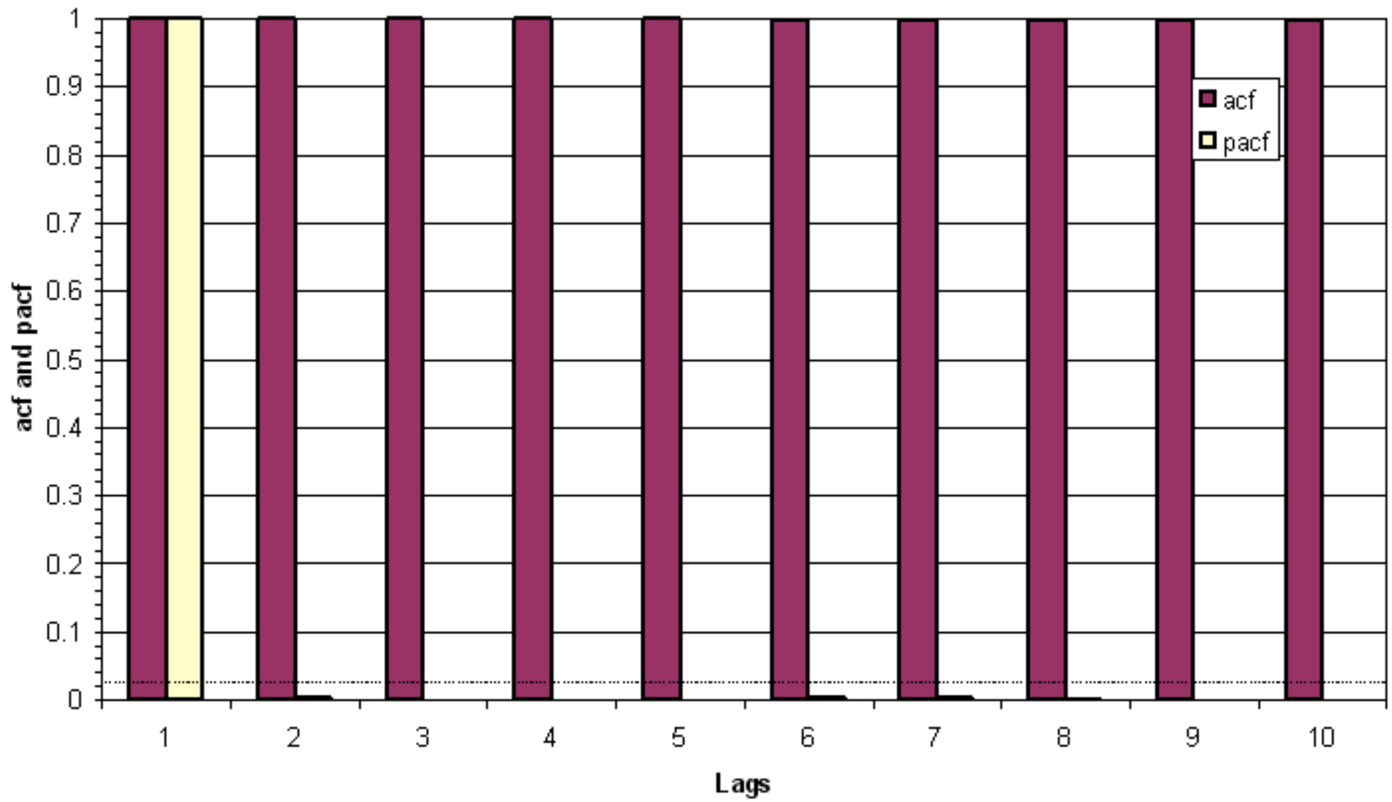
ACF and PACF plots for standard processes



ACF and PACF plots for standard processes

the shock persists indefinitely
↳ the model is not stable

Non-stationary model: $y_t = y_{t-1} + u_t$

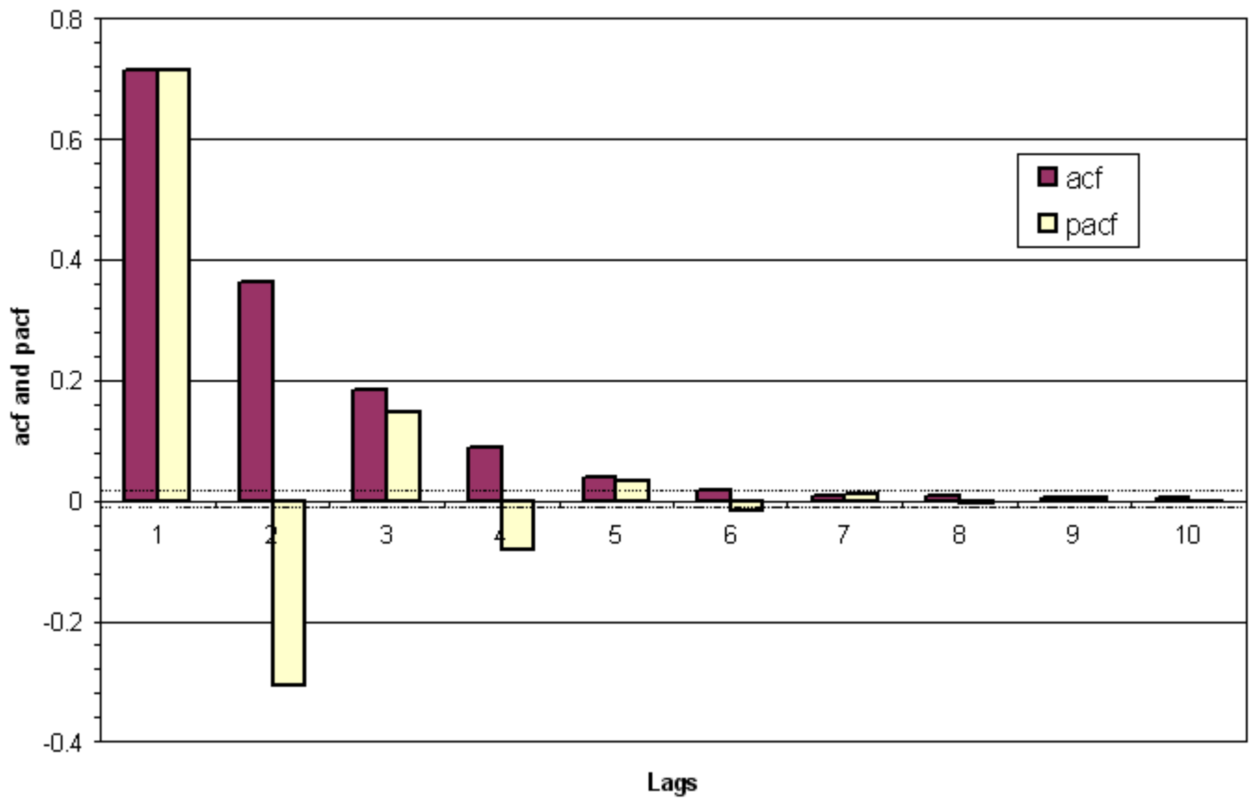


random walk process

ACF and PACF plots for standard processes

we need to have certain tools to decide if the model is ARMA model.

ARMA(1, 1): $y_t = 0.5y_{t-1} + 0.5u_{t-1} + u_t$



Box–Jenkins Approach

↳ very important for forecasting

Box and Jenkins (1970) were the first to approach the task of estimating an ARMA model in a systematic manner. There are three steps to their approach:

1. Identification
2. Estimation
3. Model diagnostic checking

Step 1: Identification

- Involves determining the order of the model required to capture the dynamic features of the data. ⇒ p, q
- Initially: use of graphical procedures (time plots, ACFs, PACFs).
- Nowadays: use of information criteria (AIC, SBIC).

Box–Jenkins Approach

Step 2: Estimation

- Estimation of the model parameters (regression coefficients and variances).
- Can be done using the least squares estimator or the maximum likelihood estimator, depending on the model.

Step 3: Model diagnostic checking

Box and Jenkins suggested two methods:

- Overfitting: deliberately fitting a larger model than that required to capture the dynamics of the data and checking for statistically insignificant AR and MA terms;
- Residual diagnostics: checking the residuals of the model for evidence of linear dependence (autocorrelation tests).

Box–Jenkins Approach

Box – Jenkins approach
Slide 39:
Parsimony of the model:
$k\downarrow$, $df: (n-k)\uparrow$, $Var(b_j)\downarrow$ and vice versa.
less coef. more degrees of freedom variance of regression coef. estimates decreases

- We want to construct a parsimonious model, because:
  - Variance of estimators is inversely proportional to the number of degrees of freedom (used by AR and MA terms);
  - Models that are too complex might be inclined to fit to data specific features, which would not be replicated out of sample. => then it is not good for forecasting
    ↳ model is tailored for past data
- In identification, graphical procedures are nowadays used as a starting point, focusing instead on information criteria.
- Information criteria embody two factors: a term that is a function of the *RSS* and some penalty for adding extra parameters.
- The information criteria vary according to how stiff the penalty term is.
- The objective is to choose the number of parameters that minimises an information criterion.

Box–Jenkins Approach

- The three most popular criteria are Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (SBIC), and the Hannan-Quinn criterion (HQIC):

$$AIC = \ln(\hat{\sigma}^2) + 2k / T$$

$$SBIC = \ln(\hat{\sigma}^2) + \frac{k}{T}\ln T$$

$$HQIC = \ln(\hat{\sigma}^2) + \frac{2k}{T}\ln(\ln(T))$$

where $k = p + q + 1$, T = sample size. So we minimize IC s.t. $p \leq \bar{p}, q \leq \bar{q}$.

- Thus, *SBIC* embodies a stiffer penalty term than *AIC*.
- Which IC should be preferred if they suggest different model orders?
  - *AIC* is not consistent, and will typically pick "bigger" models;
  - *SBIC* is strongly consistent, though efficient only asymptotically.

If we have a large sample we use SBIC,
if we have a small sample we use AIC

University *of Ljubljana*

SEB SCHOOL OF ECONOMICS AND BUSINESS

Extensions to ARMA Models

- **ARMAX(p, q)** is a multivariate extension to ARMA models, where exogenous explanatory variables are added to the model specification.
- An ARMAX model is estimated analogously to classical ARMA models in Stata, stating explicitly the exogenous explanatory variables.
- **ARIMA(p, d, q)** is another extension to ARMA models, where the "I" stands for an "integrated" autoregressive moving average process.
- An integrated autoregressive moving average process is one with a characteristic root on the unit circle (non-stationary).
- Typically researchers difference the variable as necessary and then build an ARMA model on those differenced variables.
- However, an ARMA(p, q) model in the variable differenced d–times is equivalent to an ARIMA(p, d, q) model on the original data.

ARMA does not allow for non-stationary variable in the model, while ARIMA does.

EFMD EQUIS ACCREDITED

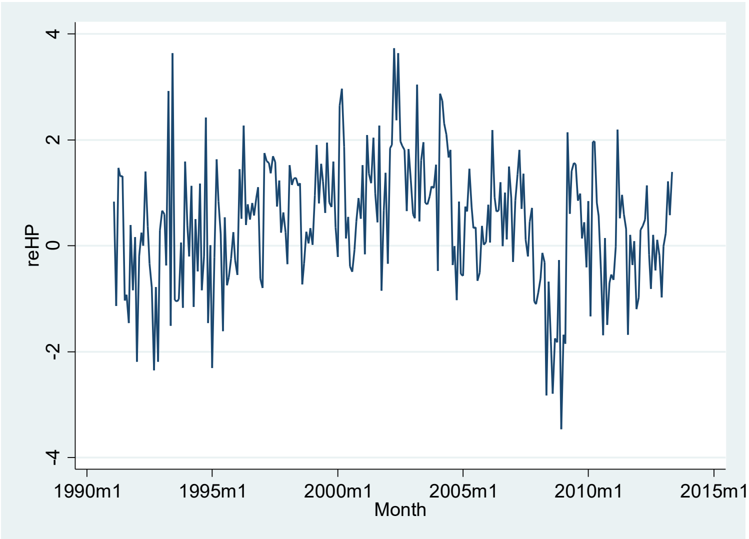
AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

```
. dfuller d.HP, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =       263

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.044            -3.459            -2.879            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Generating continuously compounded percentage returns (percentage changes) and repeating the ADF test page 41 in chapter 8

```
. generate reHP=100*(ln(HP/L.HP))
(1 missing value generated)

. twoway line reHP Month
```



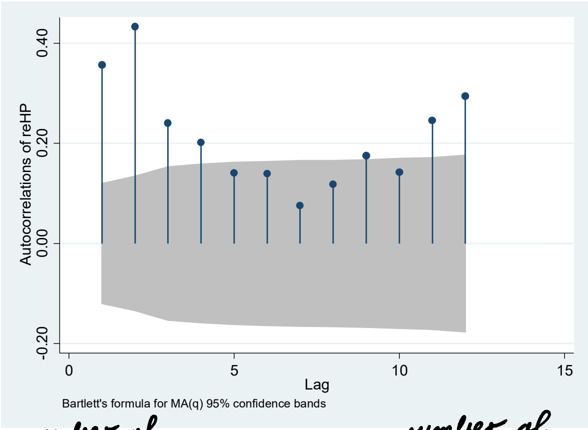

```
. dfuller reHP, lags(4)

Augmented Dickey-Fuller test for unit root         Number of obs   =       263

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.152            -3.459            -2.879            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

Constructing an ARMA model

```
. corrgram reHP, lags(12)

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.3566   0.3575   34.468  0.0000          |--                |--
2        0.4329   0.3504   85.444  0.0000          |---               |--
3        0.2407   0.0179   101.27  0.0000          |-                 |
```

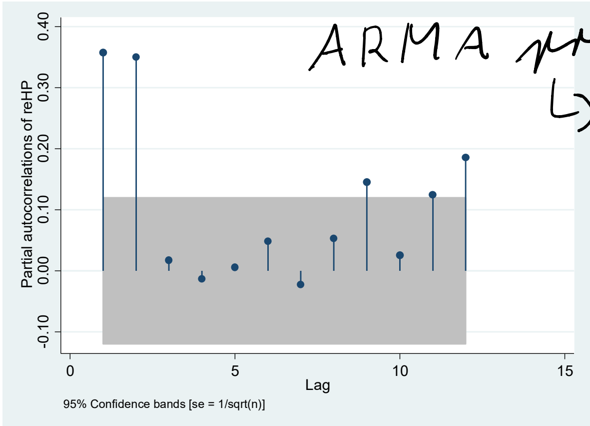
```
4          0.2018   -0.0132   112.43  0.0000          |-                  |
5          0.1406    0.0058   117.86  0.0000          |-                  |
6          0.1396    0.0486   123.24  0.0000          |-                  |
7          0.0755   -0.0225   124.82  0.0000          |                   |
8          0.1182    0.0533   128.71  0.0000          |                   |
9          0.1754    0.1457   137.31  0.0000          |-                  |-
10         0.1421    0.0260   142.97  0.0000          |-                  |
11         0.2457    0.1244   159.97  0.0000          |-                  |
12         0.2946    0.1855   184.51  0.0000          |--                 |-
```

. ac reHP, lags(12)



. pac reHP, lags(12)

