Lecture: Stationary Time Series Analysis

222061-1617: Time Series Econometrics

Jacek Suda

Website

Class website: http://www.jaceksuda.com/teach/

Email: jacek.suda@sgh.waw.pl

Office hours: by appointment

Description

The purpose of this course is to familiarize students with current techniques used in macroeconomic time series models with applications in macroeconomics, international finance, and finance; with the ultimate aim of providing necessary tools to conduct original research in the area.

The focus is on implementation of the models presented in the course.

Topics include ARMA models, (B/S)VARs and impulse response functions; local projection; unit roots, and structural breaks; spurious regressions; cointegration and VECM; ARCH models of volatility, and trend/cycle decomposition methods, including Kalman filtering.

We will mostly work with the classical framework in the time domain but will also talk about Bayesian approach.

Prerequisites

I will assume some familiarity with matrix algebra and introductory statistics and econometrics.

The course is the continuation of “Econometrics” course but we will also review the univariate time series analysis in this course.

Requirements

Attending lectures and participating in classroom discussions is essential to the learning process.

There will be two homeworks and a paper-reading assignment.

The class ends with a take-home exam.

The assignments and exam will require the use of econometrics software.

The weights in determining your grade are given as follows:

Homework assignments	30%
Paper presentation	30%
Final exam	40%

Readings

The book that covers most of the material is

- *Time Series Analysis* by James D. Hamilton, Princeton University Press, 1994.

Other texts are

- *State-Space Models with Regime Switching* by Chiang-Jin Kim and Charles R. Nelson, MIT Press, 1999.
- *New Introduction to Multiple Time Series Analysis* by Helmut Lütkepohl, Springer-Verlag, 2005.
- *Introduction to Bayesian Econometrics* by Edward Greenberg, Cambridge University Press, 2007.
- *Time Series and Panel Data Econometrics* by M. Hashem Pesaran, Oxford University Press, 2015.
- *Applied Econometric Time Series* by Walter Enders, Wiley, 2010.

The readings include journal articles and chapters from the above books.

Outline

1. Stationary Time Series Analys
    - Overview of ARMA models
    - State-Space Representation
    - Kalman Filter
2. Structural Analysis
   Granger Causality, VAR, IRFs, Estimation, Variance decomposition
    - Reduced-form and structural VAR models
    - Bayesian VAR models
    - Jordà's local projection
3. Unit Roots and Structural Breaks
    - Unit root tests
    - Structural break tests
    - Trend/cycle decomposition
    - Cointegration
    - VEC models
4. Nonlinearity *(time permitting)*
    - ARCH/GARCH models
    - Markov switching
    - Time-varying parameters
    - Gibbs sampling
    - Threshold models

Today

Outline:

1. Covariance-Stationary Processes
2. Wold Decomposition Theorem
3. ARMA Models
    1. Auto-Correlation Function (ACF)
    2. Partial Auto-Correlation Function (PACF)
    3. Model Selection
    4. Estimation
    5. Forecasting

STATIONARITY

Stochastic Process

- Stochastic process: a collection of random variables

$$\{\ldots, Y_{-1}, Y_0, Y_1, Y_2, \ldots, Y_T, \ldots\} = \{Y_t\}_{-\infty}^{\infty}.$$

- Observed series $\{y_1, y_2, \ldots, y_T\}$ – realizations of a stochastic process.

  We want a model for $\{Y_t\}_{\infty}^{-\infty}$ to explain observed realizations $\{y_t\}_1^T$.

Covariance-Stationary

Definition

$\{Y_t\}$ *is covariance-stationary (weak stationary) if*

(i) $E[Y_t] = \mu \quad \forall t$

(ii) $Cov(Y_t, Y_{t-j}) = E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \quad \forall t, j$

Note: mean is time-invariant

Covariance-Stationary

Definition

$\{Y_t\}$ *is covariance-stationary (weak stationary) if*

(i) $E[Y_t] = \mu \quad \forall t$

(ii) $Cov(Y_t, Y_{t-j}) = E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \quad \forall t, j$

Note: covariance doesn't depend on t

Covariance-Stationary

Definition

$\{Y_t\}$ is covariance-stationary (weak stationary) if

(i) $E[Y_t] = \mu \quad \forall t$

(ii) $Cov(Y_t, Y_{t-j}) = E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \quad \forall t, j$

<u>Note:</u> $Var(Y_t) = \gamma_0$ – variance is also constant.

Covariance-Stationary

Definition

$\{Y_t\}$ *is covariance-stationary (weak stationary) if*

(i) $E[Y_t] = \mu \quad \forall t$

(ii) $Cov(Y_t, Y_{t-j}) = E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \quad \forall t, j$

- It is weak stationarity because it only relates to the first two moments. Higher moments can be time-variant.

Examples

1. $\{Y_t\}$ with $E(Y_t) = 0$, $Var(Y_t) = \sigma^2$, and $E(Y_t Y_\tau) = 0 \Rightarrow$ white noise (WN).
2. $Y_t \sim iid(0, \sigma^2) \Rightarrow$ independent white noise.
3. $Y_t \sim iidN(0, \sigma^2) \Rightarrow$ Gaussian white noise.

Strict (Strong) Stationary

Definition

$\{Y_t\}$ is (strictly/strongly) stationary if for any values of $j_1, j_2, \ldots, j_n$ the joint distribution of $(Y_t, Y_{t+j_1}, Y_{t+j_2}, \ldots, Y_{t+j_n})$ depends only on the intervals separating the dates $(j_1, j_2, \ldots, j_n)$ and not on a date itself (t).

- For all τ, $t_1, t_2, \ldots, t_n$:

$$F_Y(y_{t_1}, y_{t_2}, \ldots, y_{t_n}) = F_Y(y_{t_1+\tau}, y_{t_2+\tau}, \ldots, y_{t_n+\tau})$$

- If a process is strictly stationary with a finite second moment it is also covariance-stationary.
- Normality ⇒ strong stationarity: whole distribution depends on the first two moments.

Nonstationary Processes

Examples:

1. $Y_t = \beta \cdot t + \varepsilon_t, \qquad \varepsilon_t \sim WN$

Nonstationary Processes

Examples:

1. $Y_t = \beta \cdot t + \varepsilon_t, \qquad \varepsilon_t \sim WN$

t - time dummy

Nonstationary Processes

Examples:

1. $Y_t = \beta \cdot t + \varepsilon_t, \qquad \varepsilon_t \sim WN$

deterministic part

Nonstationary Processes

Examples:

1. $Y_t = \beta \cdot t + \varepsilon_t, \qquad \varepsilon_t \sim WN$

stochastic component

Nonstationary Processes

Examples:

1. $Y_t = \beta \cdot t + \varepsilon_t, \qquad \varepsilon_t \sim WN$
    - $E[Y_t] = \beta \cdot t$ depends on t.
    - But $X_t = Y_t - \beta \cdot t$ is covariance stationary.

Nonstationary Processes

Examples:

1. $Y_t = \beta \cdot t + \varepsilon_t, \qquad \varepsilon_t \sim WN$
    - $E[Y_t] = \beta \cdot t$ depends on t.
    - But $X_t = Y_t - \beta \cdot t$ is covariance stationary.
2. $Y_t = Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN, Y_0$ constant
    - Random walk.

Nonstationary Processes

Examples:

1. $Y_t = \beta \cdot t + \varepsilon_t, \qquad \varepsilon_t \sim WN$
   - $E[Y_t] = \beta \cdot t$ depends on t.
   - But $X_t = Y_t - \beta \cdot t$ is covariance stationary.

2. $Y_t = Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN, Y_0$ constant
   - Random walk.
   - Solving recursively :

$$Y_t = \sum_{j=1}^{t} \varepsilon_j + Y_0.$$

   - $E[Y_t] = Y_0$, time-invariant mean.
   - But $Var(Y_t) = t \cdot \sigma^2$ depends on t.
   - $X_t = Y_t - Y_{t-1}$ is covariance stationary.

Wold's Decomposition Theorem

Any covariance stationary $\{Y_t\}$ has infinite order, moving-average representation:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t, \qquad \psi_0 = 1, \varepsilon_t \sim WN.$$

- $\psi_0 = 1, \sum_{j=0}^{\infty} \psi_j^2 < \infty,$
- $\varepsilon_t \sim WN(0, \sigma^2), E(\varepsilon_t \kappa_s) = 0 \quad \forall_{t,s},$
- κ_t deterministic term (perfectly forecastable).
- Example:
    - $\kappa_t = \mu$, constant mean.
    - $\kappa = a\cos(\lambda t) + b\sin(\lambda t)$, λ fixed number, $Ea = Eb = Eab = 0$.

Wold's Decomposition Theorem

Any covariance stationary $\{Y_t\}$ has infinite order, moving-average representation:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t, \qquad \psi_0 = 1, \varepsilon_t \sim WN.$$

- $\psi_0 = 1, \sum_{j=0}^{\infty} \psi_j^2 < \infty,$
- $\varepsilon_t \sim WN(0, \sigma^2), E(\varepsilon_t \kappa_s) = 0 \quad \forall_{t,s},$
- κ_t deterministic term (perfectly forecastable).

- Linear combination of ε_s (innovations over time),
- $\varepsilon_t = Y_t - E(Y_t | Y_{t-1}, Y_{t-2}, \dots),$
- Weights does not depend on time t, they only depend on j, i.e. how long ago the shock ε occurred.

Wold's Decomposition Theorem - Illustration

Let $X_t = Y_t - \kappa_t$ (purely indeterministic process). Then,

$$\begin{aligned}
E[X_t] &= \sum_{j=0}^{\infty} \psi_j E[\varepsilon_{t-j}] = 0, \\
E[X_t^2] &= \sum_{j=0}^{\infty} \psi_j^2 E[\varepsilon_{t-j}^2] = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty, \\
&\quad \text{as } \varepsilon_t \text{ are uncorrelated; we have constant finite variance.} \\
E[X_t \cdot X_{t-j}] &= E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots)(\varepsilon_{t-j} + \psi_1 \varepsilon_{t-j-1} + \psi_2 \varepsilon_{t-j-2} + \ldots)] \\
&= \sigma^2 (\psi_j + \psi_{j+1}\psi_1 + \psi_{j+2}\psi_2 + \ldots) \\
&= \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+j}, \qquad \text{depends on } j \text{ not } t.
\end{aligned}$$

So we have a covariance stationary process in mean and variance.

