Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• **Introduction**
	• Identification
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Stochastic Time Series Models

- series has been generated by a **stochastic** process
- stochastic process is a sequence of random variables $Y_1, Y_2, \ldots, Y_t, \ldots, Y_T$
- these random variables have a **joint distribution**
- the observed series $y_1, y_2, \ldots, y_T$ is assumed to be drawn from this joint distribution
- i.e., it is just one realization of an infinite number of possible realizations

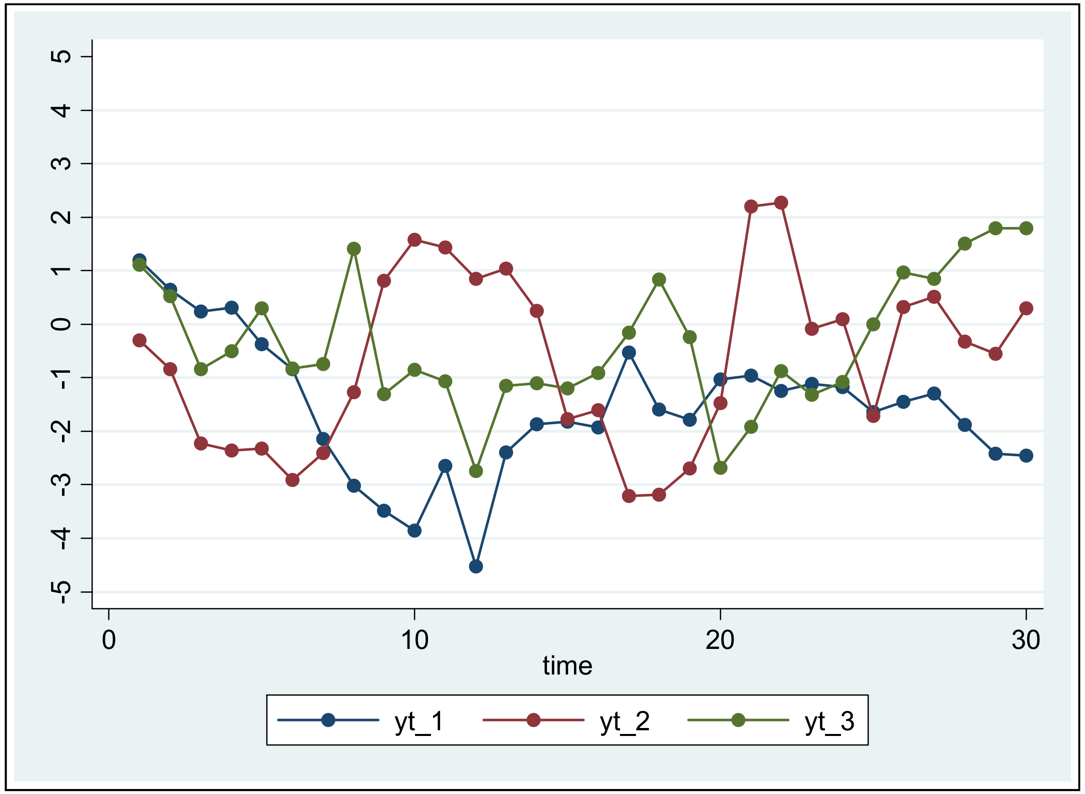
Stochastic Time Series Models

- often we imagine, that we could have observed earlier (y_0, y_{-1}, y_{-2} …) or later observations (y_{T+1} , y_{T+2}, y_{T+3} …)
- the observed series y_1, y_2, …, y_T could be viewed as a finite segment of an infinite process

$$\{y_t\}_{t=-\infty}^{\infty} = \left\{ \dots, y_{-1}, y_0, \underbrace{y_1, y_2, \dots\dots, y_{T-1}, y_T}_{\text{observed series}}, y_{T+1}, y_{T+2}, \dots \right\}$$

- but even if we imagine having **observed** the infinite **series** $\{y_t\}_{t=-\infty}^{\infty}$ it is still viewed from the perspective of Stochastic Time Series Models as a **single realization** of the underlying process.

Three realizations of a time series



Stochastic Time Series Models

A future observation y_{T+1} can be thought of as being generated by a conditional probability distribution function $F(y_{T+1} | y_1, y_2, \ldots, y_T)$.

Example: **marginal, conditional and joint density**

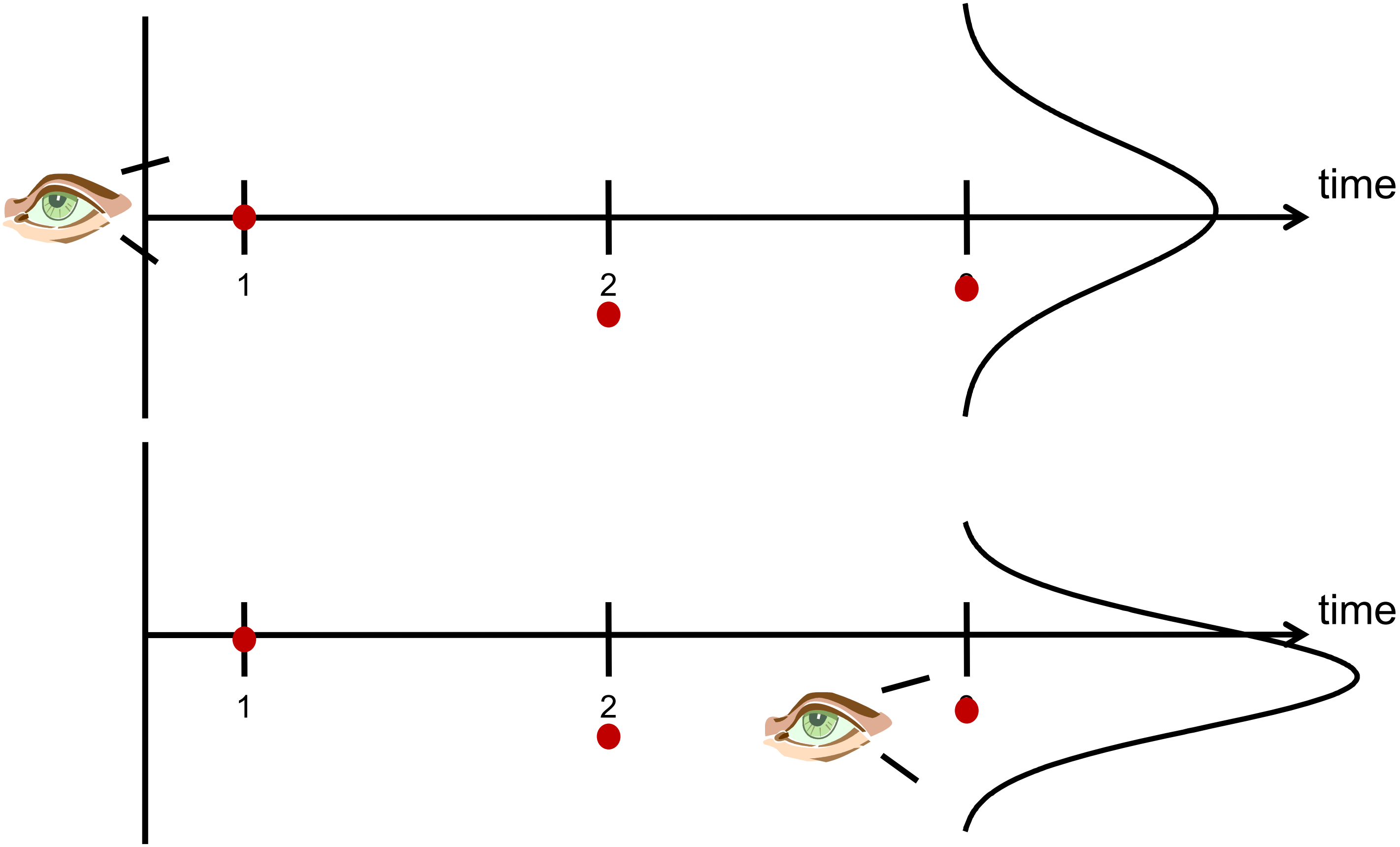
$$f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1;\boldsymbol{\theta}) \cdot f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\theta}) \cdot f_{Y_1}(y_1;\boldsymbol{\theta})$$

In general,

$$f_{Y_3|Y_2,Y_1}(y_3 \mid y_2,y_1;\boldsymbol{\theta}) \neq f_{Y_3}(y_3;\boldsymbol{\theta})$$

$$E[Y_3|Y_2 = y_2, Y_1 = y_1] \neq E[Y_3]$$

Stochastic Time Series Models



Stationary Time Series

The underlying stochastic process is assumed to be **invariant with respect to time**.

The set of data points $y_1, y_2, \ldots, y_T$ represents a particular outcome of the joint probability distribution function $F(y_1, y_2, \ldots, y_T)$.

A future observation y_{T+1} can be thought of as being generated by a conditional probability distribution function $F(y_{T+1} | y_1, y_2, \ldots, y_T)$.

Stationary Time Series

A stochastic process is stationary when its joint distribution and conditional distribution are **invariant with respect to time**.

$$F(y_t,\dots,y_{t+k}) = F(y_{t+m},\dots,y_{t+k+m}) \quad \text{and}$$
$$F(y_t) = F(y_{t+m}) \quad \text{for any } t, k, \text{and } m$$

Note that if the series y_t is stochastic, also the mean, the variance and the covariance must be stationary.

Mean, Variance, and Covariance of a Stationary Time Series

$$\mu_y = E(y_t) = E(y_{t+m})$$

$$\sigma_y^2 = E\left[(y_t - \mu_y)^2\right] = E\left[(y_{t+m} - \mu_y)^2\right]$$

$$\gamma_k = Cov(y_t, y_{t+k}) = E\left[(y_t - \mu_y)(y_{t+k} - \mu_y)\right] = Cov(y_{t+m}, y_{t+m+k})$$

for any t, k, and m

Implications of Stationarity for time series data

$$\mu_y = E(y_t) = E(y_{t+m}) \quad \rightarrow \quad \text{no trend}$$

$$\sigma_y^2 = E\left[(y_t - \mu_y)^2\right] = E\left[(y_{t+m} - \mu_y)^2\right]$$

→variance (magnitude of fluctuations) constant

$$\gamma_k = Cov(y_t, y_{t+k}) = E\left[(y_t - \mu_y)(y_{t+k} - \mu_y)\right] = Cov(y_{t+m}, y_{t+m+k})$$

for any *t, k,* and *m*

→pattern of serial correlation does not change (hard to spot)

Implications of Stationarity for estimation

$$\mu_y = E(y_t) = E(y_{t+m}) \qquad \hat{\mu}_y = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

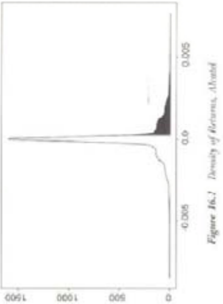
$$\sigma_y^2 = E\left[(y_t - \mu_y)^2\right] = E\left[(y_{t+m} - \mu_y)^2\right] \qquad \hat{\sigma}_y^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2$$

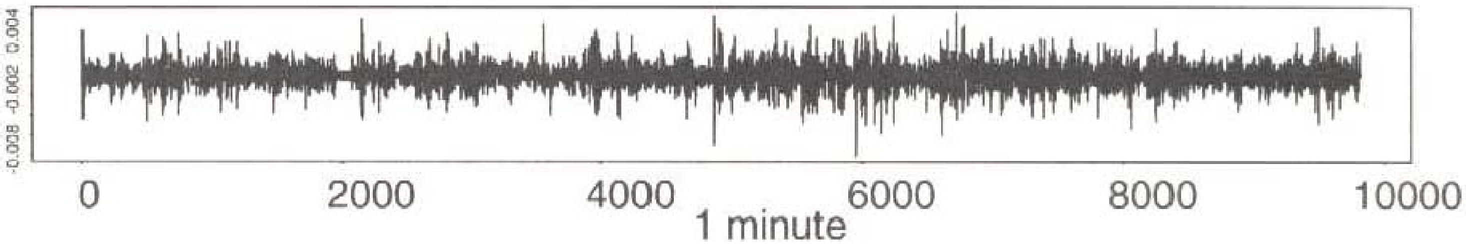
$$\gamma_1 = Cov(y_t, y_{t+1}) = E\left[(y_t - \mu_y)(y_{t+1} - \mu_y)\right] = Cov(y_{t+m}, y_{t+m+1})$$

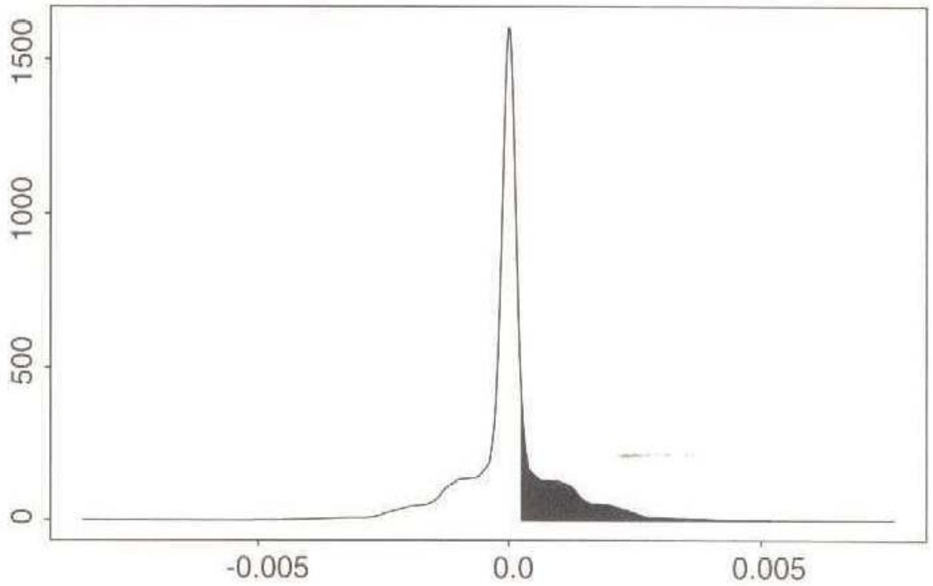
$$\hat{\gamma}_1 = \hat{C}ov(y_t, y_{t+1}) = \frac{1}{T-1}\sum_{t=1}^{T-1}(y_t - \bar{y})(y_{t+1} - \bar{y})$$

Implications of stationarity

(minute-by-minute) returns on Alcatel stock







$$\mu_y = E(y_t) \quad \Rightarrow \quad \hat{\mu}_y = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

$$\sigma_y^2 = E\left[(y_t - \mu_y)^2\right]$$

$$\Rightarrow \quad \hat{\sigma}_y^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2$$

Figure 16.1 Density of Returns, Alcatel

Goutieroux and Jasiak (2001), p. 31 and p. 429

Example: Random walk process $y_t = y_{t-1} + \varepsilon_t$

$$\begin{aligned}
\gamma_0 &= E\left(y_t^2\right) = E\left[\left(y_{t-1} + \varepsilon_t\right)^2\right] = E\left(y_{t-1}^2\right) + \sigma_\varepsilon^2 \\
&= E\left(y_{t-2}^2\right) + 2\sigma_\varepsilon^2 \\
&= \qquad \vdots \\
&= E\left(y_{t-n}^2\right) + n\sigma_\varepsilon^2
\end{aligned}$$

The variance is infinite and hence undefined.

Making a nonstationary process stationary

Example:

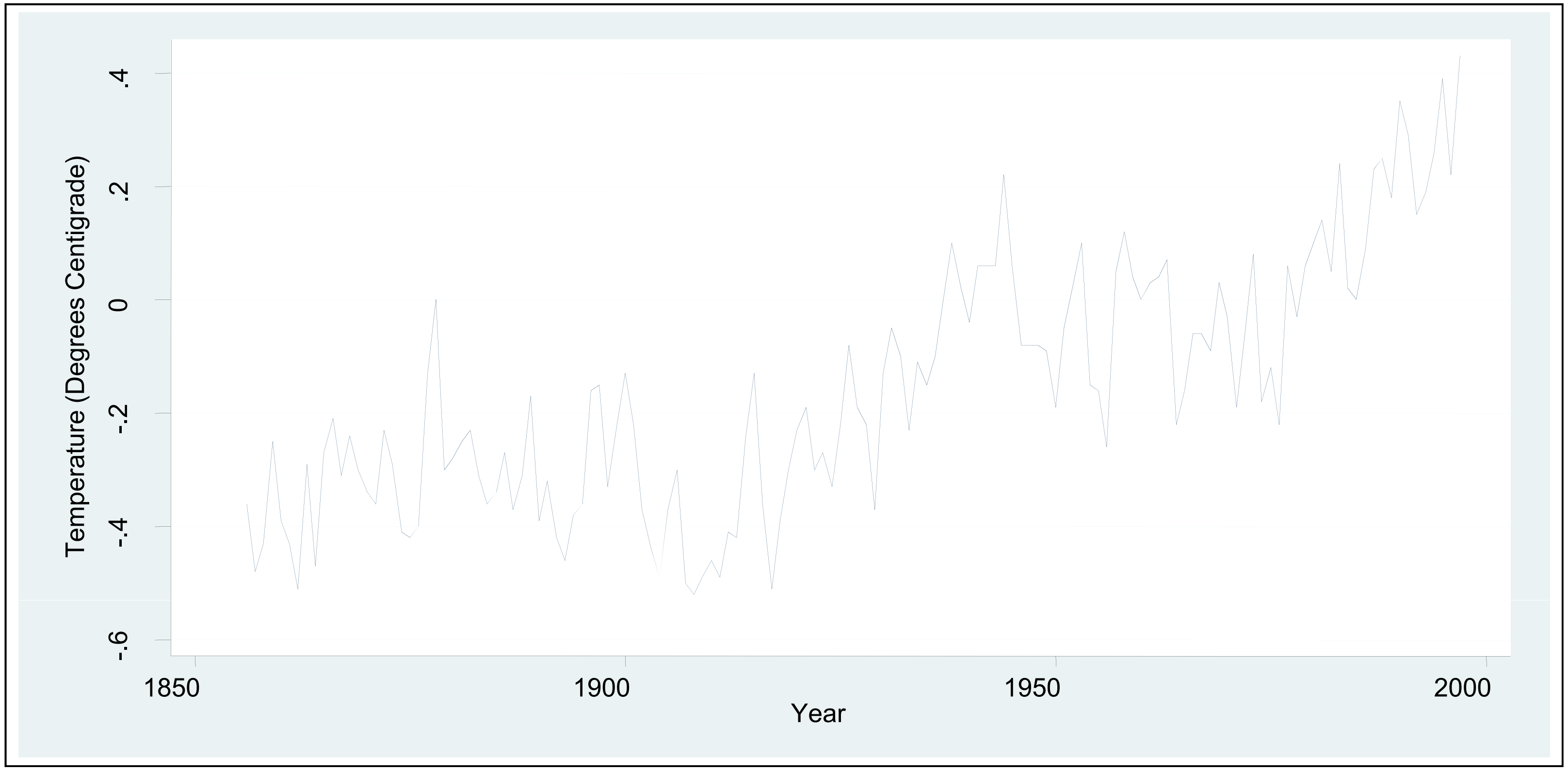
Differencing the Random walk process

$$w_t = \Delta y_t = y_t - y_{t-1} = \varepsilon_t$$

Since the ε_t are assumed to be independent over time, w_t is a stationary process (white noise).