AR(2) process or ARMA process
↳ we don't know yet



number of lags AR —— p d q – number of lags MA

Useful models for forecasting have up to 5 lags (not more)

```
. arima reHP, arima(0,0,0)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -427.80328
Iteration 1:   log likelihood = -427.80328

ARIMA regression

Sample:  1991m2 - 2013m5                        Number of obs     =        268
                                                Wald chi2(.)      =          .
Log likelihood = -427.8033                      Prob > chi2       =          .

------------------------------------------------------------------------------
             |                 OPG
        reHP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reHP         |
       _cons |   .4299142   .0732932     5.87   0.000     .2862623    .5735662
-------------+----------------------------------------------------------------
      /sigma |    1.19404   .0484286    24.66   0.000     1.099122    1.288958
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        268         .  -427.8033       2    859.6066   866.7885
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.

// We continue to estimate all possible ARMA models manually. //
```

```
// Alternatively, we can use the following program: //

. generate variable=reHP
(1 missing value generated)

. scalar pset=4
. scalar qset=4
```

We plug in the variable

```
. forvalues p=0/`=pset' {
  2. forvalues q=0/`=qset' {
  3. qui arima variable, arima(`p',0,`q')
  4. scalar aic_`p'_`q'=-2*e(ll)+2*e(rank)
  5. scalar bic_`p'_`q'=-2*e(ll)+e(rank)*ln(e(N))
  6. }
  7. }

. forvalues p=0/`=pset' {
  2. forvalues q=0/`=qset' {
  3. display "AIC(`p',`q') = "aic_`p'_`q'
  4. }
  5. }

AIC(0,0) = 859.60656
AIC(0,1) = 840.82476
AIC(0,2) = 804.35416
AIC(0,3) = 801.8014
AIC(0,4) = 800.01893
AIC(1,0) = 825.17453
AIC(1,1) = 802.16762
AIC(1,2) = 795.50644
AIC(1,3) = 794.0435
AIC(1,4) = 795.2866
AIC(2,0) = 792.09006
AIC(2,1) = 794.01043
AIC(2,2) = 795.95615
AIC(2,3) = 794.75456
AIC(2,4) = 796.10834
AIC(3,0) = 794.00444
AIC(3,1) = 793.38271
AIC(3,2) = 797.94062
AIC(3,3) = 796.4548
AIC(3,4) = 795.86876
AIC(4,0) = 795.95778
AIC(4,1) = 795.26527
AIC(4,2) = 791.60913
AIC(4,3) = 791.43349
AIC(4,4) = 795.39434
```

this one fits best
minimum

```
. forvalues p=0/`=pset' {
  2. forvalues q=0/`=qset' {
  3. display "SBIC(`p',`q') = "bic_`p'_`q'
  4. }
  5. }

SBIC(0,0) = 866.78854
SBIC(0,1) = 851.59773
SBIC(0,2) = 818.71811
SBIC(0,3) = 819.75633
SBIC(0,4) = 821.56485
SBIC(1,0) = 835.94749
SBIC(1,1) = 816.53157
SBIC(1,2) = 813.46137
SBIC(1,3) = 815.58942
SBIC(1,4) = 820.42351
SBIC(2,0) = 806.45401
```

this one fits best

```
SBIC(2,1) = 811.96537
SBIC(2,2) = 817.50207
SBIC(2,3) = 819.89146
SBIC(2,4) = 824.83624
SBIC(3,0) = 811.95937
SBIC(3,1) = 814.92863
SBIC(3,2) = 823.07753
SBIC(3,3) = 825.1827
SBIC(3,4) = 828.18765
SBIC(4,0) = 817.5037
SBIC(4,1) = 820.40218
SBIC(4,2) = 816.74604
SBIC(4,3) = 823.75237
SBIC(4,4) = 831.30421

// SBIC is consistent, but requires a large sample, which we seem to have (T = 268) //
```

that is why we use the model which has the lowest value of SBIC criterion (2,0)

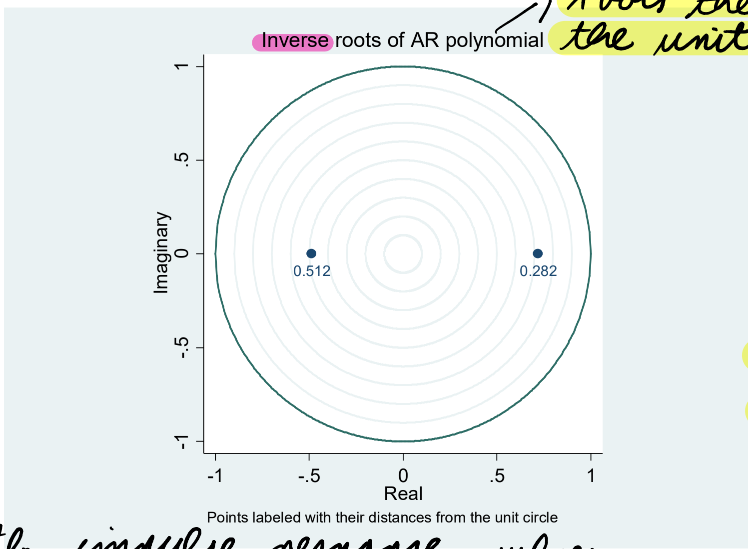
```
. arima reHP, ar(1/2)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -392.04832
Iteration 1:   log likelihood = -392.04526
Iteration 2:   log likelihood = -392.04505
Iteration 3:   log likelihood = -392.04503

ARIMA regression

Sample:  1991m2 - 2013m5                        Number of obs     =        268
                                                Wald chi2(2)      =      79.64
Log likelihood =  -392.045                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
        reHP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reHP         |
       _cons |   .4324219   .1542394     2.80   0.005     .1301183    .7347255
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .2299696   .0508839     4.52   0.000      .130239    .3297002
         L2. |   .3508321   .0488959     7.18   0.000     .2549978    .4466663
-------------+----------------------------------------------------------------
      /sigma |   1.044089   .0415722    25.12   0.000     .9626087    1.125569
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.

. arima reHP, arima(2,0,0)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -392.04832
Iteration 1:   log likelihood = -392.04526
Iteration 2:   log likelihood = -392.04505
Iteration 3:   log likelihood = -392.04503

ARIMA regression

Sample:  1991m2 - 2013m5                        Number of obs     =        268
                                                Wald chi2(2)      =      79.64
Log likelihood =  -392.045                      Prob > chi2       =     0.0000
```

test for testing statistical significance of the model as a whole

```
------------------------------------------------------------------------------
             |                 OPG
        reHP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reHP         |
       _cons |   .4324219   .1542394     2.80   0.005     .1301183    .7347255
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .2299696   .0508839     4.52   0.000      .130239    .3297002
         L2. |   .3508321   .0488959     7.18   0.000     .2549978    .4466663
-------------+----------------------------------------------------------------
      /sigma |   1.044089   .0415722    25.12   0.000     .9626087    1.125569
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

```
. estat aroots, dlabel

   Eigenvalue stability condition
  +----------------------------------------+
  |        Eigenvalue         |   Modulus  |
  |---------------------------+------------|
  |   .7183533                |    .718353 |
  |  -.4883837                |    .488384 |
  +----------------------------------------+
   All the eigenvalues lie inside the unit circle.
   AR parameters satisfy stability condition.
```

because they are inverse roots they need to be in the unit circle



the roots should be outside of the inner model

create a file with impulse response values

```
. irf create AR2, set(myirf, replace) step(12)
(file myirf.irf created)
(file myirf.irf now active)
(file myirf.irf updated)

. irf describe
Contains irf results from myirf.irf

       irfname | model    endogenous variables and order (*)
---------------+-----------------------------------------------------------------
           AR2 | arima    reHP
---------------------------------------------------------------------------------
  (*) order is relevant only when model is var
```

Example 1 continued

Ad g) PDF, pp. 8–10.

$u_t \sim N(0, \sigma_u^2)$

"sigma" is $\hat{\sigma}_u$

Impulse response (IR) function:

We apply a one-time unit increase (shock) to the disturbance term u_t and observe the effects over time on y_t.

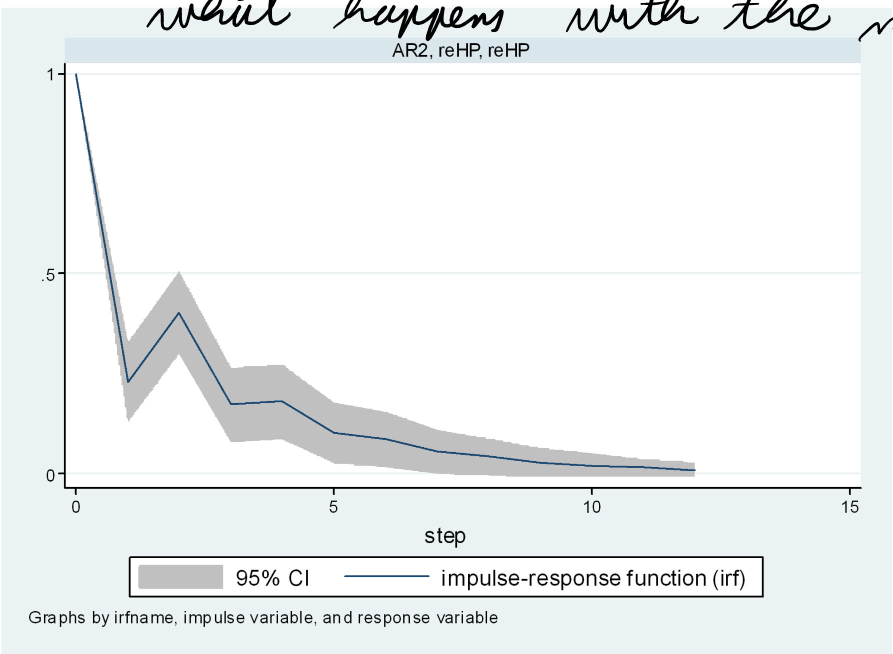
we apply a shock once and we check what happens

function IRF

the effect of the shock is supposed to die out if the model is stationary.

```
. irf graph irf
```

if you shock the model this is what happens with the model




```
. irf table irf
```

Results from AR2

step	(1) irf	(1) Lower	(1) Upper
0	1	1	1
1	.22997	.130239	.3297
2	.403718	.301934	.505502
3	.173524	.081296	.265751
4	.181542	.088293	.274792
5	.102627	.027752	.177502
6	.087292	.018908	.155676
7	.056079	.002009	.11015
8	.043521	-.002335	.089378
9	.029683	-.006318	.065684
10	.022095	-.007103	.051293
11	.015495	-.007249	.038239
12	.011315	-.00664	.02927

95% lower and upper bounds reported
(1) irfname = AR2, impulse = reHP, and response = reHP

Forecasting using an ARMA model

→ it is gonna estimate the same AR(2) model with fewer time periods in the past

```
. arima reHP if Month<=tm(2010m12), arima(2,0,0)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -355.03462
Iteration 1:   log likelihood = -355.02874
Iteration 2:   log likelihood = -355.02815
Iteration 3:   log likelihood = -355.02809
Iteration 4:   log likelihood = -355.02808
```

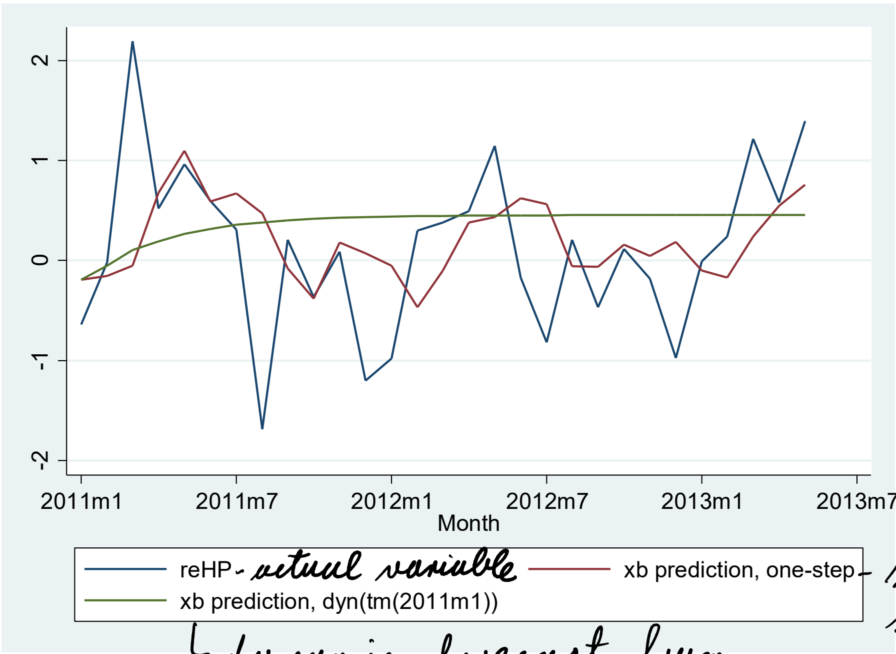
```
ARIMA regression

Sample:  1991m2 - 2010m12                       Number of obs     =        239
                                                Wald chi2(2)      =      73.84
Log likelihood = -355.0281                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
        reHP |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reHP         |
       _cons |   .4540309   .1702274     2.67   0.008     .1203913    .7876704
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .2255672   .0535307     4.21   0.000      .120649    .3304854
         L2. |   .3631374   .0516035     7.04   0.000     .2619963    .4642784
-------------+----------------------------------------------------------------
      /sigma |    1.06785   .0460784    23.17   0.000     .9775385    1.158163
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

```
. predict reHP_fstatic, xb
. predict reHP_fdynamic, xb dynamic(tm(2011m1))
```

static forecast

dynamic forecast

```
. twoway (tsline reHP) (tsline reHP_fstatic) (tsline reHP_fdynamic) if
  Month>tm(2010m12)
```



static forecast from AR(2) model

dynamic forecast from AR(2) model

Static forecast should mimic the actual function.
Dynamic forecast should follow the equilibrium
It shows long term movement.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

8.3 Vector Autoregression

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Basic definitions

- A vector autoregressive or vector autoregression model (VAR) is a **multivariate econometric model** used to capture linear interdependencies among multiple time series.
- Each time-series variable has an equation explaining its evolution based on its own lags and lags of the other model variables.
- Simplest case is a bivariate VAR(k) model:

2 equations

$$y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \ldots + \beta_{1k} y_{1t-k} + \alpha_{11} y_{2t-1} + \ldots + \alpha_{1k} y_{2t-k} + u_{1t}$$
$$y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \ldots + \beta_{2k} y_{2t-k} + \alpha_{21} y_{1t-1} + \ldots + \alpha_{2k} y_{1t-k} + u_{2t}$$

where u_{it} is an IID disturbance term with $E(u_{it})=0$, $i=1,2$ and $E(u_{1t}\, u_{2t})=0$.

- The analysis could be extended to a g–variate VAR(k) model, so that there are g variables and thus g equations.

Basic definitions

- One important feature of VAR models is the compactness with which we can write the notation. For example, consider the case where $k = 1$.

- We can write this as:
$$y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + u_{1t}$$
$$y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + u_{2t}$$

or

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

or even more compactly as: – matrix notation

$$\underset{g\times 1}{\mathbf{y}_t} = \underset{g\times 1}{\boldsymbol{\beta}_0} + \underset{g\times g}{\boldsymbol{\beta}_1} \underset{g\times 1}{\mathbf{y}_{t-1}} + \underset{g\times 1}{\mathbf{u}_t}$$

Basic definitions

- This model can be extended to the case where there are k lags of each variable in each equation:

$$\underset{g\times 1}{\mathbf{y}_t} = \underset{g\times 1}{\beta_0} + \underset{g\times g}{\beta_1}\,\underset{g\times 1}{\mathbf{y}_{t-1}} + \underset{g\times g}{\beta_2}\,\underset{g\times 1}{\mathbf{y}_{t-2}} + \ldots + \underset{g\times g}{\beta_k}\,\underset{g\times 1}{\mathbf{y}_{t-k}} + \underset{g\times 1}{\mathbf{u}_t}$$

- These formulations represent the so-called **standard form VAR model**.
- We can also extend this to the case where the model includes first difference terms and cointegrating relationships (the so-called vector error-correction model – VECM, examined later).