ARMA

ARMA Models

- Approximate Wold form with finite number of parameters.

  Wold Form:

$$Y_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \varepsilon_t \sim WN,$$

  ARMA(p,q):

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \ldots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}.$$

Lag Operator

Define the operator L as

$$\begin{aligned} LX_t &\equiv X_{t-1}, \\ L^2 X_t &= L \cdot LX_t = X_{t-2}. \end{aligned}$$

In general,

$$L^k X_t = X_{t-k}.$$

If c is a constant,

$$Lc = c.$$

Also,

$$\begin{aligned} L^{-1} X_t &\equiv X_{t+1}, \\ \Delta X_t &= (1 - L) X_t = X_t - X_{t-1}. \end{aligned}$$

Lag Operator

It satisfies

$$\begin{aligned} L(\alpha X_t + \beta Y_t) &= \alpha X_{t-1} + \beta Y_{t-1} \\ (a\,L + b\,L^2)\,X_t &= aX_{t-1} + bX_{t-2}, \end{aligned}$$

and, when $|\phi| < 1$,

$$\lim_{j\to\infty} (1 + \phi L + \phi^2 L^2 + \ldots + \phi^j L^j) = (1 - \phi L)^{-1}$$

ARMA Models in Lag Notaion

- ARMA(p,q):

  $$Y_t-\mu=\phi_1(Y_{t-1}-\mu)+\ldots+\phi_p(Y_{t-p}-\mu)+\varepsilon_t+\theta_1\varepsilon_{t-1}+\ldots+\theta_q\varepsilon_{t-q},$$

  or

  $$Y_t-\mu-\phi_1(Y_{t-1}-\mu)-\ldots-\phi_p(Y_{t-p}-\mu)=\varepsilon_t+\theta_1\varepsilon_{t-1}+\ldots+\theta_q\varepsilon_{t-q},$$

- With lag operator:

  $$\phi(L)(Y_t-\mu) \quad = \quad \theta(L)\varepsilon_t,$$

  where

  $$\begin{aligned}\phi(L) &= 1-\phi_1L-\phi_2L^2-\ldots-\phi_pL^p,\\ \theta(L) &= 1+\theta_1L+\theta_2L^2+\ldots+\theta_qL^q.\end{aligned}$$

Stochastic Difference Equation (SDE) Representation

Let $X_t = Y_t - \mu$ and $w_t = \theta(L)\varepsilon_t$.

Then

$$\phi(L)X_t = w_t,$$

or

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + w_t,$$

is a pth-order stochastic difference equation.

SDE Representation - AR(1) Example

Example: First-order SDE (AR(1)):

$$X_t = \phi X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN$$

- Solve for Wold Form (recursive substitution)

$$\begin{aligned} X_t &= \phi^{t+1} X_{-1} + \phi^t \varepsilon_0 + \phi^{t-1} \varepsilon_1 + \ldots + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \phi^{t+1} X_{-1} + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}. \end{aligned}$$

  where X_{-1} is an initial condition and $\psi_i = \phi^i$.
- We approximated Wold form with 1 parameter form for AR(1).

Dynamic Multiplier

The dynamic multiplier measurers the effect of ε_t on subsequent values of X_τ:

$$\frac{\partial X_{t+j}}{\partial \varepsilon_t} = \frac{\partial X_j}{\partial \varepsilon_0} = \psi_j.$$

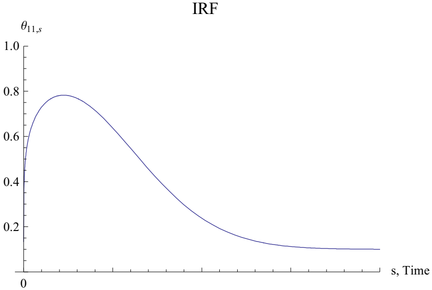
For the X_t being AR(1) process

$$\frac{\partial X_{t+j}}{\partial \varepsilon_t} = \psi_j = \phi^j.$$

The dynamic multiplier for any linear difference equation depends only on the length of time j, not on time t.

Impulse Response Function

The impulse-response function is a sequence of dynamic multipliers as a function of time from the one time impulse to ε_t



Cumulative impact

- Permanent increase in ε at time t, i.e. $\varepsilon_t = 1, \varepsilon_{t+1} = 1, \varepsilon_{t+2} = 1, \ldots$

$$\frac{\partial X_{t+j}}{\partial \varepsilon_t} + \frac{\partial X_{t+j}}{\partial \varepsilon_{t+1}} + \frac{\partial X_{t+j}}{\partial \varepsilon_{t+2}} + \ldots + \frac{\partial X_{t+j}}{\partial \varepsilon_{t+j}} = \psi_j + \psi_{j-1} + \ldots + \psi + 1$$

or

$$\frac{\partial X_t}{\partial \varepsilon_t} + \frac{\partial X_{t+1}}{\partial \varepsilon_t} + \frac{\partial X_{t+2}}{\partial \varepsilon_t} + \ldots + \frac{\partial X_{t+j}}{\partial \varepsilon_t} = \psi_j + \psi_{j-1} + \ldots + \psi + 1$$

- In the limit, as $j \to \infty$

$$\lim_{j\to\infty} \left[\frac{\partial X_{t+j}}{\partial \varepsilon_t} + \frac{\partial X_{t+j}}{\partial \varepsilon_{t+1}} + \ldots + \frac{\partial X_{t+j}}{\partial \varepsilon_{t+j}} \right] = \sum_{j=0}^{\infty} \psi_j = \psi(1),$$

where

$$\psi(1) = \psi(L = 1) = 1 + \psi_1 + \psi_2 + \ldots$$

AUTOREGRESSIVE PROCESS

AR(1) Model

Recall

$$\begin{aligned} X_t &= \phi X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN, \\ &= \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \end{aligned}$$

Wold coefficients

$$\psi_j = \phi \psi_{j-1}, \qquad \psi_j = \frac{\partial Y_{t+j}}{\partial \varepsilon_t}$$

- If $|\phi| < 1$ X_t is a stable and stationary solution to first-order SDE.
- If $\phi = 1$ then $\psi_j = 1\ \forall_j$, $X_t = X_{-1} + \sum_{j=0}^{t} \varepsilon_j$ is neither stationary nor stable solution, and $\psi(1)$ is infinite.

AR(1): Lag notation

AR(1):

$$\begin{aligned} X_t &= \phi X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN, \\ (1-\phi L)X_t &= \varepsilon_t \end{aligned}$$

Multiply both sides by $(1-\phi L)^{-1}$:

$$\begin{aligned} X_t &= (1-\phi L)^{-1}\varepsilon_t \\ &= (1+\phi L+\phi^2 L^2+\phi^3 L^3+\ldots)\varepsilon_t \\ &= \sum_{j=0}^{\infty}\phi^j\varepsilon_{t-j} = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}, \\ \psi(L) &= (1-\phi L)^{-1}. \end{aligned}$$

AR(1): Long-run effects

For AR(1), if $|\phi| < 1$

- the permanent increase in ε_t equals

$$\frac{\partial X_{t+j}}{\partial \varepsilon_t} + \frac{\partial X_{t+j}}{\partial \varepsilon_{t+1}} + \ldots + \frac{\partial X_{t+j}}{\partial \varepsilon_{t+j}} = 1 + \phi + \phi^2 + \phi^3 + \ldots + \phi^j.$$

  and as $j \to \infty$

$$\psi(1) = 1 + \phi + \phi^2 + \ldots = 1/(1-\phi)$$

- the cumulative consequences for X of a one-time change in ε,

$$\sum_{j=0}^{\infty} \frac{\partial X_{t+j}}{\partial \varepsilon_t} = 1/(1-\phi)$$

AR(1): Mean

- Intercept representation for $Y_t \equiv X_t + \mu$

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t.$$

- Mean

$$E[Y_t] = c + \phi E[Y_{t-1}] + E[\varepsilon_t],$$

Since we have covariance stationary process, $E[Y_t] = E[Y_{t-1}]$ and

$$E[Y_t] = \frac{c}{1-\phi} \equiv \mu, \qquad c = \mu(1-\phi).$$

AR(1): Variance

- Variance

$$\begin{aligned} Var(Y_t) &= E[(Y_t - \mu)^2] \\ &= E[(\phi(Y_{t-1} - \mu) + \varepsilon_t)^2] \\ &= \phi^2 E[(Y_{t-1} - \mu)^2] + 2\phi E[(Y_{t-1} - \mu)\varepsilon_t] + E[\varepsilon_t^2]. \end{aligned}$$

Since Y_t is covariance stationary and ε_t is independently distributed,

$$Var(Y_t) = Var(Y_{t-1})$$

and

$$E[Y_{t-1}\varepsilon_t] = 0,$$

so

$$Var(Y_t) = \frac{\sigma^2}{1 - \phi^2} \equiv \gamma_0.$$

AR(1): Covariance

- Covariance

$$\begin{aligned} Cov(Y_t, Y_{t-j}) &= E[(Y_t - \mu)(Y_{t-j} - \mu)] \\ &= \phi E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + E[\varepsilon_t(Y_{t-j} - \mu)], \\ Cov(Y_t, Y_{t-j}) &\equiv \gamma_j = \phi\gamma_{j-1}. \end{aligned}$$

AR(1): Moments

- Summing up

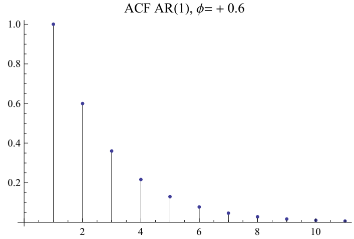
$$\begin{aligned} E[Y_t] &= \frac{c}{1-\phi} \equiv \mu, \\ Var(Y_t) &= \frac{\sigma^2}{1-\phi^2} \equiv \gamma_0, \\ Cov(Y_t, Y_{t-j}) &\equiv \gamma_j = \phi\gamma_{j-1}. \end{aligned}$$

Auto-correlation Function (ACF)

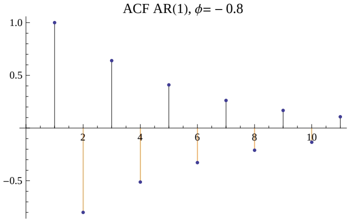
- Define
$$\rho_j \equiv \frac{\gamma_j}{\gamma_0} \equiv j^{\text{th}} \text{ autocorrelation } \equiv corr(Y_t, Y_{t-j})$$
- For AR(1), $\rho_j = \phi\rho_{j-1}$.
  - For AR(1) ACF and IRF are the same. In general it not true.
  - ACF $\in < -1, 1 >$.