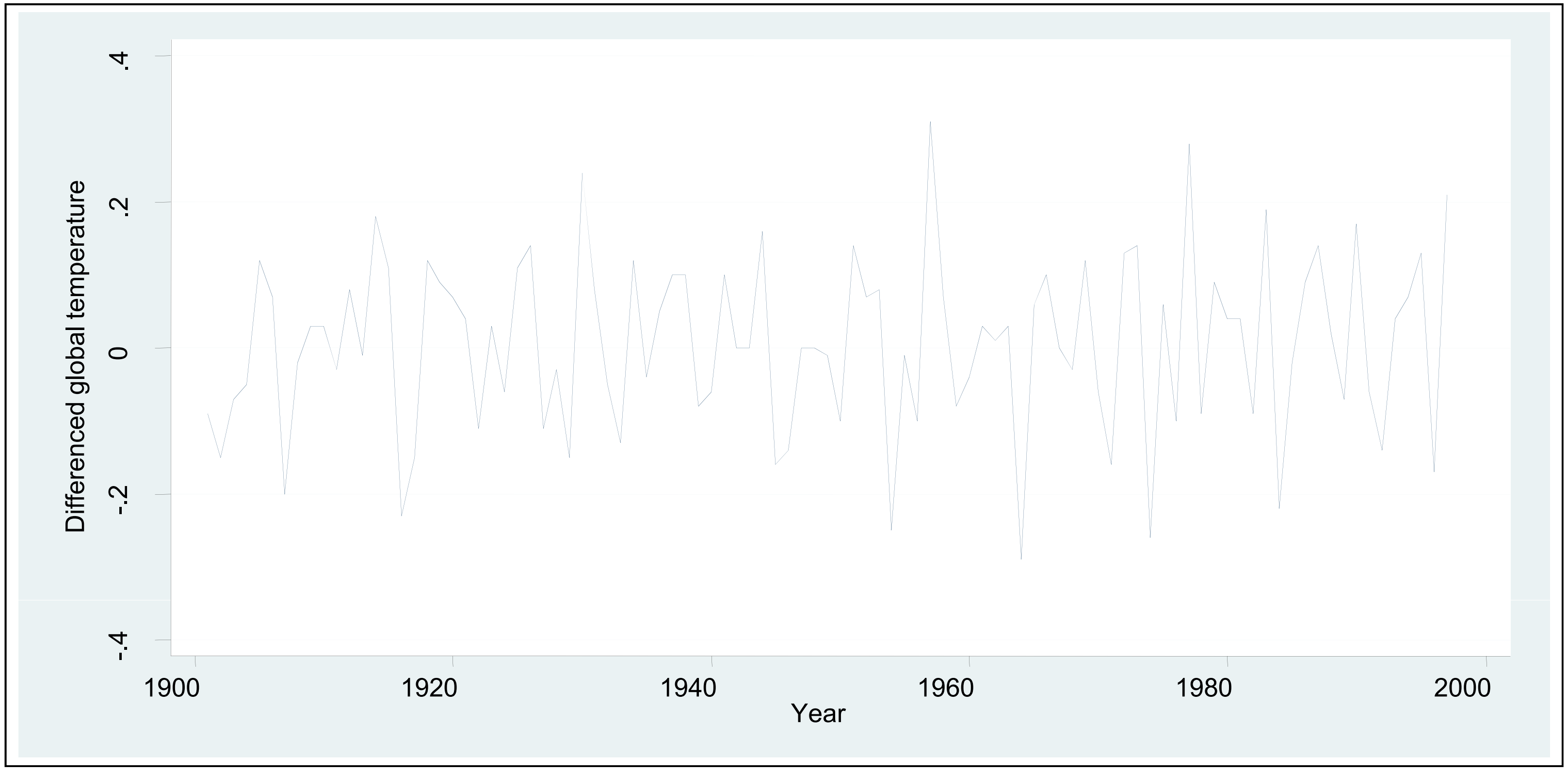
So the random walk process is first-order homogeneous.

Global warming data



Shumway/Stoffer (2000) “Time Series Analysis and its Applications”

Differenced Global warming data



Shumway/Stoffer (2000) “Time Series Analysis and its Applications”

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• **Identification**
	    • **Autocorrelation Function**
	    • Moving Average and Autoregressive Models
	    • Partial Autocorrelation Function
	    • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

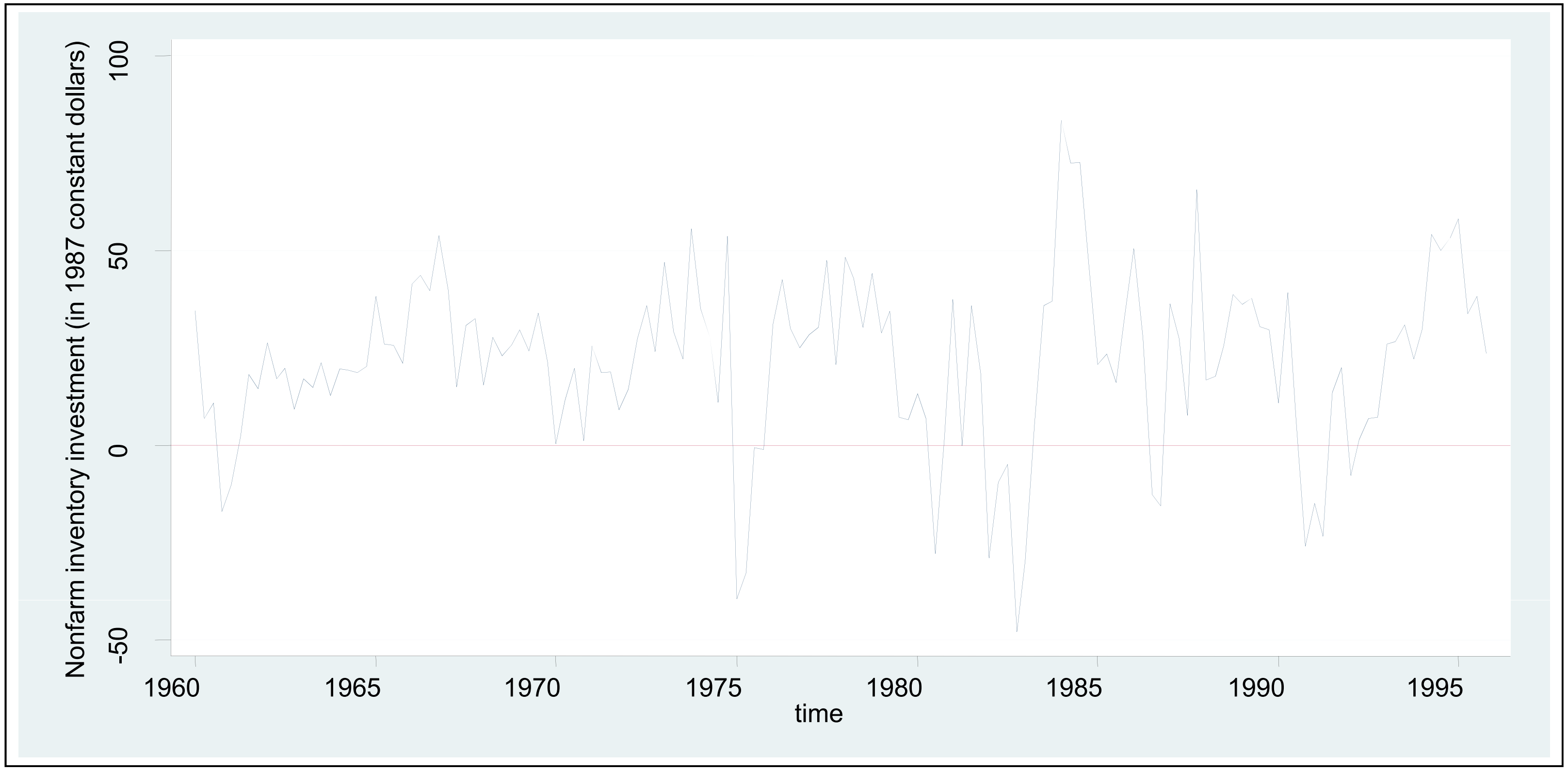
Introduction to Autocorrelation Functions

The basic idea is that the observations from different time periods may be related to each other.

First the observations (expressed in deviations from the mean) are arranged to create ordered pairs so that each observation is paired to the corresponding observation that occurred k periods earlier.

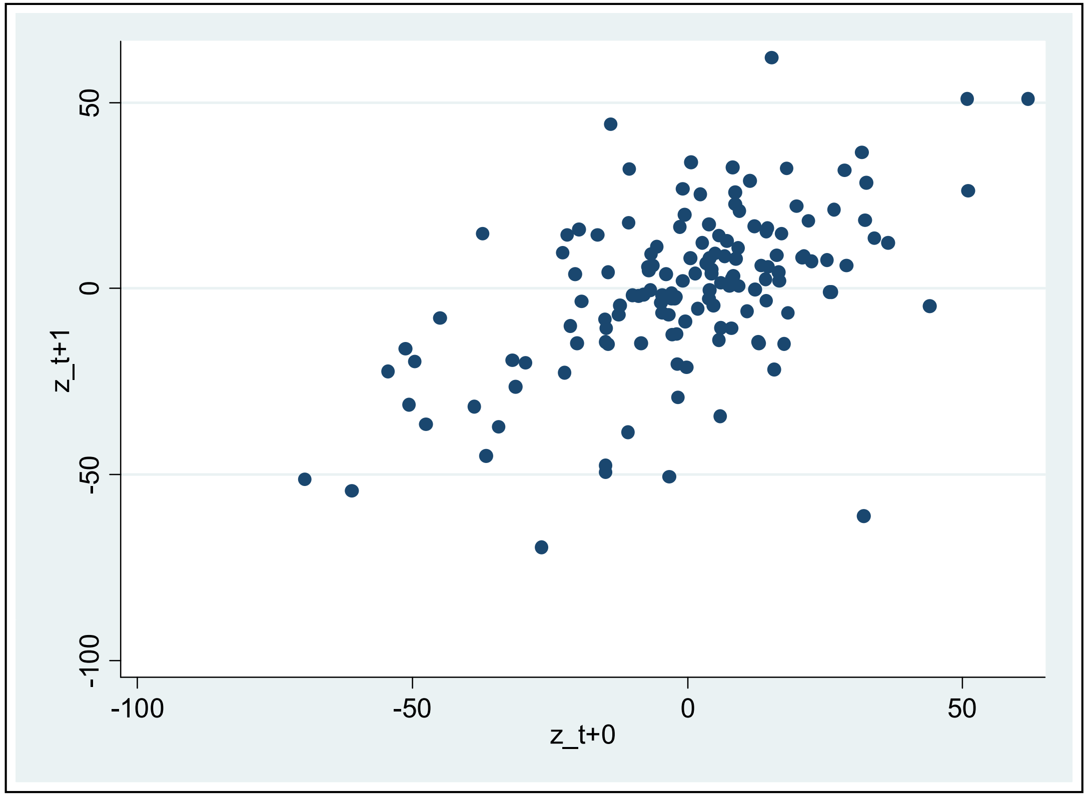
Then the ordered pairs are plotted in a two-space graph to see how the observations occurred in period *t* are related, on average, to the observations occurred in period *t* + *k*.

Nonfarm Inventory Investment data



Pindyck/Rubinfeld (1998) “Econometric Models and Economic Forecasts”

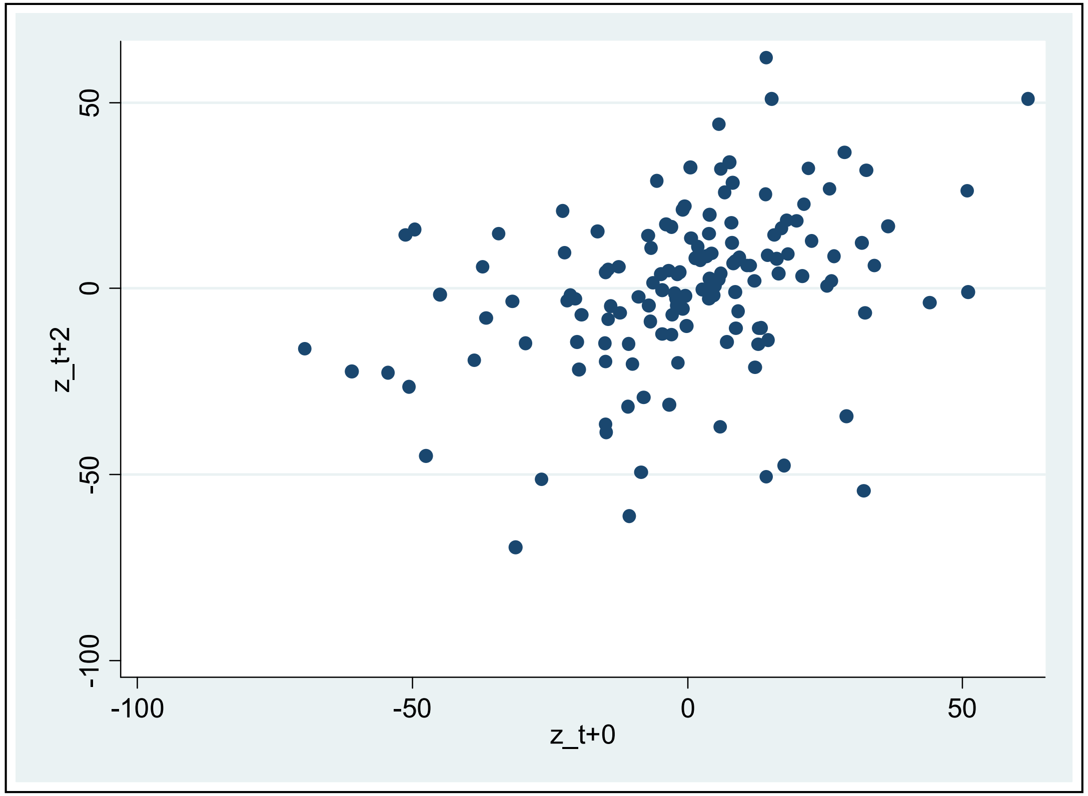
Plot of ordered pairs with *k* = 1 of Nonfarm Inventory Investment data:



On average there seems to be a positive relationship between these ordered pairs.

Pindyck/Rubinfeld (1998) "Econometric Models and Economic Forecasts"

Plot of ordered pairs with *k* = 2 of Nonfarm Inventory Investment data:



On average there seems to be a positive relationship between these ordered pairs.

Pindyck/Rubinfeld (1998) “Econometric Models and Economic Forecasts”

Introduction to Autocorrelation Functions

The autocorrelation function summarizes this information of many of these ordered pair plots by calculating a correlation coefficient for each set of ordered pairs.

The correlation between sets of values that are part of the same series is called **autocorrelation**.

Autocorrelation Function with lag *k*

$$\rho_k = \frac{E\left[\left(y_t - \mu_y\right)\left(y_{t+k} - \mu_y\right)\right]}{\sqrt{E\left[\left(y_t - \mu_y\right)^2\right]E\left[\left(y_{t+k} - \mu_y\right)^2\right]}} = \frac{Cov\left(y_t, y_{t+k}\right)}{\sigma_{y_t}\sigma_{y_{t+k}}}$$

$$-1 \leq \rho_k \leq 1$$

If we can predict y_{t+k} perfectly from y_t through a linear relationship, $y_{t+k} = \beta_0 + \beta_1 y_t$, then the correlation will be 1 when $\beta_1 > 0$, -1 when $\beta_1 < 0$, and 0 when $\beta_1 = 0$.

So the autocorrelation function can be seen as a measure of the ability to forecast the series at time *t+k* from the value at time *t*.

Autocorrelation Function with lag *k* of a stationary time series

Note, that for a stationary process the variance at time *t* is the same as at time *t* + *k*; thus the denominator is just the variance of the stochastic process.

$$\rho_k = \frac{E\left[\left(y_t - \mu_y\right)\left(y_{t+k} - \mu_y\right)\right]}{\sigma_y^2}$$

$\rho_k = \dfrac{\gamma_k}{\gamma_0}$ so for any stoachstic process $\rho_0 = 1$

Sample Autocorrelation Function with lag *k*

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

This estimation of the autocorrelation function may suggest which of the time series models is suitable to reflect the dependence in the data.

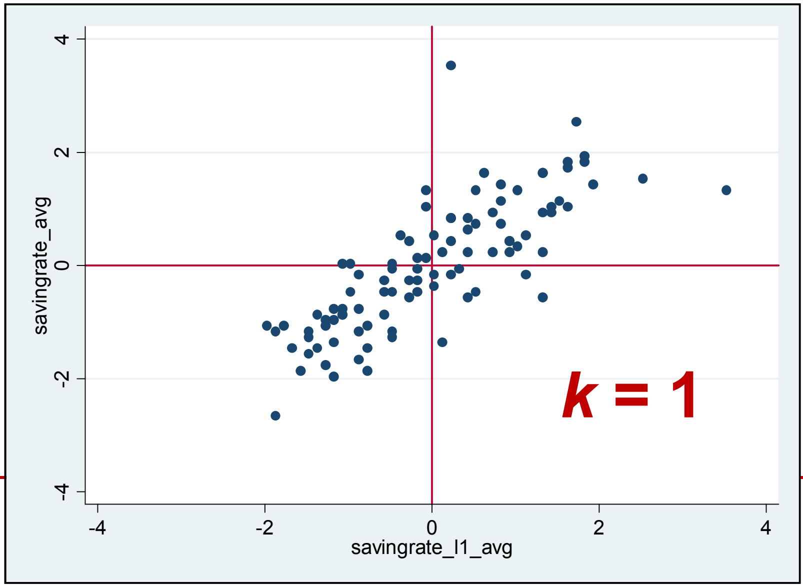
Note, that the autocorrelation function is **symmetrical** and ρ_k is plotted only for different **positive** values of k.