Comparison to structural equations models

Advantages of VAR modelling:

- Do not need to specify which variables are endogenous or exogenous (the latter is often violated); in a VAR setting all variables are endogenous.
- Allows the value of a variable to depend on more than just its own lags or combinations of white noise terms, so it offers a very rich structure.
- Provided that there are no contemporaneous terms on the right hand side of the equations (all RHS variables are predetermined), we can apply the least squares estimator separately on each equation.
- VAR forecasts are often better than those of "traditional structural" models, as they contain less ad-hoc identification restrictions (e.g. on exogeneity).
- We can examine how a shock to one variable affects all other variables and how important one variable is in affecting movements of the other variables.

Comparison to structural equations models

Problems with VAR models: / they are ad-hoc

- VAR models are atheoretical, as they use little theoretical information about the relationship between the variables.
- The is no single approach to decide the appropriate lag length of a VAR model.
- There are many parameters to be estimated, thus affecting efficiency. If we have g equations for g variables and we have k lags of each of the variables in each equation, we have to estimate $g + kg^2$ parameters (e.g. when $g = 3$ and $k = 3$, we have 30 parameters).
- We need to ensure all variables of the standard VAR model are stationary in order to have valid statistical inference (stability condition).
- As VAR model is not a structural model, it is difficult to interpret the regression coefficients.

Choosing the optimal lag length

- Theory offers very little insight into this issue.
- Two possible approaches: cross-equation restrictions and information criteria.

Cross-equation restrictions:

- In the spirit of (unrestricted) VAR modelling, each equation should have the same lag length.
- General-to-simple approach: suppose that a bivariate VAR(8) model estimated using quarterly data has 8 lags of the two variables in each equation, and we want to examine a restriction that the coefficients on lags 5 through 8 are jointly zero. This can be done using a likelihood ratio test.
- Denote the variance-covariance matrix of residuals (given by $\hat{u}\hat{u}'/T$), as $\hat{\Sigma}$. The likelihood ratio test for this joint hypothesis is given by:

$$\boxed{LR = T\left[\log\left|\hat{\Sigma}_r\right| - \log\left|\hat{\Sigma}_u\right|\right]}$$

Lagrange test statistic

Choosing the optimal lag length

- $\hat{\Sigma}_r$ is the variance-covariance matrix of the residuals for the restricted model (with 4 lags), $\hat{\Sigma}_u$ is the variance-covariance matrix of residuals for the unrestricted VAR (with 8 lags), and T is the sample size.
- The test statistic is asymptotically distributed as a χ^2 with degrees of freedom equal to the total number of restrictions. In the VAR case above, we are restricting 4 lags of two variables in each of the two equations, thus we have a total of $4 \cdot 2 \cdot 2 = 16$ restrictions.
- In the general case, where we have a VAR with g equations, and we want to impose the restriction that the last q lags have zero coefficients, there would be g^2q restrictions altogether.
- Disadvantages: conducting the LR-test is cumbersome and requires a normality assumption for the disturbances.

Choosing the optimal lag length

the other approach

Information criteria:

Multivariate versions of the information criteria are required. These can be defined as:

$$MAIC = \ln\left|\hat{\Sigma}\right| + 2k'/T$$

$$MSBIC = \ln\left|\hat{\Sigma}\right| + \frac{k'}{T}\ln(T)$$

$$MHQIC = \ln\left|\hat{\Sigma}\right| + \frac{2k'}{T}\ln(\ln(T))$$

where all notation is as above and k' is the total number of regressors in all equations, which will be equal to $g^2k + g$ for g equations, each with k lags of the g variables, plus a constant term in each equation. The values of the information criteria are constructed for 0, 1, … $\bar{k}$ lags (i.e. up to some pre-specified maximum $\bar{k}$).

Structural versus standard form VARs

- So far, we have assumed the VAR model is of the form:

there are no contemporaneous effects => there is only lagged effect

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + u_{1t}$$
$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + u_{2t}$$

- But what if the equations had a contemporaneous feedback term:

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + \boxed{\alpha_{12}y_{2t}} + u_{1t}$$
$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + \boxed{\alpha_{22}y_{1t}} + u_{2t}$$

we have simultaneity α_{12} affects y_{1t} and y_{1t} affects y_{2t}

- We can write this as: to make it solvable one value has to be 0

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{12} & 0 \\ 0 & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{2t} \\ y_{1t} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

- This VAR model is the so-called structural or primitive form.

Structural versus standard form VARs

- We can take the contemporaneous terms over to the LHS and write:

$$\begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{22} & 1 \end{pmatrix}\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix}\begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

or

$$\underbrace{\mathbf{A}^{-1} \cdot / \mathbf{A}\, \mathbf{y}_t}_{\mathbf{I} \cdot \mathbf{y}_t = \mathbf{y}_t} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{y}_{t\text{-}1} + \mathbf{u}_t$$

- We can then pre-multiply both sides by $\mathbf{A}^{-1}$ to get:

$$\mathbf{y}_t = \mathbf{A}^{-1}\boldsymbol{\beta}_0 + \mathbf{A}^{-1}\boldsymbol{\beta}_1 \mathbf{y}_{t\text{-}1} + \mathbf{A}^{-1}\mathbf{u}_t$$

we cannot distinguish between A and β_0

or

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{y}_{t\text{-}1} + \mathbf{v}_t$$

↳ identification issue

- This is known as the standard or reduced form VAR, which was already discussed before and can be estimated by using OLS.

EFMD EQUIS ACCREDITED
AACSB ACCREDITED
ASSOCIATION OF MBAS ACCREDITED

Structural versus standard form VARs

- The regression coefficients of a structural or primitive form VAR model have more firm theoretical interpretations, but this comes at a cost.
- First of all, due to simultaneity between y_1 and y_2, such models are not estimable by using the OLS estimator.
- In addition, the structural VAR model cannot be identified without *a priori* restrictions, as different structural models give rise to the same reduced form.
- Most commonly, the identification restrictions are set by imposing one of the coefficients on the contemporaneous terms (either α_{12} or α_{22}) to zero.
- This choice is ideally made on theoretical grounds (i.e. considering whether $\mathbf{y}_1$ should primarily affect $\mathbf{y}_2$ or the other way around) and is non-testable within the VAR model.

↳ we have to determine this upfront

Interpretation of VAR models

- The focus of a VAR model is usually on forecasting, but may also be on interpretation.
- Due to its atheoretical nature, VAR models exhibit difficulties in interpretation of the regression coefficients (e.g. some lagged variables may have coefficients that change sign across the lags).
- In order to alleviate the interpretation, three sets of statistics are usually constructed for an estimated VAR model:
  - Block significance (Granger) tests;
  - Impulse responses and
  - Variance decompositions.

Block significance and causality tests

- It is likely that, when a VAR includes many lags of variables, it will be difficult to see which sets of variables have significant effects on each dependent variable and which do not.
- For illustration, consider the following bivariate VAR(3):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{10} \\ \alpha_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}\begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}\begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}\begin{pmatrix} y_{1t-3} \\ y_{2t-3} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

- This VAR could be written out to express the individual equations as:

$$y_{1t} = \alpha_{10} + \beta_{11} y_{1t-1} + \beta_{12} y_{2t-1} + \gamma_{11} y_{1t-2} + \gamma_{12} y_{2t-2} + \delta_{11} y_{1t-3} + \delta_{12} y_{2t-3} + u_{1t}$$
$$y_{2t} = \alpha_{20} + \beta_{21} y_{1t-1} + \beta_{22} y_{2t-1} + \gamma_{21} y_{1t-2} + \gamma_{22} y_{2t-2} + \delta_{21} y_{1t-3} + \delta_{22} y_{2t-3} + u_{2t}$$

lags of y_1

Block significance and causality tests

- We might be interested in testing the following (sets of) hypotheses, and their implied restrictions on the parameter matrices:

Hypothesis	Implied Restriction
1. Lags of y_1 do not explain current y_2	$\beta_{21} = 0$ and $\gamma_{21} = 0$ and $\delta_{21} = 0$
2. Lags of y_1 do not explain current y_1	$\beta_{11} = 0$ and $\gamma_{11} = 0$ and $\delta_{11} = 0$
3. Lags of y_2 do not explain current y_1	$\beta_{12} = 0$ and $\gamma_{12} = 0$ and $\delta_{12} = 0$
4. Lags of y_2 do not explain current y_2	$\beta_{22} = 0$ and $\gamma_{22} = 0$ and $\delta_{22} = 0$

↳ we test several coefficients to be equal to zero
⇒ we use F-test or Lagrange multiplier test

Block significance and causality tests

- Each of these four joint hypotheses can be tested within the F-test framework, since each set of restrictions contains only parameters drawn from one equation.
- These tests are often referred to as Granger causality tests.
- Granger causality tests seek to answer questions such as: "Do changes in y_1 cause changes in y_2?"
  - If y_1 causes y_2, lags of y_1 should be significant in the equation for y_2. If this is the case, we say that y_1 "Granger-causes" y_2.
  - If y_2 causes y_1, lags of y_2 should be significant in the equation for y_1. If this is the case, we say that y_2 "Granger-causes" y_1.
  - If both sets of lags are statistically significant, we say that there is bi-directional "Granger causality" between y_1 and y_2.
  - If neither set of lags is statistically significant in the equation for the other variable, then there is no "Granger causality" between y_1 and y_2.

Block significance and causality tests

- When the analysis involves more than two variables, this is extended to the so-called block significance or block Granger causality.
- However, block significance tests cannot, by construction, explain the signs of relationships or the duration of effects in time.
- This is only possible by examining impulse responses and variance decompositions of a VAR model.

Impulse responses

↳ you impose a shock to the disturbance term and see how the dependent variable changes => we do this with every equation

- Impulse responses trace out the responsiveness of the dependent variables in the VAR model to shocks to a disturbance term.
- A **one-time unit increase** (shock) is applied separately to each variable (in particular, to the disturbance term of the respective equation) and its effects over time are noted in the VAR system.
- Thus, for g variables in a system g^2 impulse responses are generated.
- If the system is stable, the shock should gradually die away.
- Consider for example a simple bivariate VAR(1) model:

$$y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + u_{1t}$$
$$y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + u_{2t}$$

- A change in u_{1t} will immediately change y_1. It will change y_2 and also y_1 during the next time period etc.

Impulse responses

- We can examine how long and to what degree a shock to a given equation affects all of the variables in the system.
- Consider the above simple bivariate VAR(1) model with the following parameters:

$$y_t = A_1 y_{t-1} + u_t \quad \text{where} \quad A_1 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix}$$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Impulse responses

- The first couple of impulse responses are as follows:

we impose a unit shock to the first equation

$$y_0 = \begin{bmatrix} u_{10} \\ u_{20} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad y_0 = \begin{bmatrix} u_{10} \\ u_{20} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$y_1 = A_1 y_0 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} \qquad y_1 = A_1 y_0 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix}$$

$$y_2 = A_1 y_1 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0 \end{bmatrix} \qquad y_2 = A_1 y_1 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.21 \\ 0.04 \end{bmatrix}$$

- Impulse responses are often presented graphically for a given number of time periods as impulse response functions (IRFs).
- If normality of VAR residuals is ensured, asymptotic confidence intervals can be computed for impulse responses.

Variance decompositions

- Variance decompositions offer a slightly different method of examining the VAR dynamics.
- They give the decomposition of variability in the dependent variables that is due to their "own" shocks and shocks to the other variables.
- Namely, a shock to a variable will directly affect that variable, but it will also be transmitted to all other variables in the system through the dynamic structure of the VAR model.
- This is calculated by determining how much of the forecast-error variance for each variable is explained by shocks to each explanatory variable for a given time period.

Variance decompositions

- The variance decomposition gives information about the relative importance of each shock to the variables in the VAR model.
- In practice, own series shocks usually explain most of the forecast-error variance of the series.
- Variance decompositions are often presented graphically for a given number of time periods as forecast-error variance decompositions (FEVDs).
- If normality of VAR residuals is ensured, asymptotic confidence intervals can be computed for variance decompositions.

The ordering of the variables

- For calculating impulse responses and variance decompositions, the ordering of the variables is important.
- Namely, the impulse responses refer to a unit shock to the disturbances of only one VAR equation at a time, whereas the disturbance terms of all other equations are held constant.
- This is coherent, as we assumed that the VAR disturbance terms were statistically independent of one another.
- However, this is generally not the case in practice. The disturbance terms will typically be correlated to some degree.
- Therefore, the notion of examining the effect of the disturbances separately has little meaning (it misrepresents the system dynamics). We say that the disturbances have a common component that cannot be associated with a single variable.

we assume that disturbance terms are not correlated

The ordering of the variables

To resolve the issue of correlation between the disturbance terms

- What is done is to orthogonalise the disturbances, most often by applying the **Cholesky decomposition**. This gives us the orthogonalised impulse response functions (OIRFs).
- Similarly to IRFs, a **one-time unit standard deviation increase** (shock) is applied separately to each variable (to the disturbance term of the respective equation) and its effects over time are noted in the VAR system.
- In a bivariate VAR, the orthogonalisation is done by attributing all of the effect of the common component to the first of the two variables in the VAR model.
- In the general case of a g–variate VAR, the calculations are more complex, but the basic idea is that the first variable is the only one with a potential immediate impact on all other g – 1 variables, the second variable may have (all 4 variables) immediate impact just on the remaining g – 2 variables etc. (remaining 3) The last variable cannot affect other variables.

We do the orthogonalization in order to make the disturbance terms independent

The ordering of the variables

- The higher the correlation among the residuals from estimated equations, the more the variable ordering will be important.
- This choice of ordering is ideally made on theoretical grounds (i.e. considering whether movement of a certain variable precedes that of another or not).
- Empirically, sensitivity analysis can be performed by assuming different orderings and comparing the resulting OIRFs and FEVDs.

Orthogonalization process

- The correlation of VAR disturbance terms that appears in practice is being accounted for by orthogonalization of disturbances.
- In principle, orthogonalization can be done in at least two ways:
  - Orthogonalization matrix is constructed as the **Cholesky decomposition** of the estimated variance-covariance matrix (already described), which imposes a recursive structure on the model. The ordering of the recursive structure is that in which the endogenous variables appear in the VAR estimation. Such a model is also called *recursive form* VAR.
  - Alternativelly, **restrictions** are placed on the orthogonalization matrix, either in terms of short-run restrictions on the contemporaneous covariances between shocks or in terms of restrictions on the long-run accumulated effects of the shocks. This gives us the *structural form* VAR (**SVAR**, already discussed).