AR(1): Auto-correlation Function (ACF)

- $0 < \phi < 1$



- $-1 < \phi < 0$



*P*th-order SDE: AR(p)

An AR(p) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + v_t,$$

can be rewritten as a 1^{st} order system

In the state-space, companion form notation

$$\beta_t = F \cdot \beta_{t-1} + \varepsilon_t$$

*P*th-order SDE: AR(p)

An AR(p) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + v_t,$$

can be rewritten as a 1^{st} order system

$$\begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{bmatrix} = \qquad\qquad\qquad \begin{bmatrix} Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ Y_{t-3} - \mu \\ \vdots \\ Y_{t-p} - \mu \end{bmatrix}$$

In the state-space, companion form notation

$$\beta_t = F \cdot \beta_{t-1} + \varepsilon_t$$

*P*th-order SDE: AR(p)

An AR(p) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + v_t,$$

can be rewritten as a 1^{st} order system

$$\begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{bmatrix} = \qquad\qquad\qquad\qquad \begin{bmatrix} Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ Y_{t-3} - \mu \\ \vdots \\ Y_{t-p} - \mu \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In the state-space, companion form notation

$$\beta_t = F \cdot \beta_{t-1} + \varepsilon_t$$

*P*th-order SDE: AR(p)

An AR(p) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + v_t,$$

can be rewritten as a 1^{st} order system

$$\begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \begin{bmatrix} Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ Y_{t-3} - \mu \\ \vdots \\ Y_{t-p} - \mu \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In the state-space, companion form notation

$$\beta_t = F \cdot \beta_{t-1} + \varepsilon_t$$

*P*th-order SDE: AR(p)

An AR(p) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + v_t,$$

can be rewritten as a 1^{st} order system

$$\begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \begin{bmatrix} Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ Y_{t-3} - \mu \\ \vdots \\ Y_{t-p} - \mu \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In the state-space, companion form notation

$$\beta_t = F \cdot \beta_{t-1} + \varepsilon_t$$

*P*th-order SDE: AR(p)

An AR(p) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + v_t,$$

can be rewritten as a 1^{st} order system

$$\begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & .. & 1 & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ Y_{t-3} - \mu \\ \vdots \\ Y_{t-p} - \mu \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In the state-space, companion form notation

$$\beta_t = F \cdot \beta_{t-1} + \varepsilon_t$$

and we are back in 1^{st} order system.

AR(p): Stability

Consider a state space form

$$\beta_{t+j} = F^{j+1}\beta_{t-1} + F^{j}\varepsilon_t + \ldots + F\varepsilon_{t+j-1} + \varepsilon_{t+j}.$$

AR(p) is stable and stationary if

$$\lim_{j\to\infty} F^{j} = 0$$

,

- i.e. when eigenvalues of F are inside unit circle (have modulus < 1).
- Shocks die out.

Eigenvalues

- Consider equation

  $$Fx = \lambda x.$$

  x is eigenvector and λ is a corresponding eigenvalue.
  To compute the eigenvalue, write it as

  $$(F - \lambda \mathbb{I})x = 0.$$

  If x is a non-zero vector then

  $$F - \lambda \mathbb{I} \text{ is singular} \qquad \Rightarrow \qquad \det(F - \lambda \mathbb{I}) = 0.$$

Example

Consider the AR(2)

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + v_t$$

Then, in a matrix notation

$$\beta_t = F\beta_{t-1} + \varepsilon_t,$$

where

$$\beta_t = \begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \end{bmatrix}, \qquad F = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$$

Eigenvalues λ of the matrix F solve

$$\Rightarrow \det \begin{pmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - \lambda\phi_1 - \phi_2 = 0,$$

$$\lambda_i = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

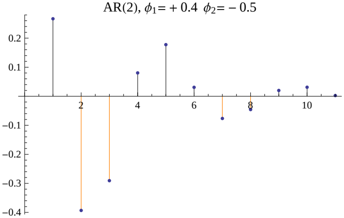
If $|\lambda_i|<1$ for $i = 1, 2$, the AR(2) is stable .

AR(p): Stability

- For pth-order SDE, solve

$$\lambda^p - \phi_1 \lambda^{p-1} - \ldots - \phi_{p-1}\lambda - \phi_p = 0$$

- In general, the solution involves complex and real roots.
- The AR(p) system is stable if all eigenvalues are inside the unit circle. Note: complex eigenvalues imply periodic behavior.



Characteristic Equation

Recall that we could write down AR(p) process

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t,$$

in the lag polynomial form

$$\phi(L)(Y_t - \mu) = \varepsilon_t,$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p.$$

The stability can be then studied through the characteristic equation

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0.$$

Roots of Characteristic Equation

- The roots (z's) of characteristic equation, $\phi(z) = 0$, are inverse of eigenvalues (λ's) of companion matrix F:

  $$z = 1/\lambda.$$

  as

  $$\begin{aligned} 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p &= (1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z) \\ &= z^p (1/z - \lambda_1)(1/z - \lambda_2) \cdots (1/z - \lambda_p) \end{aligned}$$

- For $|z| > 1$ stochastic difference equation is stable and stationary.

Stability

$\phi(L)$ can be decomposed into

$$(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p)=(1-\lambda_1 L)(1-\lambda_2 L)\ldots(1-\lambda_p L)$$

Factor the polynomial into AR(1) elements.

- Since $\forall i \; |\lambda_i|<1$ then $(1-\lambda_i L)^{-1}$ exists and so $\phi(L)^{-1}$ exists

$$\Longrightarrow \qquad \psi(L)=\phi(L)^{-1} \qquad \text{for AR(p).}$$

Yule-Walker Equations

Variance, covariances, autocorrelations, and dynamic multipliers have the same p^{th}-order form:

$$\begin{aligned}
Var(Y_t) = \gamma_0 &= \phi_1\gamma_1 + \phi_2\gamma_2 + \ldots + \phi_p\gamma_p + \sigma^2 \\
Cov(Y_t, Y_{t-j}) = \gamma_j &= \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \ldots + \phi_p\gamma_{j-p} \\
corr(Y_t, Y_{t-j}) = \rho_j &= \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \ldots + \phi_p\rho_{j-p} \\
\psi_j &= \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \ldots + \phi_p\psi_{j-p}
\end{aligned}$$

Yule-Walker Equations: AR(2) Example

- AR(2) Example:

$$\phi(L)^{-1} = \psi(L) \Rightarrow \phi(L)\psi(L) = 1$$
$$(1 - \phi_1 L - \phi_2 L^2)(1 + \psi_1 L + \psi_2 L^2 + \ldots) = 1$$
$$1 + (\psi_1 - \phi_1)L + (\psi_2 - \phi_1\psi_1 - \phi_2)L^2 + \ldots = 1$$

Then

$$\begin{aligned} \psi_1 &= \phi_1, \\ \psi_2 &= \phi_1\psi_1 + \phi_2, \\ &\vdots \\ \psi_j &= \phi_1\psi_{j-1} + \phi_2\psi_{j-2} \qquad . \end{aligned}$$

Example: AR(2) Moments

AR(2) Process:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

- Mean

$$\mu = c/(1 - \phi_1 - \phi_2)$$

- Covariance

$$\begin{aligned}
\gamma_j &= E[(Y_t - \mu)(Y_{t-j} - \mu)] \\
&= E\left[\{\phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \varepsilon_t\}(Y_{t-j} - \mu)\right] \\
&= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \\
\gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 \\
&= \gamma_0 \frac{\phi_1}{1 - \phi_2} \\
\gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0 \\
&= \gamma_0 \left(\frac{\phi_1^2}{1 - \phi_2} + \phi_2 \right)
\end{aligned}$$

Example: AR(2) Moments

- Variance

$$\begin{aligned}
\gamma_0 &= E[(Y_t-\mu)(\phi_1(Y_{t-1}-\mu)+\phi_2(Y_{t-2}-\mu)+\varepsilon_t)] \\
&= \phi_1\gamma_1+\phi_2\gamma_2+\sigma^2 \\
&= \left[\frac{\phi_1^2}{1-\phi_2}+\frac{\phi_2\phi_1^2}{1-\phi_2}+\phi_2^2\right]\gamma_0+\sigma^2 \\
\gamma_0 &= \frac{1-\phi_2}{(1+\phi_2)\left[(1-\phi_2)^2-\phi_1^2\right]}
\end{aligned}$$

- Autocorrelation

$$\begin{aligned}
\rho_j &= \phi_1\rho_{j-1}+\phi_2\rho_{j-2} \\
\rho_1 &= \phi_1+\phi_2\rho_1 \\
&= \phi_1/(1-\phi_2) \\
\rho_2 &= \phi_1\rho_1+\phi_2
\end{aligned}$$

MOVING AVERAGE (MA) PROCESSES

Moving Average (MA) Processes

Recall Wold Form:

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \varepsilon_t \sim WN,$$

MA(q) process : truncated Wold form.

$$\begin{aligned} Y_t &= \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \\ Y_t &= \mu + \theta(L)\varepsilon_t, \\ & \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q. \end{aligned}$$

- "Moving average" as Y_t is constructed as a weighed sum of the q most recent values of ε

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1 + \theta L)\varepsilon_t$$

Moments

- Mean

$$E[Y_t] = \mu,$$

- Variance

$$Var(Y_t) = \gamma_0 = E[(Y_t - \mu)^2]$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1 + \theta L)\varepsilon_t$$

Moments

- Mean

$$E[Y_t] = \mu,$$

- Variance

$$\begin{aligned} Var(Y_t) = \gamma_0 &= E[(Y_t - \mu)^2] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] \end{aligned}$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1+\theta L)\varepsilon_t$$

Moments

- Mean

$$E[Y_t] \quad = \quad \mu,$$

- Variance

$$\begin{aligned} Var(Y_t) = \gamma_0 &= E[(Y_t - \mu)^2] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] \\ &= E[\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2] \end{aligned}$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1+\theta L)\varepsilon_t$$

Moments

- Mean

$$E[Y_t] \quad = \quad \mu,$$

- Variance

$$\begin{aligned} Var(Y_t) = \gamma_0 \quad &= \quad E[(Y_t - \mu)^2] \\ &= \quad E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] \\ &= \quad E[\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2] \\ &= \quad \sigma^2 + 0 + \theta^2\sigma^2 \end{aligned}$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1+\theta L)\varepsilon_t$$

