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 2025

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:

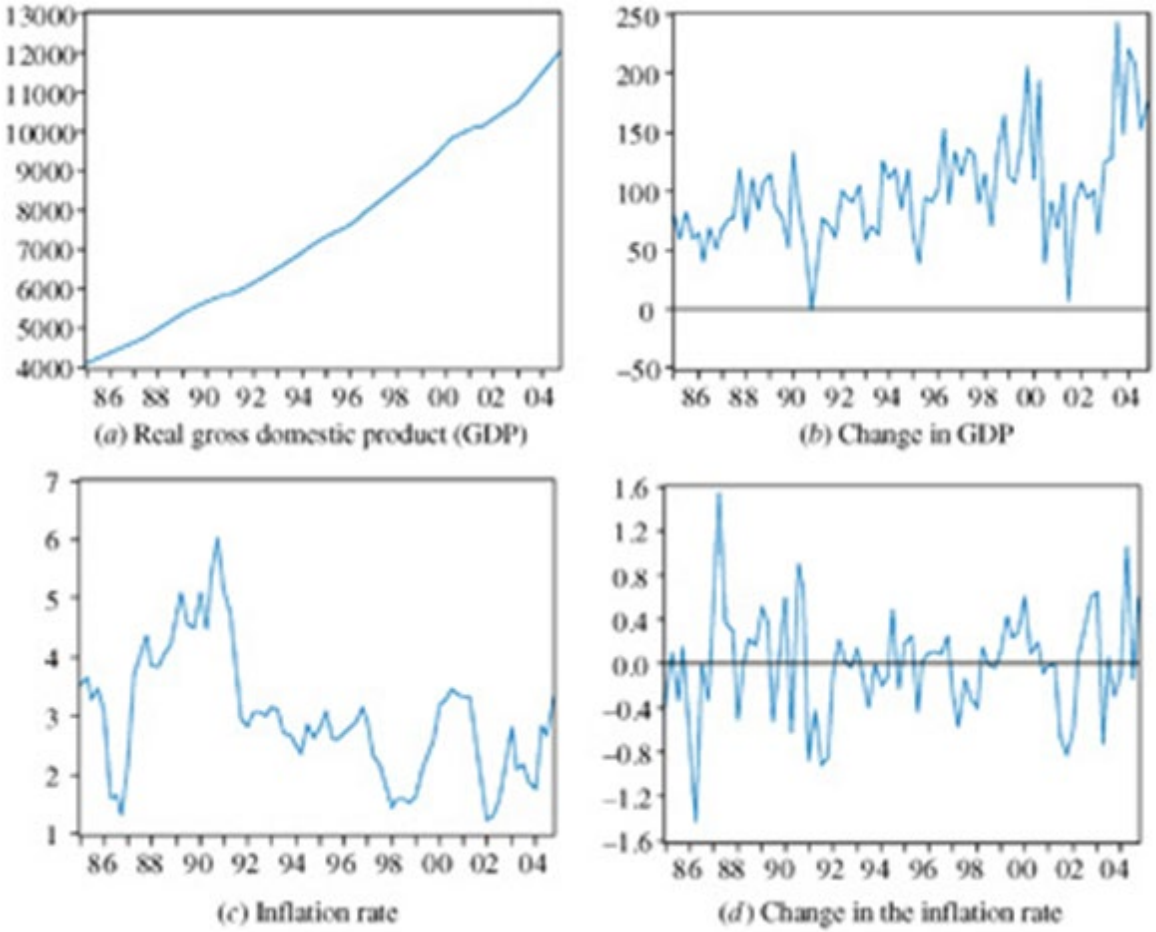
$$E(y_t) = \mu$$

$$Var(y_t) = \sigma^2$$

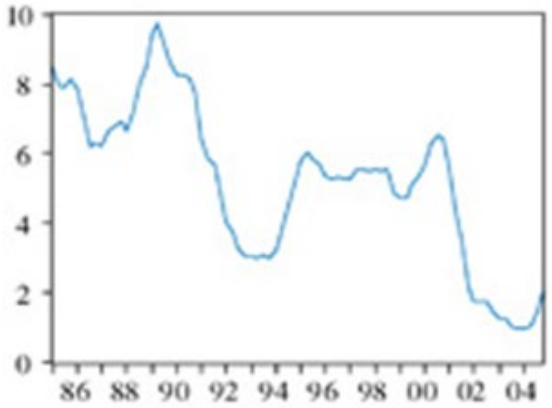
$$Cov(y_t, y_{t+s}) = Cov(y_t, y_{t-s}) = \gamma_s$$