Short-run SVAR model

A **short-run SVAR model** without exogenous variables is derived from the standard form VAR (with lag operator notation, L):

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{v}_t \\
\mathbf{y}_t - \mathbf{A}_1\mathbf{y}_{t-1} - \mathbf{A}_2\mathbf{y}_{t-2} - \ldots - \mathbf{A}_p\mathbf{y}_{t-p} &= \mathbf{v}_t \\
\mathbf{y}_t - \mathbf{A}_1 L\mathbf{y}_t - \mathbf{A}_2 L^2\mathbf{y}_t - \ldots - \mathbf{A}_p L^p\mathbf{y}_t &= \mathbf{v}_t \\
\left(\mathbf{I}_g - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \ldots - \mathbf{A}_p L^p\right)\mathbf{y}_t &= \mathbf{v}_t
\end{aligned}
$$

L - lag operator notation

which we premultiply by the matrix $\mathbf{A}$ to impose orthogonalization:

$$
\mathbf{A}\left(\mathbf{I}_g - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \ldots - \mathbf{A}_p L^p\right)\mathbf{y}_t = \mathbf{A}\mathbf{v}_t = \mathbf{B}\tilde{\mathbf{v}}_t
$$

The vector $\mathbf{v}_t$ refers to the original disturbances in the model (with covariance matrix $\mathbf{\Sigma}$), while the vector $\tilde{\mathbf{v}}_t$ represents the orthogonalized disturbances (with covariance matrix $\mathbf{I}_g$).

Short-run SVAR model

The form of the **A** matrix imposes the recursive structure, while the (diagonal) matrix **B** orthogonalizes the effects of the disturbances.

Identification is obtained by placing restrictions on the matrices **A** and **B**, which are assumed to be nonsingular. The orthgonalization matrix $\mathbf{A}^{-1}\mathbf{B}$:

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{v}_t = \mathbf{A}^{-1}\mathbf{B}\tilde{\mathbf{v}}_t$$
$$\mathbf{v}_t = \mathbf{A}^{-1}\mathbf{B}\tilde{\mathbf{v}}_t$$

any variance covariance matrix is symmetric

is then related to the variance-covariance matrix $\mathbf{\Sigma}$.

As there are $g(g+1)/2$ free parameters in $\mathbf{\Sigma}$, given its symmetric nature, only that many parameters may be estimated in the **A** and **B** matrices. As there are $2g^2$ parameters altogether in **A** and **B**, the order condition for identification requires that (at least) $2g^2 - g(g+1)/2$ restrictions be placed on the elements of these matrices.

EFMD EQUIS ACCREDITED
AACSB ACCREDITED
AMBA ACCREDITED

Vector autoregression

Short-run SVAR model

Slide 69

For $g = 3$: , in this case we impose 18 restrictions

- at most $3 \cdot 4 / 2 = 6$ parameters are estimated in matrices A and B combined;
- at least $2 \cdot 3^2 - 6 = 12$ restrictions are needed.

Long-run SVAR model

A short-run SVAR model without exogenous variables can also be written as:

$$\mathbf{A}\left(\mathbf{I}_g - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \ldots - \mathbf{A}_p L^p\right)\mathbf{y}_t = \mathbf{A}\overline{\mathbf{A}}\mathbf{y}_t = \mathbf{B}\tilde{\mathbf{v}}_t$$

where $\overline{\mathbf{A}}$ is the parenthesized expression. If we set $\mathbf{A} = \mathbf{I}$, we can write this equation as the **long-run SVAR model**:

$$\overline{\mathbf{A}}\mathbf{y}_t = \mathbf{B}\tilde{\mathbf{v}}_t$$
$$\overline{\mathbf{A}}^{-1}\overline{\mathbf{A}}\mathbf{y}_t = \overline{\mathbf{A}}^{-1}\mathbf{B}\tilde{\mathbf{v}}_t$$
$$\mathbf{y}_t = \overline{\mathbf{A}}^{-1}\mathbf{B}\tilde{\mathbf{v}}_t$$
$$\mathbf{y}_t = \mathbf{C}\tilde{\mathbf{v}}_t$$

When the Δ (delta) is zero we are in the long run. => change in the long run is 0.

In a long-run SVAR, constraints are placed on elements of the **C** matrix. These constraints are often exclusion restrictions. For instance, constraining $\mathbf{C}_{1,2} = 0$ forces the long-run response of the first variable to a shock to (the equation for) the second variable to zero.

the response of the first variable to a shock is equal to zero

$$\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

These values are the same $\Rightarrow$ because the matrix is symmetrical

Long-run SVAR model

Again, as there are $g(g + 1)/2$ free parameters in $\boldsymbol{\Sigma}$, only that many parameters may be estimated in the **C** matrix. As there are g^2 parameters in **C**, the order condition for identification requires that (at least) $g^2 - g(g + 1)/2$ restrictions be placed on the elements of the matrix.

Two limitations have to be emphasized related to hypothesis testing:

- We cannot place constraints on the elements of matrix **A** in terms of the elements of matrix **B**, or vice versa (at least not without additional constraint matrices);
- We cannot mix short-run constraints (those on matrices **A** and **B**) and long-run constraints (those on the matrix **C**).

↳ they need to be done seperately

Long-run SVAR model Slide 71

For $g=3$:

- at most $3 \cdot 4/2 = 6$ parameters are estimated in matrix C.
- at least $3^2 - 6 = 3$ restrictions are needed.

Example 2: We will examine whether there are lead-lag relationships between the returns to three exchange rates against the US dollar: the euro, the British pound and the Japanese yen. The data are daily and run from 7 July 2002 to 6 June 2013, giving a total of 3,988 observations. The data are contained in Stata data file `currencies.dta`. The programming code is given in Stata Do file `currencies-commands-l18.do`.

a) Inspect the three time series visually. Do they look stationary? Perform the Dickey-Fuller GLS test (DFGLS) on each of the three time series. What do you find?
b) Construct a set of continuously compounded percentage returns (percentage changes). Check for stationarity of the newly generated time series. What do you find?
c) Now, we will focus on a vector autoregressive (VAR) model based on the returns to three exchange rates. Choose the optimal lag length based on multivariate information criteria and estimate the suggested VAR model. Interpret the Stata output, but keep in mind that because the construction of these models is usually not based on economic or financial theory, it is often best not to interpret the individual parameter estimates, but rather to examine the plausibility of the model as a whole, and to determine whether it describes the data well and produces accurate forecasts.
d) After fitting a VAR model, one hypothesis of interest is that all the variables at a given lag are jointly zero. In order to test this hypothesis, perform the Wald lag-exclusion tests for different lag orders. What do you find?
e) Check for stability of the chosen model by examining roots of the companion matrix (whether the eigenvalues lie inside the unit circle). What do you find?
f) Check for residual autocorrelation of at least the first two lag orders with an appropriate Lagrange-multiplier (LM) test. What do you find?
g) Run Granger causality Wald tests and figure out whether the "excluded" variable Granger-causes the "equation" variable. Do you find evidence of lead-lag interactions between the series? Which exchange rate reacts the fastest to market information?
h) Obtain the impulse responses (IRs), orthogonalised impulse responses (OIRs) and forecast-error variance decompositions (FEVDs) for the estimated model. Interpret the results. Could these results have been expected based on previous results?
i) Remember that the ordering of variables has an effect on the impulse responses and variance decompositions, and when, as in this case, theory does not suggest an obvious ordering of the series, sensitivity analysis should be undertaken. Reverse the ordering of variables and compare the new results with the initial ones. What do you find?
j) Finally, employ the estimated VAR model in order to compute dynamic forecasts of the returns to three exchange rates for the following 20 days. Present the forecasts graphically, together with 95% confidence intervals.

Computer printout of the results in Stata:

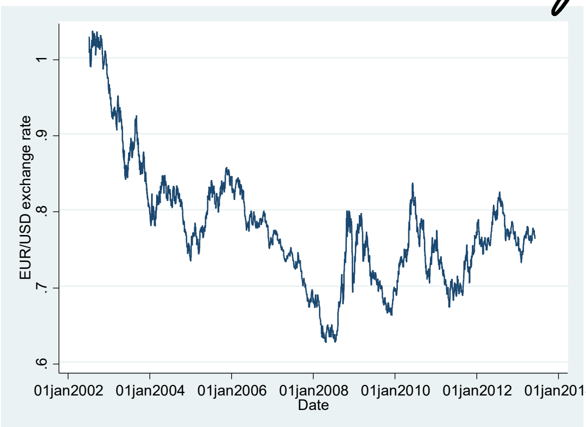
Data exploration

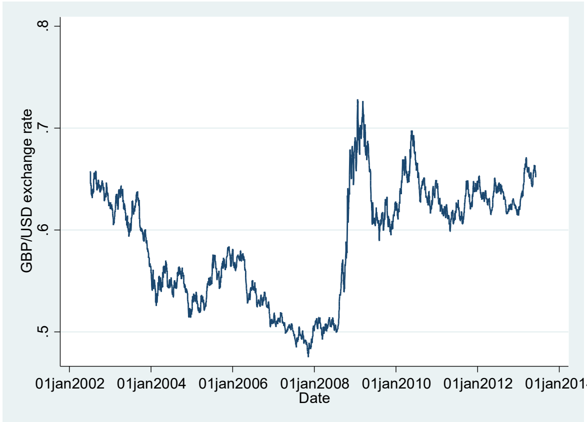
```
. tsset Date
        time variable:  Date, 07jul2002 to 06jun2013
                delta:  1 day
```

```
. twoway line EUR Date
```

it isn't stationary

```
. twoway line GBP Date
```






```
. twoway line JPY Date
```

this one isn't stationary either

Performing the Dickey-Fuller GLS test

```
. dfgls EUR, maxlag(5)

DF-GLS for EUR                                          Number of obs =   3982

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    5            -0.985           -3.480            -2.841            -2.554
    4            -0.960           -3.480            -2.842            -2.554
    3            -1.000           -3.480            -2.842            -2.554
    2            -1.016           -3.480            -2.842            -2.554
    1            -1.026           -3.480            -2.843            -2.555

Opt Lag (Ng-Perron seq t) =  1 with RMSE  .0037332
Min SC   = -11.17684 at lag  1 with RMSE  .0037332
Min MAIC = -11.17997 at lag  1 with RMSE  .0037332
```

we don't reject Ho
(we don't reject non stationarity

both information criterions
suggest use of 1 lag

```
. dfgls GBP, maxlag(5)

DF-GLS for GBP                                          Number of obs =   3982

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    5            -1.185           -3.480            -2.841            -2.554
    4            -1.204           -3.480            -2.842            -2.554
```

```
     3           -1.237           -3.480            -2.842            -2.554
     2           -1.223           -3.480            -2.842            -2.554
     1           -1.306           -3.480            -2.843            -2.555

Opt Lag (Ng-Perron seq t) =  2 with RMSE   .002617
Min SC    = -11.88519 at lag  2 with RMSE   .002617
Min MAIC = -11.88968 at lag  2 with RMSE   .002617
```

We can't reject H_0 at $\alpha = 0,05$.

```
. dfgls JPY, maxlag(5)

DF-GLS for JPY                                        Number of obs =   3982

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic       Value             Value             Value
------------------------------------------------------------------------------
     5           -1.625           -3.480            -2.841            -2.554
     4           -1.644           -3.480            -2.842            -2.554
     3           -1.771           -3.480            -2.842            -2.554
     2           -1.788           -3.480            -2.842            -2.554
     1           -1.764           -3.480            -2.843            -2.555

Opt Lag (Ng-Perron seq t) =  4 with RMSE  .4787147
Min SC    =  -1.46636 at lag  1 with RMSE   .47938
Min MAIC = -1.469925 at lag  4 with RMSE  .4787147
```

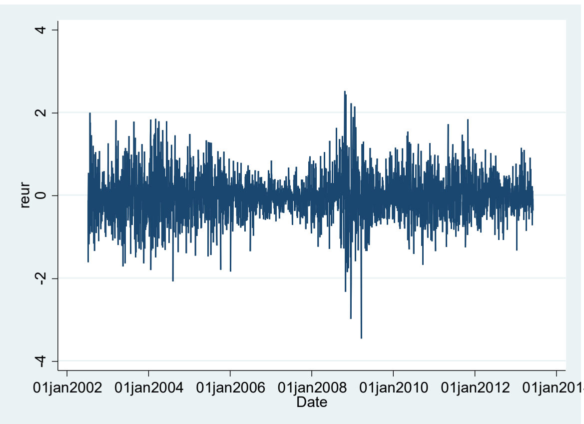
We can't reject H_0.

we use 1 lag because we have a large sample

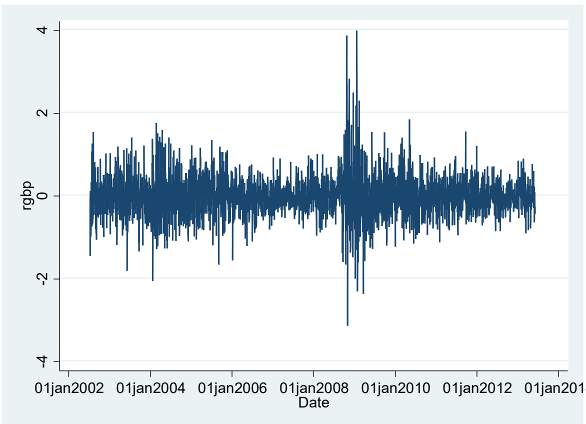
Generating continuously compounded percentage returns (percentage changes) and repeating the Dickey-Fuller GLS test

```
. generate reur=100*(ln(EUR/L.EUR))
(1 missing value generated)

. generate rgbp=100*(ln(GBP/L.GBP))
(1 missing value generated)

. generate rjpy=100*(ln(JPY/L.JPY))
(1 missing value generated)
```

. twoway line reur Date



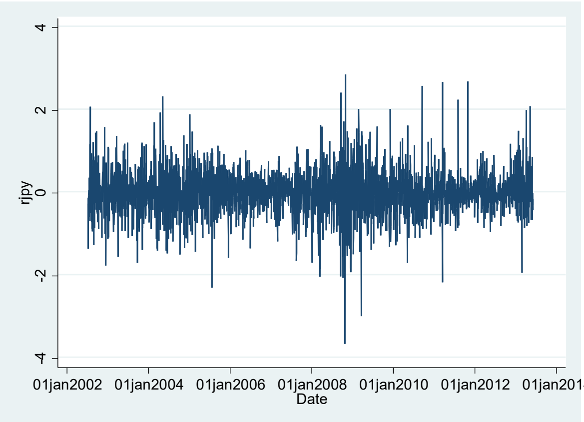
. twoway line rgbp Date



there is no trend
it looks better

. **twoway line rjpy Date**

there is no trend



. **dfgls reur, maxlag(5)**

```
DF-GLS for reur                                          Number of obs =   3981

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    5           -24.487           -3.480            -2.841            -2.554
    4           -26.944           -3.480            -2.842            -2.554
    3           -30.815           -3.480            -2.842            -2.554
    2           -34.634           -3.480            -2.842            -2.554
    1           -41.158           -3.480            -2.843            -2.555

Opt Lag (Ng-Perron seq t) =  4 with RMSE  .4722431
Min SC   = -1.495368 at lag  1 with RMSE  .4724769
Min MAIC = -.0863627 at lag  5 with RMSE  .4722112
```

We reject Ho

. **dfgls rgbp, maxlag(5)**

```
DF-GLS for rgbp                                          Number of obs =   3981

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    5           -25.393           -3.480            -2.841            -2.554
    4           -27.246           -3.480            -2.842            -2.554
    3           -30.256           -3.480            -2.842            -2.554
    2           -33.928           -3.480            -2.842            -2.554
    1           -40.914           -3.480            -2.843            -2.555

Opt Lag (Ng-Perron seq t) =  1 with RMSE  .4333588
Min SC   = -1.668214 at lag  1 with RMSE  .4333588
Min MAIC = -.3197095 at lag  2 with RMSE  .4333307
```

We reject Ho.