Moments

- Mean

$$E[Y_t] = \mu,$$

- Variance

$$\begin{aligned} Var(Y_t) = \gamma_0 &= E[(Y_t - \mu)^2] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] \\ &= E[\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2] \\ &= \sigma^2 + 0 + \theta^2\sigma^2 \\ &= (1+\theta^2)\sigma^2, \end{aligned}$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1 + \theta L)\varepsilon_t$$

Moments

- Covariance

$$Cov(Y_t, Y_{t-1}) = \gamma_1 = E[(Y_t - \mu)(Y_{t-1} - \mu)]$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1 + \theta L)\varepsilon_t$$

Moments

- Covariance

$$\begin{aligned} Cov(Y_t, Y_{t-1}) &= \gamma_1 = E[(Y_t - \mu)(Y_{t-1} - \mu)] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \end{aligned}$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1 + \theta L)\varepsilon_t$$

Moments

- Covariance

$$\begin{aligned} Cov(Y_t, Y_{t-1}) &= \gamma_1 = E[(Y_t - \mu)(Y_{t-1} - \mu)] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \\ &= \theta\sigma^2, \end{aligned}$$

MA(1)

Example MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} = \mu + (1+\theta L)\varepsilon_t$$

Moments

- Covariance

$$\begin{aligned} Cov(Y_t, Y_{t-1}) &= \gamma_1 = E[(Y_t-\mu)(Y_{t-1}-\mu)] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1}+\theta\varepsilon_{t-2})] \\ &= \theta\sigma^2, \\ Cov(Y_t, Y_{t-j}) &= \gamma_j = 0 \qquad \forall j > 1. \end{aligned}$$

- Autocorrelation Function

$$\begin{aligned} \rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2} \\ \rho_j &= 0, \qquad \forall j > 1. \end{aligned}$$

Invertibility of MA Processes

Consider MA(1)

$$Y_t = \mu + (1 + \theta L)\varepsilon_t$$

- θ and $1/\theta$ give the same ACF

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta}{1+\theta^2} = \frac{1/\theta}{1+\frac{1}{\theta^2}} = \frac{1}{\theta}\frac{1}{\frac{\theta^2+1}{\theta^2}} = \frac{\theta}{1+\theta^2}$$

Normalize θ by using invertible MA representation.

- Example: Both

  $$Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$$

  and

  $$Y_t = \varepsilon_t + 2\varepsilon_{t-1}$$

  have the same autocorrelation function:

  $$\rho_1 = \frac{2}{1+4} = \frac{0.5}{1+(0.5)^2} = 0.4$$

Invertibility of MA Processes

But for MA(1) with $|\theta| < 1$, there exists ∞-order AR representation:

$$\begin{aligned} Y_t - \mu &= (1 + \theta L)\varepsilon_t, & |\theta| < 1, \\ &= (1 - \theta^* L)\varepsilon_t, & \text{for } \theta^* = -\theta \end{aligned}$$

It can rewritten as

$$\begin{aligned} & (1 - \theta^* L)^{-1}(Y_t - \mu) = \varepsilon_t \\ & (1 + \theta^* L + {\theta^*}^2 L^2 + \ldots)(Y_t - \mu) = \varepsilon_t \\ Y_t - \mu &= \theta(Y_{t-1} - \mu) - \theta^2(Y_{t-2} - \mu) + \theta^3(Y_{t-3} - \mu) \ldots + \varepsilon_t \end{aligned}$$

i.e. AR(∞) with $\phi(L) = 1 + \theta - \theta^2 + \theta^3 - \theta^4 + \ldots$.

$|\theta| < 1$ is not a stability requirement, MA system is always stable. It allows invertibility and AR(∞) representation.

Partial Autocorrelation Function (PACF)

Definition

k^{th}-order partial autocorrelation is the regression coefficient (for the population) ϕ_{kk} in k^{th}-order autoregression

$$Y_t = c + \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \ldots + \phi_{kk}Y_{t-k} + \varepsilon_t$$

It measures how important is the last lag.

MODEL SELECTION

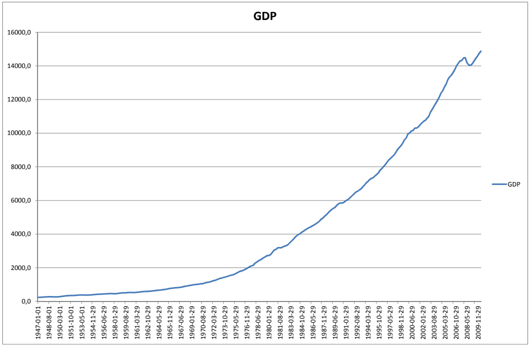
Box-Jenkins Approach

Matching model with actual data

- **Transform data to “appear” covariance stationary**
  We may have data that is not covariance stationary, e.g. GDP.

  Box-Jenkins approach: maybe GDP is not covariance-stationary but some transformation of it is and we can get forecast of it from transformed series.
  - take logs (natural)
  - differences
  - detrend

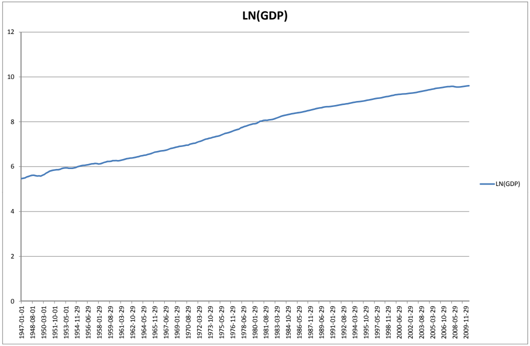
GDP



We can’t use ARMA model as the graph shows it’s not covariance stationary series.
Also, not linear.

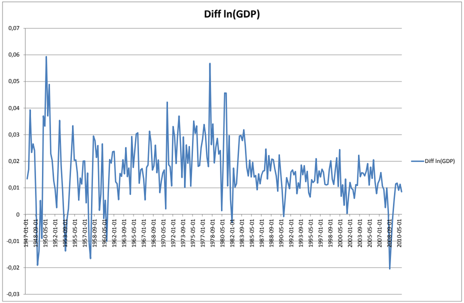
Log of GDP

Take natural logarithm.



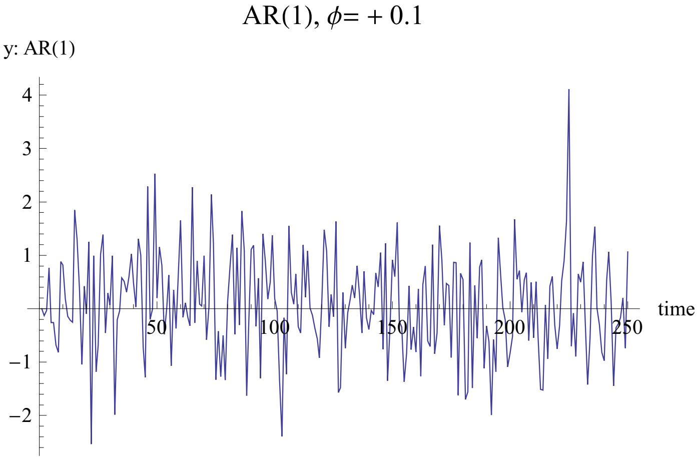
Difference in log of GDP = Growth rate

If we take first difference of the log series, we get growth rate of level series.

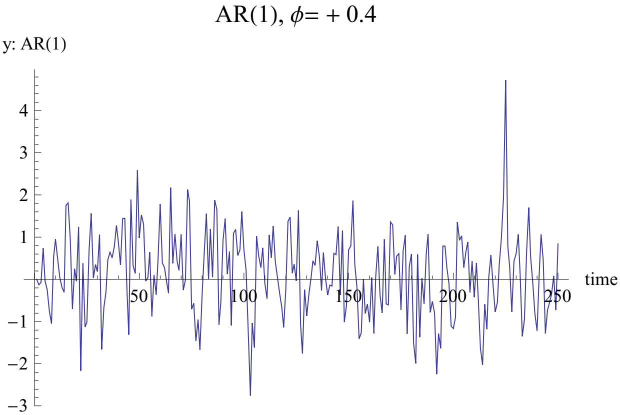


It looks much more like AR or MA process.
Just looking at graph may not be enough.

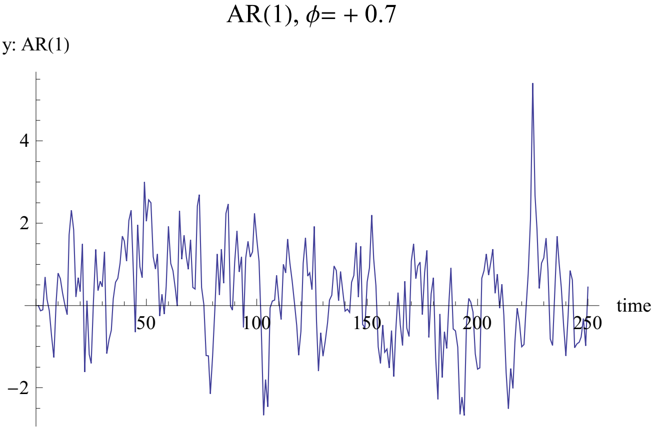
AR(1)



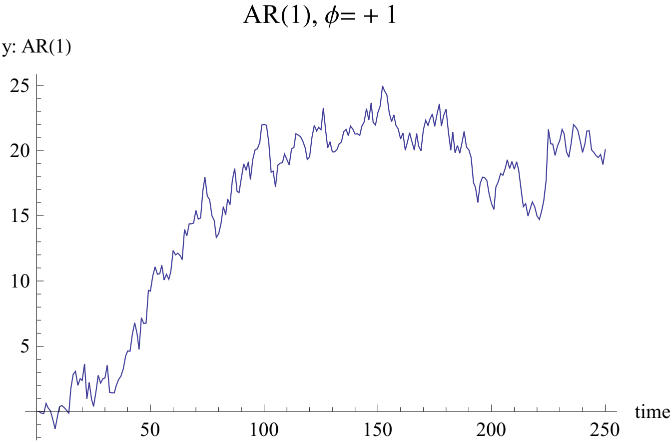
AR(1)