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

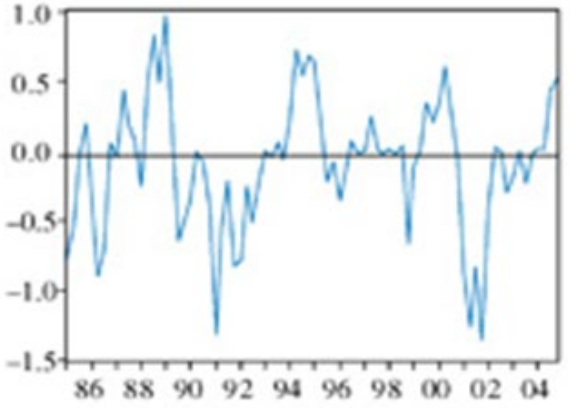
Basic definitions



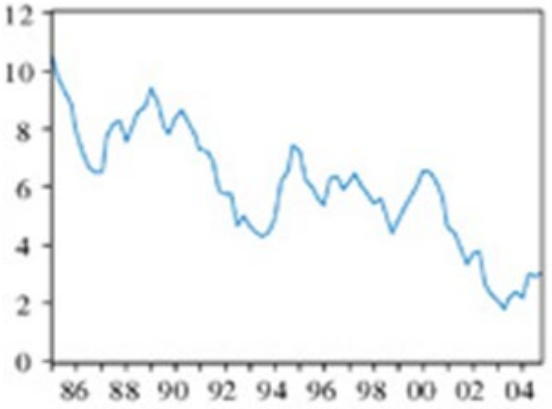
Basic definitions



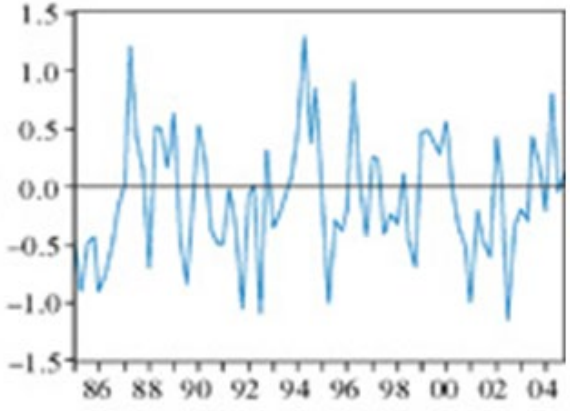
(e) Federal Funds rate



(f) Change in the Federal Funds rate



(g) 3-year Bond rate



(h) Change in the Bond rate

Spurious regression

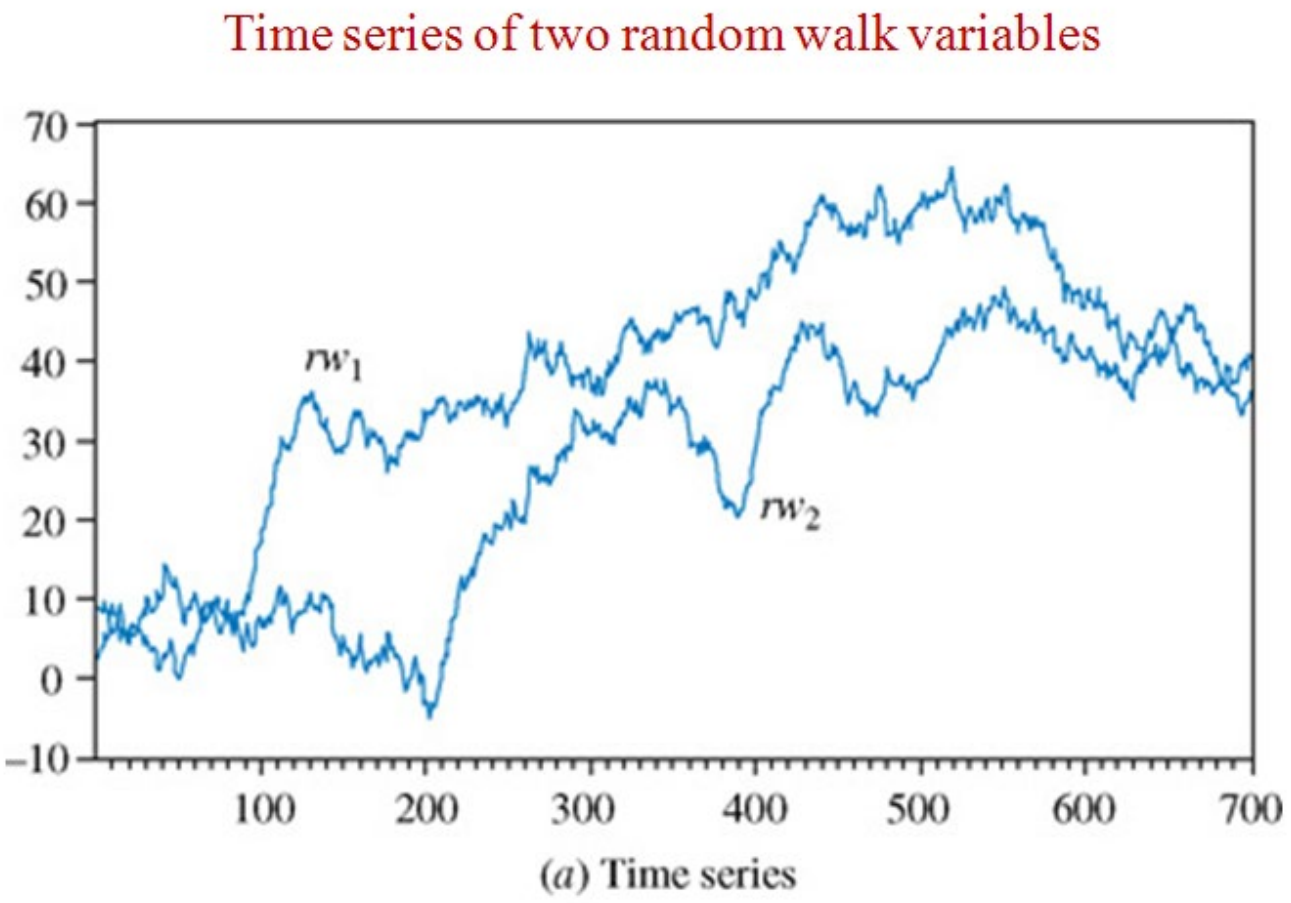
- If (weak) stationary is *not satisfied* for the variables included in a time-series regression model, we are likely to obtain the so-called spurious regression and invalid statistical inference (test statistics do *not* follow their respective statistical distributions).
- **Spurious regression** can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

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

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

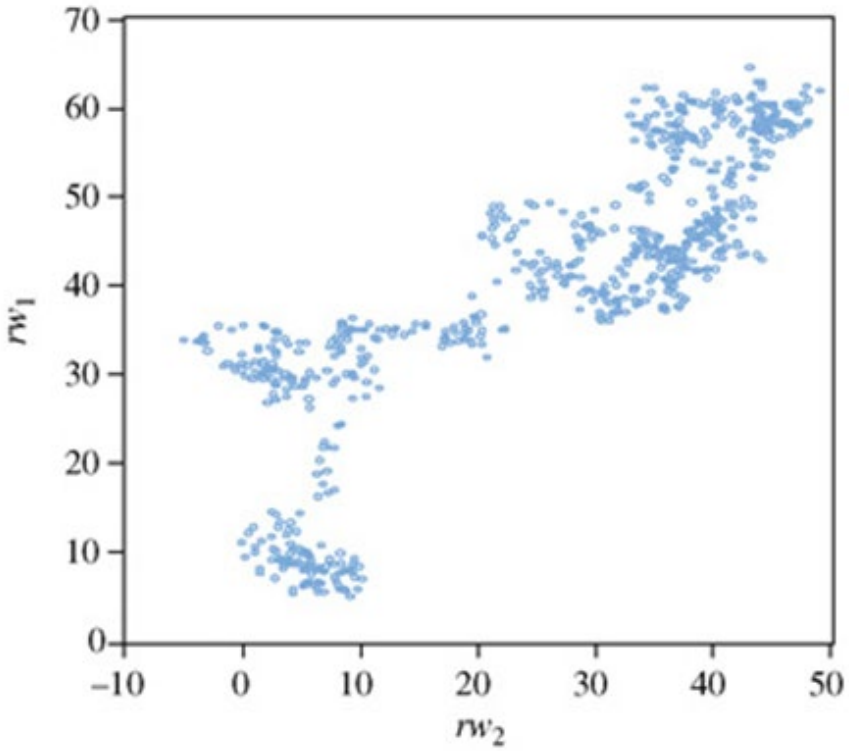
$$\widehat{rw_{1t}} = 17.818 + 0.842\, rw_{2t}, \quad R^2 = 0.70$$
$$(t) \qquad\qquad (40.837)$$

Spurious regression



Spurious regression

Scatter plot of two random walk variables



(*b*) Scatter plot

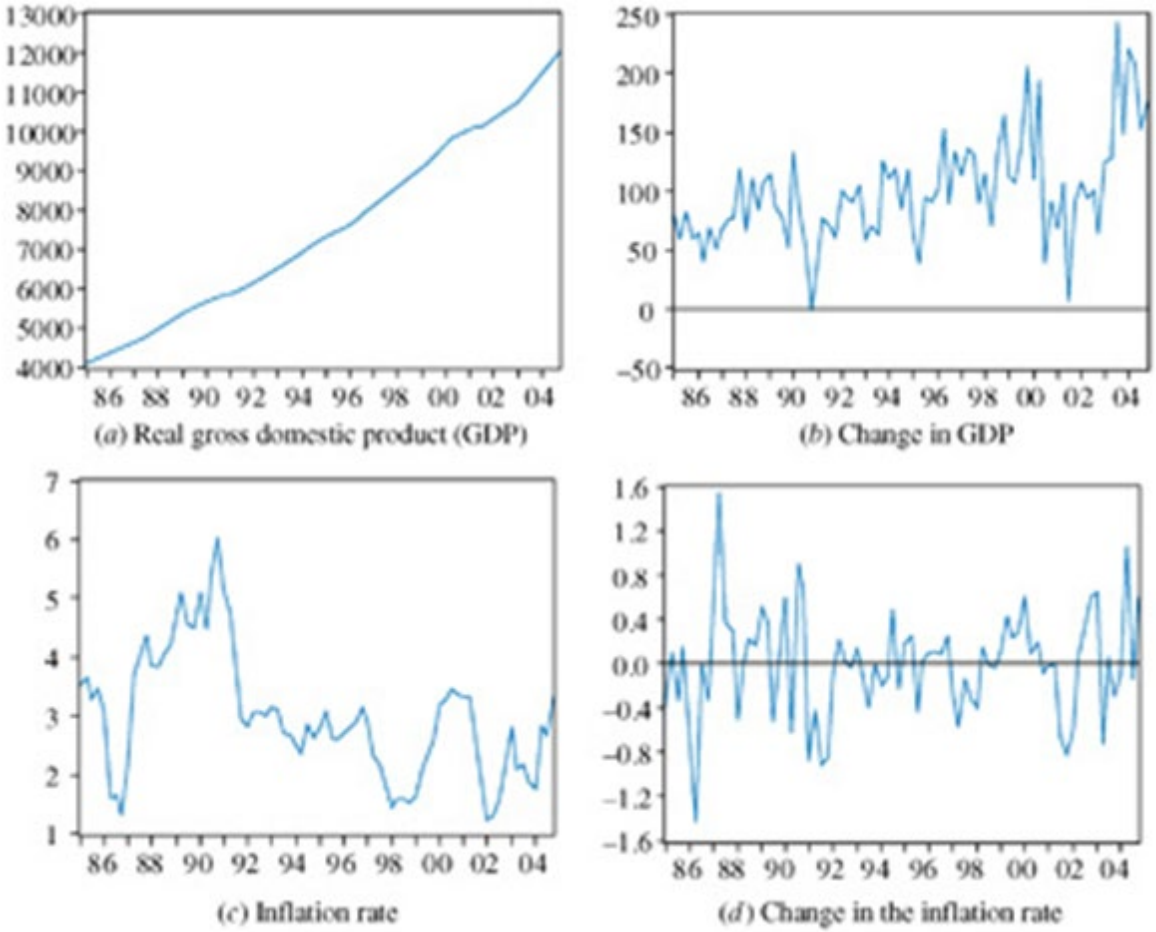
Stationarity and integration

- 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):