. **dfgls rjpy, maxlag(5)**

```
DF-GLS for rjpy                                          Number of obs =   3981

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    5           -24.263           -3.480            -2.841            -2.554
    4           -27.220           -3.480            -2.842            -2.554
    3           -30.605           -3.480            -2.842            -2.554
    2           -33.610           -3.480            -2.842            -2.554
    1           -40.031           -3.480            -2.843            -2.555
```

We reject Ho.

```
Opt Lag (Ng-Perron seq t) =  5 with RMSE  .4701428
Min SC   = -1.503243 at lag  1 with RMSE  .4706203
Min MAIC = -.1675009 at lag  5 with RMSE  .4701428
```

Estimating a VAR model

HQIC - Hannan Quinn information criterion

↳ mix between AIC and SBIC

```
. varsoc reur rgbp rjpy, maxlag(10)

   Selection-order criteria
   Sample:  18jul2002 - 06jun2013                   Number of obs      =      3977
  +---------------------------------------------------------------------------+
  |lag |    LL        LR       df     p       FPE       AIC      HQIC      SBIC    |
  |----+----------------------------------------------------------------------|
  |  0 | -6324.33                            .004836   3.18196   3.18364   3.18671  |
  |  1 | -6060.26  528.13     9  0.000  .004254   3.05369   3.06042   3.07266* |
  |  2 | -6034.87  50.784     9  0.000  .004219*  3.04545*  3.05722*  3.07865  |
  |  3 | -6030.96  7.8286     9  0.552   .00423     3.048    3.06482   3.09544  |
  |  4 | -6022.94   16.04     9  0.066  .004232    3.0485   3.07036   3.11016  |
  |  5 | -6015.11  15.655     9  0.074  .004234   3.04909     3.076   3.12498  |
  |  6 | -6009.17  11.881     9  0.220  .004241   3.05063   3.08258   3.14075  |
  |  7 | -6000.17  17.998*    9  0.035  .004241   3.05063   3.08763   3.15498  |
  |  8 | -5992.97  14.408     9  0.109  .004245   3.05153   3.09358   3.17012  |
  |  9 | -5988.13  9.6673     9  0.378  .004254   3.05362   3.10072   3.18644  |
  | 10 | -5984.25  7.7658     9  0.558  .004264    3.0562   3.10834   3.20325  |
  +---------------------------------------------------------------------------+
   Endogenous:  reur rgbp rjpy
    Exogenous:  _cons
```

when AIC and SBIC recommend different values we take the HQIC recommendation

```
. var reur rgbp rjpy, lags(1/2)

Vector autoregression

Sample:  10jul2002 - 06jun2013                   Number of obs     =      3,985
Log likelihood =   -6043.54                      AIC               =   3.043684
FPE            =   .0042115                      HQIC              =   3.055437
Det(Sigma_ml)  =   .0041673                      SBIC              =   3.076832

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
reur                 7      .470301   0.0255   104.1884   0.0000
rgbp                 7      .430566   0.0522   219.6704   0.0000
rjpy                 7      .466151   0.0243   99.23658   0.0000
----------------------------------------------------------------
```

model fit of the model (all three equations)

```
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reur         |
        reur |
         L1. |   .2001552   .0226876     8.82   0.000     .1556883    .2446221
         L2. |  -.0334134   .0225992    -1.48   0.139     -.077707    .0108802
             |
        rgbp |
         L1. |  -.0615658   .0240862    -2.56   0.011    -.1087738   -.0143578
         L2. |    .024656   .0240581     1.02   0.305    -.0224971    .0718091
             |
        rjpy |
         L1. |  -.0201509   .0166431    -1.21   0.226    -.0527707    .0124689
         L2. |    .002628   .0166676     0.16   0.875    -.0300398    .0352958
             |
       _cons |  -.0058355   .0074464    -0.78   0.433    -.0204301    .0087591
-------------+----------------------------------------------------------------
```

for Euro

```
rgbp         |
        reur |
         L1. |  -.0427769   .0207708    -2.06   0.039    -.0834868   -.0020669
         L2. |   .0567707   .0206898     2.74   0.006     .0162194     .097322
             |
        rgbp |
         L1. |   .2616429   .0220512    11.87   0.000     .2184234    .3048623
         L2. |  -.0920986   .0220255    -4.18   0.000    -.1352678   -.0489294
             |
        rjpy |
         L1. |  -.0566386   .0152369    -3.72   0.000    -.0865024   -.0267747
         L2. |   .0029643   .0152593     0.19   0.846    -.0269435    .0328721
             |
       _cons |   .0000454   .0068172     0.01   0.995    -.0133161    .0134069
-------------+----------------------------------------------------------------
rjpy         |
        reur |
         L1. |   .0241862   .0224874     1.08   0.282    -.0198883    .0682607
         L2. |  -.0313338   .0223997    -1.40   0.162    -.0752365    .0125689
             |
        rgbp |
         L1. |  -.0679786   .0238736    -2.85   0.004    -.1147701   -.0211872
         L2. |   .0324034   .0238458     1.36   0.174    -.0143336    .0791404
             |
        rjpy |
         L1. |   .1508446   .0164962     9.14   0.000     .1185127    .1831766
         L2. |   .0007184   .0165205     0.04   0.965    -.0316611    .0330979
             |
       _cons |  -.0036822   .0073807    -0.50   0.618    -.0181481    .0107836
------------------------------------------------------------------------------
```

for GBP

for yen

VAR model diagnostics

```
. varwle
```

Wald lag exclusion

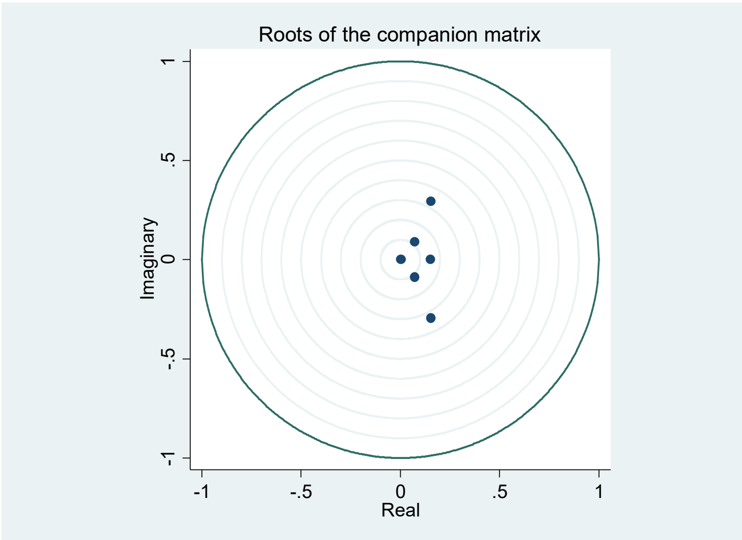
```
  Equation: reur
 +-----------------------------------+
 | lag |    chi2       df   Prob > chi2 |
 |-----+-----------------------------|
 |   1 |  104.0216      3      0.000    |
 |   2 |  2.222079      3      0.528    |
 +-----------------------------------+

  Equation: rgbp
 +-----------------------------------+
 | lag |    chi2       df   Prob > chi2 |
 |-----+-----------------------------|
 |   1 |   214.893      3      0.000    |
 |   2 |  17.60756      3      0.001    |
 +-----------------------------------+

  Equation: rjpy
 +-----------------------------------+
 | lag |    chi2       df   Prob > chi2 |
 |-----+-----------------------------|
 |   1 |  95.98737      3      0.000    |
 |   2 |  2.298777      3      0.513    |
 +-----------------------------------+

  Equation: All
 +-----------------------------------+
 | lag |    chi2       df   Prob > chi2 |
 |-----+-----------------------------|
 |   1 |  592.6466      9      0.000    |
```

In general we don't allow for different number of lags between equations.
↳ they have to have the same number of lags

We reject that all the coef. from lag 1 are equal to 0 ⇒ we need lag 1 we reject H_0 at $\alpha = 0,05$. This means that we need lag 2 as well.

```
|    2 |  51.14752     9     0.000     |
+--------------------------------------+
```

```
. varstable, graph

   Eigenvalue stability condition
  +----------------------------------------+
  |        Eigenvalue          |  Modulus  |
  |----------------------------+-----------|
  |   .1556644 +  .2942196i    |  .332861  |
  |   .1556644 -  .2942196i    |  .332861  |
  |   .1531685                 |  .153168  |
  |  .07266967 + .08855483i    |  .114555  |
  |  .07266967 - .08855483i    |  .114555  |
  |  .00280605                 |  .002806  |
  +----------------------------------------+
   All the eigenvalues lie inside the unit circle.
   VAR satisfies stability condition.
```



We have a stable system

```
. varlmar
```

testing correlation

```
   Lagrange-multiplier test
  +--------------------------------------+
  | lag  |      chi2    df   Prob > chi2 |
  |------+-------------------------------|
  |   1  |    6.8459     9     0.65316   |
  |   2  |    9.7778     9     0.36877   |
  +--------------------------------------+
   H0: no autocorrelation at lag order
```

this is good news

we don't reject that there is no autocorrelation of order 1 and order 2

this could be missing on the exam

12 PIT

Performing Granger causality tests

g)

```
. vargranger

   Granger causality Wald tests
  +------------------------------------------------------------------+
  |          Equation           Excluded |   chi2     df Prob > chi2 |
  |--------------------------------------+---------------------------|
  |              reur               rgbp |  6.6529     2    0.036    |
  |              reur               rjpy |  1.4668     2    0.480    |
  |              reur                ALL |  7.9253     4    0.094    |
  |--------------------------------------+---------------------------|
  |              rgbp               reur |  9.8352     2    0.007    |
```

reject H_0
Pound rate affects Euro rate

can't reject H_0

reject H_0

H_0: There is no Granger causality

Example 2

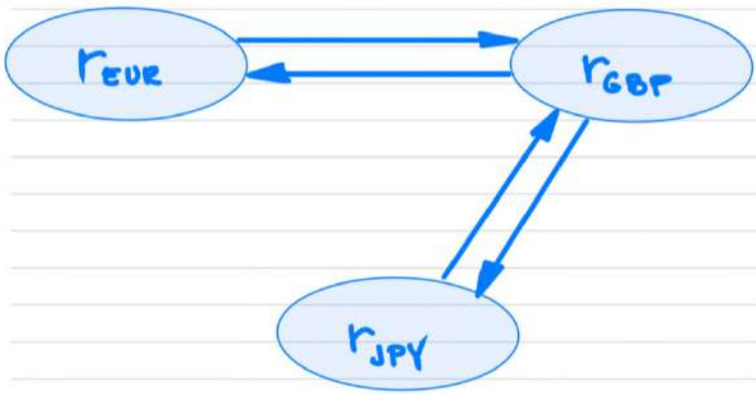
Ad g) PDF, pp. 18–19.

Granger causality Wald test:

H_0: "excluded" variable does **not** Granger cause the "equation" variable

H_1: "excluded" variable Granger causes the "equation" variable

Scheme of Granger causality test results:



Consequently, the pound-dollar rate (r_{GBP}) reacts the fastest.

```
|              rgbp                   rjpy |  13.963     2     0.001    |
|              rgbp                    ALL |  28.095     4     0.000    |
|------------------------------------------+---------------------------|
|              rjpy                   reur |  2.5974     2     0.273    |
|              rjpy                   rgbp |  8.4808     2     0.014    |
|              rjpy                    ALL |  10.905     4     0.028    |
+----------------------------------------------------------------------+
```

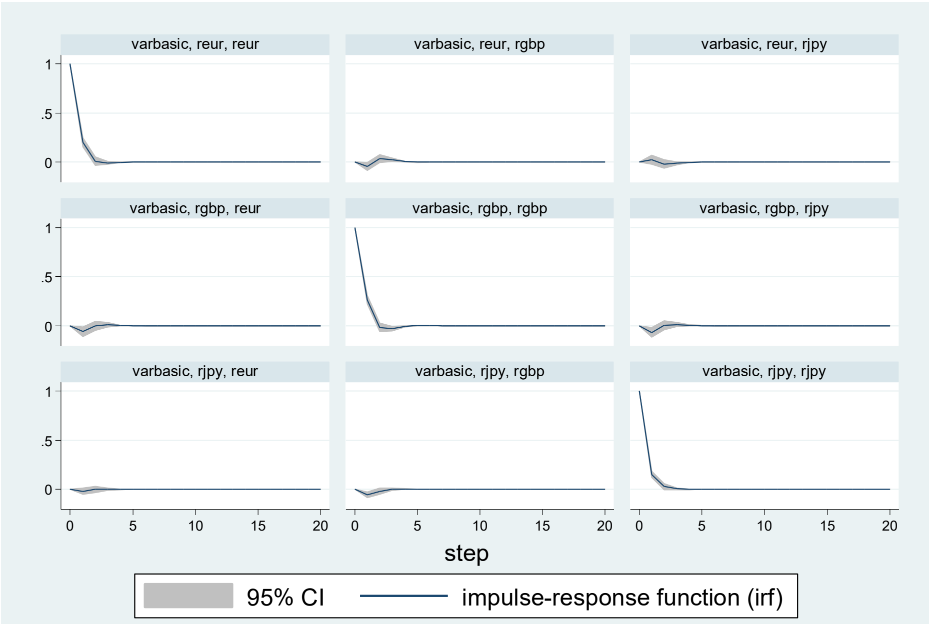
reject H_0

can't reject H_0

Impulse responses and variance decompositions for 20 time periods (days)

```
. varbasic reur rgbp rjpy, lags(1/2) step(20) irf
  // Impulse response functions, IRFs //

Vector autoregression

Sample:  10jul2002 - 06jun2013                  Number of obs     =      3,985
Log likelihood =  -6043.54                      AIC               =   3.043684
FPE            =  .0042115                      HQIC              =   3.055437
Det(Sigma_ml)  =  .0041673                      SBIC              =   3.076832

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
reur                  7     .470301   0.0255   104.1884   0.0000
rgbp                  7     .430566   0.0522   219.6704   0.0000
rjpy                  7     .466151   0.0243   99.23658   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reur         |
        reur |
         L1. |   .2001552   .0226876     8.82   0.000     .1556883    .2446221
         L2. |  -.0334134   .0225992    -1.48   0.139     -.077707    .0108802
             |
        rgbp |
         L1. |  -.0615658   .0240862    -2.56   0.011    -.1087738   -.0143578
         L2. |    .024656   .0240581     1.02   0.305    -.0224971    .0718091
             |
        rjpy |
         L1. |  -.0201509   .0166431    -1.21   0.226    -.0527707    .0124689
         L2. |    .002628   .0166676     0.16   0.875    -.0300398    .0352958
             |
       _cons |  -.0058355   .0074464    -0.78   0.433    -.0204301    .0087591
-------------+----------------------------------------------------------------
rgbp         |
        reur |
         L1. |  -.0427769   .0207708    -2.06   0.039    -.0834868   -.0020669
         L2. |   .0567707   .0206898     2.74   0.006     .0162194     .097322
             |
        rgbp |
         L1. |   .2616429   .0220512    11.87   0.000     .2184234    .3048623
         L2. |  -.0920986   .0220255    -4.18   0.000    -.1352678   -.0489294
             |
        rjpy |
         L1. |  -.0566386   .0152369    -3.72   0.000    -.0865024   -.0267747
         L2. |   .0029643   .0152593     0.19   0.846    -.0269435    .0328721
             |
       _cons |   .0000454   .0068172     0.01   0.995    -.0133161    .0134069
-------------+----------------------------------------------------------------
rjpy         |
        reur |
         L1. |   .0241862   .0224874     1.08   0.282    -.0198883    .0682607
         L2. |  -.0313338   .0223997    -1.40   0.162    -.0752365    .0125689
```