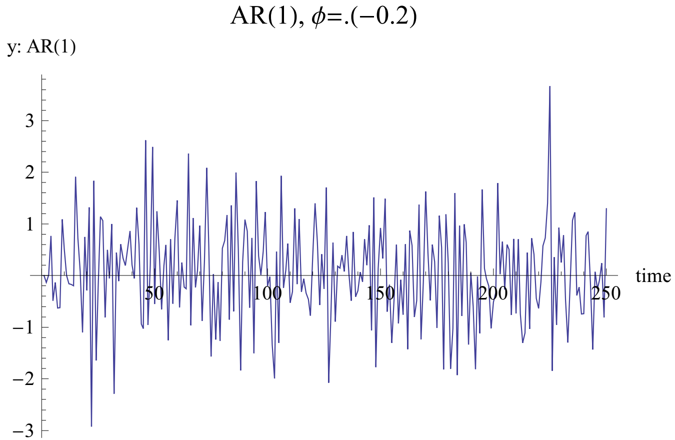
AR(1)



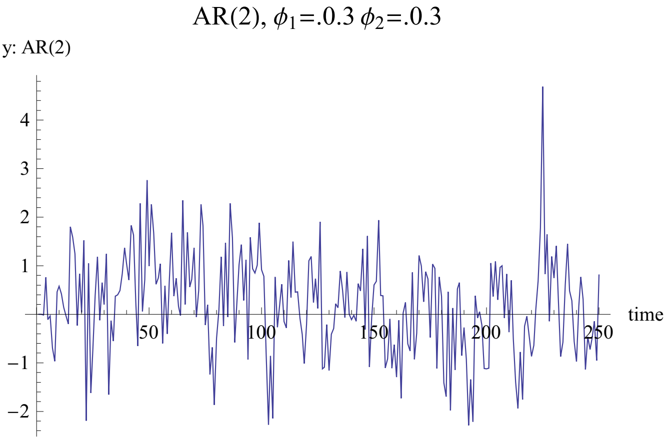
AR(1)



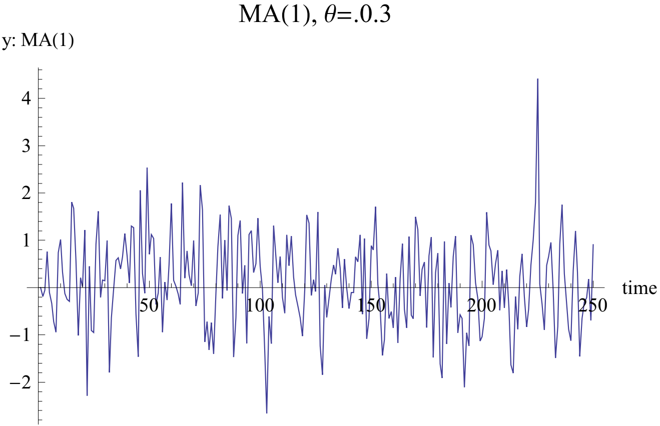
AR(1)



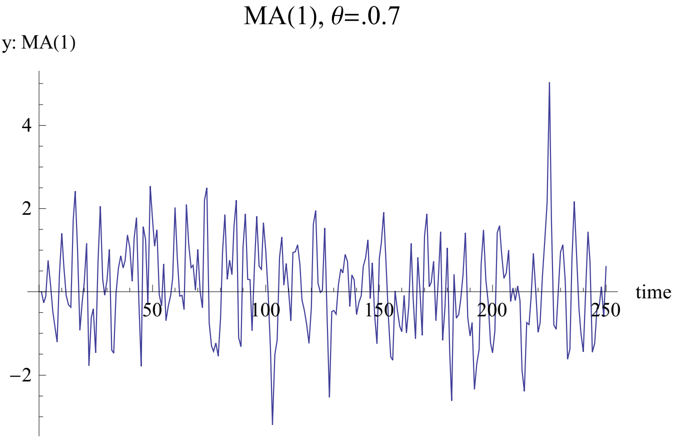
AR(2)



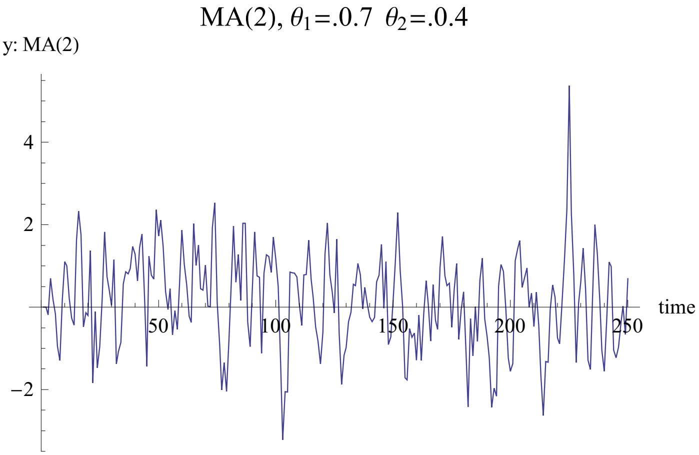
MA(1)



MA(1)

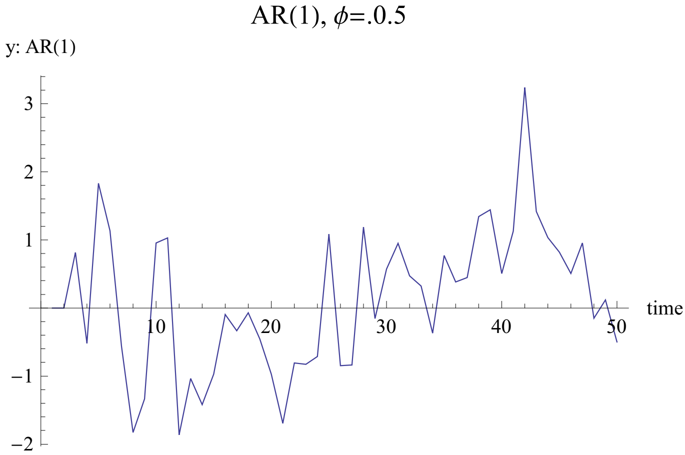


MA(2)



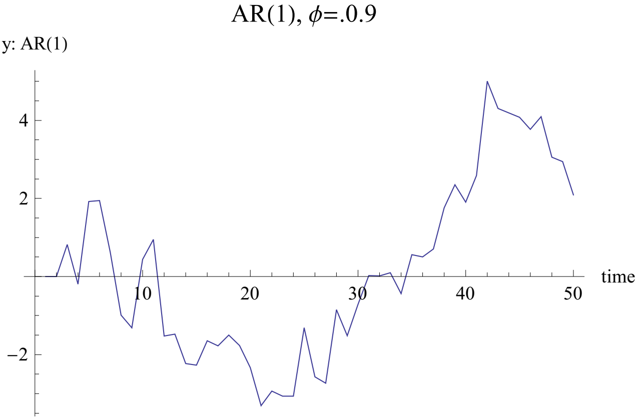
AR(1)

It's not that bad, though:



AR(1)

It’s not that bad, though:

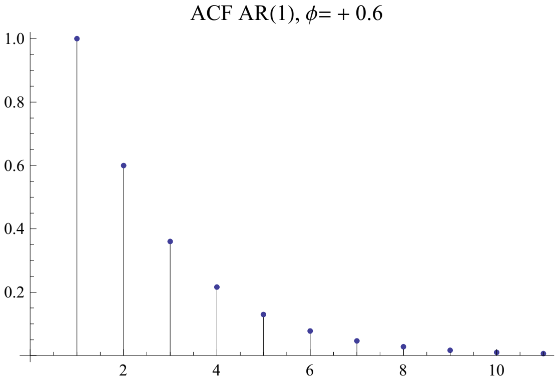


Box-Jenkins Approach

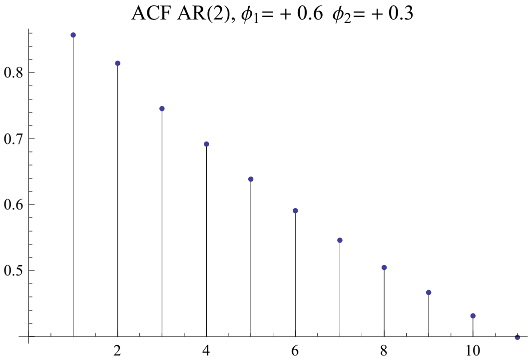
Matching model with actual data

- Transform data to “appear” covariance stationary
  - take logs (natural)
  - differences
  - detrend
- **Examine the sample ACF and PACF**
  We know that there is a 1-1 mapping between ACF and series

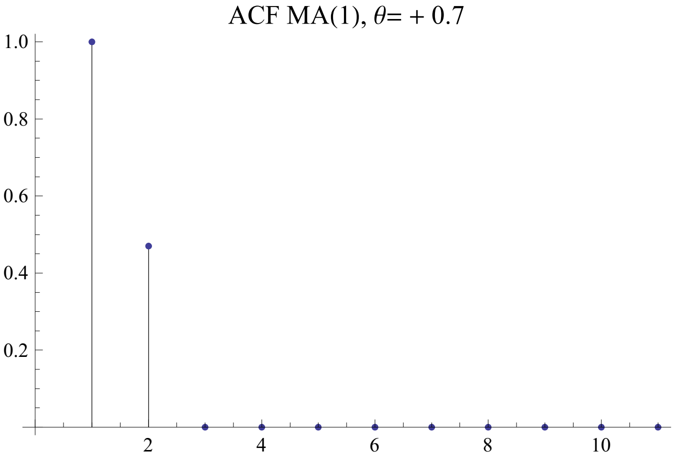
ACF: AR(1)



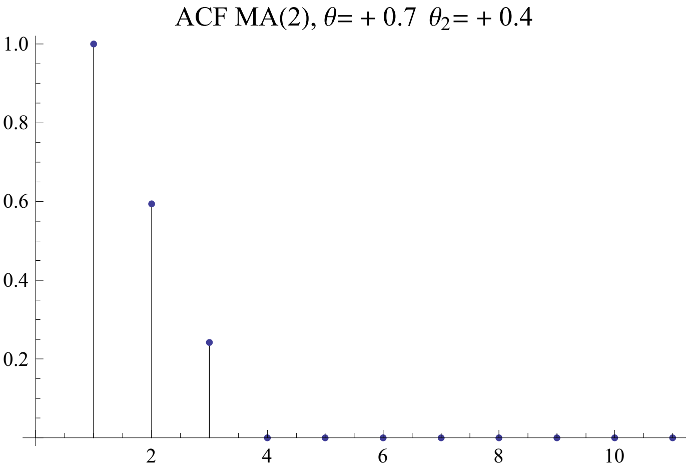
ACF: AR(2)



ACF: MA(1)



ACF: MA(2)



Sample ACF

$$\begin{aligned} \bar{y} &= \frac{1}{T}\sum_{t=1}^{T} y_t, \\ \hat{\gamma}_j &= \frac{1}{T}\sum_{t=j+1}^{T}(y_t-\bar{y})(y_{t-j}-\bar{y}), \qquad \textit{sample auto covariance estimate} \\ \bar{\rho}_j &= \frac{\hat{\gamma}_j}{\hat{\gamma}_0} \qquad \textit{sample auto correlation} \end{aligned}$$

Sample PACF

- Use OLS

$$y_t = c + \phi_{1k} y_{t-1} + \ldots + \phi_{kk} y_{t-k} + \varepsilon_t, \qquad k = 1, 2, \ldots$$

- If $y_t \sim iid(\mu, \sigma^2)$, $\rho_j = 0, \quad \forall j \neq 0 \quad .$
  (no correlation across observation)

$$\hat{var}(\hat{\rho}_j) \approx \frac{1}{T} \qquad \text{and} \qquad \hat{var}(\hat{\phi}_{kk}) \approx \frac{1}{T}$$

This values are used when reporting bounds in ACF and PACF: 5%-95%: $\pm 1.96 \hat{var}(\hat{\rho}_j)$ *or* $\pm 1.96 \hat{var}(\hat{\phi}_{kk})$.