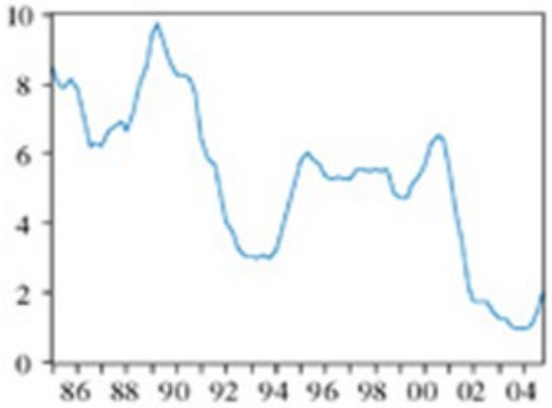
$$y_t \sim I(d) \text{ if and only if } \Delta^d y_t \sim I(0),$$
$$\text{whereas } \Delta^{d-1} y_t \text{ is still non-stationary.}$$

- 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**.

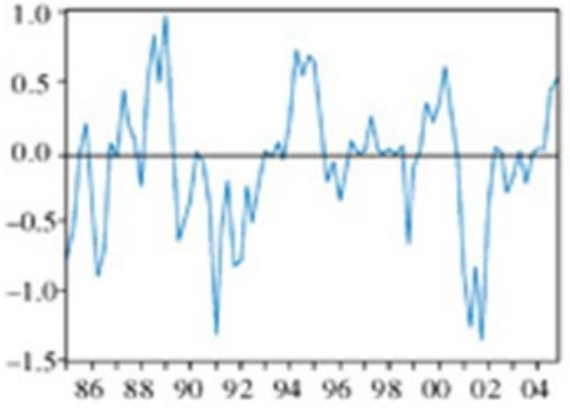
Examples of taking first differences



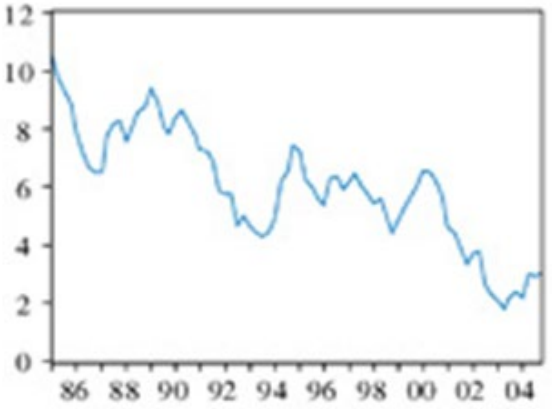
Examples of taking first differences



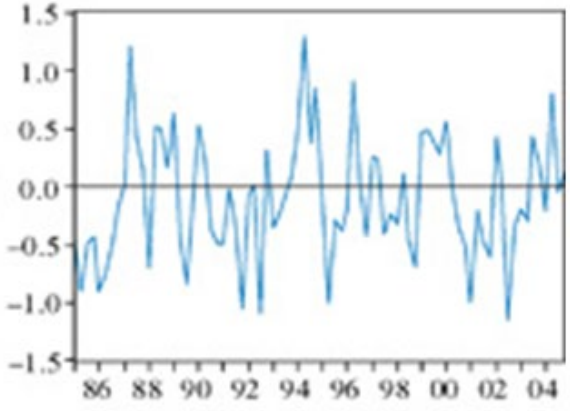
(e) Federal Funds rate



(f) Change in the Federal Funds rate



(g) 3-year Bond rate



(h) Change in the Bond rate

Dickey–Fuller test

Version 1 (no constant and no trend):

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

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

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

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

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

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

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):

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

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).
- 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:

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

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

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

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 is often facilitated by using the *Schwarz criterion* (SC) or the *modified AIC criterion* (MAIC).
- If the sample size is sufficient, the SC criterion is the preferred choice. Otherwise, the MAIC criterion is most often used.

Additional considerations

- 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 should in principle be ensured for every (dependent and explanatory) variable of a time-series regression model.**

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:

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

- 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:

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

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:

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

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

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

Autoregressive (AR) Processes

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

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

  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.

Autoregressive (AR) Processes

- **Wold's decomposition theorem** states that any stationary series can be decomposed into a pair of uncorrelated 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: $E(y_t) = \mu$

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

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

- 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.

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

- We shall introduce an *additional statistical concept*, particularly useful for distinguishing among "mixed" 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} , …, 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.

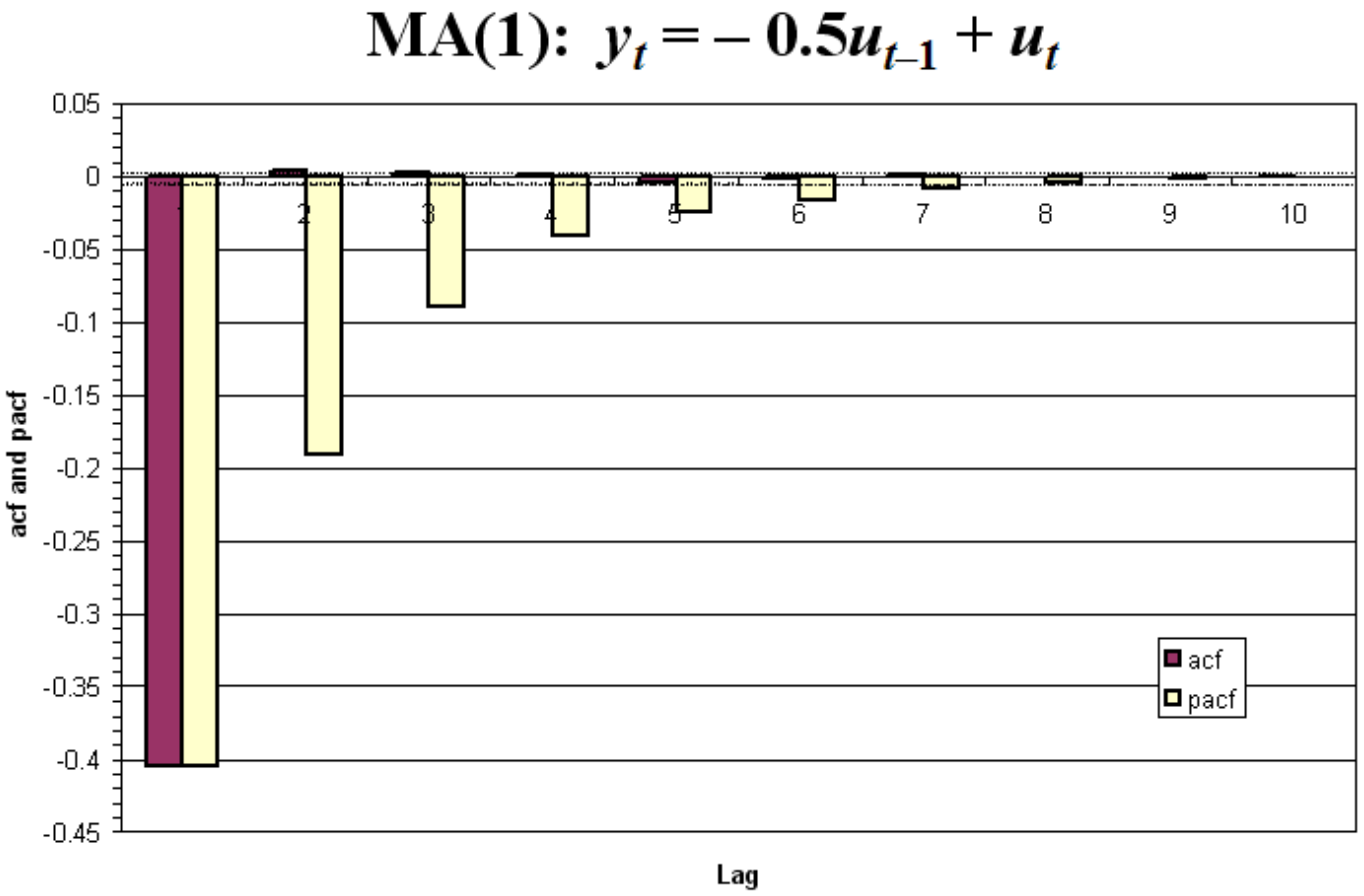
A moving average (MA) process has:

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

An autoregressive moving average (ARMA) process has:

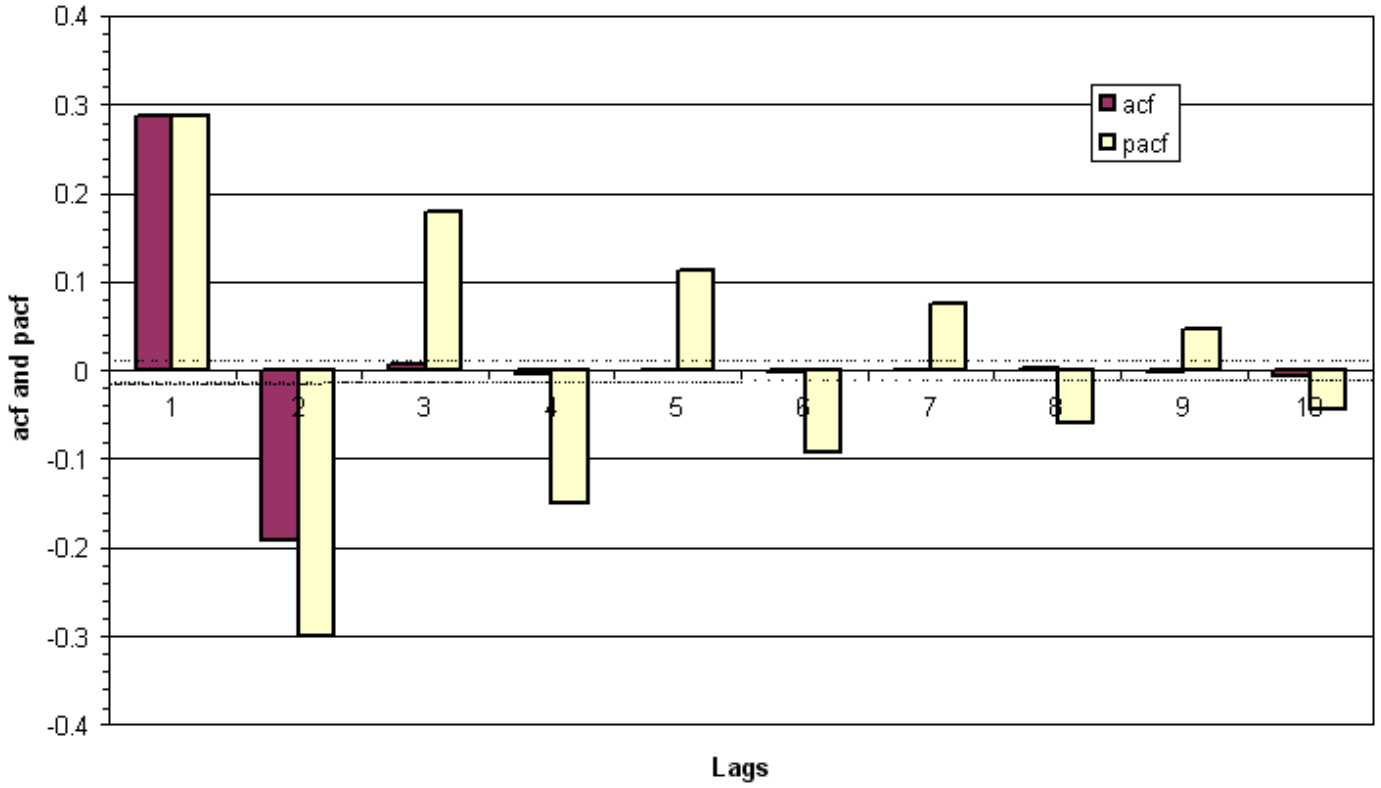
- a geometrically decaying ACF and
- a geometrically decaying PACF.

ACF and PACF plots for standard processes



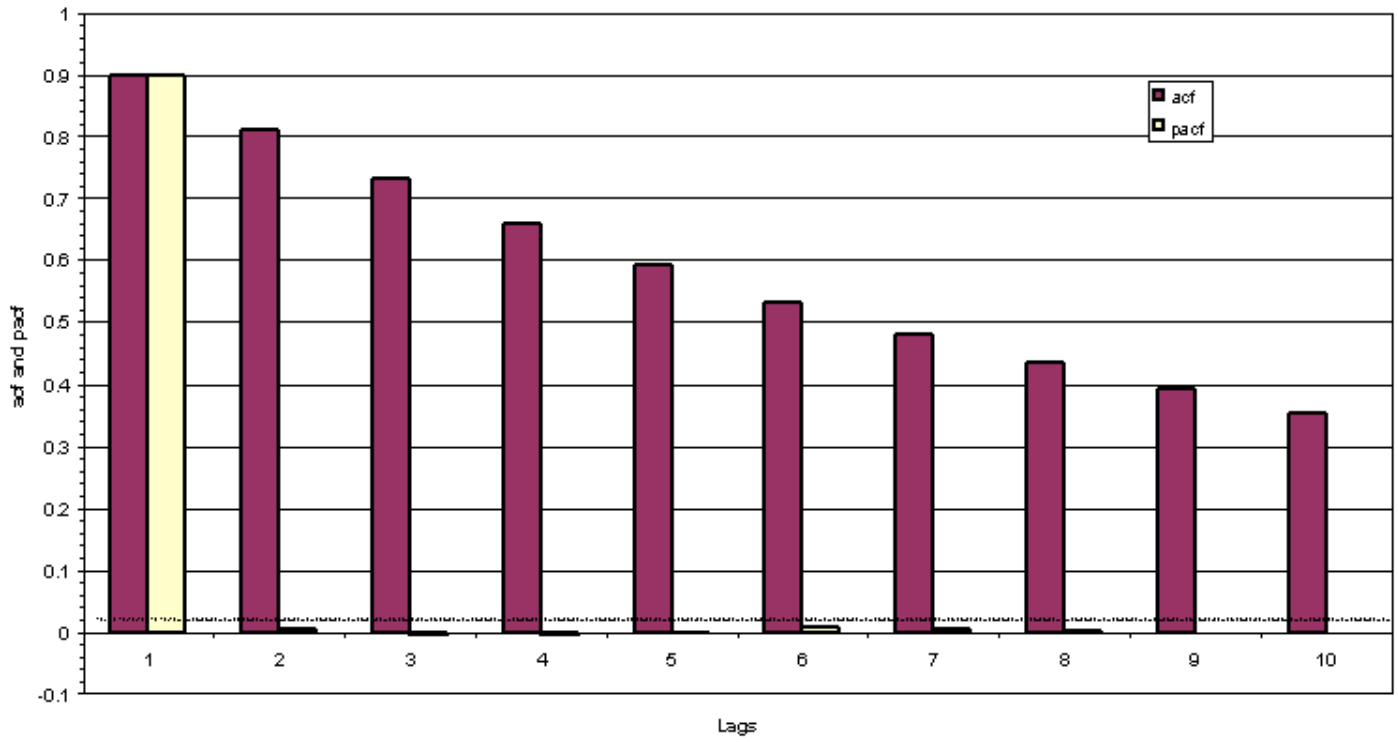
ACF and PACF plots for standard processes