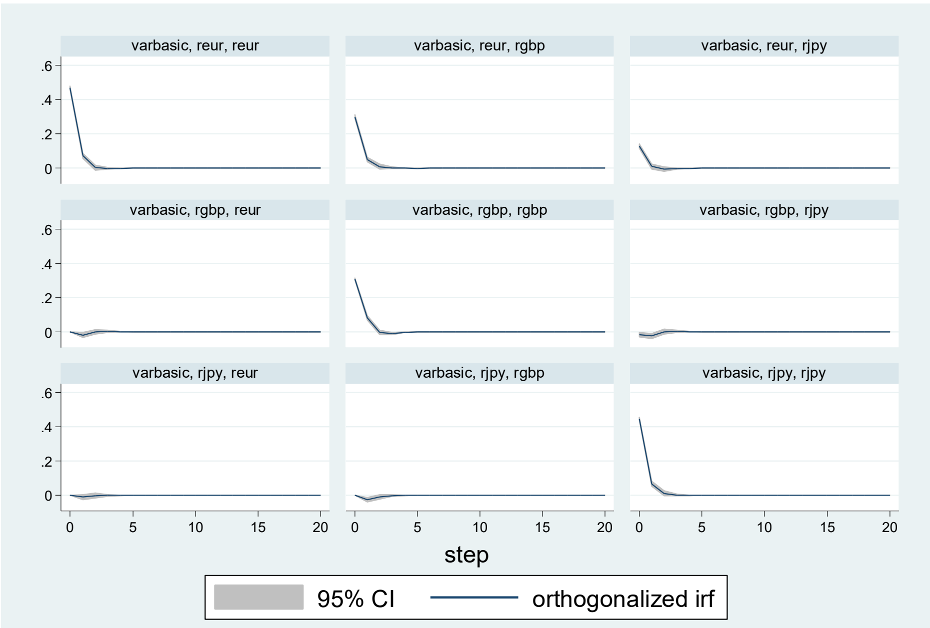
```
             |
        rgbp |
         L1. |  -.0679786   .0238736    -2.85   0.004    -.1147701   -.0211872
         L2. |   .0324034   .0238458     1.36   0.174    -.0143336    .0791404
             |
        rjpy |
         L1. |   .1508446   .0164962     9.14   0.000     .1185127    .1831766
         L2. |   .0007184   .0165205     0.04   0.965    -.0316611    .0330979
             |
       _cons |  -.0036822   .0073807    -0.50   0.618    -.0181481    .0107836
------------------------------------------------------------------------------
```



Graphs by irfname, impulse variable, and response variable

```
. varbasic reur rgbp rjpy, lags(1/2) step(20)
  // Orthogonalised impulse response functions, OIRFs //

Vector autoregression

Sample:  10jul2002 - 06jun2013                  Number of obs     =      3,985
Log likelihood =  -6043.54                      AIC               =   3.043684
FPE            =   .0042115                     HQIC              =   3.055437
Det(Sigma_ml)  =   .0041673                     SBIC              =   3.076832

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
reur                  7     .470301   0.0255   104.1884   0.0000
rgbp                  7     .430566   0.0522   219.6704   0.0000
rjpy                  7     .466151   0.0243   99.23658   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reur         |
        reur |
         L1. |   .2001552   .0226876     8.82   0.000     .1556883    .2446221
         L2. |  -.0334134   .0225992    -1.48   0.139     -.077707    .0108802
             |
        rgbp |
         L1. |  -.0615658   .0240862    -2.56   0.011    -.1087738   -.0143578
         L2. |    .024656   .0240581     1.02   0.305    -.0224971    .0718091
```

rjpy						
L1.	-.0201509	.0166431	-1.21	0.226	-.0527707	.0124689
L2.	.002628	.0166676	0.16	0.875	-.0300398	.0352958
_cons	-.0058355	.0074464	-0.78	0.433	-.0204301	.0087591
rgbp						
reur						
L1.	-.0427769	.0207708	-2.06	0.039	-.0834868	-.0020669
L2.	.0567707	.0206898	2.74	0.006	.0162194	.097322
rgbp						
L1.	.2616429	.0220512	11.87	0.000	.2184234	.3048623
L2.	-.0920986	.0220255	-4.18	0.000	-.1352678	-.0489294
rjpy						
L1.	-.0566386	.0152369	-3.72	0.000	-.0865024	-.0267747
L2.	.0029643	.0152593	0.19	0.846	-.0269435	.0328721
_cons	.0000454	.0068172	0.01	0.995	-.0133161	.0134069
rjpy						
reur						
L1.	.0241862	.0224874	1.08	0.282	-.0198883	.0682607
L2.	-.0313338	.0223997	-1.40	0.162	-.0752365	.0125689
rgbp						
L1.	-.0679786	.0238736	-2.85	0.004	-.1147701	-.0211872
L2.	.0324034	.0238458	1.36	0.174	-.0143336	.0791404
rjpy						
L1.	.1508446	.0164962	9.14	0.000	.1185127	.1831766
L2.	.0007184	.0165205	0.04	0.965	-.0316611	.0330979
_cons	-.0036822	.0073807	-0.50	0.618	-.0181481	.0107836



Graphs by irfname, impulse variable, and response variable

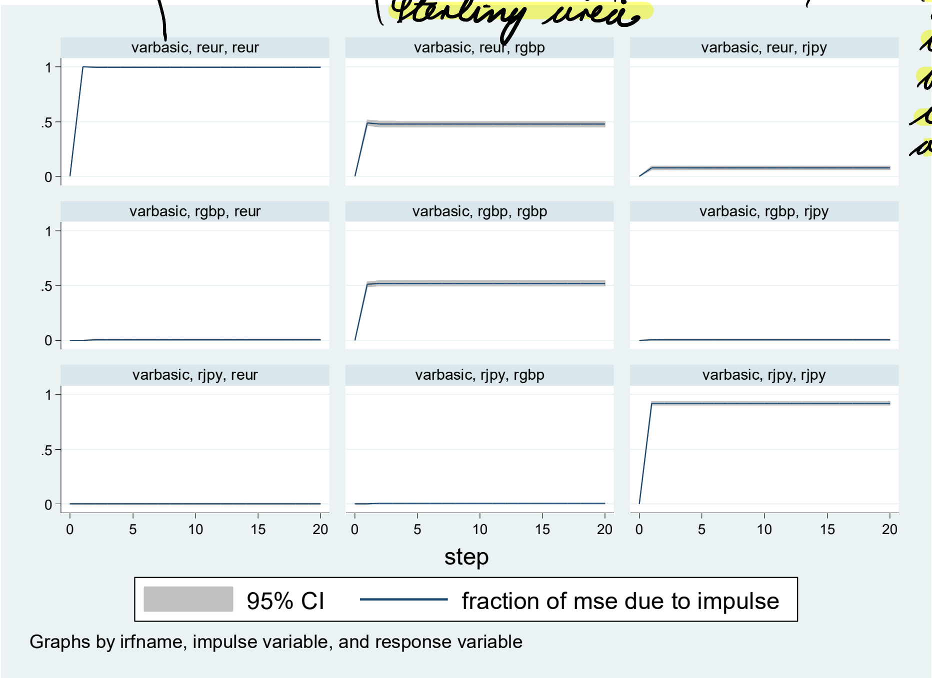
- now a shock is one unit of standard deviation

```
. varbasic reur rgbp rjpy, lags(1/2) step(20) fevd
  // Forecast-error variance decompositions, FEVDs //

Vector autoregression

Sample:  10jul2002 - 06jun2013                 Number of obs     =      3,985
Log likelihood =   -6043.54                    AIC               =   3.043684
FPE            =   .0042115                    HQIC              =   3.055437
Det(Sigma_ml)  =   .0041673                    SBIC              =   3.076832

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
reur                  7     .470301   0.0255   104.1884   0.0000
rgbp                  7     .430566   0.0522   219.6704   0.0000
rjpy                  7     .466151   0.0243   99.23658   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
reur         |
        reur |
         L1. |   .2001552   .0226876     8.82   0.000     .1556883    .2446221
         L2. |  -.0334134   .0225992    -1.48   0.139     -.077707    .0108802
             |
        rgbp |
         L1. |  -.0615658   .0240862    -2.56   0.011    -.1087738   -.0143578
         L2. |    .024656   .0240581     1.02   0.305    -.0224971    .0718091
             |
        rjpy |
         L1. |  -.0201509   .0166431    -1.21   0.226    -.0527707    .0124689
         L2. |    .002628   .0166676     0.16   0.875    -.0300398    .0352958
             |
       _cons |  -.0058355   .0074464    -0.78   0.433    -.0204301    .0087591
-------------+----------------------------------------------------------------
rgbp         |
        reur |
         L1. |  -.0427769   .0207708    -2.06   0.039    -.0834868   -.0020669
         L2. |   .0567707   .0206898     2.74   0.006     .0162194     .097322
             |
        rgbp |
         L1. |   .2616429   .0220512    11.87   0.000     .2184234    .3048623
         L2. |  -.0920986   .0220255    -4.18   0.000    -.1352678   -.0489294
             |
        rjpy |
         L1. |  -.0566386   .0152369    -3.72   0.000    -.0865024   -.0267747
         L2. |   .0029643   .0152593     0.19   0.846    -.0269435    .0328721
             |
       _cons |   .0000454   .0068172     0.01   0.995    -.0133161    .0134069
-------------+----------------------------------------------------------------
rjpy         |
        reur |
         L1. |   .0241862   .0224874     1.08   0.282    -.0198883    .0682607
         L2. |  -.0313338   .0223997    -1.40   0.162    -.0752365    .0125689
             |
        rgbp |
         L1. |  -.0679786   .0238736    -2.85   0.004    -.1147701   -.0211872
         L2. |   .0324034   .0238458     1.36   0.174    -.0143336    .0791404
             |
        rjpy |
         L1. |   .1508446   .0164962     9.14   0.000     .1185127    .1831766
         L2. |   .0007184   .0165205     0.04   0.965    -.0316611    .0330979
             |
       _cons |  -.0036822   .0073807    -0.50   0.618    -.0181481    .0107836
------------------------------------------------------------------------------
```

almost all the variability of EURO comes from Euro area

about half of the variability of GBP comes from Euro area and the other half from Sterling area

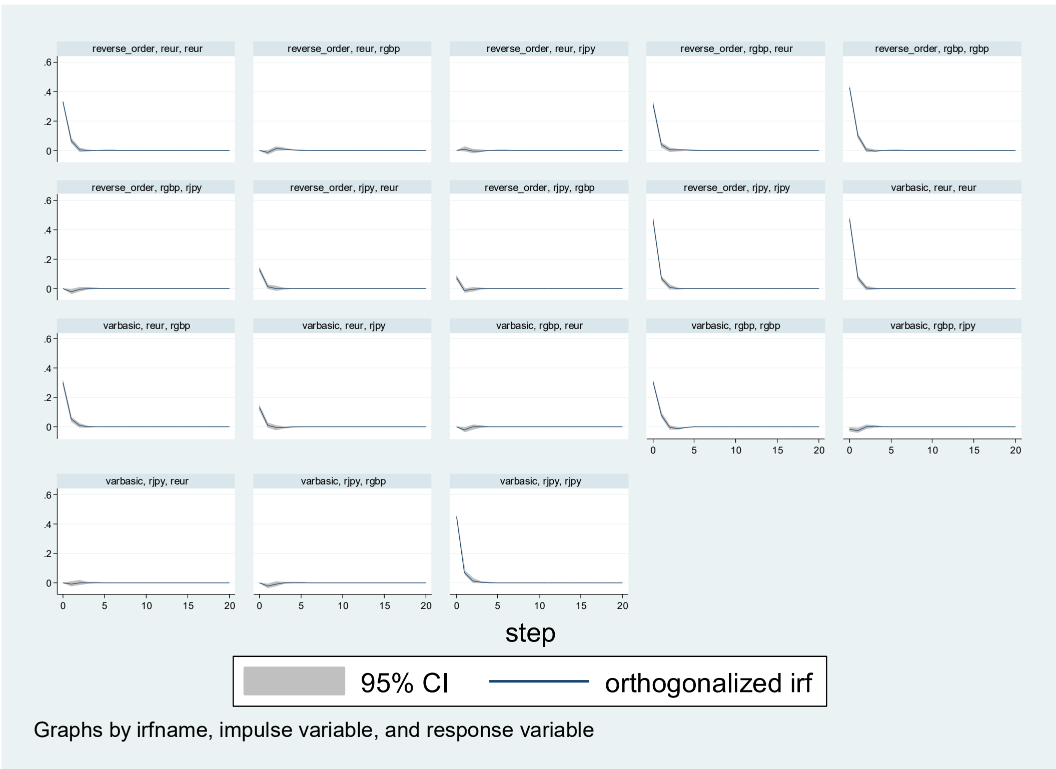
Almost all of Yen's variability comes from yen-area and a little bit of variability comes from Euro area



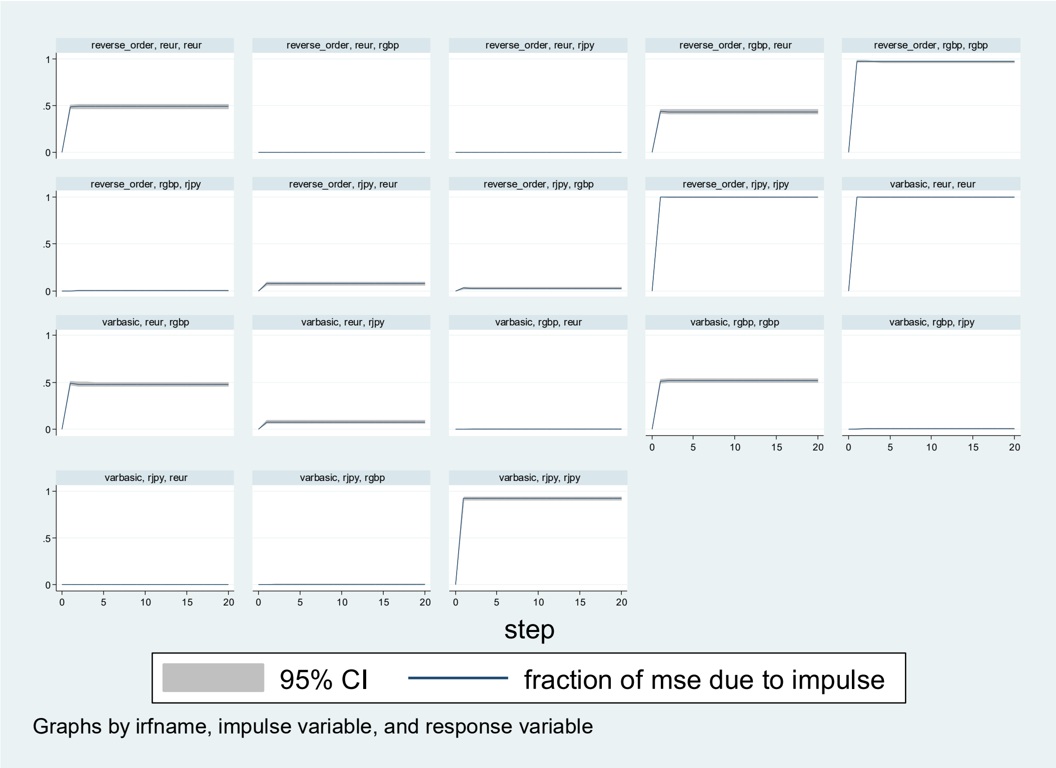
, a shock is one standard deviation

Changing the ordering of variables

```
. irf create reverse_order, replace step(20) order(rjpy rgbp reur)
irfname reverse_order not found in _varbasic.irf
(file _varbasic.irf updated)

. irf graph oirf
```



```
. irf graph fevd
```