Model selection

- Assume we can reject iid, i.e. PACF and ACF show some significant lags. How to determine which model is correct?

Process	ACF	PACF
AR(p)	Exponential or oscillatory decay	$\phi_{kk} = 0, \ k > p$
MA(q)	$\rho_k = 0, \ k > q$	Exponential or oscillatory decay
ARMA(p,q)	decay begins at lag q	decay begins at lag p

- *But we have many models we can't really discriminate between just by looking.*

Box-Jenkins

Matching model with actual data

- Transform data to “appear” covariance stationary
  - take logs (natural)
  - differences
  - detrend
- Examine the sample ACF and PACF
- Estimate ARMA models
- **Perform diagnostic analysis to confirm that model is consistent with data**

Box-Ljung Statistics

- Use modified Box-Ljung Statistics (Q-Stat) to test joint statistical significance of $\hat{\rho}_1, \ldots, \hat{\rho}_k$:

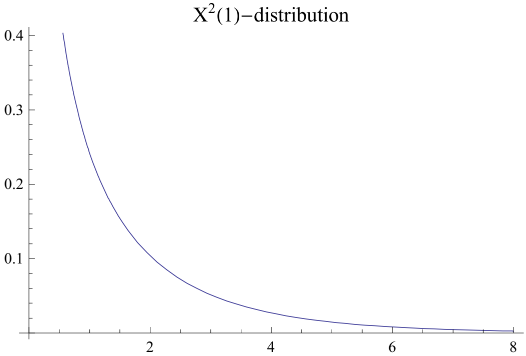
$$Q^*(k) = T(T+2)\sum_{i=1}^{k} \frac{1}{T-i}\hat{\rho}_i^2.$$

- Under $H_0 : Y_t \sim iid(\mu, \sigma^2)$, $Q^*(k) \overset{A}{\sim} \chi^2(k)$.
- For significance level α, the critical region for rejection of the hypothesis of randomness (i.e. reject $\rho_1 = \ldots = \rho_k = 0$) is

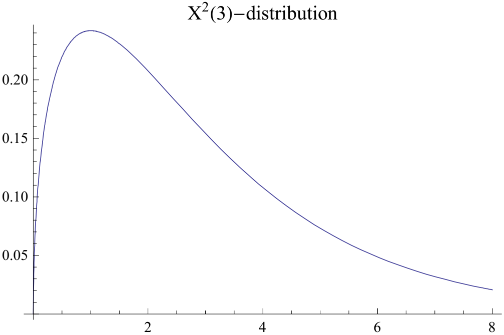
$$Q^*(k) > \chi^2_{1-\alpha}(k).$$

- *For smaller sample, empirical distribution has fatter tail and one needs to move the critical point to the right .*

$\chi^2(1)$



$\chi^2(3)$



Bootstrapping

Bootstrapping:

- takes available data and sample directly from them (randomly),
- uses assumption that $Y_t \sim iid(\mu, \sigma^2)$: independence.

Bootstrapping: example

Say we want to find small sample distribution of GDP growth rates

$$\begin{bmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_T \end{bmatrix} \qquad Q^*(1) = 13.133$$

- *Based on these data we can get* $Q^*(1) = 13.133$.
- *We know it has asymptotically* $\chi^2(1)$ *distribution.*
- *If we want small sample distribution, however, we can create new data from old data set.*

Bootstrapping: example

- *Use random number generator that will pick a number from* $i = 1, \ldots, T$.
- *Take this number and take respective* Δy_i*; repeat it* T *times.*

$$\begin{bmatrix} \Delta y_{34} \\ \Delta y_{100} \\ \vdots \\ \Delta y_{27}, \end{bmatrix} \begin{matrix} i = 34 \\ i = 100 \\ \vdots \\ i = 27, \end{matrix} \quad , \qquad \textit{bootstrap sample}$$

- *The reshuffling breaks any correlation between observations*

  $->$ *it is just a random draw.*
- *Even if* $\Delta y_1, \Delta y_2$ *have some correlation, once bootstrapped (simulated) they don't have it anymore.*

Bootstrapping: example

- *It's a draw from original distribution: has the same histogram as original one.*
- *Now, for this new series, calculate* $Q_1^*(1) = 3.21$*.*
- *Do it again, get 2nd series (the same histogram). Get* $Q_2^* = 10.21$
- *Repeat it* 10000 *times to obtain* 10000 $Q^*(1)$*s.*
- *All under the assumption of independence.*

Bootstrapping: example

- *Do a histogram of all* $Q^*(1)$.
- *It will have* $\chi^2(1)$ *type of distribution, since each individual draw should be from this distribution.*

$$Q^*(1) = \begin{bmatrix} Q_1^*(1) \\ \vdots \\ \vdots \\ Q_{10000}^*(1) \end{bmatrix} \overset{\textit{sort them}}{\Longrightarrow} \begin{bmatrix} Q_{1024}^*(1) = 30.11 \\ \vdots \\ \vdots \\ Q_{2758}^*(1) = 0.11 \end{bmatrix} \begin{array}{l} \textit{largest value} \\ \vdots \\ \vdots \\ \textit{smallest value} \end{array}$$

- *Look at the value of* $Q^*(1)$ *at place 500 and get a critical value for the sample.*
- *Compare with the original one* $Q^*(1) = 13.133$.

Bootstrapping

Bootstrap vs Monte Carlo:

- *Bootstrap: use the original data to construct empirical distribution.*
- *Monte Carlo: instead of using data, make assumption for it, e.g.* $\ln Y_t \sim iid(\mu, \sigma^2)$ *and make simulations using this assumption (don't have to use actual data).*
- *Monte Carlo gives more power if* H_0 *is true, but without any assumption on distribution of the series, bootstrap can be applied.*

Why small sample may have fatter tails than the asymptotic one?

- *Recall*

$$Q^*(k) = T(T+2)\sum_{i=1}^{k}\frac{1}{T_i}\hat{\rho}_i^2 \qquad \hat{\rho}_i = \frac{\hat{\gamma}_i}{\hat{\gamma}_0};\ \hat{\gamma}_i = \frac{1}{T}\sum_{t=i+1}^{T}(y_t-\bar{y})(y_{t-i}-\bar{y}),$$

  where $\bar{y}$ *is estimated.*

- *The other estimate* $\bar{y}$ *will affect* Q^* *so for the finite sample we might be off from the asymptotic distribution.*

Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC):

$$AIC(p,q) = \ln(\hat{\sigma}^2) + \frac{1}{T}2(p+q),$$

where the first term is responsible for the fit of the model and the second one is a penalty for number of parameters.

Schwarz (Bayesian) Information Criterion

Schwarz (Bayesian) Information Criterion (BIC):

$$BIC(p,q) = \ln(\hat{\sigma}^2) + \frac{1}{T}\ln(T)(p+q),$$

where now penalty is related to sample size.

Selection: AIC vs BIC

- Calculate selection criterium for a set of estimated models and try to find the one that minimize the AIC or BIC.
- Result:
  If p, q considered are larger than "true" orders
  (i) BIC picks "true" ARMA(p,q) model as $T \rightarrow \infty$.
  (ii) AIC picks overparameterized model as $T \rightarrow \infty$.
- For smaller sample it might be a a problem with choice BIC:
  - Trade-off between efficiency and consistency:
- *Better to overparameterize than to have incorrect result.*

Residual Diagnostic

- Use sample ACF and PACF of sample residuals.
- Compute Box-Ljung Statistics for sample of residuals:

$$Q^*(k) \overset{A}{\sim} \chi^2(k-(p+q)),$$

where (p+q) is a degree of freedom adjustment

- Use LM test (it has higher power)

$$T \cdot R^2 \sim \chi^2(k),$$

with R^2 computed from the regression

$$\hat{\varepsilon}_t = c + \alpha_1 \hat{\varepsilon}_{t-1} + \ldots + \alpha_k \hat{\varepsilon}_{t-k} + v_t.$$

ESTIMATION

ARMA

- Wold representation:

$$Y_t = \kappa_t + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \qquad \varepsilon_t \sim WN(0, \sigma^2), \qquad \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

- AR(p):

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + \ldots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t, \qquad \varepsilon_t \sim WN$$

- MA(q):

$$Y_t - \mu = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}, \qquad \varepsilon_t \sim WN$$

- ARMA(p,q):

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \ldots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

Estimation

- OLS
- MLE
- AR(1)
- MA(1)

OLS

For AR(p), OLS is equivalent to Conditional MLE

- Model:

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2). \\ &= x_t' \beta + \varepsilon_t, \quad \beta = (c, \phi_1, \phi_2, \ldots, \phi_p), \quad x_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p}) \end{aligned}$$

- OLS:

$$\begin{aligned} \hat{\beta} &= \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t, \\ \hat{\sigma}^2 &= \frac{1}{T - (p+1)} \sum_{t=1}^{T} (y_t - x_t' \hat{\beta})^2. \end{aligned}$$