$$\text{MA(2): } y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$$



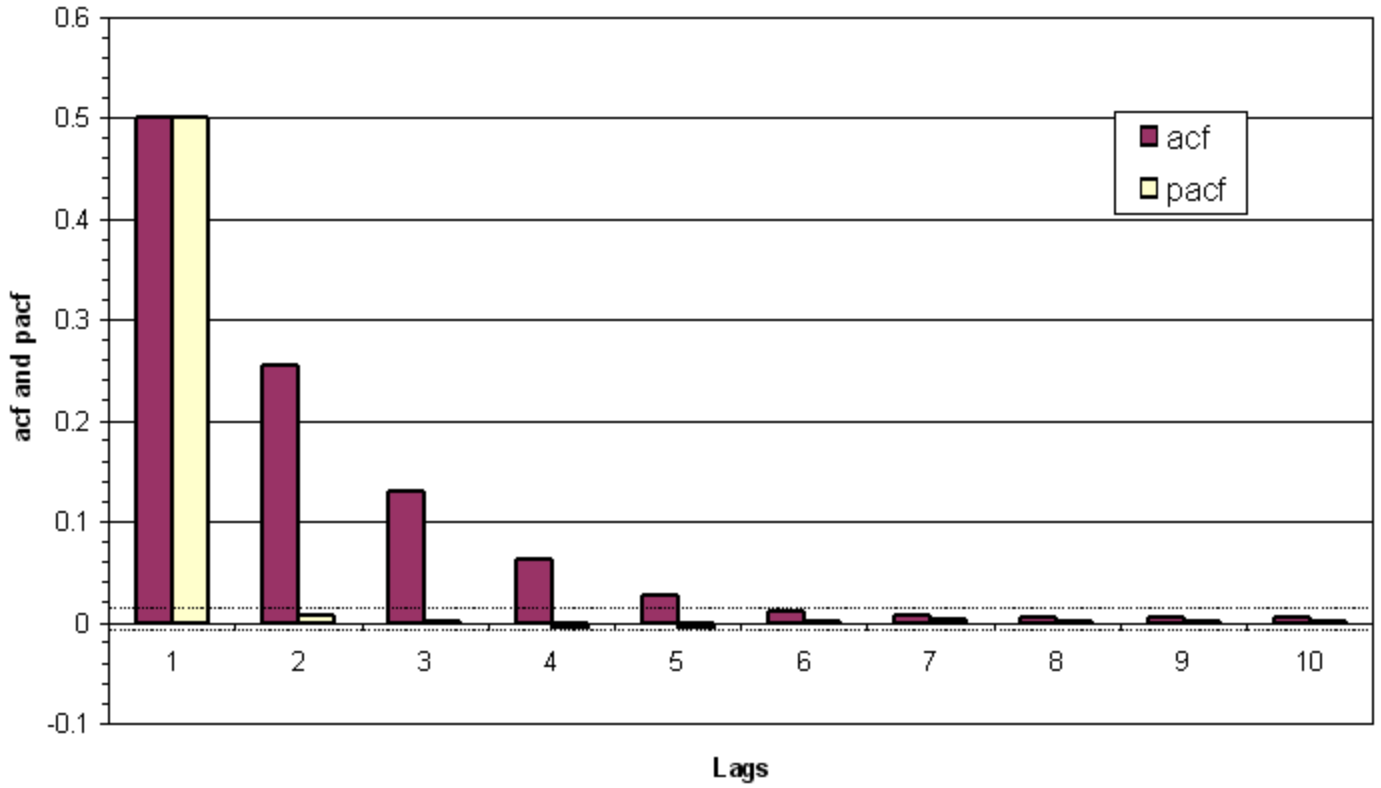
ACF and PACF plots for standard processes