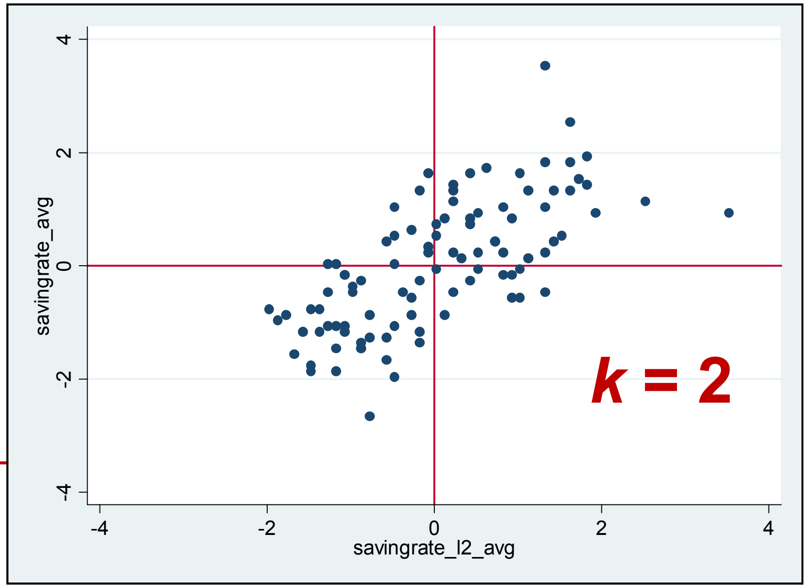
Example: 100 quarterly observations of the US saving rate for years 1955-1979

Note that the data are seasonally adjusted prior to publication by the U.S. Department of Commerce.

Pankratz (1983) "Forecasting with univariate Box-Jenkins models"

Plots of ordered pairs

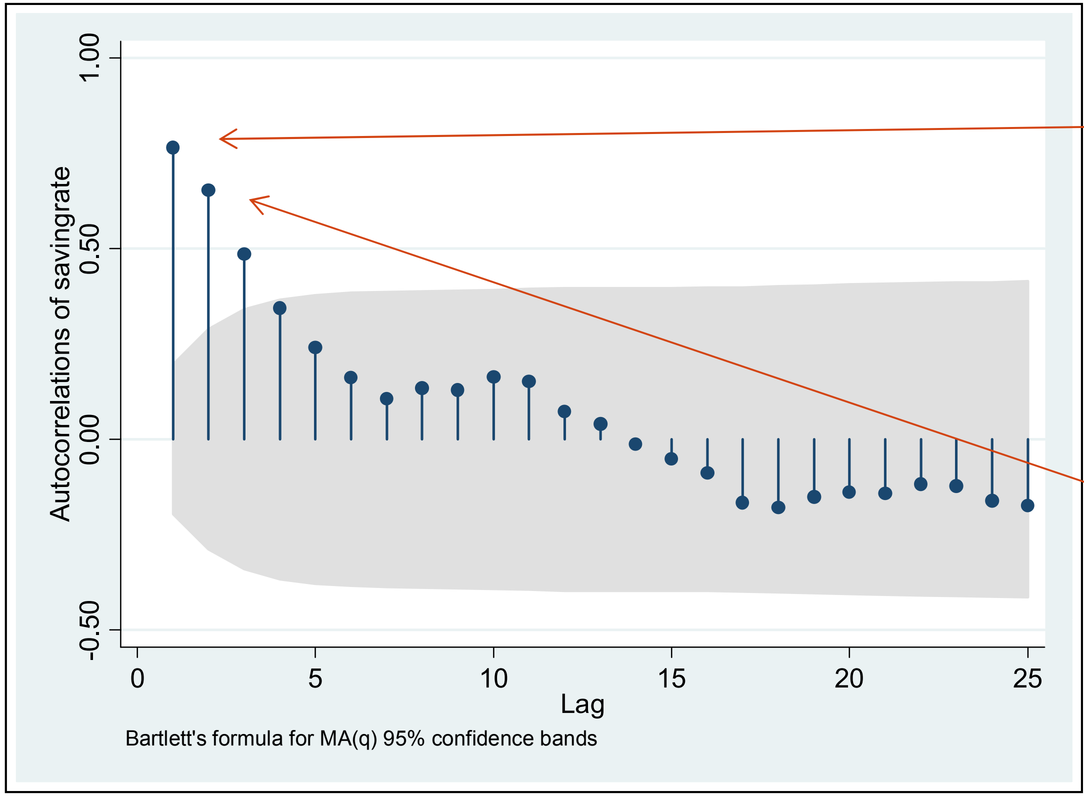


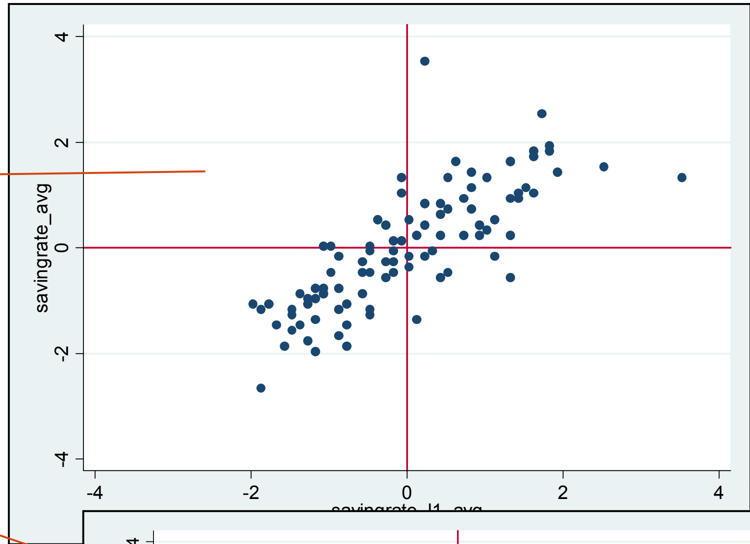


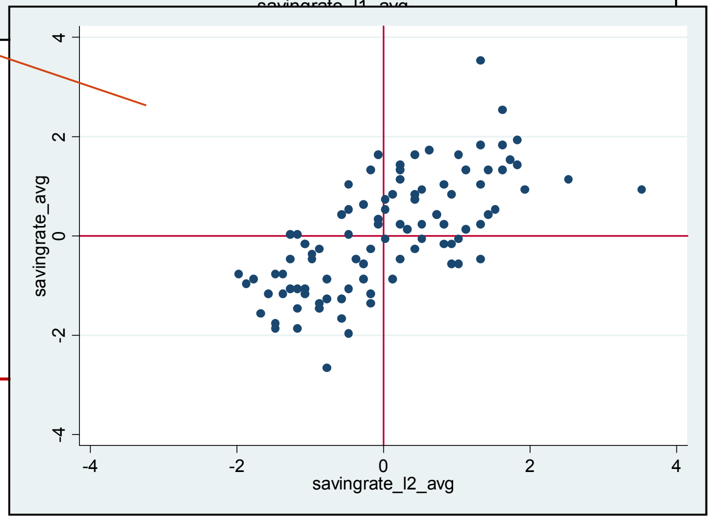
$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Sample Autocorrelation Function

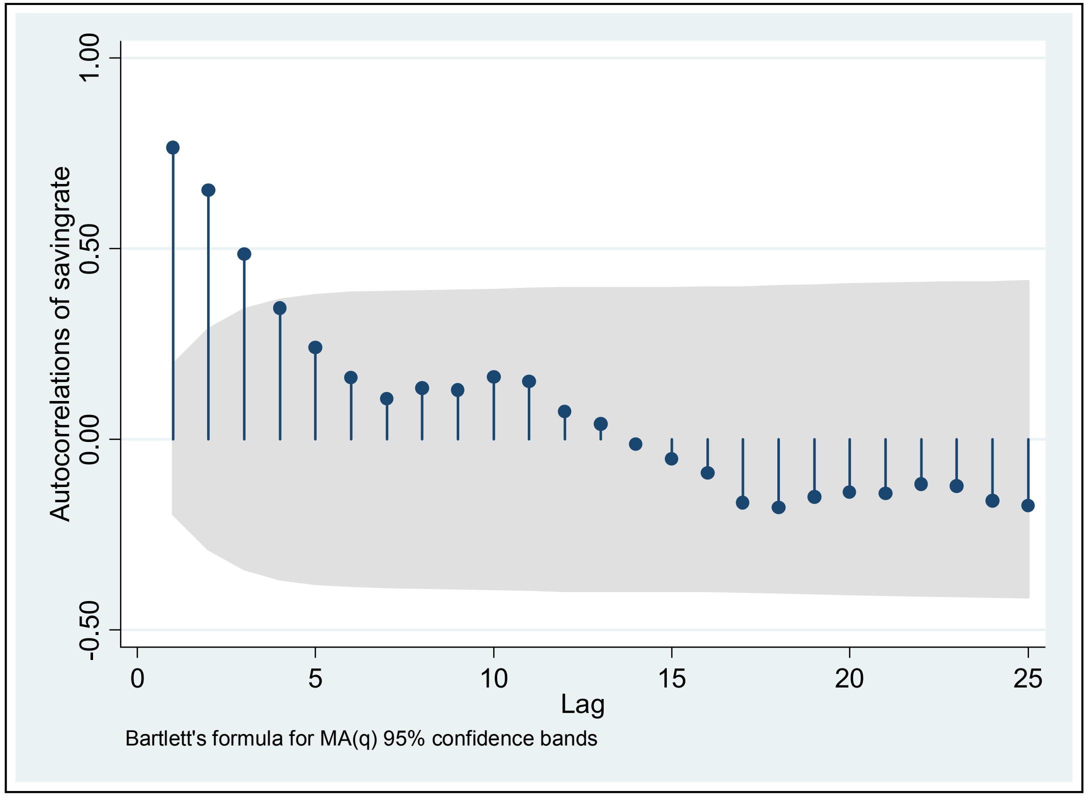
(Quarterly observations of the US saving rate for years 1955-1979)







Example: ACF of the saving rate



Pankratz (1983) “Forecasting with univariate Box-Jenkins models”

Schedule

Deterministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

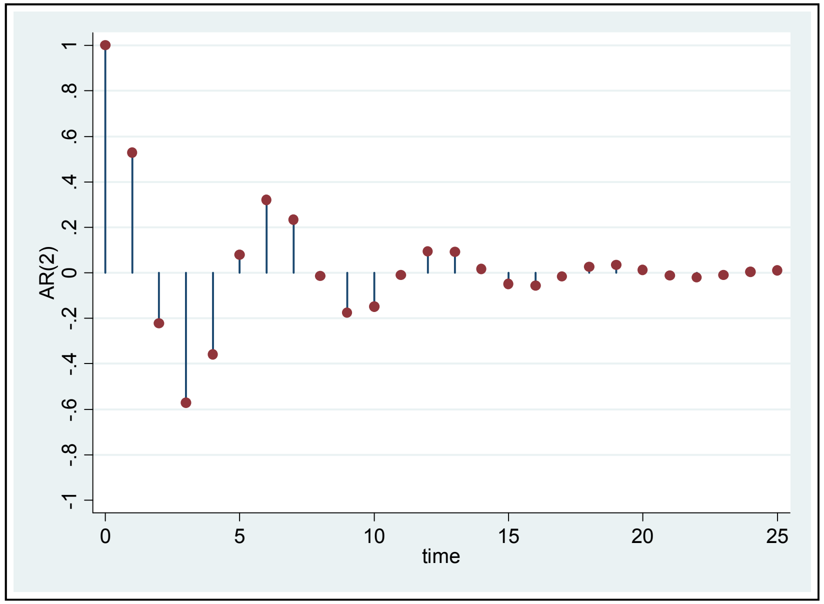
Stationary Stochastic Processes	• Introduction
	• **Identification**
	    • Autocorrelation Function
	    • **Moving Average and Autoregressive Models**
	    • Partial Autocorrelation Function
	    • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

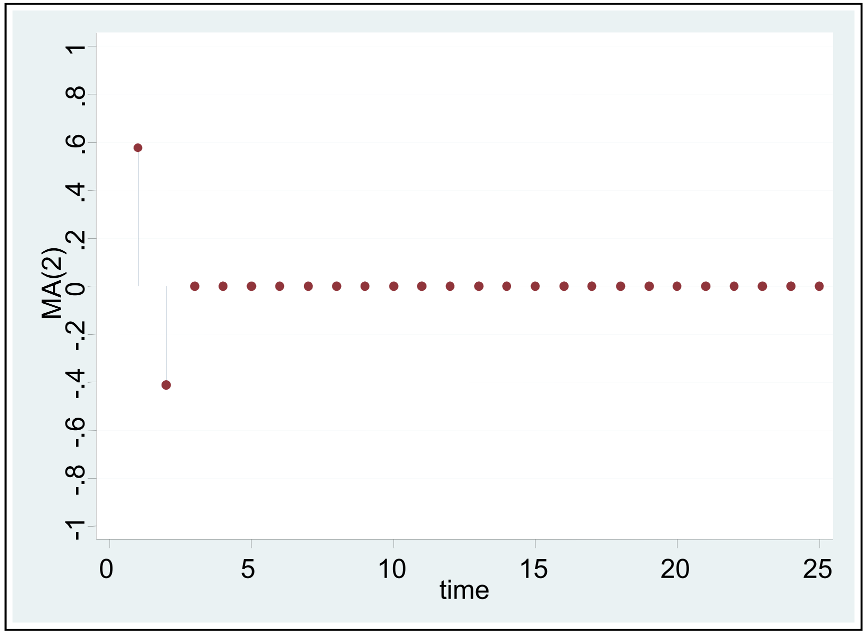
Examples: Models and their ACFs

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \delta + \varepsilon_t$$

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$



Theoretical ACF for φ_1 = 0.9 and φ_2 = -0.7 :



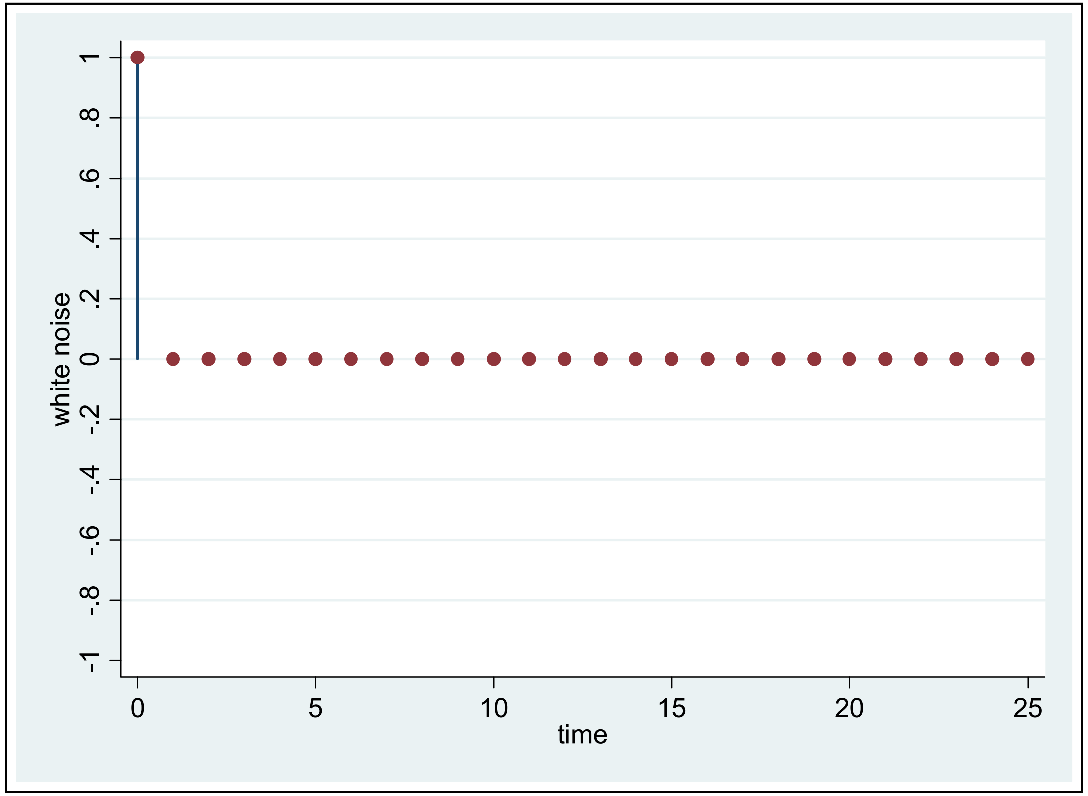
Theoretical ACF for θ_1 = -0.6, θ_2 = 0.3

White Noise

$$y_t = \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim i.i.d. \quad \text{and} \quad E(\varepsilon_t) = 0$$

The autocorrelation function of white noise is given by $\rho_0 = 1$ and $\rho_k = 0$ for $k > 0$. No model provides a better forecast than $\hat{y}_{T+l} = 0$ for all l.