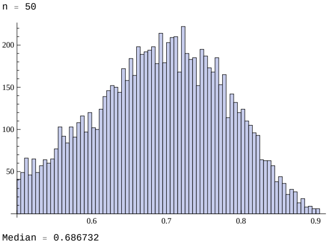
Properties

Note:

- $E[\hat{\beta}] \neq \beta$ because x_t is random and $E(\varepsilon|Y) \neq 0$. But, if $|z| > 1$ for $\phi(z) = 0$ (or $|\lambda| < 1$ for F) then,

$$\hat{\beta} \xrightarrow{p} \beta, \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

- *Estimator might be biased but consistent, it converges in probability.*
- Illustration:
  $\phi = 0.7$



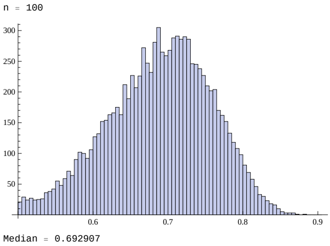
Properties

Note:

- $E[\hat{\beta}] \neq \beta$ because x_t is random and $E(\varepsilon|Y) \neq 0$. But, if $|z| > 1$ for $\phi(z) = 0$ (or $|\lambda| < 1$ for F) then,

$$\hat{\beta} \xrightarrow{p} \beta, \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

- *Estimator might be biased but consistent, it converges in probability.*
- Illustration:
  $\phi = 0.7$



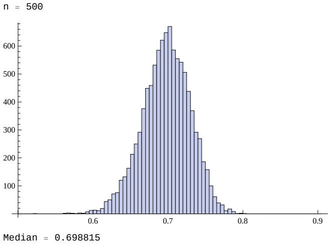
Properties

Note:

- $E[\hat{\beta}] \neq \beta$ because x_t is random and $E(\varepsilon|Y) \neq 0$. But, if $|z| > 1$ for $\phi(z) = 0$ (or $|\lambda| < 1$ for F) then,

$$\hat{\beta} \xrightarrow{p} \beta, \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

- *Estimator might be biased but consistent, it converges in probability.*
- Illustration:
  $\phi = 0.7$



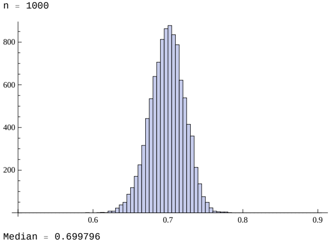
Properties

Note:

- $E[\hat{\beta}] \neq \beta$ because x_t is random and $E(\varepsilon|Y) \neq 0$. But, if $|z| > 1$ for $\phi(z) = 0$ (or $|\lambda| < 1$ for F) then,

$$\hat{\beta} \xrightarrow{p} \beta, \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

- *Estimator might be biased but consistent, it converges in probability.*
- Illustration:
  $\phi = 0.7$



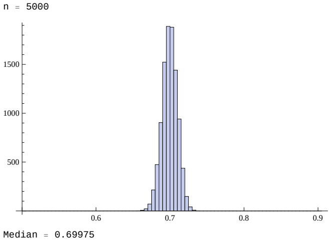
Properties

Note:

- $E[\hat{\beta}] \neq \beta$ because x_t is random and $E(\varepsilon|Y) \neq 0$. But, if $|z| > 1$ for $\phi(z) = 0$ (or $|\lambda| < 1$ for F) then,

$$\hat{\beta} \xrightarrow{p} \beta, \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

- *Estimator might be biased but consistent, it converges in probability.*
- Illustration:
  $\phi = 0.7$



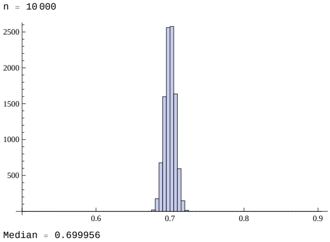
Properties

Note:

- $E[\hat{\beta}] \neq \beta$ because x_t is random and $E(\varepsilon|Y) \neq 0$. But, if $|z| > 1$ for $\phi(z) = 0$ (or $|\lambda| < 1$ for F) then,

$$\hat{\beta} \xrightarrow{p} \beta, \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

- *Estimator might be biased but consistent, it converges in probability.*
- Illustration:
  $\phi = 0.7$



Hypothesis testing

Hypothesis testing:

$$\sqrt{T}(\hat{\beta}-\beta) \xrightarrow{d} N(0,\sigma^2 V^{-1}), \qquad V = plim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} x_t x_t'.$$

- *If we have enough data $(T \to \infty)$ then t-student will converge to Normal distribution.*

$$t^{\beta=\beta_0} = \frac{\hat{\beta}-\beta_0}{\hat{SE}(\hat{\beta})} \sim t - student(T-1) \underset{T\to\infty}{\longrightarrow} N(0,1).$$

- *See Hamilton for a downward bias in a finite sample, i.e. $E[\hat{\beta}] < \beta$.*

MLE

If $Y_t \sim iid$, joint likelihood is the product of marginal likelihoods (pdf)

$$\begin{aligned} L(\tilde{\theta}|y_1,\ldots,y_T) &= f(y_1,\ldots,y_T|\tilde{\theta}) \\ &= \prod_{t=1}^{T} f(y_t|\tilde{\theta}), \\ &= \prod_{t=1}^{T} L(\tilde{\theta}|y_t), \end{aligned}$$

Different parameters have different probability of generating $\{y_1,\ldots,y_T\}$.

MLE

If $Y_t \sim iid$, joint likelihood is the product of marginal likelihoods (pdf)

$$\begin{aligned} L(\tilde{\theta}|y_1,\ldots,y_T) &= f(y_1,\ldots,y_T|\tilde{\theta}) \\ &= \prod_{t=1}^{T} f(y_t|\tilde{\theta}), \\ &= \prod_{t=1}^{T} L(\tilde{\theta}|y_t), \end{aligned}$$

But if Y_t is not independent such factorization is invalid. Instead, factor into conditional distributions.

$$\begin{aligned} f(Y_1,Y_2|\tilde{\theta}) &= f(Y_2|Y_1,\tilde{\theta})f(Y_1|\tilde{\theta}), \\ f(Y_1,Y_2,Y_3|\tilde{\theta}) &= f(Y_3|Y_2,Y_1,\tilde{\theta})f(Y_2,Y_1|\tilde{\theta}), \\ L(\tilde{\theta}|y_1,\ldots,y_T) = f(Y_1,\ldots,Y_T|\tilde{\theta}) &= \prod_{t=2}^{T} f(Y_t|Y_{t-1},\ldots,Y_2,Y_1,\tilde{\theta})f(Y_1|\tilde{\theta}) \end{aligned}$$

MLE

- Conditional MLE assumes Y_1 fixed (not random). *It's just a simplifying assumption.*

$$L^c(\tilde{\theta}|y_1,\ldots,y_T) = \prod_{t=2}^{T} f(Y_t|Y_{t-1},\ldots,Y_1,\tilde{\theta}).$$

- Given normality, first order conditions of maximization L^c are linear in $\tilde{\theta}$.
- For AR models $\hat{\phi}^{CMLE} \Leftrightarrow OLS$.
- Conditional MLE is consistent. (*It's not so efficient as we ignore randomness of Y_1.*)
- Exact MLE requires non-linear optimization.
- *Often we don't know distribution of the the data. One thing that is assumed in OLS is the $\varepsilon \sim WN$.*
- *Assuming normality in CMLE is not a bad assumption as we still get consistent estimator: quasi MLE. (see Davidson and MacKinnon.)*
- *We don't get unbiasness in any of MLE. What can be done is to compute what the bias is and correct for it.*

Estimation AR(1)

Recall: AR(1)

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon \sim iidN(0, \sigma^2), \quad |\phi| < 1.$$

- If we don't believe in normality of ε_t , we have quasi-MLE.
- If we do: MLE.

If $\varepsilon \sim N$ so is $Y_t | Y_{t-1} \sim N$:

$$Y_t | Y_{t-1} \sim N(c + \phi Y_{t-1}, \sigma^2).$$

Conditional MLE

- If we know Y_{t-1} the only term that is random is ε_t.

$$f(y_t|y_{t-1},\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_t - c - \phi y_{t-1})^2},$$

  where $\theta = (c, \phi, \sigma^2)$.

- What about Y_1?

$$\begin{aligned} Y_1 &\sim N(E[Y_t], var(Y_t)) \\ &\sim N(\mu, \frac{\sigma^2}{1-\phi^2}), \qquad \mu = \frac{c}{1-\phi}. \end{aligned}$$

- If $\phi = 1$ no unconditional mean or variance exist.

Exact MLE

Maximum likelihood

$$L(\tilde{\theta}|y_1,\dots,y_T) = f(y_1|\tilde{\theta}) \times \prod_{t=2}^{T} f(y_t|y_{t-1},\tilde{\theta}),$$

i.e.

$$L(\tilde{\theta}|y_1,\dots,y_T) = \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{1-\phi^2}\right)}} e^{-\frac{1}{2\sigma^2/(1-\phi^2)}\left(y_1-\frac{c}{1-\phi}\right)^2} \times \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(y_t-c-\phi y_{t-1}\right)^2}$$

- Need to solve non-linear optimization problem.

MA(1)

Recall

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \qquad \varepsilon_t \sim iidN(0, \sigma^2), \ |\theta| < 1.$$

$|\theta| < 1$ *is assumed for invertible representation only.*

Figure: Bi-model likelihood function for MA process

Estimation MA(1)

$$\begin{aligned} Y_t|\varepsilon_{t-1} &\sim N(\mu+\theta\varepsilon_{t-1},\sigma^2), \\ f(y_t|\varepsilon_{t-1},\tilde{\theta}) &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(y_t-\mu-\theta\varepsilon_{t-1})^2}, \end{aligned}$$

$\tilde{\theta} = (\mu,\theta,\sigma^2)$.