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

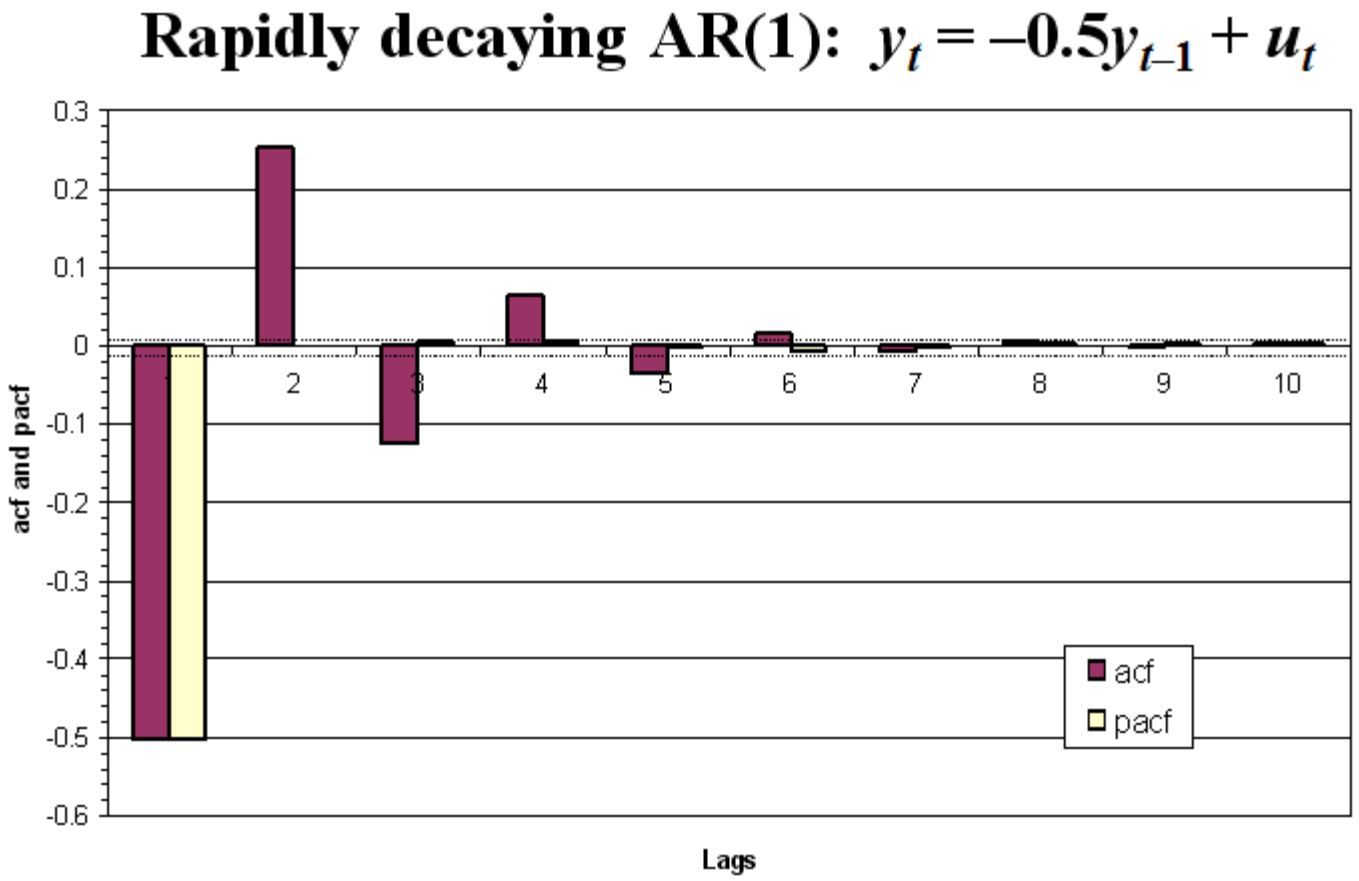


ACF and PACF plots for standard processes

Rapidly decaying AR(1): $y_t = 0.5y_{t-1} + u_t$

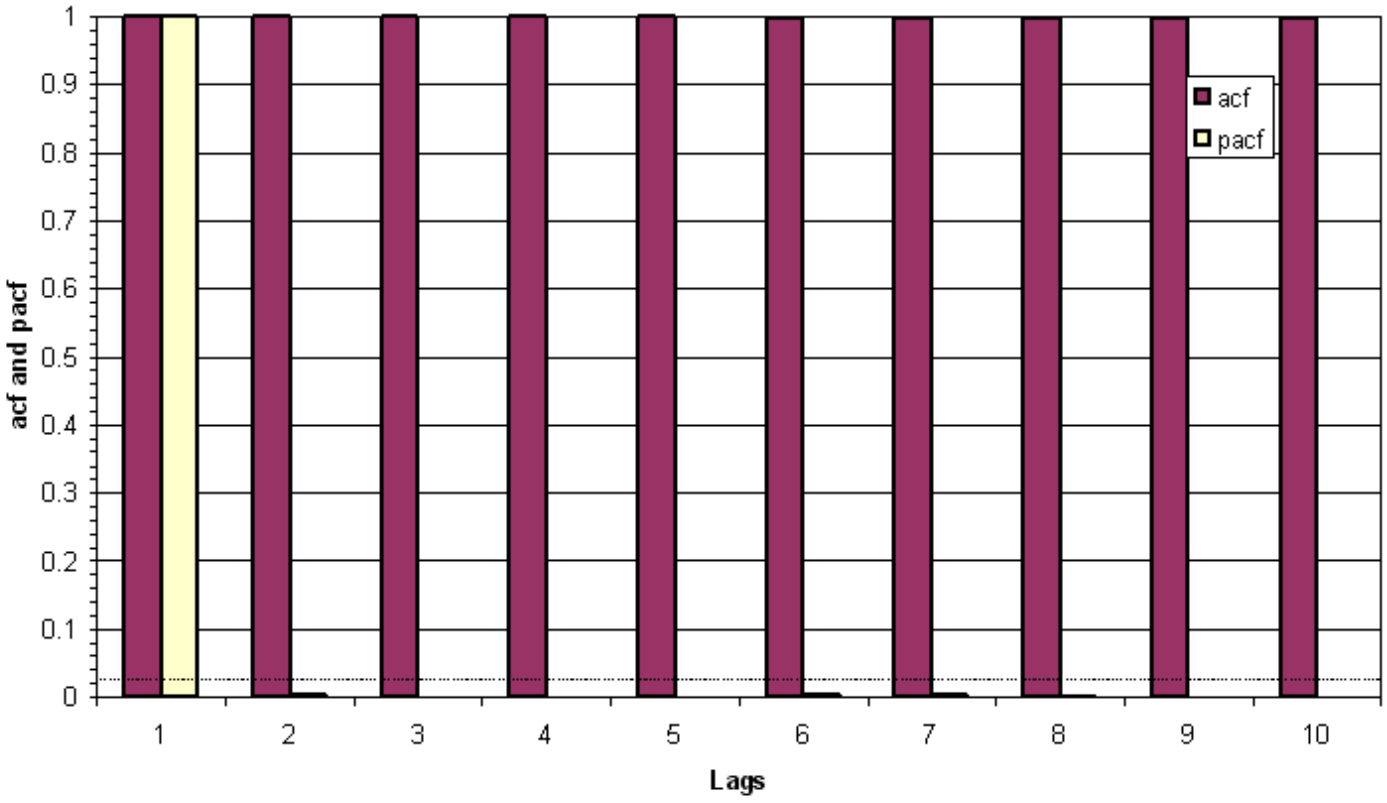


ACF and PACF plots for standard processes



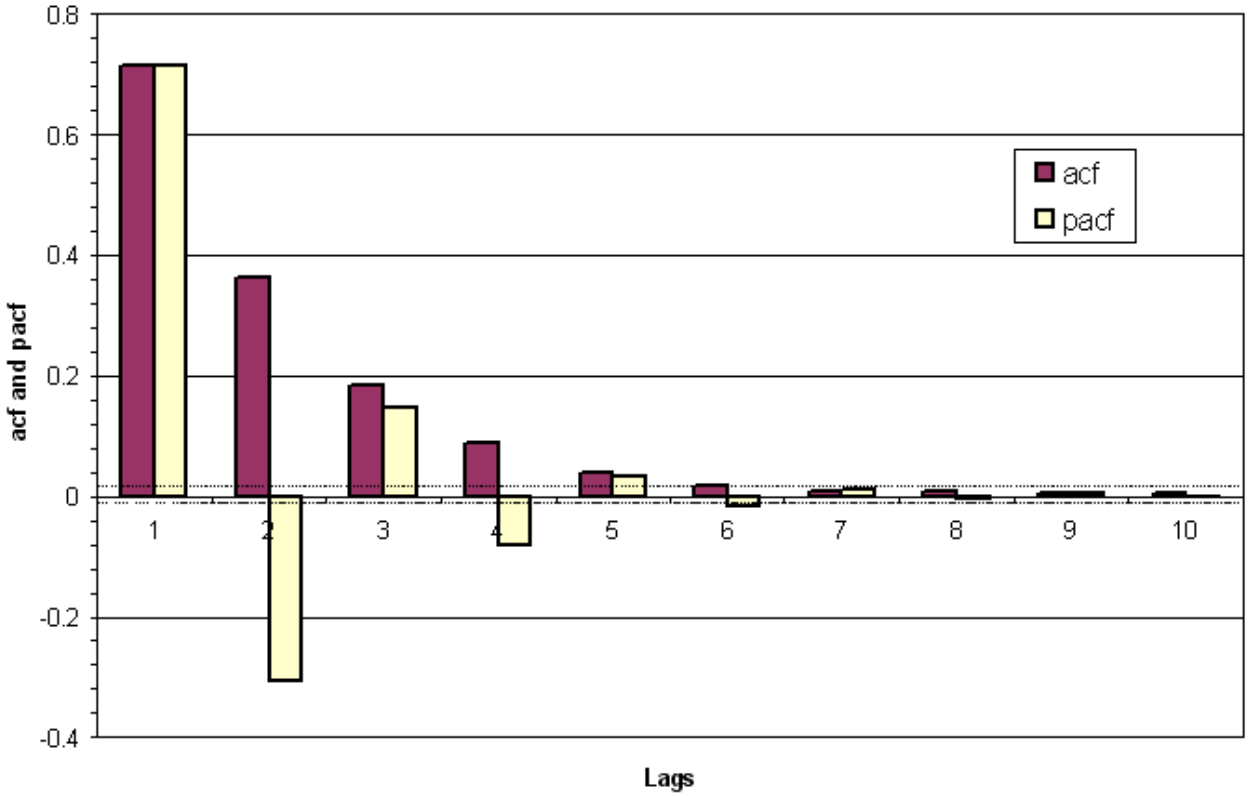
ACF and PACF plots for standard processes

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



ACF and PACF plots for standard processes

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



Box–Jenkins Approach

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.
- 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

- 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.
- 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. $p \le \bar{p}, q \le \bar{q}$.
- 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.

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, 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.

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:

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

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:

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

Basic definitions

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

$$\underset{g\times 1}{\mathbf{y}_t} = \underset{g\times 1}{\boldsymbol{\beta}_0} + \underset{g\times g}{\boldsymbol{\beta}_1}\,\underset{g\times 1}{\mathbf{y}_{t-1}} + \underset{g\times g}{\boldsymbol{\beta}_2}\,\underset{g\times 1}{\mathbf{y}_{t-2}} + \ldots + \underset{g\times g}{\boldsymbol{\beta}_k}\,\underset{g\times 1}{\mathbf{y}_{t-k}} + \underset{g\times 1}{\mathbf{u}_t}$$

- 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, not considered here).

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:

- 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:

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

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

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:

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

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

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

- We can write this as:

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

- 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

$$\underbrace{\mathbf{A}^{-1} \cdot / \mathbf{A}\, \mathbf{y}_t}_{\mathbf{I} \cdot \mathbf{y}_t = \mathbf{y}_t} = \beta_0 + \beta_1 \mathbf{y}_{t-1} + \mathbf{u}_t$$

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

$$\mathbf{y}_t = \mathbf{A}^{-1}\beta_0 + \mathbf{A}^{-1}\beta_1 \mathbf{y}_{t-1} + \mathbf{A}^{-1}\mathbf{u}_t$$

or

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

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

Structural versus standard form VARs

- First of all, due to simultaneity between y_1 and y_2, structural VAR models are not (directly) 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.
- The regression coefficients of a structural or primitive form VAR model thus have more firm theoretical interpretations, but this comes at a cost.

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:

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

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

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 $\delta_{22} = 0$

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

- 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:

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

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

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

- 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** of a variable is explained by shocks to each explanatory variable (to the disturbance term of the respective equation) 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.

The ordering of the variables

- What is done is to orthogonalise the disturbances, most often by applying the **Cholesky decomposition** on the covariance matrix of the model. 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 (contemporaneous) impact on all other $g - 1$ variables, the second variable may have immediate impact just on the remaining $g - 2$ variables etc.