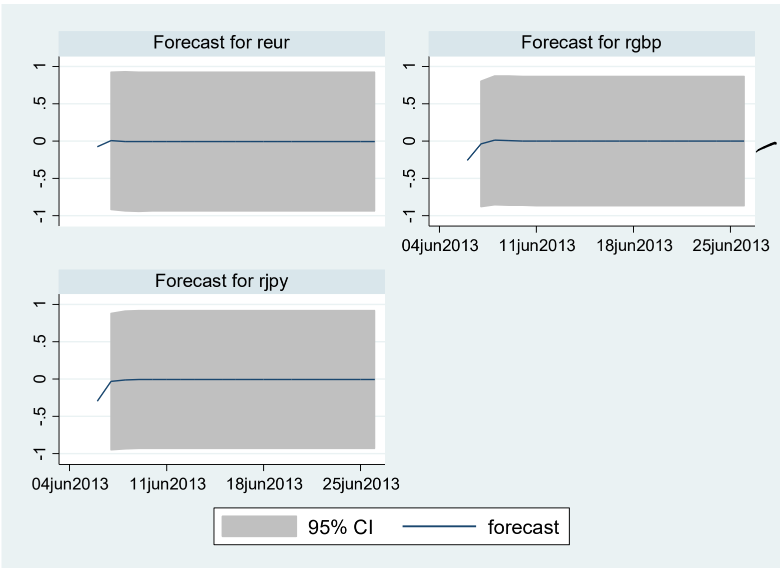


Forecasting based on VAR(2) for the following 20 time periods (days)

```
. qui var reur rgbp rjpy, lags(1/2)
```

new prefix

```
. fcast compute fc_, step(20)
  // generates level, lower bound, upper bound, and std error //

. fcast graph fc_reur fc_rgbp fc_rjpy
```



we don't have exchange rates on the y-axis

■

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

8.4 Modelling Volatility

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

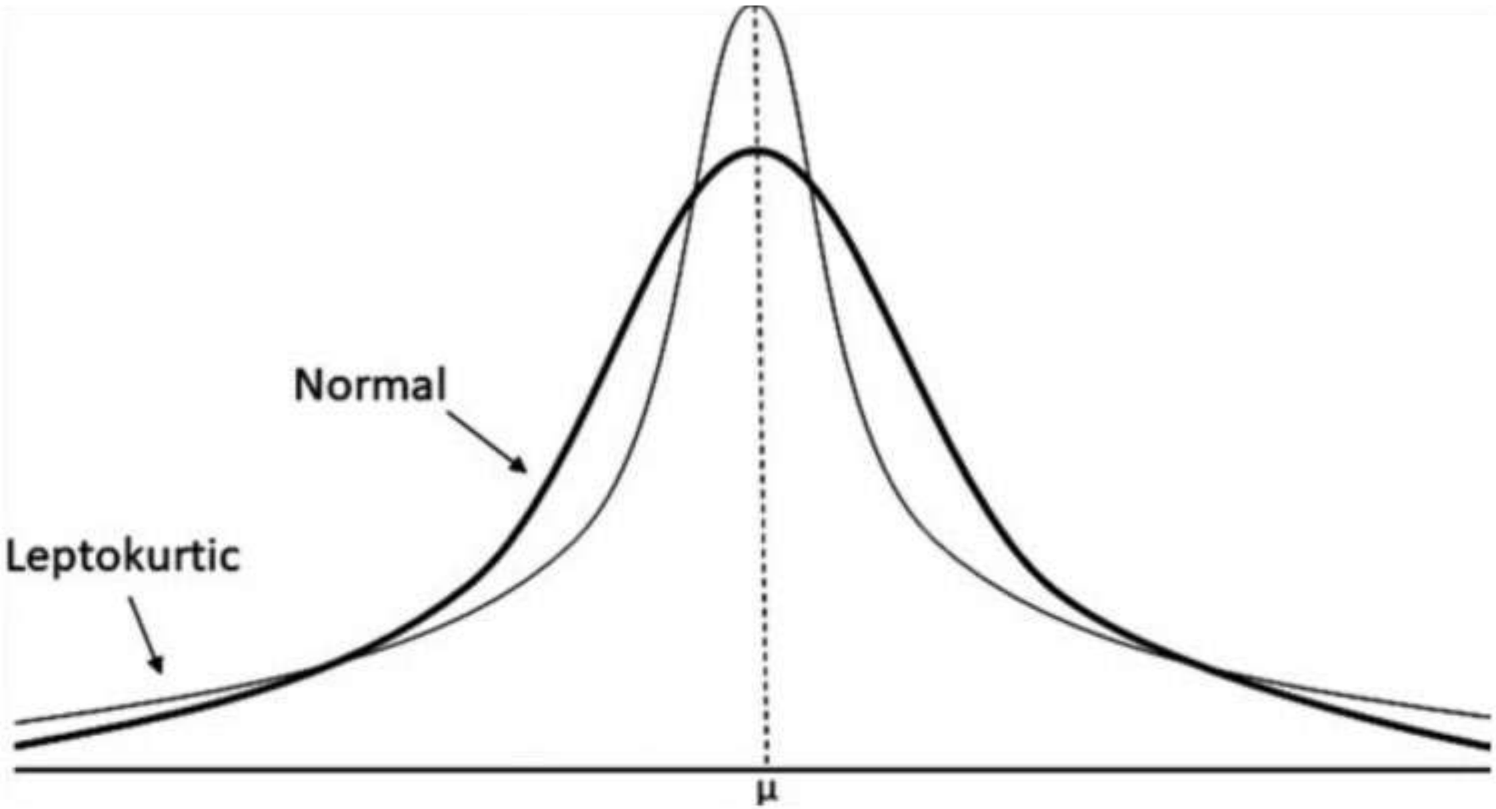
Basic definitions

- So far, most of the studied regression models were linear, at least in parameters (regression coefficients).
- However, linear models *cannot* explain a number of important features common to some economic and much of financial data:
  - **leptokurtosis**: tendency for time series to have distributions that exhibit fat tails and excess peakedness at the mean (issue with *heteroscedasticity*);
  - **volatility clustering** or **volatility pooling**: the tendency for volatility to appear in *bursts* – thus large returns (of either sign) are expected to follow large returns, and small returns to follow small returns;
  - **leverage effects**: the tendency for volatility to rise more following a large variable decrease than following a variable increase of the same magnitude.

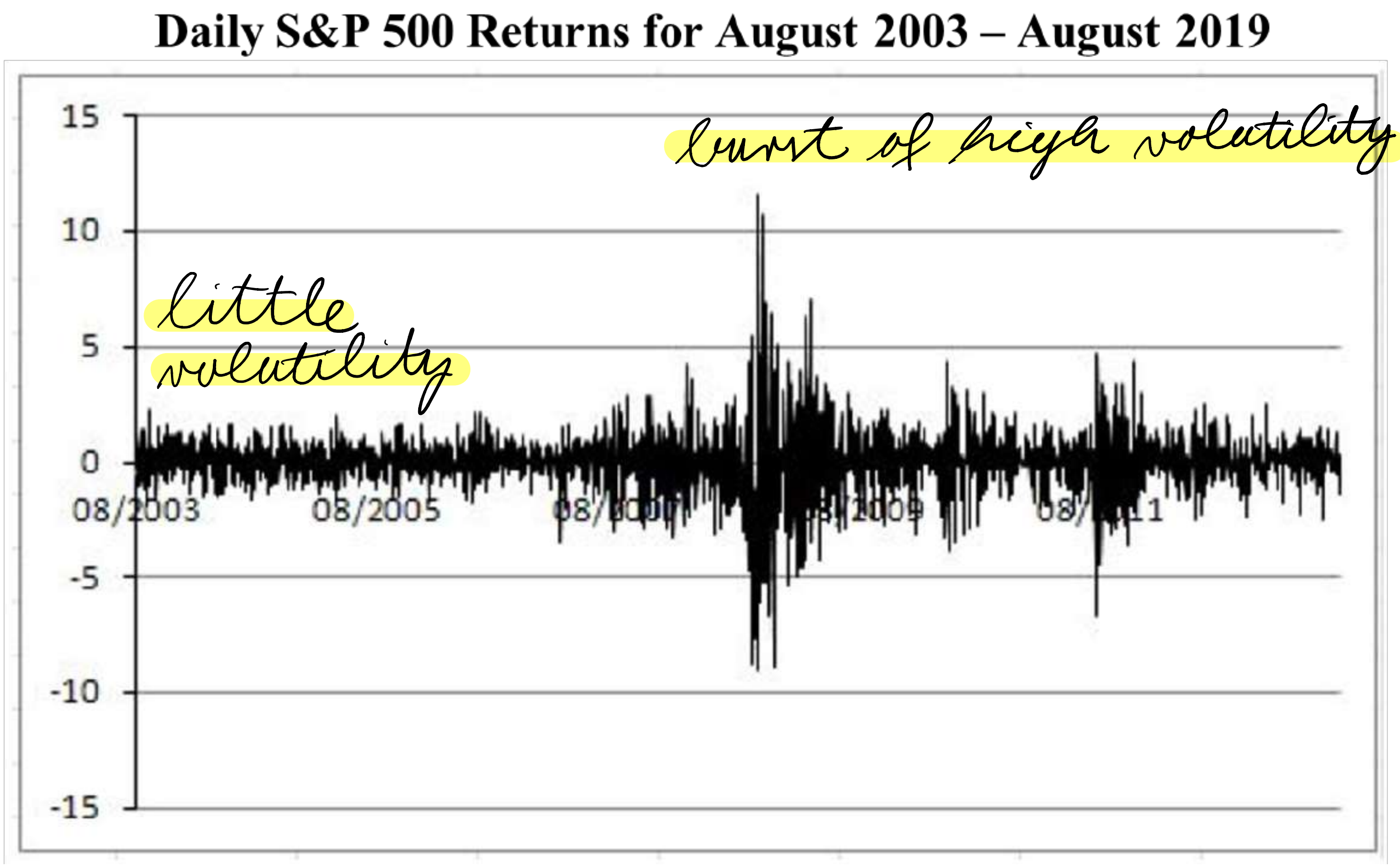
↳ if the yield is negative then the volatility is risen more than if the yield is positive

University of Ljubljana
SEB School of Economics and Business

Basic definitions



Basic definitions



Basic definitions

- Therefore, many relationships economics and finance are intrinsically non-linear and there is a need for models that can cover these features.
- We will first address the family of **univariate volatility models** that started developing with the so-called *autoregressive conditionally heteroscedastic (ARCH) models*, which cover particularly well heteroscedasticity and volatility clustering or volatility pooling.
- *Heteroscedasticity* in our case refers to the specific case of leptokurtosis, where the density function exhibits "fatter" tails and a higher peak at the mean than the normal distribution.

ARCH model

this model is able to take leptokurtosis and volatility clustering into account

- Lets us start with one of the key deviations from the classical linear regression model – heteroscedasticity, where $\mathrm{Var}(u_t) \neq \sigma_u^2$.
- What could the current value of the variance of the disturbances plausibly depend upon? *Previous squared disturbance terms.*
- This leads to the **autoregressive conditionally heteroscedastic (ARCH) model** for the variance of the disturbances, e.g.:

it changes with observations - sigma t squared

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

we go 1 unit to the past

- This simple process is known as an ARCH(1) model and is completely *deterministic* (no disturbance term is present in the variance equation).
- The ARCH model due to Engle (1982) has proved very useful in economics and especially in finance.

ARCH model

- The full model would be:

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad u_t \sim N(0, \sigma_t^2)$$ (mean equation)

where $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$. (variance equation)

- We can easily extend this to the *general case* where the variance of the disturbances depends on q lags of squared disturbances:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2 .$$

- This is an ARCH(q) model. $\sigma_t^2 = h_t$
- Instead of calling the variance σ_t^2, in the literature it is usually called h_t, so the model is:

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad u_t \sim N(0, h_t)$$

where $h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2$.

ARCH model

- Consider an ARCH(1). Instead of the above, we can write:

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad u_t = v_t \sigma_t,$$
$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}, \quad v_t \sim \mathrm{N}(0, 1).$$

another way of writing the model

- The two are *different ways of expressing exactly the same model*. The first form is easier to understand, while the second form is required for simulating from an ARCH model.
- As $h_t = \sigma_t^2$ is a (conditional) variance, it *cannot be negative*. To ensure this in an ARCH(q) model, a sufficient but not necessary condition would be: $\alpha_i \geq 0, \forall i = 0, 1, 2, \ldots, q$.

↳ all the alphas must be positive

Testing for ARCH effects

1. First, run any postulated linear regression, e.g. $y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$ (mean equation), and save the residuals, $\hat{u}_t$.
   $\hat{u} = e$
2. Then square the residuals, and regress them on q own lags to test for ARCH of order q, i.e. run the auxiliary regression:

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2 + \ldots + \gamma_q \hat{u}_{t-q}^2 + v_t$$

   where v_t is IID. Obtain R^2 from this auxilliary regression.

3. The test statistic is defined as $T{\cdot}R^2$ (the number of observations multiplied by the multiple determination coefficient from the last regression) and is distributed as a $\chi^2(q)$.

↳ Lagrange multiplier test

Testing for ARCH effects

4. The null and alternative hypotheses are:

$H_0 : \gamma_j = 0, \forall j = 1, \ldots, q;$ → there are no arch effects

$H_1 : \gamma_j \neq 0, \exists j = 1, \ldots, q.$ → there are arch effects in the model

If the value of the test statistic is greater than the critical value from the χ^2 distribution, then we reject the null hypothesis.

- The test can also be thought of as a test for autocorrelation in the *squared* residuals.
- Note that the ARCH test is also sometimes applied directly to the dependent variable instead of the residuals from step 1 above.

↳ it is more correct to test residuals

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION OF MBAs ACCREDITED

Issues with the ARCH model

- ARCH(q) models are rarely used in practice due to their shortcomings:
  - How do we decide on q?
  - The required value of q might be very large.
  - Non-negativity constraints might be violated (implying negative estimated variances) if not explicitly imposed. $\Rightarrow$ α need to be positive
- A widely used natural extension of an ARCH(q) model, which gets around some of these problems, is the so-called *generalised ARCH (GARCH) model*.

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

GARCH model

- The **generalised ARCH (GARCH) model** is due to Bollerslev (1986) and allows the conditional variance to be dependent upon previous own lags.
- The variance equation is now:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2$$

variance is lagged

- The current fitted variance, h_t, is interpreted as a *weighted function* of:
  - a long-term average value (dependent on α_0),
  - information about volatility during the previous period ($\alpha_1 u_{t-1}^2$) and
  - the fitted variance from the model during the previous period ($\beta \sigma_{t-1}^2$).
- This is a GARCH(1,1) model, which is like an ARMA(1,1) model, but for the variance equation. In fact, the GARCH(1,1) model can be written as an infinite order ARCH model.