Theoretical autocorrelation function of white noise



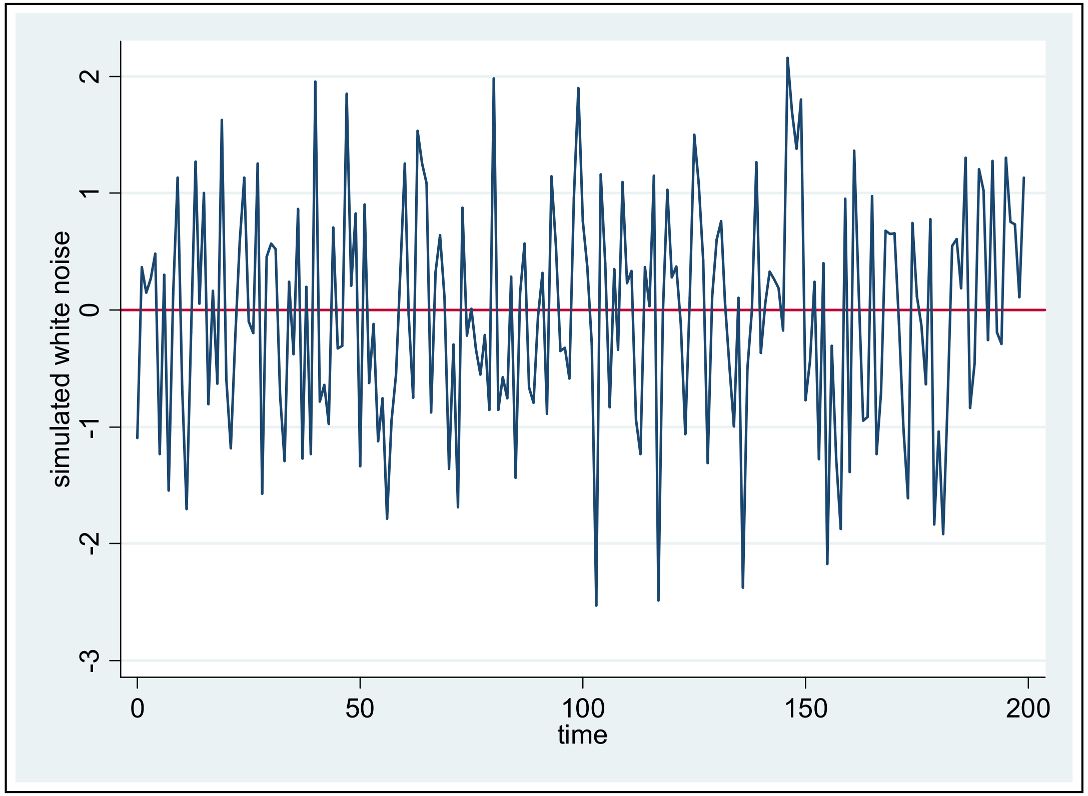
Gaussian White Noise

$$y_t = \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim \text{i.i.d. } N(0, 1)$$

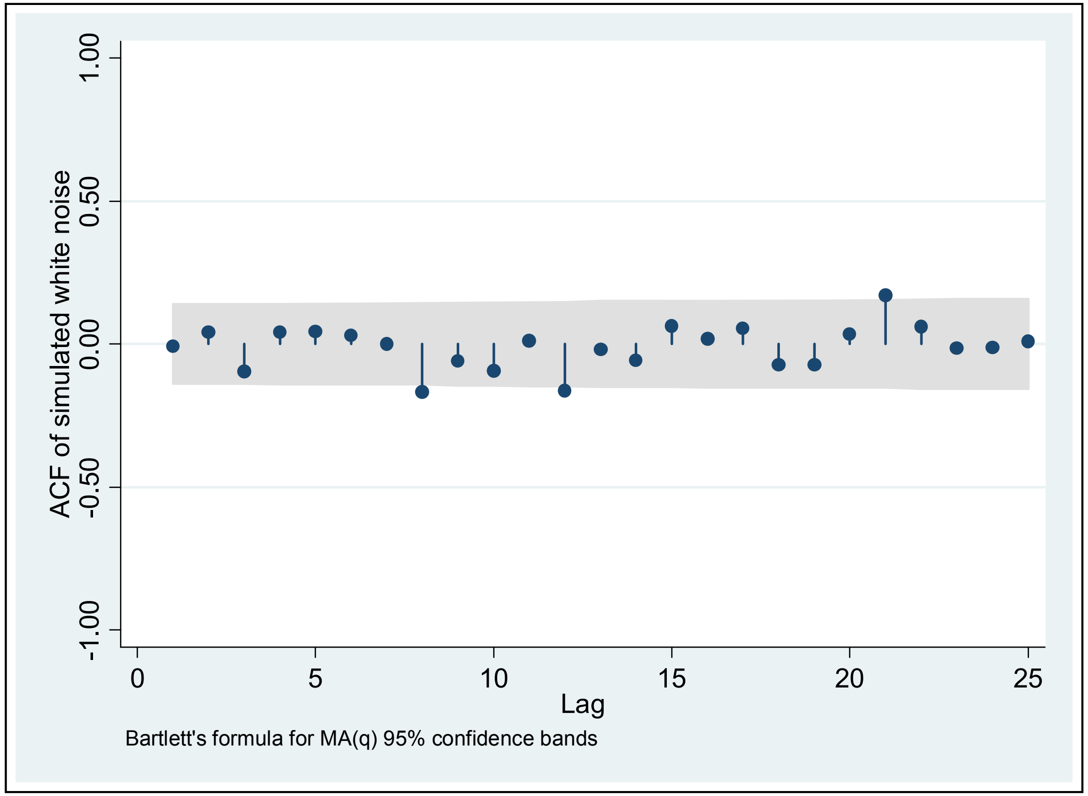
Unconditional density of y_t :

$$\mathrm{f}_{\mathrm{y}_\mathrm{t}}(y_t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[\frac{-y_t^2}{2\sigma^2}\right]$$

Simulated values of N(0,1), i.e. Gaussian white noise



Autocorrelation Function of the simulated values of Gaussian white noise



Wolds Decomposition Theorem

Any purely nondeterministic stationary process can be written as the linear combination of a sequence of uncorrelated random variables:

$$y_t - \mu_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \qquad \text{with } \psi_0 = 1 \text{ and } \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

where $E(\varepsilon_t) = 0, Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2,$

$Cov(\varepsilon_t, \varepsilon_{t-k}) = E(\varepsilon_t, \varepsilon_{t-k}) = 0$ for all $k \neq 0$

The deterministic part (if there is one) can be perfectly forecasted from ist own past values, for instance the constant mean around which the process fluctuates

Wolds Decomposition Theorem

$$y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \textit{with } \psi_0 = 1$$

Any stationary time series can be written as a moving average of (all) past ‘shocks’!

This representation of y_t is also useful for computing moments (mean, variance, covariance) and prediction intervals.

Wolds Decomposition Theorem

$$y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \textit{with } \psi_0 = 1$$

Examples:

$$\text{AR}(1): \quad y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$$

$$= \delta / (1 - \varphi_1) + \sum_{j=0}^{\infty} \varphi_1^j \cdot \varepsilon_{t-j}$$

$$\Rightarrow \psi_1 = \varphi_1, \psi_2 = \varphi_1^2, \; \ldots \; , \psi_j = \varphi_1^j, \ldots \; ,$$

$$\text{MA}(2): \quad y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

$$\Rightarrow \psi_1 = -\theta_1, \psi_2 = -\theta_2; \quad \psi_j = 0 \text{ for all } j > 2$$

Mean, Variance and Covariance

$$y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{with } \psi_0 = 1$$

$$E(\varepsilon_t) = 0, Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2, Cov(\varepsilon_t, \varepsilon_{t-k}) = E(\varepsilon_t, \varepsilon_{t-k}) = 0 \text{ for all } k \neq 0$$

Hence, $E(y_t) = \mu$ and

$$\gamma_0 = Var[\tilde{y}_t] = E\left[(y_t - \mu)^2\right] = E\left[\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right)^2\right]$$

$$= E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots)(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots)]$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2$$

Mean, Variance and Covariance

$$y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{with } \psi_0 = 1$$

$$E(\varepsilon_t) = 0, Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2, Cov(\varepsilon_t, \varepsilon_{t-k}) = E(\varepsilon_t, \varepsilon_{t-k}) = 0 \text{ for all } k \neq 0$$

$$\begin{aligned}\gamma_k &= E[(y_t - \mu)(y_{t-k} - \mu)] \\ &= E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots)(\varepsilon_{t-k} + \psi_1 \varepsilon_{t-k-1} + \cdots + \psi_k \varepsilon_{t-k-k} + \cdots)] \\ &= E\left[\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right)\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-k-j}\right)\right] = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}\end{aligned}$$

Moving Average Model

Moving Average Process of order q:

$$\text{MA}(q): y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

Each observation is generated by a weighted average of random disturbances going back q periods.

Example of an MA process in the real world

"There are a number of ways that a moving average model might arise in the real world. Consider a small commodity market, say the Prickly Pear Commodity Exchange of Uganda, if such a thing existed. A commodity market is a place where buyers and sellers of a metal or an agriculture crop meet and trade. The current selling price might be thought to arise from the classical supply and demand arguments, but in practice there is great uncertainty about future supply by the miners or farmers, and so the users of the market have to maintain inventories of the good. The presence of uncertainty leads to the presence of speculators on the market, which sometimes complicates matters. However, the Uganda Prickly Pear market will probably not have attracted the attention of speculators and suppose that it is currently in equilibrium, so that everyone involved agrees on what should be the correct price at this time.

Granger (1989) "Forecasting in Business and Economics", 2nd ed., p. 55-56

Example of an MA process in the real world

If now some extra, unexpected piece of information reaches the market, for example, that a sudden tropical storm has wiped out half of the entire crop of prickly pears in Cuba or that a research doctor in Baltimore has found that rats regularly washed in prickly pear juice developed skin cancer, then this news item will very likely induce a change in price.

If P_t is the price series, the next price might be determined by

$$P_{t+1} = P_t + \varepsilon_{t+1}$$

where ε_{t+1} is the effect on price of the unexpected news item. This may be written

$$y_{t+1} = \varepsilon_{t+1}$$

where y_t is the price change series.

Granger (1989) “Forecasting in Business and Economics”, 2nd ed., p. 55-56

Example of an MA process in the real world

However, if the full impact of the news item is not immediately absorbed by the market, the price change on the next day might be formed by

$$y_{t+2} = \varepsilon_{t+2} + b\varepsilon_{t+1} \, ,$$

where ε_{t+2} is the effect of yet a further news item and $b\varepsilon_{t+1}$ reflects the reassessment of the earlier piece of news. If a sequence of unexpected news items keep affecting the price on the market and their complete impact takes several days to work out, the price change series will be well represented by a moving average model."

Granger (1989) "Forecasting in Business and Economics", 2nd ed., p. 55-56

Moving Average Model

$$\text{MA}(q): y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

The random disturbances are assumed to have mean zero, finite constant variance and to be uncorrelated:

$$E(\varepsilon_t) = 0 \text{ and } E(\varepsilon_t^2) = \sigma_\varepsilon^2 \text{ and } E(\varepsilon_t \varepsilon_{t-k}) = 0 \text{ for } k \neq 0$$

A stronger assumption, used below (estimation), is that the ε_ts are **Gaussian white noise** :

$$\varepsilon_t \sim \text{i.i.d. } N(0, \sigma_\varepsilon^2)$$

Moving Average Model

$$\mathrm{MA}(q): y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$$

The MA(q) process can be written in lag-operator notation which we will be using below

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$$

$$y_t = \left(1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q\right)\varepsilon_t$$

$$y_t = b_q(L)\varepsilon_t$$

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

$$L^2 y_t = L(L y_t) = L(y_{t-1}) = y_{t-2}$$

$$L^k y_t = y_{t-k} \qquad k = 1,2,3,\ldots$$

$$L^0 y_t = y_t$$

Moving Average Process of order 1, MA(1)

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

where μ and θ_1 are any constants

$$E(y_t) = \mu$$

$$Var(y_t) = \gamma_0 = \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \theta_1^2 = (1 + \theta_1^2)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-1}) = \gamma_1 = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2})] = -\theta_1 \sigma_\varepsilon^2$$

The MA(1) process has a covariance of zero when the displacement is more than one period.

Using

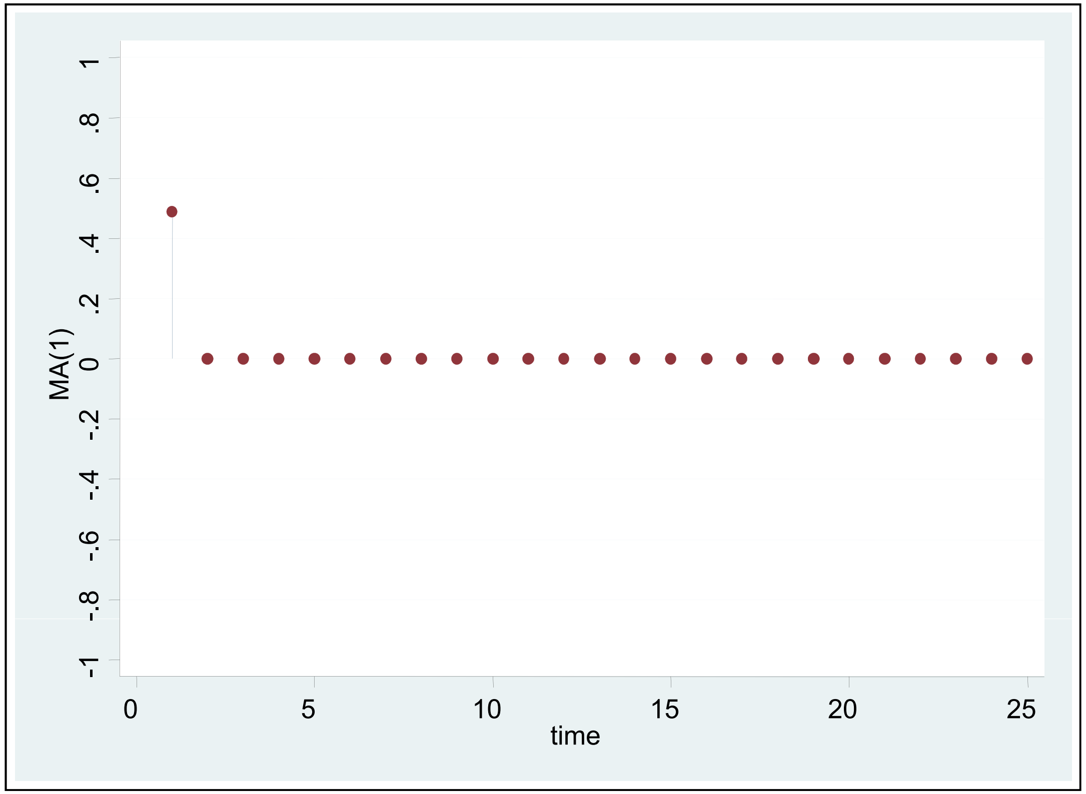
$$Var(y_t) = \gamma_0 = \sigma_\varepsilon^2 + \sigma_\varepsilon^2\theta_1^2 = (1+\theta_1^2)\sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-1}) = \gamma_1 = E[(\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-2})] = -\theta_1\sigma_\varepsilon^2$$

we obtain the **Autocorrelation function** for the **MA(1)** process

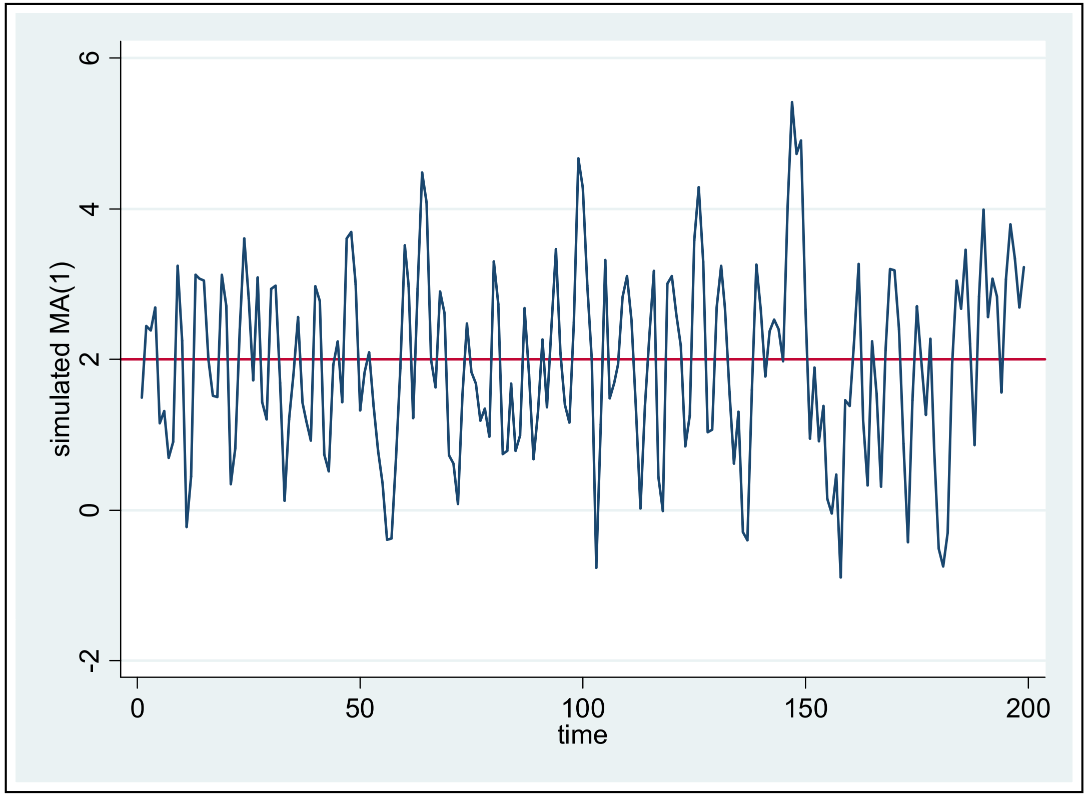
$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