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 the disturbances.
- In principle, orthogonalization can be done in at least two ways:
  - Orthogonalization matrix is constructed as the **Cholesky decomposition** of the estimated 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.
  - Alternatively, **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):

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{v}_t$$
$$\mathbf{y}_t - \mathbf{A}_1\mathbf{y}_{t-1} - \mathbf{A}_2\mathbf{y}_{t-2} - \ldots - \mathbf{A}_p\mathbf{y}_{t-p} = \mathbf{v}_t$$
$$\mathbf{y}_t - \mathbf{A}_1L\mathbf{y}_t - \mathbf{A}_2L^2\mathbf{y}_t - \ldots - \mathbf{A}_pL^p\mathbf{y}_t = \mathbf{v}_t$$
$$\left(\mathbf{I}_g - \mathbf{A}_1L - \mathbf{A}_2L^2 - \ldots - \mathbf{A}_pL^p\right)\mathbf{y}_t = \mathbf{v}_t$$

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

$$\mathbf{A}\left(\mathbf{I}_g - \mathbf{A}_1L - \mathbf{A}_2L^2 - \ldots - \mathbf{A}_pL^p\right)\mathbf{y}_t = \mathbf{A}\mathbf{v}_t = \mathbf{B}\tilde{\mathbf{v}}_t$$

The vector $\mathbf{v}_t$ refers to the disturbances of the reduced-form model (with covariance matrix $\boldsymbol{\Sigma}$), while the vector $\tilde{\mathbf{v}}_t$ represents the orthogonalized, structural disturbances (with covariance matrix $\mathbf{I}_g$).

Short-run SVAR model

The matrix **A** represents the contemporanouos (short-run) effects among $\mathbf{y}_t$, while the (diagonal) matrix **B** scales the structural disturbances $\tilde{\mathbf{v}}_t$ to the disturbances of the reduced-form model $\mathbf{v}_t$.

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

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

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

is then related to the covariance matrix $\boldsymbol{\Sigma}$, such that $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1}\mathbf{B})^T$.

As there are $g(g+1)/2$ free parameters in $\boldsymbol{\Sigma}$, 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.

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**:

$$\begin{aligned}
\overline{\mathbf{A}}\,\mathbf{y}_t &= \mathbf{B}\tilde{\mathbf{v}}_t \\
\overline{\mathbf{A}}^{-1}\overline{\mathbf{A}}\,\mathbf{y}_t &= \overline{\mathbf{A}}^{-1}\mathbf{B}\tilde{\mathbf{v}}_t \\
\mathbf{y}_t &= \overline{\mathbf{A}}^{-1}\mathbf{B}\tilde{\mathbf{v}}_t \\
\mathbf{y}_t &= \mathbf{C}\tilde{\mathbf{v}}_t
\end{aligned}$$

In a long-run SVAR, constraints are placed on elements of the **C** matrix, which is a matrix of long-run effects.

Long-run SVAR model

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.

Again, as there are $g(g + 1)/2$ free parameters in $\mathbf{\Sigma}$, only that many parameters may be estimated in the $\mathbf{C}$ matrix. As there are g^2 parameters in $\mathbf{C}$, 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 $\mathbf{A}$ in terms of the elements of matrix $\mathbf{B}$, or vice versa (at least not without additional constraint matrices);
- We cannot mix short-run constraints (those on matrices $\mathbf{A}$ and $\mathbf{B}$) and long-run constraints (those on the matrix $\mathbf{C}$).

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

8.4 Modelling Volatility

EFMD EQUIS ACCREDITED

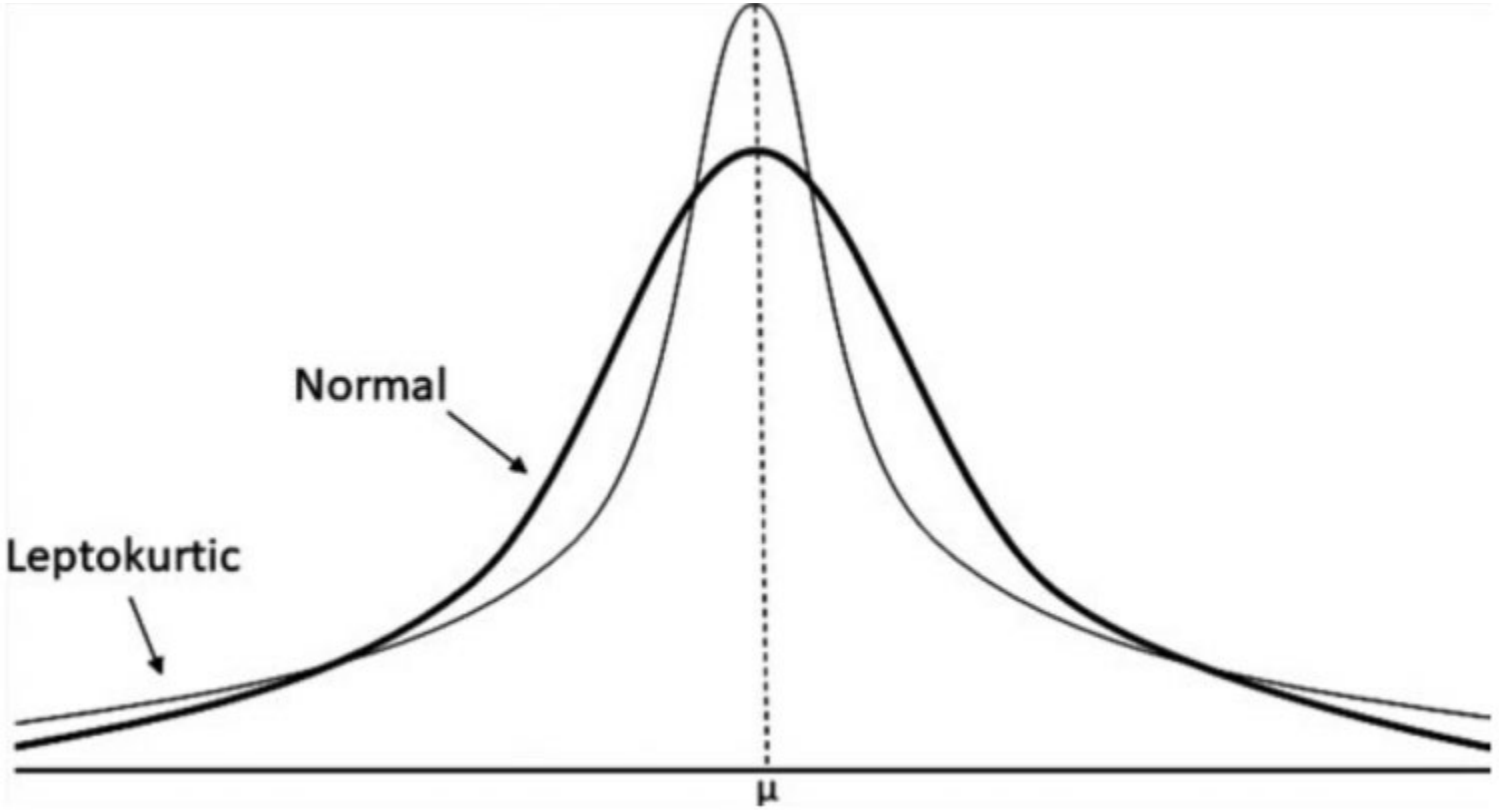
AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Basic definitions