ARMA model was modeling of the mean.
GARCH model is modeling of the variance.

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

GARCH model

- Why is GARCH *better* than ARCH?
  - ➢ more parsimonious – avoids overfitting;
  - ➢ less likely to breech non-negativity constraints.
- We can extend the GARCH(1,1) model to a GARCH(p,q):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \cdots + \beta_p \sigma_{t-p}^2$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

- But in general a GARCH(1,1) model will be sufficient to capture the volatility clustering in the data.

↳ Generally speaking GARCH(1,1) is good enough for finance

GARCH model

- In the GARCH(1,1) model, the unconditional variance of u_t is given by:

$$\boxed{\mathrm{Var}(u_t) = \frac{\alpha_0}{1-(\alpha_1+\beta)}}$$

as long as $\boxed{\alpha_1 + \beta < 1}$, whereas:

  - for $\boxed{\alpha_1 + \beta \geq 1}$ we obtain non-stationarity in variance and
  - for $\boxed{\alpha_1 + \beta = 1}$ we obtain the integrated GARCH (IGARCH) model.

- For non-stationarity in variance, the conditional variance forecasts will *not converge* on their unconditional value as the horizon increases.

- Since these models are no longer of the usual linear form, we *cannot* use the OLS. We use the **maximum likelihood (ML) estimator** instead.

- As the disturbances are often not normally distributed, but instead *leptokurtic*, we can use the ML with a robust variance-covariance estimator. This is called the quasi-maximum likelihood (QML).

robust version of the ML estimator

Extensions to the basic GARCH model

- Two key issues with GARCH(p,q) models:
  - *non-negativity constraints* may still be violated; you would still get negative variance ⇒ not okay
  - GARCH models cannot account for *leverage effects*.
- Since the GARCH model was developed, a large number of **extensions** and variants have been proposed.
- Plausible solutions include the GJR model and the exponential GARCH (EGARCH) model, which (unlike the basic symmetric GARCH model) are *asymmetric* GARCH models, i.e. they allow an asymmetric effect of "news" (disturbances or unanticipated changes).
- In addition, we also consider the GARCH–M model. ↳ u

GJR model

- The **GJR model** is due to Glosten, Jaganathan and Runkle (1993). The variance equation is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma u_{t-1}^2 I_{t-1}$$

dummy variable

where $I_{t-1} = 1$ if $u_{t-1} < 0$; – if the disturbance was negative
$= 0$ otherwise.

- For a leverage effect, we would see $\gamma > 0$.
- We *still* require $\alpha_1 + \gamma \geq 0$ and $\alpha_1 \geq 0$ for non-negativity.

variance can still be negative if we have the "right" combination of α and γ

EGARCH model

- The **exponential GARCH (EGARCH) model** was suggested by Nelson (1991). The variance equation is given by:

$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$$

- Advantages of the model:
  - Since we model the $\log(\sigma_t^2)$, then even if the parameters are negative, σ_t^2 will be *positive* and non-negativity is satisfied.
  - We can account for the leverage effect, as if the relationship between volatility and the dependent variable in the mean equation (e.g. returns) is negative, γ will be *negative*.

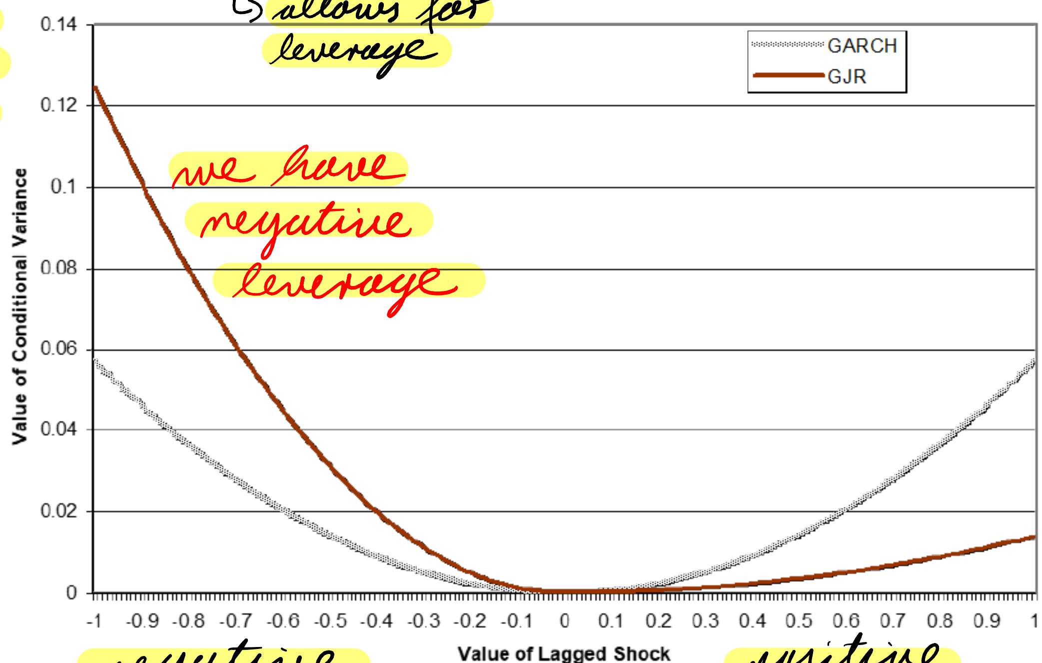
Asymmetry in GARCH models

The **news-impact curve** or **news-response curve** plots the next period volatility (h_t) that would arise from various positive and negative values of u_{t-1}, given an estimated model.

An example of news-impact curves for S&P 500 returns using the coefficients from GARCH (symmetric) and GJR (asymmetric) model estimates:

↳ doesn't include leverage

↳ allows for leverage



negative shocks

positive shocks

GARCH–M model

- In certain situations, we can expect that volatility σ or σ^2 affects the dependent variable y in the mean equation.
- E.g., we expect a risk to be compensated by a higher return. So why not let the return of a security be partly determined by its risk?
- This leads to another extension of the basic GARCH model, called the **GARCH-in-mean (GARCH–M) model**:

volatility can now affect yield

$$y_t = \mu + \delta\sigma_{t-1} + u_t,\ u_t \sim N(0, \sigma_t^2)$$ mean equation

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2$$ variance equation

- The parameter δ measures the *feedback* from the conditional variance to the conditional mean.
- Above, if δ is positive and statistically significant, then increased risk, given by an increase in the conditional variance, leads to a rise in the mean return. Then, δ can be interpreted as a *risk premium*.

Volatility forecasting

- GARCH can model the volatility clustering effect, as the conditional variance is autoregressive. Such models can be used to **forecast volatility**.
- We could show that Var $(y_t \mid y_{t-1}, y_{t-2}, \ldots)$ = Var $(u_t \mid u_{t-1}, u_{t-2}, \ldots)$, which indicates that modelling σ_t^2 will give us models and **forecasts for the variance of y_t as well**.
- Variance forecasts are *additive over time*, which is very useful for calculating **periodic forecasts**, but standard deviation forecasts are not.
- If standard deviations are analysed instead of variances, they need to be squared for forecasting purposes before adding up, and then the square root should be taken to obtain a periodic standard deviation.
- Some *typical examples* of volatility forecasting in finance include option pricing, conditional betas, and dynamic hedge ratios.

Limitations of univariate volatility models

- A major limitation of univariate volatility models is that they model the conditional variance of each series entirely *independently* of all other series (conditional *covariances* are assumed to be zero).
- This is potentially important for **two reasons**: ↳ that might not be the case
  - ➢ If there are volatility spillovers between markets or assets (non-zero covariances), the univariate model will be *mis-specified*;
  - ➢ It is often the case that the covariances between series are *of interest* (in finance, such examples include the calculation of hedge ratios, portfolio value at risk estimates, and CAPM betas).
- In addition, modelling the volatilities together may increase efficiency.
- These deficiencies can be overcome by introducing the family of **multivariate volatility models**, in particular the *multivariate GARCH models* and the *direct correlation models*.

Multivariate volatility models

- **Multivariate GARCH models** are used to estimate and to forecast (also) covariances and correlations of multiple series over time.
- The basic formulation is similar to that of the GARCH model, but where the covariances as well as the variances are permitted to be *time-varying*.
- There are **two main classes** of multivariate GARCH formulation that are widely used and will be discussed here:
  - the VECH (and diagonal VECH) model, and
  - the BEKK (and diagonal BEKK) model.
- In addition, we will also discuss the CCC model, which – strictly speaking – belongs to the family of **direct correlation models**.

most often used

VECH model

- Let us e.g. suppose that there are *two variables* (*series*) used in the model, i.e. $g = 2$. The conditional covariance matrix is denoted H_t:

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{bmatrix}$$

$h_{21t} = h_{12t}$
↳ this usually holds

- The **VECH model** includes half-vectorization (*vech*, thus the name) of matrices, which are assumed here to be *symmmetric*. First, we half-vectorize the presumably symmetric matrix H_t ($h_{ij} = h_{ji}, \forall i \neq j$):

$$VECH(H_t) = \begin{bmatrix} h_{11t} \\ h_{22t} \\ h_{12t} \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{variances} \\ - \text{covariance} \end{matrix}$$

where h_{iit} represent the conditional variances at time t ($i = 1, 2$) and h_{ijt} ($i \neq j$) represent the conditional covariances between the variables.

EFMD EQUIS ACCREDITED
AACSB ACCREDITED
ASSOCIATION AMBA ACCREDITED

VECH model

Xi

- Similarly, if Ξ_t represents the disturbance (innovation) vector:

$$\Xi_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

and

$$\Xi_t \Xi_t' = \begin{bmatrix} u_{1t}^2 & u_{1t}u_{2t} \\ u_{1t}u_{2t} & u_{2t}^2 \end{bmatrix}$$

the model includes half-vectorization of the latter:

$$VECH(\Xi_t \Xi_t') = \begin{bmatrix} u_{1t}^2 \\ u_{2t}^2 \\ u_{1t}u_{2t} \end{bmatrix}$$

VECH model

- The conditional variances and covariances each depend upon lagged values of all of the *variances* and *covariances* and on lags of both the *squares of disturbance terms* and *their cross products*.
- In **matrix form**, the VECH model is thus written as:

$$VECH(H_t) = C + A\,VECH(\Xi_{t-1}\Xi'_{t-1}) + B\,VECH(H_{t-1})$$

where $\Xi_t \sim N(0, H_t)$ and A, B, C represent the corresponding estimable sets of *parameters*:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}, C = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix}$$

- **Stationarity** of the model requires that the *eigenvalues* of $A + B$ are less than one in absolute value.

VECH model

- In **scalar notation**, writing out all of the elements gives us three equations as:

$$
\begin{aligned}
h_{11t} &= c_{11} + a_{11}u_{1t-1}^2 + a_{12}u_{2t-1}^2 + a_{13}u_{1t-1}u_{2t-1} + b_{11}h_{11t-1} \\
&\quad + b_{12}h_{22t-1} + b_{13}h_{12t-1} \\
h_{22t} &= c_{21} + a_{21}u_{1t-1}^2 + a_{22}u_{2t-1}^2 + a_{23}u_{1t-1}u_{2t-1} + b_{21}h_{11t-1} \\
&\quad + b_{22}h_{22t-1} + b_{23}h_{12t-1} \\
h_{12t} &= c_{31} + a_{31}u_{1t-1}^2 + a_{32}u_{2t-1}^2 + a_{33}u_{1t-1}u_{2t-1} + b_{31}h_{11t-1} \\
&\quad + b_{32}h_{22t-1} + b_{33}h_{12t-1}
\end{aligned}
$$

- Due to the large number of parameters (e.g. 21 above), the estimation of VECH models can quickly become *infeasible*.

VECH model

- We thus introduce the **diagonal VECH model**, which assumes that there are no direct volatility spillovers from one series to another:

$$h_{11t} = \alpha_0 + \alpha_1 u_{1t-1}^2 + \alpha_2 h_{11t-1}$$

the only terms that are kept are 11

$$h_{22t} = \beta_0 + \beta_1 u_{2t-1}^2 + \beta_2 h_{22t-1}$$

the only terms that are kept are 22

$$h_{12t} = \gamma_0 + \gamma_1 u_{1t-1} u_{2t-1} + \gamma_2 h_{12t-1}$$

the only terms that are kept are 12

- In other words, A and B in the matrix notation are *diagonal matrices*.
- However, *neither* the VECH *nor* the diagonal VECH ensure a *positive definite variance-covariance matrix* H_t (meaning, among other things, positive values on the diagonal and symmetry about the diagonal).

BEKK model

- An alternative approach is the **BEKK model**, named after Baba, Engle, Kraft and Kroner (1990), which uses a quadratic form for the parameter matrices to ensure a positive definite variance-covariance matrix H_t:

$$H_t = W'W + A'H_{t-1}A + B'\Xi_{t-1}\Xi'_{t-1}B$$

where A and B are $g \times g$ parameter matrices and W is an upper triangular parameter matrix.

- Recently, **asymmetric models** have become very popular, where the conditional variances and/or covariances are permitted to *react differently to negative than to positive disturbances* of the same magnitude.

BEKK model

- In the multivariate context, this is often achieved by implementing the so-called **asymmetric BEKK model**:

$$H_t = W'W + A'H_{t-1}A + B'\Xi_{t-1}\Xi'_{t-1}B + D'z_{t-1}z'_{t-1}D$$

they take out the negative values

where z_{t-1} is a column vector with elements taking the value $-u_{t-1}$ if the corresponding element of u_{t-1} is negative and zero otherwise.

- To allow faster, but less flexible estimation in higher dimensions, a **diagonal BEKK model** is also used, where the parameter matrices A, B and D are reduced to *diagonal matrices*.
- Model estimation for all classes of multivariate GARCH models is again performed using the maximum likelihood estimator.

ML estimator

CCC model

- The above models specify the dynamics of the conditional *covariances* between a set of series, whereas the *correlations* between any given pair of series at each point in time can be constructed later on.
- A subtly different approach would be to model the dynamics for the *correlations directly*.
- This can be done by the family of **direct correlation models**.
- The most often used is the so-called **constant conditional correlation (CCC) model**, where the *correlations* between the disturbances u_t (or, equivalently, between the variables y_t themselves) are *assumed to be fixed through time*.
- Thus, although the *conditional covariances* are not fixed, they are *tied to the variances*.

CCC model

- The **conditional variances** are identical to those of a set of univariate GARCH specifications, although they are now estimated jointly:

$$h_{ii,t} = c_i + a_i u_{i,t-i}^2 + b_i h_{ii,t-1}, \qquad i = 1, \ldots, g$$

whereas the **conditional covariances**, i.e. the off-diagonal elements of H_t, $h_{ij,t}$ $(i \neq j)$, are defined indirectly via the correlations ρ_{ij}:

$$h_{ij,t} = \rho_{ij} h_{ii,t}^{1/2} h_{jj,t}^{1/2}, \qquad i, j = 1, \ldots, g, \; i < j$$

There are tests for this!

- Is it *plausible* to assume that the **correlations are constant** through time?
- Several tests of this assumption have been developed, including a test based on the information matrix and a Lagrange Multiplier test.
- If *not*, a **dynamic conditional correlation (DCC) model** can be specified and estimated (not covered here).

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

8. Time Series Modelling and Forecasting

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ljubljana, October 2022