Theoretical Autocorrelation Function with θ_1 = -0.8

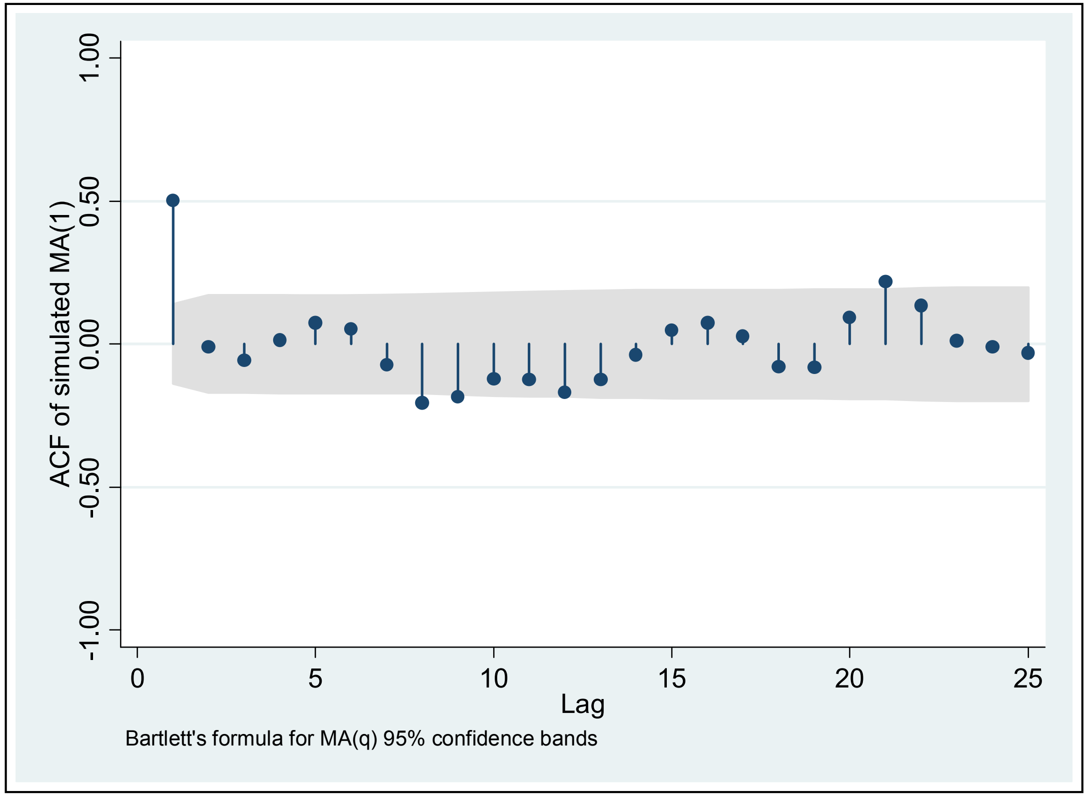


$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

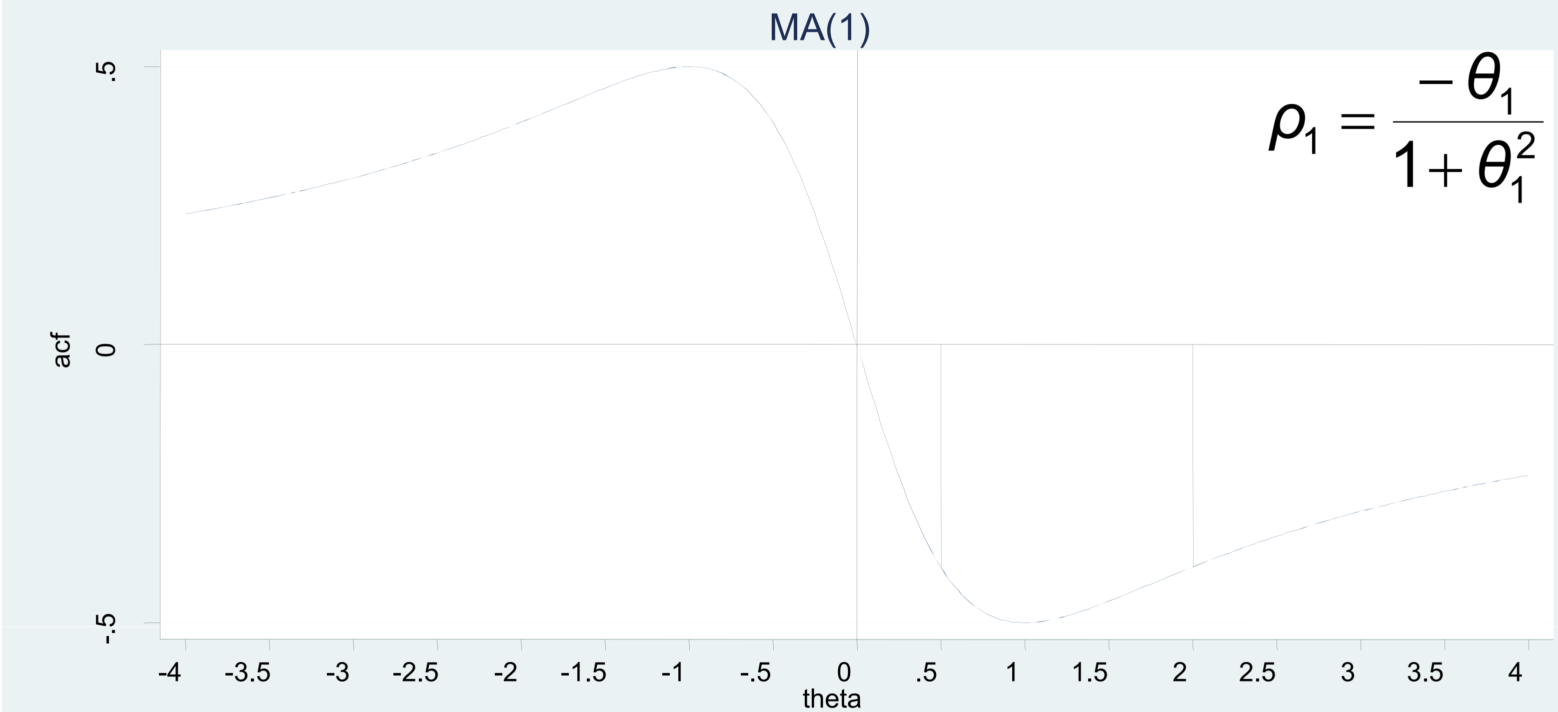
Example of a MA(1) process: $y_t = 2 + \varepsilon_t + 0.8\varepsilon_{t-1}$



Autocorrelation function for this MA(1) process



ACF MA(1)



$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$$

MA(1) **and Stationarity:**

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$E(y_t) = \mu$$
$$Var(y_t) = \gamma_0 = (1 + \theta_1^2)\sigma_\varepsilon^2$$
$$Cov(y_t, y_{t-1}) = \gamma_1 = -\theta_1 \sigma_\varepsilon^2$$
$$Cov(y_t, y_{t-k}) = \gamma_k = 0 \text{ for k} > 1$$

$$\mu_y = E(y_t) = E(y_{t+m})$$
$$\sigma_y^2 = E\left[(y_t - \mu_y)^2\right] = E\left[(y_{t+m} - \mu_y)^2\right]$$
$$\gamma_k = Cov(y_t, y_{t+k}) = Cov(y_{t+m}, y_{t+m+k})$$

for any *t, k,* and *m*

These results (i.e. weak stationarity) hold for any θ_1!

Moving Average Process of order 2, MA(2)

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

$$E(y_t) = \mu$$

$$Var(y_t) = \gamma_0 = \sigma_\varepsilon^2 \left(1 + \theta_1^2 + \theta_2^2\right)$$

$$Cov(y_t, y_{t-1}) = \gamma_1 = E\left[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3})\right]$$

$$= -\theta_1 \sigma_\varepsilon^2 + \theta_1 \theta_2 \sigma_\varepsilon^2 = -\theta_1 (1 - \theta_2) \sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-2}) = \gamma_2 = E\left[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3} - \theta_2 \varepsilon_{t-4})\right]$$

$$= -\theta_2 \sigma_\varepsilon^2$$

The MA(2) process has a covariance of zero when the displacement is more than two periods.

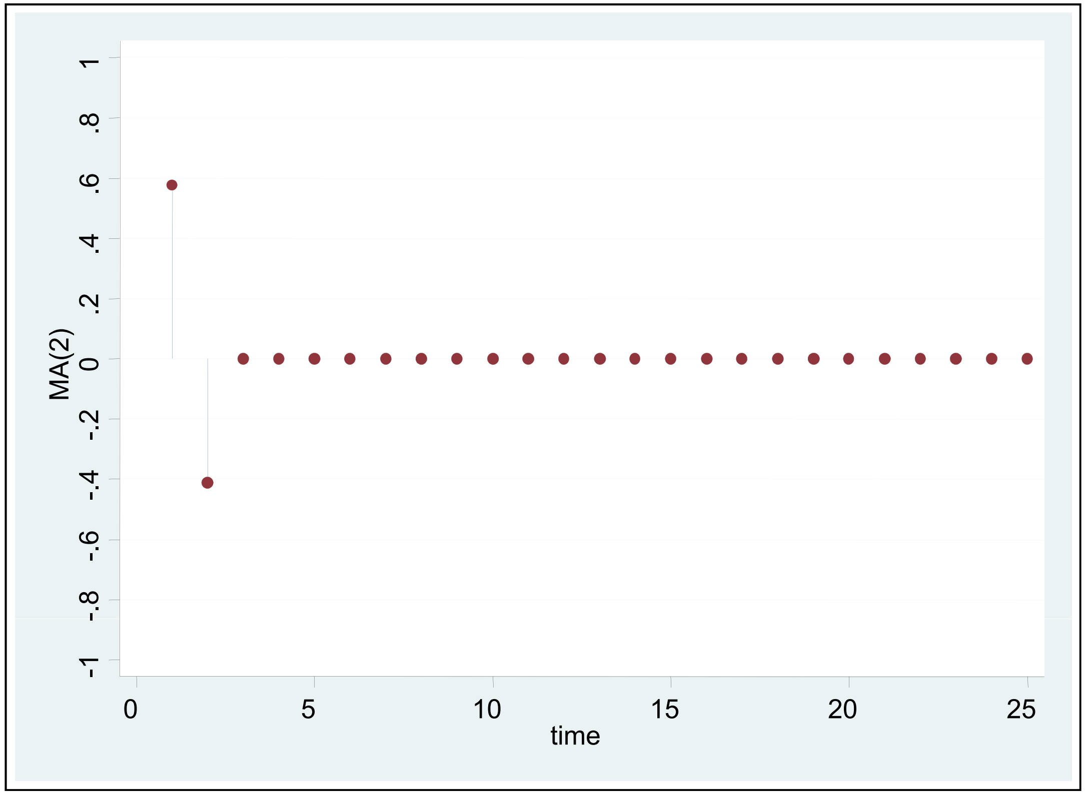
Autocorrelation function for the MA(2) process

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$$

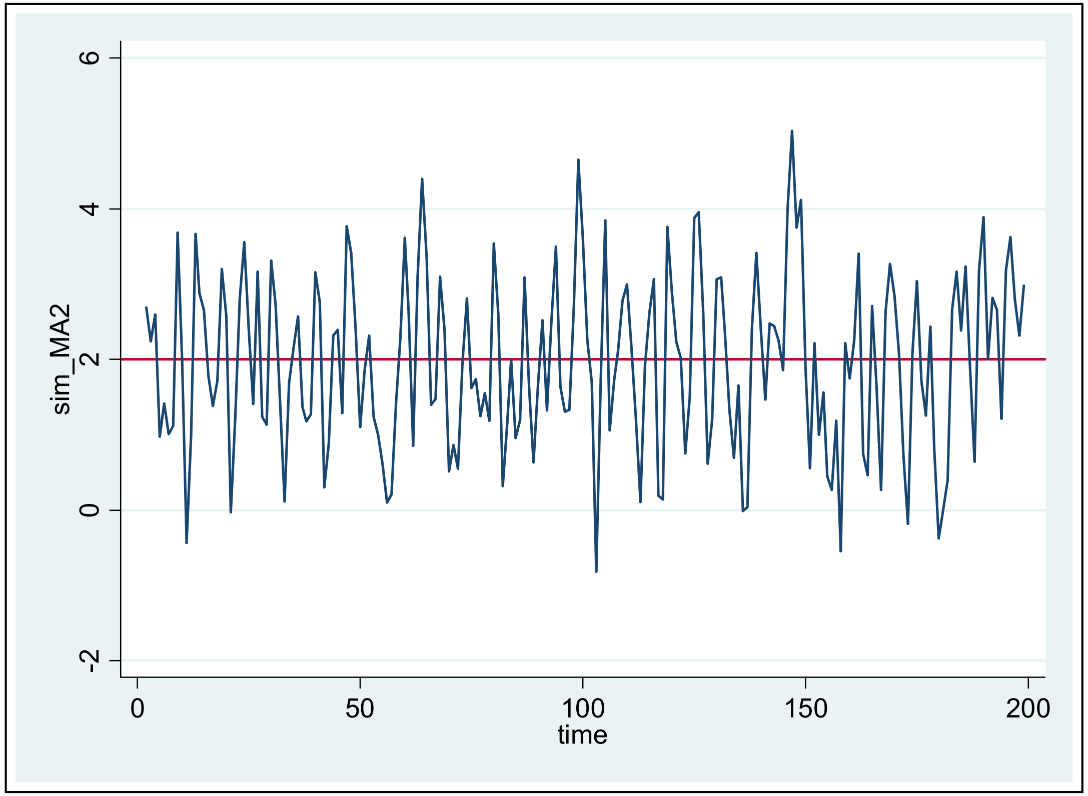
$$\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

$$\rho_k = 0 \quad \text{for } k > 2$$

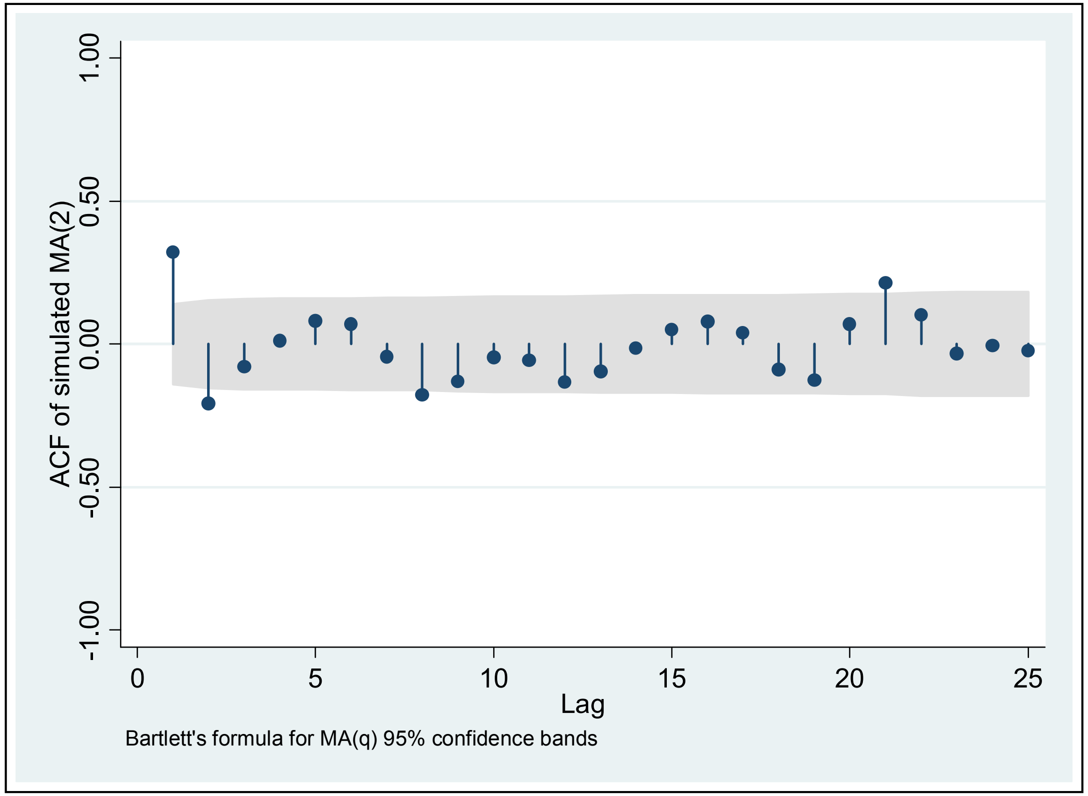
Theoretical autocorrelation function MA(2), θ_1 = -0.6, θ_2 = 0.3



Example of a MA(2) process: $y_t = 2 + \varepsilon_t + 0.6\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$



Autocorrelation function for the MA(2) process



MA(q) process

MA(q): $y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$

$$E(y_t) = \mu$$

$$Var(y_t) = \gamma_0 = E\left[(y_t - \mu)^2\right] = \sigma_\varepsilon^2\left(1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2\right)$$

$$Cov(y_t, y_{t-k}) = \gamma_k = \begin{cases} \left(-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q\right)\sigma_\varepsilon^2 & k = 1, \ldots, q \\ 0 & k > q \end{cases}$$

Autocorrelation function for the MA(*q*) process

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1+\theta_1^2 + \theta_2^2 + \ldots + \theta_q^2} & k = 1,\ldots,q \\ 0 & k > q \end{cases}$$

The autocorrelation function ρ_k for the MA(q) process has q nonzero values and is 0 for $k > q$.

Autoregressive process of order *p*

$$AR(p)\text{:}\ y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t$$

The current observation y_t is generated by a weighted average of past observations going back *p* periods, together with a random disturbance in the current period.

$\varepsilon_t \sim$ white noise

$E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_t \varepsilon_{t-k}) = 0$ for $k \neq 0$

Autoregressive process of order *p*

$$AR(p)\text{: } y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t$$

The AR(p) process can be written in lag-operator notation which we will be using below

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t$$
$$\left(1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p\right) y_t = \varepsilon_t$$
$$a_p(L) y_t = \varepsilon_t$$

$$L y_t = y_{t-1}$$
$$L^2 y_t = L(L y_t) = L(y_{t-1}) = y_{t-2}$$
$$L^k y_t = y_{t-k} \qquad k = 1,2,3,\ldots$$
$$L^0 y_t = y_t$$

Autoregressive Process of order 1, AR(1)

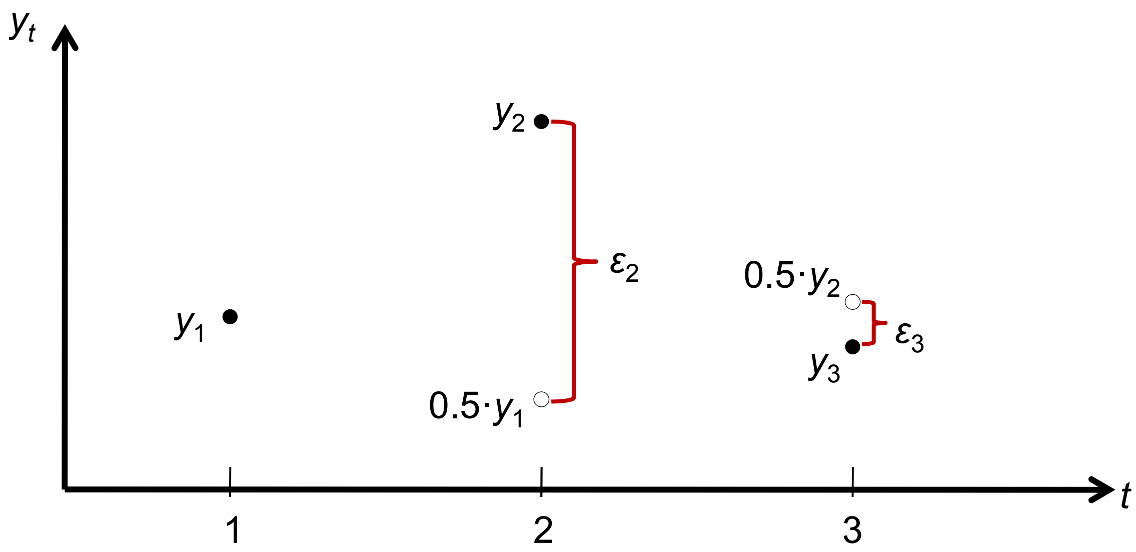
$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$$

This is the simplest AR process.

We will derive its properties (mean, variance, ACF,..).

However, unlike MA(1), it is stationary only under certain conditions.