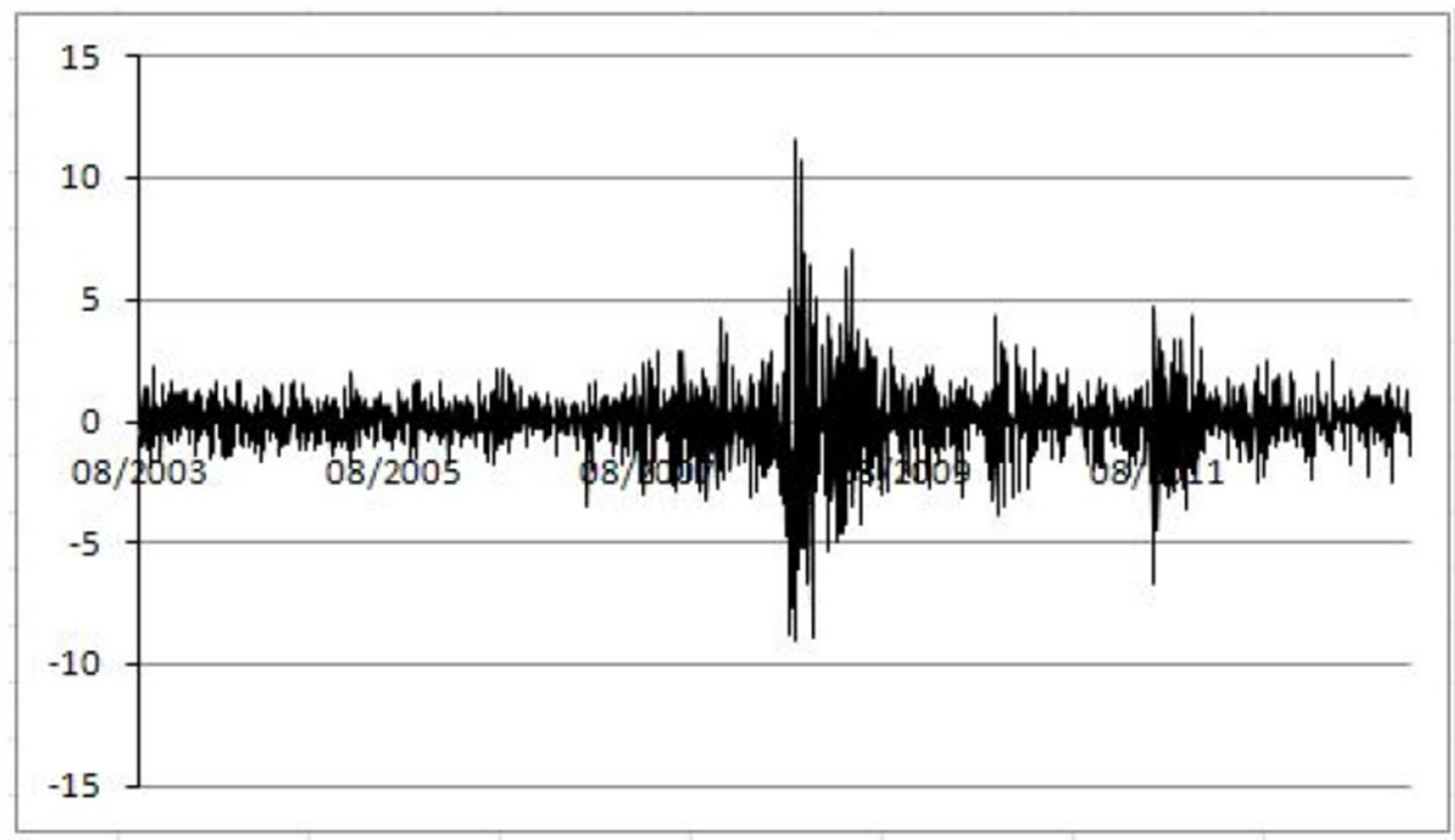
- So far, most of the studied regression models were linear, at least in parameters (regression coefficients).
- However, linear models *cannot* explain a number of important features common to some economic and much of financial data:
  - **leptokurtosis**: tendency for time series to have distributions 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 (*negative* leverage) or the opposite (*positive* leverage).

Basic definitions



Basic definitions

Daily S&P 500 Returns for August 2003 – August 2019



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 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

- Lets us start with one of the key deviations from the classical linear regression model – heteroscedasticity, where $\text{Var}(u_t) \neq \sigma_u^2$.
- 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.:

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

- 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:

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

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.
- Instead of calling the variance σ_t^2, in the literature it is usually called h_t, so the model is:

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

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:

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \quad u_t = v_t \sigma_t,$$

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}, \quad v_t \sim \mathrm{N}(0, 1).$$

- 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: $\alpha_i \geq 0$, $\forall i = 0, 1, 2, \ldots, q$.

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, $\hat{u}_t$.

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 $\chi^2(q)$.

Testing for ARCH effects

4. The null and alternative hypotheses are:

$$H_0 : \gamma_j = 0, \forall j = 1, \ldots, q;$$
$$H_1 : \gamma_j \neq 0, \exists j = 1, \ldots, q.$$

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.

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.
- 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*.

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:

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

- The current fitted variance, $\hat{h}_t = \widehat{\sigma_t^2}$, is interpreted as a *weighted function* of:
  - a long-term average value (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.

GARCH model

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

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

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

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

GARCH model

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

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

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

  - for $\alpha_1 + \beta \geq 1$ we obtain non-stationarity in variance and
  - for $\alpha_1 + \beta = 1$ 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).

Extensions to the basic GARCH model

- Two key issues with GARCH(q,p) models:
  - *non-negativity constraints* may still be violated;
  - 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.

GJR model

- The **GJR model** is due to Glosten, Jaganathan and Runkle (1993). The variance equation has a *treshold term*:

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

where $I_{t-1} = 1$ if $u_{t-1} < 0$;
$= 0$ otherwise.

- For a negative leverage effect, we would observe $\gamma > 0$, and vice versa.
- We *still* require $\alpha_1 + \gamma \geq 0$ and $\alpha_1 \geq 0$ for non-negativity.
- This is how it is programmed in the R package `rugarch`.
- In the Stata package `arch`, the indicator variable is defined in reverse: $I_{t-1} = 1$ if $u_{t-1} > 0$, and 0 otherwise. There, we would observe $\gamma < 0$ for a negative leverage effect, and vice versa.

EGARCH model

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

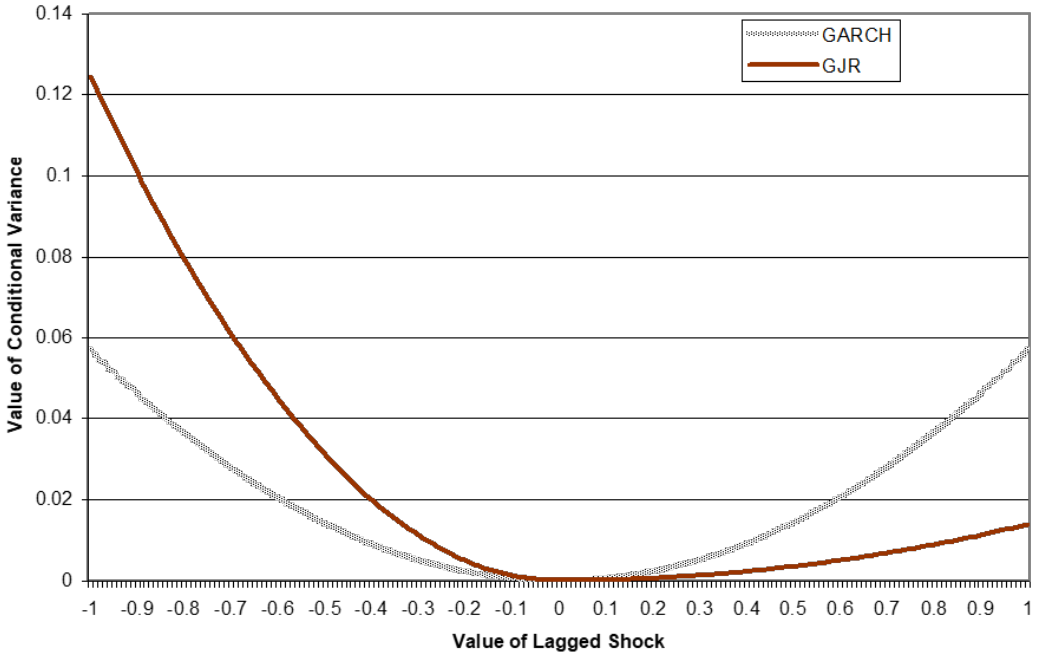
$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \alpha \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \gamma \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$$

- 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.
  - Parameter α accounts for the leverage effect, as if the relationship between volatility and the dependent variable in the mean equation (e.g. returns) is negative, α will be negative, indicating negative leverage, and vice versa for positive leverage.
  - Parameter γ accounts for the size of shocks (symmetric effect).

Asymmetry in GARCH models

The **news-impact curve** or **news-response curve** plots the 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:



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 stock be partly determined by its risk?
- This leads to another extension of the basic GARCH model, called the **GARCH-in-mean (GARCH–M) model**:

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

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

- 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 *volatility or 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 the variance of the disturbances u_t, 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**:
  - 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* (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 2025