- *Problem: without knowing ε_{t-2} we don't observe ε_{t-1}. Need to know ε_{t-2} to know $\varepsilon_{t-1} = y_{t-1} - \mu - \theta\varepsilon_{t-2}$.*
- But ε_{t-2} unobservable.
- Assume $\varepsilon_0 = 0$.
- *Make it non-random, just fix it with number 0. The trick works with any number.*

Estimation MA(1)

$$\begin{aligned} Y_1|\varepsilon_0 &\sim N(\mu, \sigma^2), \\ Y_1 &= \mu + \varepsilon_1 \Rightarrow \varepsilon_1 = Y_1 - \mu, \\ Y_2 &= \mu + \varepsilon_2 + \theta\varepsilon_1 \Rightarrow \varepsilon_2 = Y_2 - \mu - \theta(Y_1 - \mu), \\ \varepsilon_t &= Y_t - \mu - \theta(Y_{t-1} - \mu) + \ldots + (-1)^{t-1}\theta^{t-1}(Y_1 - \mu). \end{aligned}$$

- Conditional likelihood ($\varepsilon_0 = 0$).:

$$L(\tilde{\theta}|y_1, \ldots, y_T, \varepsilon_0 = 0) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\varepsilon_t^2}.$$

- If $|\theta| < 1$ *(much less)*, ε_0 doesn't matter, CMLE is consistent.
- Exact MLE requires Kalman Filter.

FORECASTING

Simple Model

- With estimated parameters of the model we can make a forecast about the future behavior of the variable of interest.
- Simple Model:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN,$$

- *Today's GDP growth depends on yesterday's GDP growth,*
- *Relationship given, for example, by economic theory.*
- *Once agents know the structure of the model they can form a forecast.*

Notation

Denote:

$\{Y_t\}$ – covariance-stationary process, e.g. $ARMA(p,q)$,

Ω_t – information available at time t,

$Y^*_{t+1|t}$ – forecast of Y_{t+1} based on Ω_t.

- *In our simple model it is Y_{t+1} given Y_t.*

Loss Function

- *We want to assess how good the forecast is.*
- *Assume people don't like mistakes in any direction.*

Rule: evaluate forecast with a quadratic loss function:

$$\min E[(Y_{t+1} - Y^*_{t+1|t})^2] = MSE(Y_{t+1}, Y^*_{t+1|t}).$$

Result: The minimum MSE forecast of Y_{t+1} based on Ω_t is $E[Y_{t+1}|\Omega_t]$.

- Y_{t+1} *is modeled as a random variable (have the whole distribution) but it's behavior is summarized by its mean.*

Linear Projection

- *Y_{t+1} may be very complicated and calculating $E[Y_{t+1}|\Omega_t]$ might be very cumbersome.*
- If $E[Y_{t+1}|\Omega_t]$ difficult to compute, use ”best linear forecast”.

  $X_{t_{p\times 1}}$ – variables in Ω_t “useful” for prediction
- Linear projection:

$$\hat{Y}_{t+1|t} = \alpha' X_t = \alpha_1 X_{1t} + \ldots + \alpha_p X_{pt},$$

where

$$E[(Y_{t+1} - \alpha' X_t) \cdot X_{it}] = 0, \qquad i = 1, \ldots, p.$$

  - *p moments conditions ensure that error is orthogonal to any information in Ω_t:*
  - *forecast errors are uncorrelated with past information.*

MSE Linear Forecast

Result: The minimum MSE linear forecast of Y_{t+1} is a linear projection.

- For Gaussian (Normal) process: $E[Y_{t+1}|\Omega_t] = \hat{Y}_{t+1|t}$.
- *Linear projection is optimal in Gaussian case.*
- *How do we deal with αs ?*
  - $\hat{Y}_{t+1|t} = \alpha' X_t$ can be thought of as computed by OLS,
  - If $\{X_t, Y_t\}$ is covariance stationary and ergodic, $b \xrightarrow{p} \alpha$, i.e. OLS estimate, b, converges in probability to the true value, α.
- *Optimality is defined in terms of quadratic loss function.*

ARMA Models

Solve Wold form

$$\begin{aligned} Y_t - \mu &= \psi(L)\varepsilon_t, \qquad \varepsilon_t \sim WN \\ &\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j, \quad \psi_0 = 1, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty \\ Y_{t+s} &= \mu + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \ldots + \psi_s \varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \ldots, \\ \hat{Y}_{t+1|t} &= \mu + \psi_s \varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \ldots . \end{aligned}$$

- *the last line uses information available at time t and $E_t[\varepsilon_{t+i}] = 0,\ i > 0$.*

ARMA Models

$$\begin{aligned} MSE(\hat{Y}_{t+s|t}, Y_{t+s}) &= E[(\varepsilon_t + \psi\varepsilon_{t+s-1} + \ldots + \psi_{s-1}\varepsilon_{t+1})^2] \\ &= \sigma^2(1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{s-1}^2) \leq var(Y_{t+s}). \end{aligned}$$

- *We are better off with linear projection than with unconditional variance.*

But,

$$\lim_{s\to\infty} \sigma^2 \sum_{k=0}^{s} \psi_k^2 = var(Y_t).$$

- *Upper limit for uncertainty is as high as the unconditional variance.*

Forecasting ARMA Models: AR(1)

E.g. AR(1):

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t$$

One period ahead forecast:

$$\hat{Y}_{t+1|t} = \mu + \phi(Y_t - \mu)$$

Using Wold formulation:

$$\begin{aligned}
\psi_j &= \phi^j, \\
\Rightarrow \hat{Y}_{t+s|t} &= \mu + \phi^s \varepsilon_t + \phi^{s+1} \varepsilon_{t-1} + \dots \\
&= \mu + \phi^s(\varepsilon_t + \phi\varepsilon_{t-1} + \dots) \\
&= \mu + \phi^s(Y_t - \mu), \\
\lim_{s\to\infty} MSE &= \sigma^2 \frac{1}{1-\phi^2} = var(Y_t).
\end{aligned}$$

Forecasting ARMA Models: MA(1)

E.g. MA(1):

$$\psi_0 = 1, \quad \psi_1 = \theta, \quad , \psi_j = 0, \quad \forall j > 1,$$

$$\begin{aligned} \hat{Y}_{t+1|t} &= \mu + \theta \hat{\varepsilon}_t, \qquad \hat{\varepsilon}_t = (Y_t - \mu) - \theta \hat{\varepsilon}_{t-1}, \\ \hat{Y}_{t+s|t} &= \mu, \qquad \forall s > 1 \\ \lim_{s \to \infty} MSE &= \sigma^2 (1 + \theta^2) = var(Y_t). \end{aligned}$$

Forecast error is just a deviation of the series from the long-run unconditional mean.

Forecasting ARMA Models: AR(2)

E.g. AR(2):

$$\underbrace{\begin{bmatrix} Y_t-\mu \\ Y_{t-1}-\mu \end{bmatrix}}_{\beta_t} = \underbrace{\begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} Y_{t-1}-\mu \\ Y_{t-2}-\mu \end{bmatrix}}_{\beta_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}}_{v_t}$$

$$\beta_t = F\beta_{t-1} + v_t$$

Then,

$$\begin{aligned}\hat{\beta}_{t+s|t} &= F^s v_t + F^{s+1} v_{t-1} + \dots \\ &= F^s(v_t + F v_{t-1} + \dots) = F^s \beta_t.\end{aligned}$$

Let

$$F^s = \begin{bmatrix} f_{11}^{(s)} & f_{12}^{(s)} \\ f_{21}^{(s)} & f_{22}^{(s)}. \end{bmatrix}$$

Then,

$$\hat{Y}_{t+s|t} = \mu + f_{11}^{(s)}(Y_t-\mu) + f_{12}^{(s)}(Y_{t-1}-\mu)$$

Finite-sample forecast

- Forecasts based on Wold form assume infinite number of observations.
- *We don't have them in reality.*
- For finite number of observation:
  - use approximation: set presample $\varepsilon_\tau = 0$,
  - do exact finite-sample forecast.
- Kalman filter calculates linear projections for finite number of observations,
  - exact finite-sample forecast,
  - allow for exact MLE of ARMA models based on prediction error decomposition.
- See Hamilton chapter 4 for alternative.

Forecast accuracy

- $\{Y_t\}$ is the series to be forecast.
- $\{Y^1_{t+h|t}\}$ and $\{Y^2_{t+h|t}\}$ are two competing forecasts of Y_{t+h} based on Ω_t. *(for example, from an AR(p) and ARMA(p,q) models, respectively)*
- Forecast errors from the two models are

$$\begin{aligned} \varepsilon^1_{t+h|t} &= y_{t+h} - y^1_{t+h|t}, \\ \varepsilon^2_{t+h|t} &= y_{t+h} - y^2_{t+h|t}, \end{aligned}$$

  producing series of serially correlated forecast errors $\{\varepsilon^1_{t+h|t}\}^T_{t_0}$, $\{\varepsilon^2_{t+h|t}\}^T_{t_0}$,
    - h-step forecasts use overlapping data.
- Forecast accuracy can be measured by a loss function

$$L(y_{t+h}, y^i_{t+h|t}) = L(\varepsilon^i_{t+h|t})$$

    - squared error loss: $L(\varepsilon^i_{t+h|t}) = (\varepsilon^i_{t+h|t})^2$,
    - absolute error loss: $L(\varepsilon^i_{t+h|t}) = \left|\varepsilon^i_{t+h|t}\right|$.

H_0

- To test if one model predicts better than another consider null hypothesis

$$H_0 : \quad E\left(L(\varepsilon^1_{t+h|t})\right) = E\left(L(\varepsilon^2_{t+h|t})\right)$$

against

$$H_1 : \quad E\left(L(\varepsilon^1_{t+h|t})\right) \neq E\left(L(\varepsilon^2_{t+h|t})\right)$$

- Define loss deferential

$$d_t = L(\varepsilon^1_{t+h|t}) - L(\varepsilon^2_{t+h|t}).$$

- The null hypothesis of equal predictive accuracy is

$$H_0 : \quad E(d_t) = 0.$$

Diebold-Mariano test

- The Diebold-Mariano test statistics is

$$S = \frac{\bar{d}}{\left(\widehat{avar}(\bar{d})\right)^{1/2}} = \frac{\bar{d}}{\left(\widehat{LRV}_{\bar{d}}/T\right)^{1/2}}$$

  where

$$\begin{aligned} \bar{d} &= \frac{1}{T_0}\sum_{t=t_0}^{T} d_t \\ LRV_{\bar{d}} &= \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j, \quad \gamma_j = cov(d_t, d_{t-j}) \end{aligned}$$

  and $\widehat{LRV}_{\bar{d}}$ is a consistent estimate of asymptotic (long-run) variance of $\sqrt{T}\bar{d}$

- Diebold-Mariano (1995) show that under the null of equal predictive accuracy

$$S \overset{A}{\sim} N(0,1).$$

  - Reject the null of equal predictive accuracy at the 5% level if $|S| > 1.96$.
  - One sided tests may also be computed.