Construction of an autoregressive process



$$\mathrm{AR}(1): y_t = \varphi_1 y_{t-1} + \varepsilon_t$$

for $\varphi_1 = 0.5$

AR(1) as a Stochastic First-order difference equation

Date	Equation
1	$y_1 = \varphi_1 \cdot y_0 + (\delta + \varepsilon_1)$
2	$y_2 = \varphi_1 \cdot y_1 + (\delta + \varepsilon_2)$
	$= \varphi_1 \cdot [\varphi_1 \cdot y_0 + (\delta + \varepsilon_1)] + (\delta + \varepsilon_2)$
	$= \varphi_1^2 \cdot y_0 + (\varphi_1 + 1)\delta + \varphi_1 \varepsilon_1 + \varepsilon_2$
3	$y_3 = \varphi_1 \cdot y_2 + (\delta + \varepsilon_3)$
	$= \varphi_1^3 \cdot y_0 + (\varphi_1^2 + \varphi_1 + 1)\delta + \varphi_1^2 \varepsilon_1 + \varphi_1 \varepsilon_2 + \varepsilon_3$
	$\vdots$
t	$y_t = \varphi_1^t \cdot y_0 + (\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1) \cdot \delta + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$

AR(1) as First-order difference equation

$$y_t = \varphi_1^t \cdot y_0 + \left(\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1\right) \cdot \delta + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$$

$$E[y_t] = \varphi_1^t \cdot y_0 + \left(\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1\right) \cdot \delta$$

$$E[y_{t+s}] = \varphi_1^{t+s} \cdot y_0 + \left(\varphi_1^{t+s-1} + \varphi_1^{t+s-2} + \ldots + \varphi_1^{t-1} + \ldots + \varphi_1 + 1\right) \cdot \delta$$

$$E[y_t] \neq E[y_{t+s}]$$

To get stationarity, we need to impose conditions.

AR(1) as First-order difference equation:

$$y_t = \varphi_1^t \cdot y_0 + \left(\varphi_1^{t-1} + \varphi_1^{t-2} + \ldots + \varphi_1 + 1\right) \cdot \delta + \sum_{j=1}^{t} \varphi_1^{t-j} \cdot \varepsilon_j$$

If $t \to \infty$ and $|\varphi_1|<1$

$$\lim y_t = \delta \big/ (1-\varphi_1) + \sum_{j=0}^{\infty} \varphi_1^j \cdot \varepsilon_{t-j} \qquad \Rightarrow E[y_t] = \delta \big/ (1-\varphi_1)$$

$t \to \infty$ means: the process has started a long time ago
"**stochastic initial conditions**"

$|\varphi_1|<1$ means: dependence can't be too strong

AR(1) and Wolds Decomposition

If $t \rightarrow \infty$ and $|\varphi_1|<1$

$$y_t = \delta \big/ (1-\varphi_1) + \sum_{j=0}^{\infty} \varphi_1^j \cdot \varepsilon_{t-j}$$

$$y_t = \mu + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1^2 \varepsilon_{t-2} + \ldots$$

$$\text{Wold}: \; y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$$

$$\Rightarrow \psi_1 = \varphi_1, \psi_2 = \varphi_1^2, \; \ldots \; , \psi_j = \varphi_1^j, \ldots \; ,$$

Hence we could calculate moments from Wolds formulas.

Preview: Nonstationarity AR

What happens if φ_1=1 ?

Random walk with drift $y_t = y_{t-1} + \delta + \varepsilon_t$

$$y_t = y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j$$

$$E(y_t) = E\left(y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j \right) = y_0 + \delta t$$

In contrast, a random walk (δ = 0) is mean stationary.

$$Var(y_t) = Var\left(y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j \right) = t \cdot \sigma_\varepsilon^2$$

$$Cov(y_t, y_{t-k}) = E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})] = (t-k) \cdot \sigma_\varepsilon^2$$

Autoregressive Process of order 1, AR(1):

$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$ with $|\varphi_1| < 1$

Standard way to derive moments (means, variances,..) for stationary AR:

assuming stationarity!

$$E[y_t] = \varphi_1 E[y_{t-1}] + \delta$$

E[y_t]=E[y_{t-1}]=μ

$$\mu = \varphi_1 \mu + \delta$$

$$\mu(1-\varphi_1) = \delta \qquad \Rightarrow E(y_t) = \mu = \frac{\delta}{1-\varphi_1}$$

Autoregressive Process of order 1, AR(1):

$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$ with $|\varphi_1| < 1$

We have derived $E(y_t) = \mu = \dfrac{\delta}{1-\varphi_1}$

This implies $\mu(1-\varphi_1) = \delta$

and the AR(1) model can be alternatively written as

$$\Rightarrow (y_t - \mu) = \varphi_1 (y_{t-1} - \mu) + \varepsilon_t$$

Autoregressive Process of order 1, AR(1):

$$E(y_t) = \mu = \frac{\delta}{1-\varphi_1}$$

$$\text{Assumption : zero mean } (\text{setting } \delta = 0)$$

$$\begin{aligned}
\gamma_0 &= Var(y_t) \\
&= E\left[(\varphi_1 y_{t-1} + \varepsilon_t)^2\right] \\
&= E\left(\varphi_1^2 y_{t-1}^2 + \varepsilon_t^2 + 2\varphi_1 y_{t-1}\varepsilon_t\right) \\
&= \varphi_1^2 \gamma_0 + \sigma_\varepsilon^2
\end{aligned}$$

$$\Rightarrow \gamma_0 = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}$$

Autoregressive Process of order 1, AR(1):

$$E(y_t) = \mu = \frac{\delta}{1-\varphi_1}$$

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}$$

$$\begin{aligned}\gamma_1 &= Cov(y_t, y_{t-1}) \\ &= E[y_{t-1}(\varphi_1 y_{t-1} + \varepsilon_t)] \\ &= \varphi_1 \gamma_0 = \frac{\varphi_1 \sigma_\varepsilon^2}{1-\varphi_1^2}\end{aligned}$$

Autoregressive Process of order 1, AR(1):

$$Cov(y_t, y_{t-2}) = \gamma_2 = E\left[y_{t-2}\left(\varphi_1^2 y_{t-2} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t\right)\right] = \varphi_1^2 \gamma_0 = \frac{\varphi_1^2 \sigma_\varepsilon^2}{1-\varphi_1^2}$$

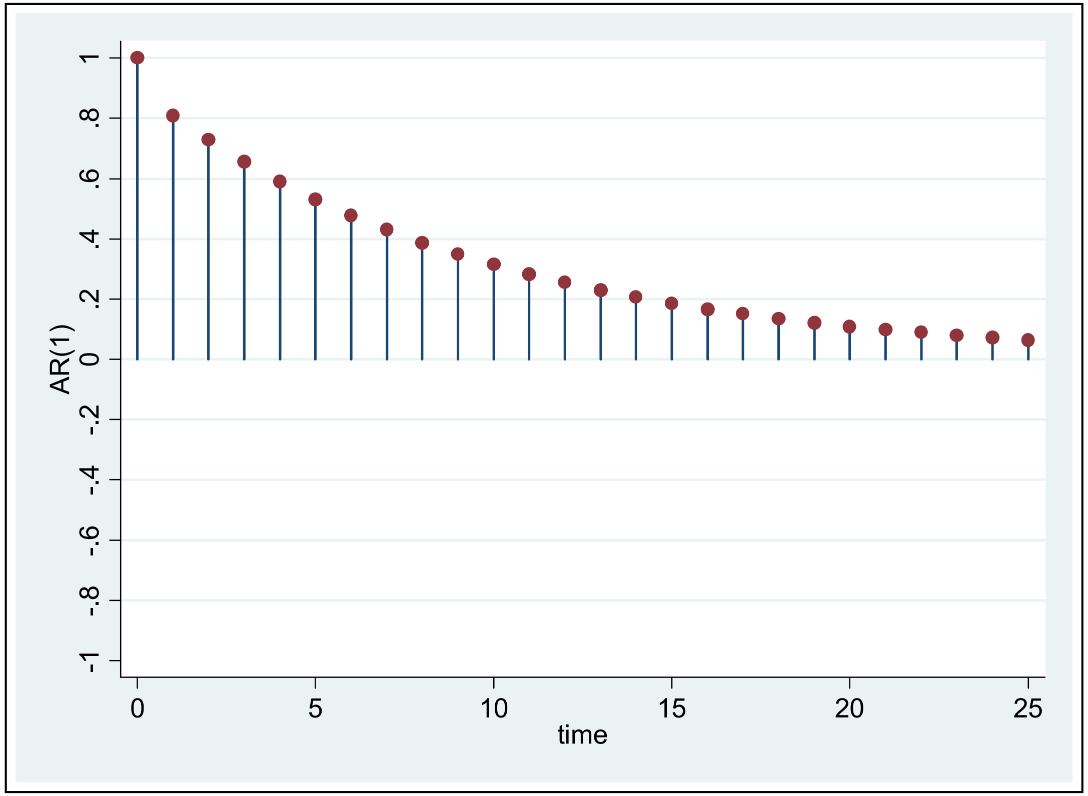
For a k - lag displacement :

$$Cov(y_t, y_{t-k}) = \gamma_k = \varphi_1^k \gamma_0 = \frac{\varphi_1^k \sigma_\varepsilon^2}{1-\varphi_1^2}$$

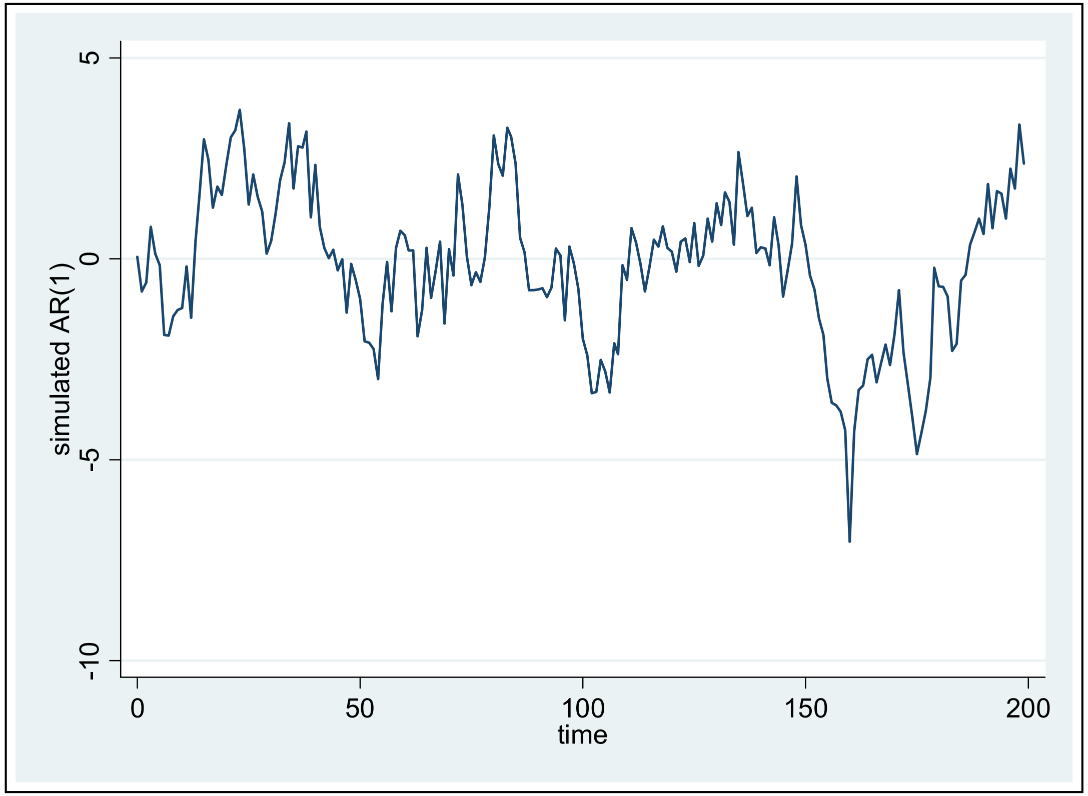
The autocorrelation function for a AR(1) process begins at ρ_0 = 1 and then declines geometrically.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1^k$$

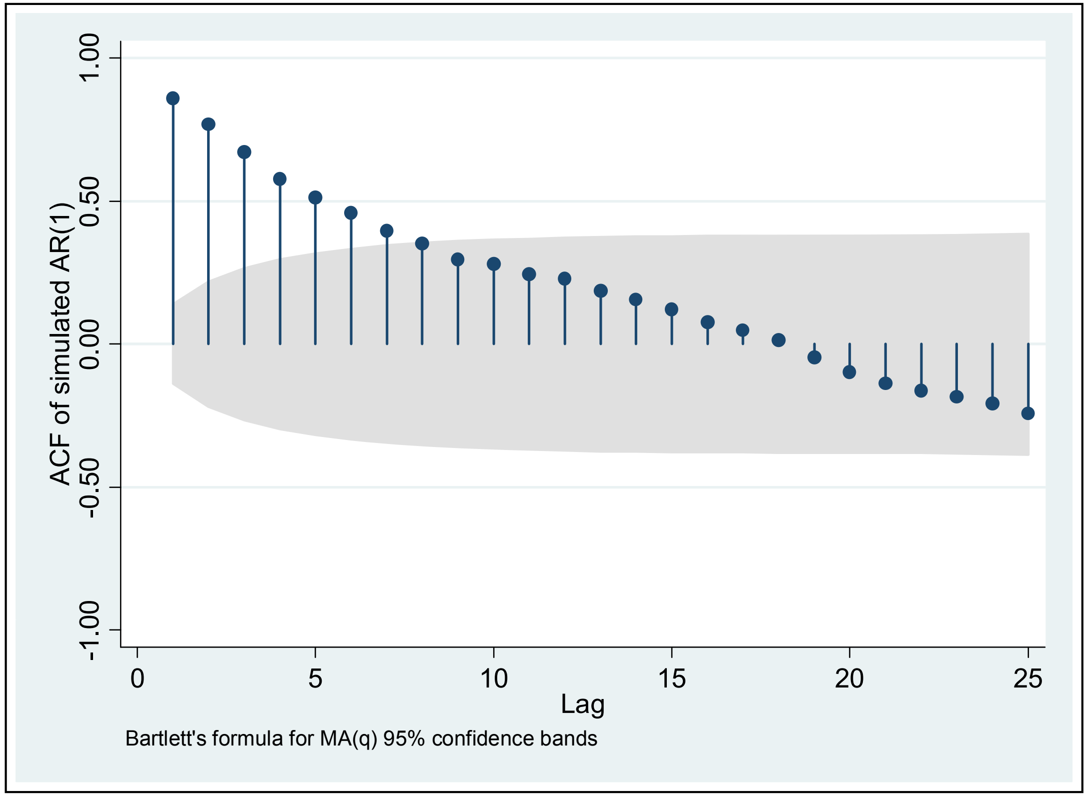
Theoretical ACF of an AR(1) process, $\varphi_1 = 0.9$:



Example of an AR(1) process: $y_t = 0.9 y_{t-1} + \varepsilon_t$



Autocorrelation function of the AR(1) process:



Autoregressive Process of order 1, AR(1):

$$\begin{aligned}
\gamma_0 &= Var(y_t) = E\left[(y_t)^2\right] - [E(y_t)]^2 = E\left[(\varphi_1 y_{t-1} + \delta + \varepsilon_t)^2\right] - \mu^2 \\
&= E\left(\varphi_1^2 y_{t-1}^2 + 2\delta\varphi_1 y_{t-1} + 2\varphi_1 y_{t-1}\varepsilon_t + \delta^2 + 2\delta\varepsilon_t + \varepsilon_t^2\right) - \mu^2 \\
&= \varphi_1^2 E\left(y_{t-1}^2\right) + 2\delta\varphi_1 E(y_{t-1}) + 2E(\varphi_1 y_{t-1})\underbrace{E(\varepsilon_t)}_{=0} + E\left(\delta^2\right) + 2\delta\underbrace{E(\varepsilon_t)}_{=0} + E\left(\varepsilon_t^2\right) - \mu^2 \\
&= \varphi_1^2 E\left(y_{t-1}^2\right) + 2\delta\varphi_1\mu + \delta^2 + \sigma_\varepsilon^2 - \mu^2 \\
&= \varphi_1^2 E\left(y_{t-1}^2\right) + 2\mu(1-\varphi_1)\varphi_1\mu + \delta^2 + \sigma_\varepsilon^2 - \mu^2 \\
&= \varphi_1^2 E\left(y_{t-1}^2\right) + 2\varphi_1\mu^2 - 2\varphi_1^2\mu^2 + \mu^2 - 2\varphi_1\mu^2 + \varphi_1^2\mu^2 - \mu^2 + \sigma_\varepsilon^2 \\
&= \varphi_1^2 E\left(y_{t-1}^2\right) - \varphi_1^2\mu^2 + \sigma_\varepsilon^2 \\
&= \varphi_1^2\left[E\left(y_{t-1}^2\right) - \mu^2\right] + \sigma_\varepsilon^2 \\
&= \varphi_1^2\gamma_0 + \sigma_\varepsilon^2 \qquad \Rightarrow \gamma_0 = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}
\end{aligned}$$

A Gaussian AR(1) process:

$$y_t = \varphi_1 \cdot y_{t-1} + \delta + \varepsilon_t$$

$$\varepsilon_t \sim \text{i.i.d. } N\left(0, \sigma_\varepsilon^2\right)$$

$$E(y_1) = \mu = \frac{\delta}{1-\varphi_1}$$

$$E[(y_1-\mu)^2] = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}$$

Density :

$$f_{Y_1}\left(y_1; \varphi_1, \delta, \sigma_\varepsilon^2\right) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2 / \left(1-\varphi_1^2\right)}} \exp\left[\frac{-\{y_1 - [\delta / 1-\varphi_1]\}^2}{2\sigma_\varepsilon^2 / \left(1-\varphi_1^2\right)}\right]$$

Distribution of the second observation conditional on observing y_1:

$$\left(Y_2 | Y_1 = y_1\right) \sim N\left(\left(\delta + \varphi_1 \cdot y_1\right), \sigma_\varepsilon^2\right)$$

$$f_{Y_2|Y_1}\left(y_2 | y_1; \varphi_1, \delta, \sigma_\varepsilon^2\right) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[\frac{-\left\{y_2 - \delta - \varphi_1 \cdot y_1\right\}^2}{2\sigma_\varepsilon^2}\right]$$

Autoregressive Process of order 2, AR(2):

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \delta + \varepsilon_t$$

$$E(y_t) = \mu = \frac{\delta}{1 - \varphi_1 - \varphi_2}$$

$$\text{Assumption : zero mean } (\text{setting } \delta = 0)$$

$$\gamma_0 = E[y_t(\varphi_1 y_{t-1} + \varphi_1 y_{t-2} + \varepsilon_t)] = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \sigma_\varepsilon^2$$

$$\gamma_1 = E[y_{t-1}(\varphi_1 y_{t-1} + \varphi_1 y_{t-2} + \varepsilon_t)] = \varphi_1 \gamma_0 + \varphi_2 \gamma_1$$

$$\gamma_2 = E[y_{t-2}(\varphi_1 y_{t-1} + \varphi_1 y_{t-2} + \varepsilon_t)] = \varphi_1 \gamma_1 + \varphi_2 \gamma_0$$

$$\gamma_k = E[y_{t-k}(\varphi_1 y_{t-1} + \varphi_1 y_{t-2} + \varepsilon_t)] = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2}$$

Autoregressive Process of order 2, AR(2):

$$\gamma_2 = \varphi_1\gamma_1 + \varphi_2\gamma_0 \text{ substituting into}$$

$$\gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \sigma_\varepsilon^2$$

$$= \varphi_1\gamma_1 + \varphi_2\varphi_1\gamma_1 + \varphi_2^2\gamma_0 + \sigma_\varepsilon^2$$

$$\text{using}: \gamma_1 = \varphi_1\gamma_0 + \varphi_2\gamma_1 \Leftrightarrow \gamma_1 = \frac{\varphi_1\gamma_0}{1-\varphi_2}$$

$$= \frac{\varphi_1^2\gamma_0}{1-\varphi_2} + \frac{\varphi_2\varphi_1^2\gamma_0}{1-\varphi_2} + \varphi_2^2\gamma_0 + \sigma_\varepsilon^2$$

$$= \frac{(1-\varphi_2)\sigma_\varepsilon^2}{(1+\varphi_2)\left[(1-\varphi_2)^2 - \varphi_1^2\right]}$$

Autocorrelation Function of a second-order Autoregressive Process:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\varphi_1}{1-\varphi_2}$$

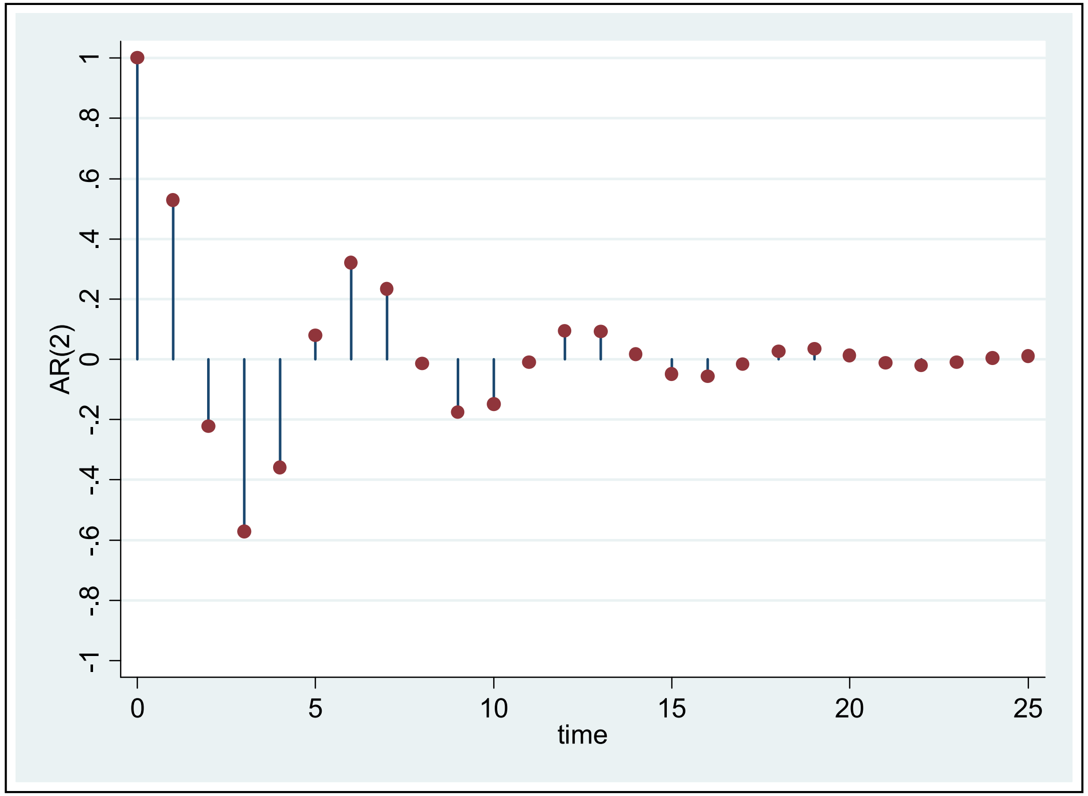
$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\varphi_1^2}{1-\varphi_2} + \varphi_2$$

Yule-Walker equations:
used to obtain estimates of the autoregressive parameters φ_1 and φ_2.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2}$$

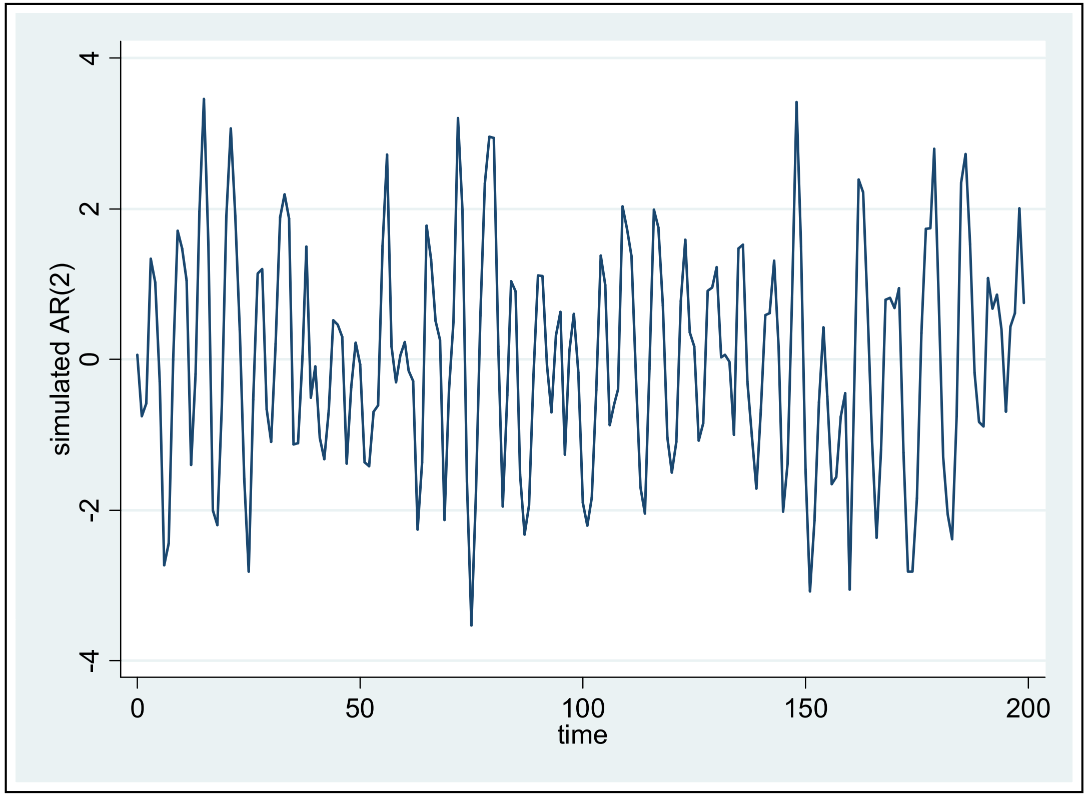
$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T-k} (y_t - \bar{y})^2}$$

Theoretical ACF of an AR(1) process, $\varphi_1 = 0.9$, $\varphi_2 = -0.7$:

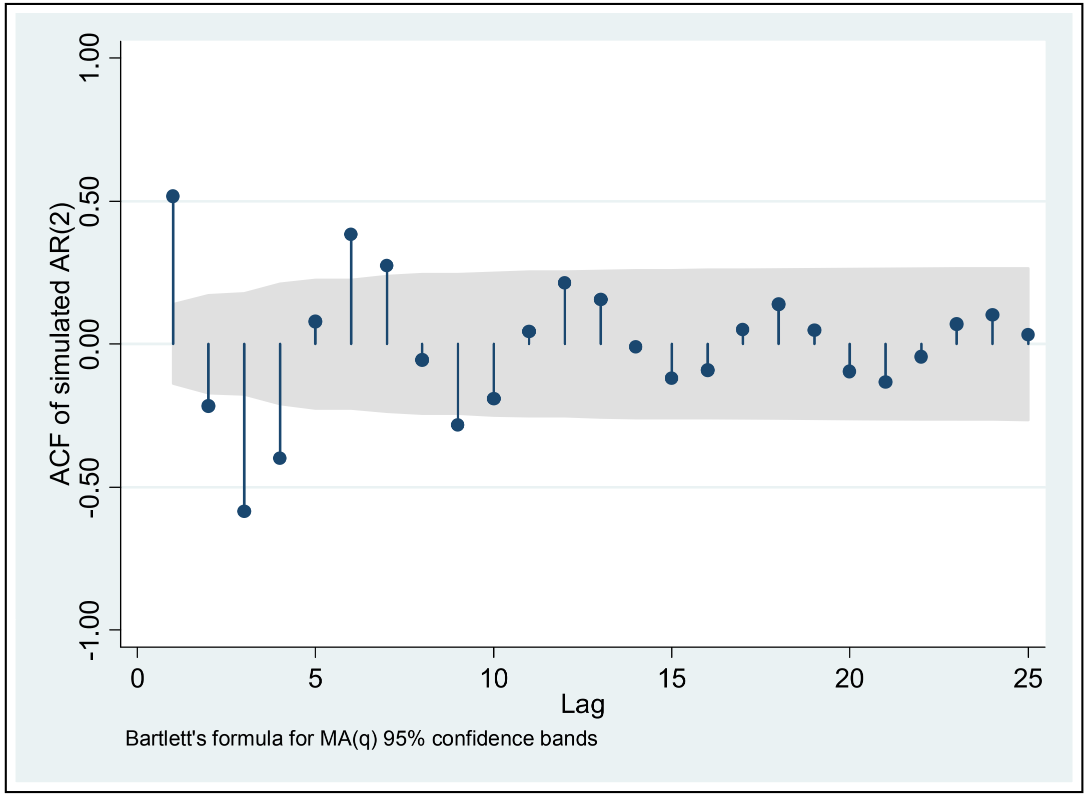


Note that this is a sinusoidal function that is geometrically dampened.

Example of a AR(2) process: $y_t = .9y_{t-1} - .7y_{t-2} + \varepsilon_t$



Autocorrelation function for the AR(2) process:



Stationary Autoregressive Process of order *p*:

$$AR(p)\text{:}\ y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t$$

$$E(y_t) = E(y_{t-1}) = E(y_{t-2}) = \ldots = \mu$$

$$\mu = \varphi_1 \mu + \varphi_2 \mu + \ldots + \varphi_p \mu + \delta$$

$$\mu = \frac{\delta}{1 - \varphi_1 - \varphi_2 - \ldots - \varphi_p}$$

Covariance and autocorrelation function for the AR(*p*):

$$\gamma_k = E\left[y_{t-k}\left(\varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t\right)\right]$$

For $k = 0,1,\ldots,p$ obtain $p+1$ difference equations

$$\gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \ldots + \varphi_p\gamma_p + \sigma_\varepsilon^2$$

$$\gamma_1 = \varphi_1\gamma_0 + \varphi_2\gamma_1 + \ldots + \varphi_p\gamma_{p-1}$$

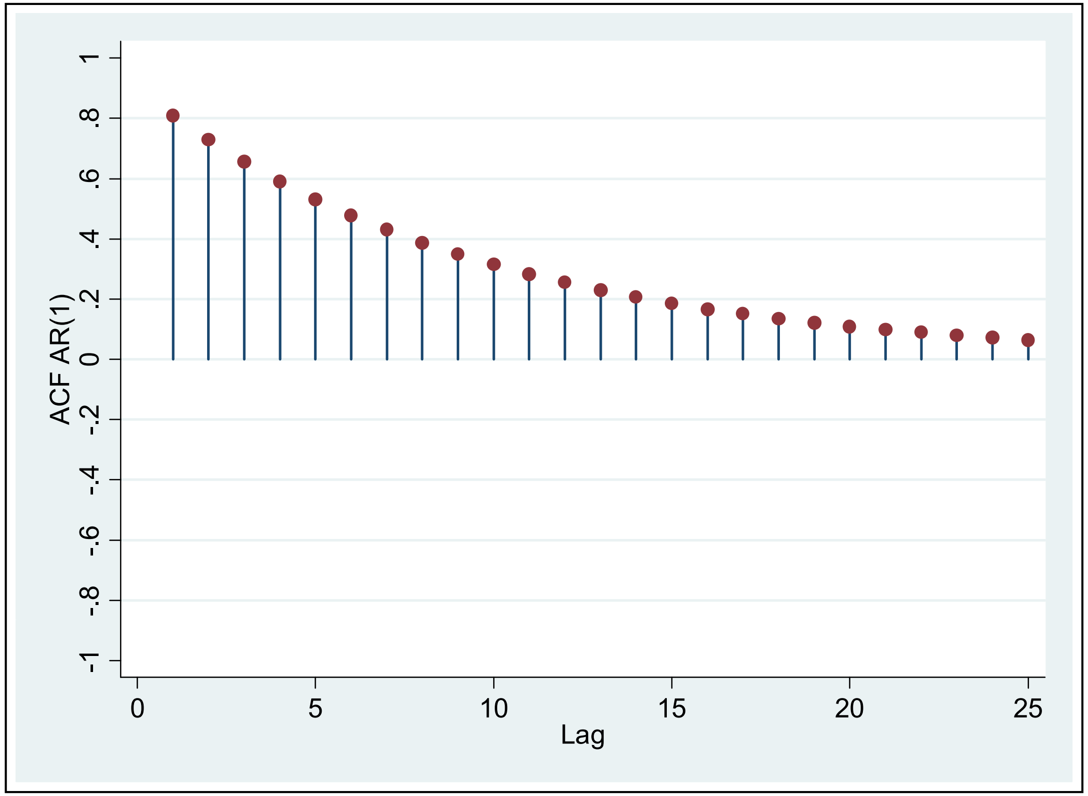
$$\vdots$$

$$\gamma_p = \varphi_1\gamma_{p-1} + \varphi_2\gamma_{p-2} + \ldots + \varphi_p\gamma_0$$

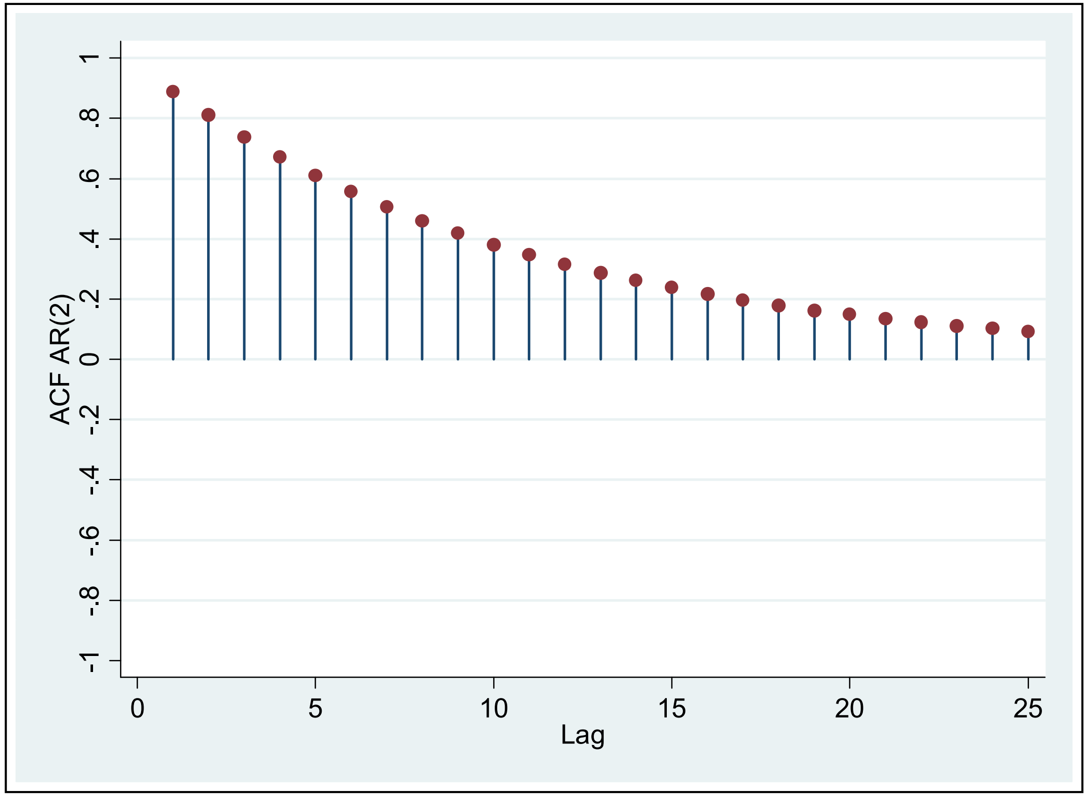
For displacements k greater than p:

$$\gamma_p = \varphi_1\gamma_{k-1} + \varphi_2\gamma_{k-2} + \ldots + \varphi_p\gamma_{k-p}$$

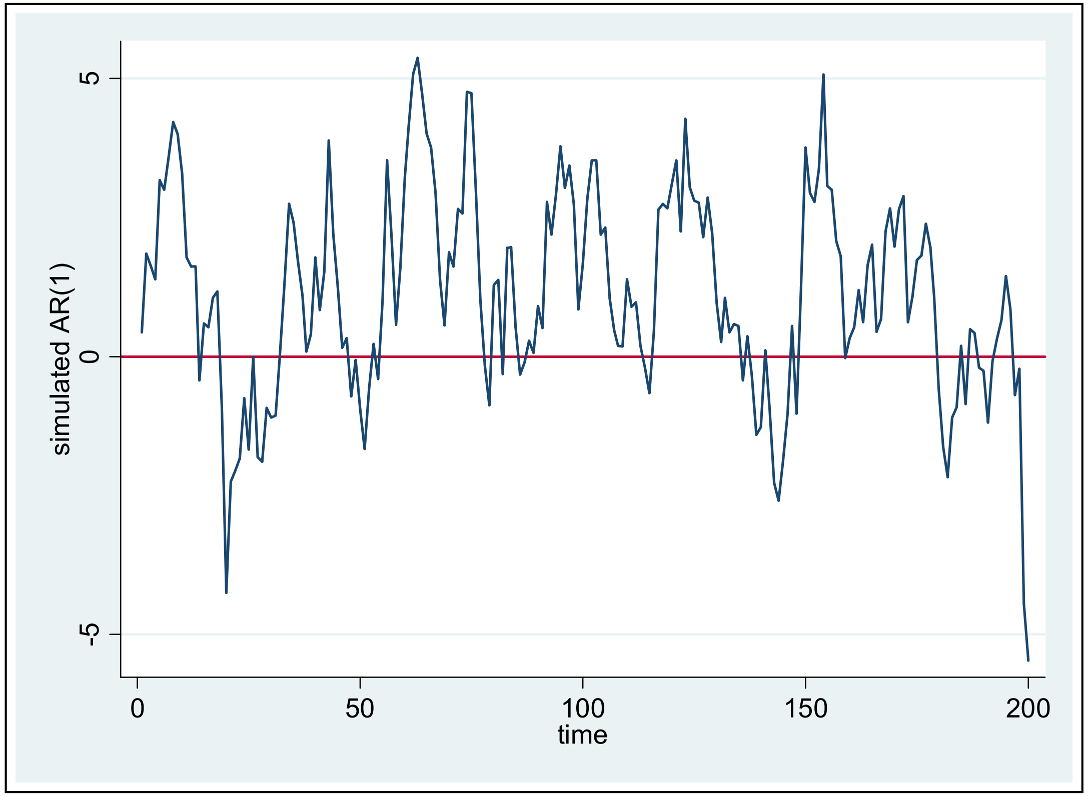
Theoretical ACF of an AR(1) process, φ_1 = 0.9:



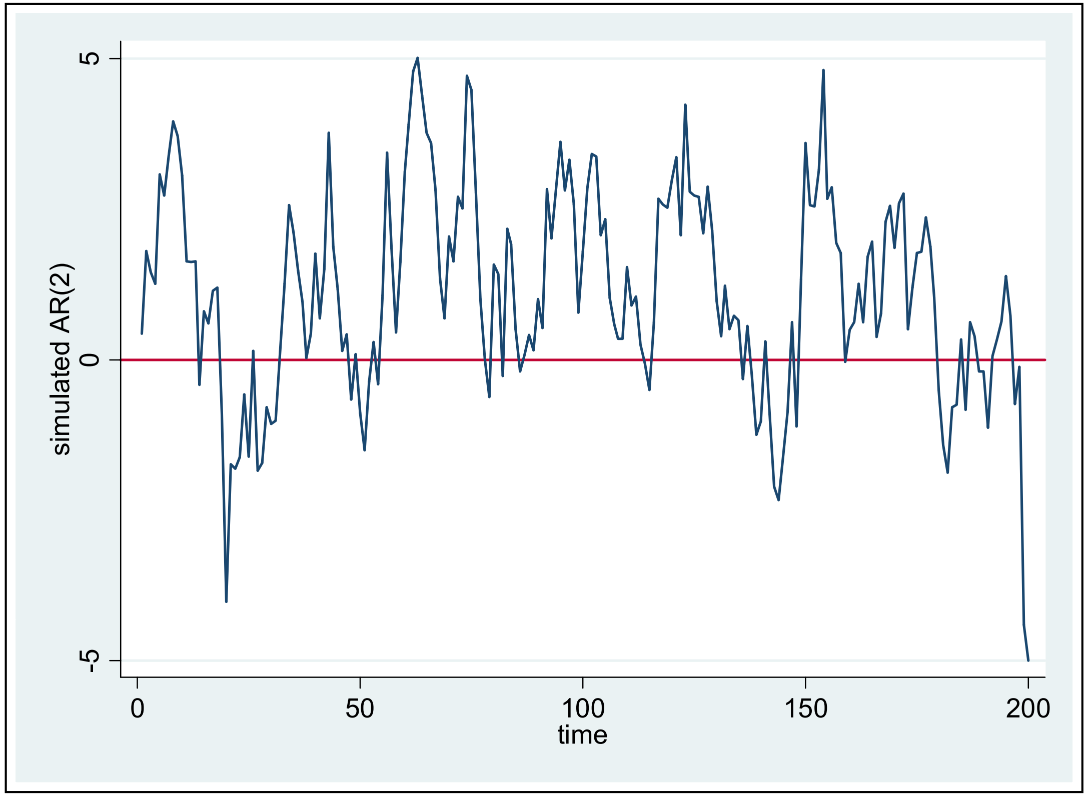
Theoretical ACF of an AR(2) process, φ_1 = 0.8 and φ_2 = 0.1:



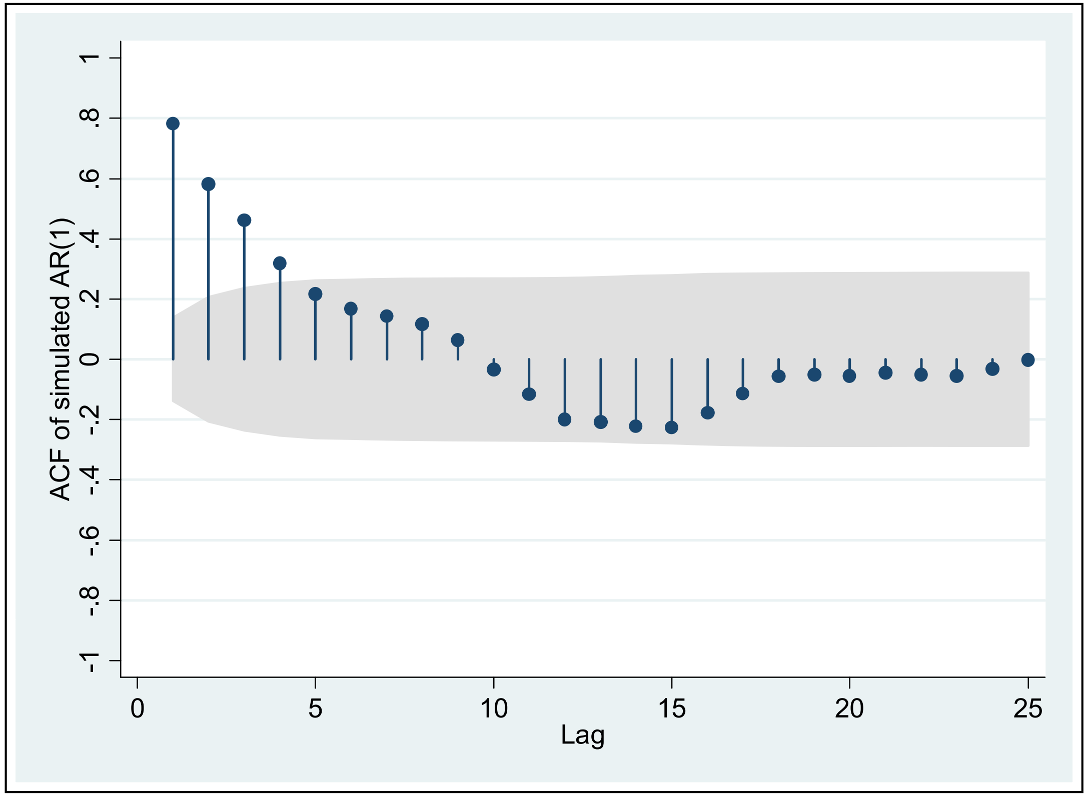
Example of an AR(1) process: $y_t = 0.9 y_{t-1} + \varepsilon_t$



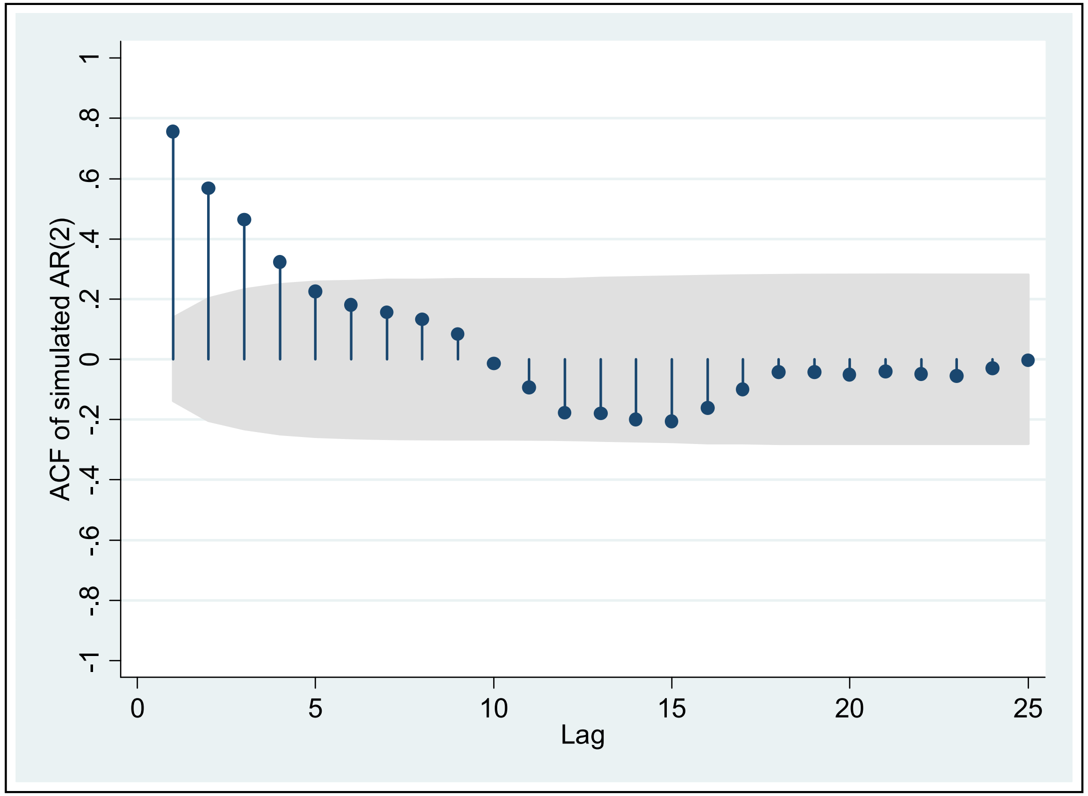
Example of a AR(2) process: $y_t = 0.8 y_{t-1} + 0.1 y_{t-2} + \varepsilon_t$



Autocorrelation function of the AR(1) process:



Autocorrelation function of the AR(2) process:



ACF is not sufficient to determine the order of AR process

→ **Partial** Autocorrelation

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

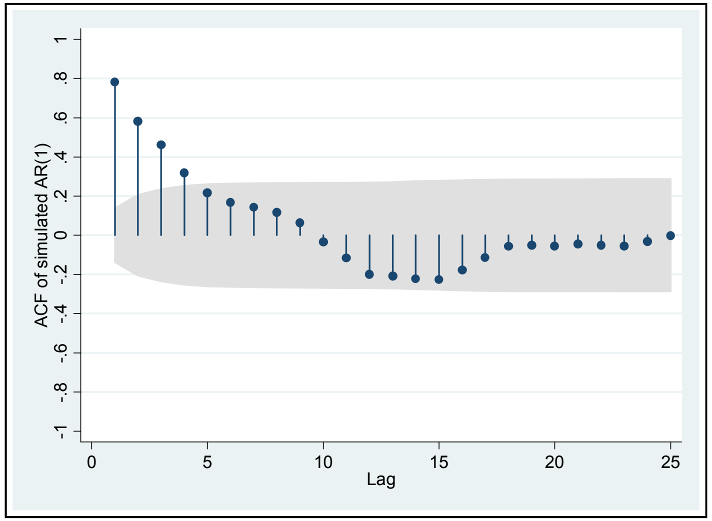
Stationary Stochastic Processes	• Introduction
	• **Identification**
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • **Partial Autocorrelation Function**
	  • ARMA Models
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Recall: “Ordinary” Autocorrelation

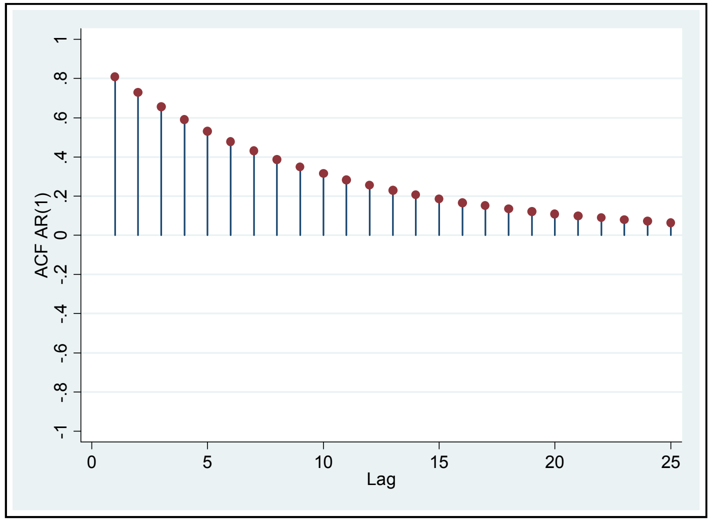
Sample ACF

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T-k}(y_t - \bar{y})^2}$$



Population ACF

$$\rho_k = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sigma_y^2}$$



Recall: “Ordinary” Autocorrelation

Sample ACF

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T-k} (y_t - \bar{y})^2}$$

Population ACF

$$\rho_k = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sigma_y^2}$$

Example: for $k = 2$

$$\hat{\rho}_2 = \frac{\sum_{t=1}^{T-2} (y_t - \bar{y})(y_{t+2} - \bar{y})}{\sum_{t=1}^{T-2} (y_t - \bar{y})^2}$$

$$\rho_2 = \frac{E[(y_t - \mu_y)(y_{t+2} - \mu_y)]}{\sigma_y^2}$$

Partial Autocorrelation

"The idea of partial autocorrelation analysis is that we want to measure how y_t and y_{t+k} are related, but *with the effects of the intervening y's accounted for*. For example, we want to show the relationship between the ordered pairs (y_t, y_{t+2}) taking into account the effect of y_{t+1} on y_{t+2}. Next, we want the relationship between the pairs (y_t, y_{t+3}), but with the effects of both y_{t+1} and y_{t+2} on y_{t+3} accounted for, and so forth, each time adjusting for the impact of any *y*'s that fall between the ordered pairs in question."

Pankratz (1983) "Forecasting with univariate Box-Jenkins models", p.39

Partial Autocorrelation

Correlation between y_t and y_{t+k} *with the effects of the intervening y's accounted for*:

Correlation between y_t and $y_{t+\mathbf{2}}$, taking into account the effect of y_{t+1} on y_{t+2}.

Correlation between y_t and $y_{t+\mathbf{3}}$, taking into account the effects of both y_{t+1} and y_{t+2} on y_{t+3}

Correlation between y_t and $y_{t+\mathbf{4}}$, taking into account the effects of y_{t+1}, y_{t+2} and y_{t+3} on y_{t+4} etc.

This will be a function of *k* again, the **PACF**!

Sample partial autocorrelation of order 2

To get the correlation between y_t and y_{t+2}, taking into account the effect of y_{t+1} on y_{t+2}, run the OLS regression

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(2)} y_{t-1} + \hat{\alpha}_2^{(2)} y_{t-2} + \hat{\varepsilon}_t$$

The last coefficient, $\hat{\alpha}_2^{(2)}$, is the desired partial autocorrelation of order 2.

Note: Don't confuse this with OLS estimation of an AR(2) model! Here, we are not assuming that y_t follows an AR(2) process or any other particular model. We are `just' estimating the linear projection of y_t on y_{t-1} and y_{t-2}.

Sample partial autocorrelation of order 3

To get the correlation between y_t and $y_{t+\mathbf{3}}$, taking into account the effects of y_{t+1} and y_{t+2}, run OLS regression

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(3)} y_{t-1} + \hat{\alpha}_2^{(3)} y_{t-2} + \hat{\alpha}_3^{(3)} y_{t-3} + \hat{\varepsilon}_t$$

The last coefficient, $\hat{\alpha}_3^{(3)}$, is the desired partial autocorrelation of order 2.

Note: Don't confuse this with OLS estimation of an AR(3) model! Here, we are not assuming that y_t follows an AR(3) process or any other particular model. We are `just' estimating the linear projection of y_t on y_{t-1}, y_{t-2} and y_{t-3}.

Estimation of the *m*th partial autocorrelation

Last coefficient in an OLS regression of y_t on a constant and its *m* most recent values.

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(m)} y_{t-1} + \hat{\alpha}_2^{(m)} y_{t-2} + \ldots + \hat{\alpha}_m^{(m)} y_{t-m} + \hat{\varepsilon}_t$$

Note: Don't confuse this with OLS estimation of an AR(m) model! Here, we are not assuming that y_t follows an AR(m) process or any other particular model. We are \`just' estimating the linear projection of y_t on y_{t-1} , y_{t-2} … y_{t-m}.

The sample partial autocorrelation function

is the collection of the last coefficients from these OLS regressions:

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(2)} y_{t-1} + \hat{\alpha}_2^{(2)} y_{t-2} + \hat{\varepsilon}_t$$

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(3)} y_{t-1} + \hat{\alpha}_2^{(3)} y_{t-2} + \hat{\alpha}_3^{(3)} y_{t-3} + \hat{\varepsilon}_t$$

$$\vdots$$

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(m)} y_{t-1} + \hat{\alpha}_2^{(m)} y_{t-2} + \ldots + \hat{\alpha}_m^{(m)} y_{t-m} + \hat{\varepsilon}_t$$

Partial autocorrelation coefficients of an AR(2) process and OLS regression

Simulated AR(2) process:

```
. corrgram sim_AR2

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.5490   0.5645   61.197  0.0000          |----              |----
2       -0.1374  -0.6483   65.052  0.0000         -|             -----|
3       -0.4636   0.0062   109.13  0.0000       ---|                  |
4       -0.4053  -0.2103   142.99  0.0000       ---|                 -|
5       -0.1435  -0.0547   147.25  0.0000         -|                  |
6        0.1084  -0.0349    149.7  0.0000          |                  |
7        0.2267   0.0071   160.46  0.0000          |-                 |
8        0.1972   0.0187   168.64  0.0000          |-                 |
9        0.0597  -0.0273   169.39  0.0000          |                  |
10      -0.0919  -0.0240   171.19  0.0000          |                  |
```

Partial autocorrelation coefficients of an AR(2) process and OLS regression

Simulated AR(2) process:

```
. regress  sim_AR2 L1.sim_AR2 L2.sim_AR2

      Source |       SS       df       MS              Number of obs =     198
-------------+------------------------------           F(  2,   195) =  141.14
       Model |  336.956872     2  168.478436           Prob > F      =  0.0000
    Residual |  232.775552   195  1.19372078           R-squared     =  0.5914
-------------+------------------------------           Adj R-squared =  0.5872
       Total |  569.732424   197  2.89204276           Root MSE      =  1.0926

------------------------------------------------------------------------------
     sim_AR2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sim_AR2 |
         L1. |   .9106452   .0551471    16.51   0.000      .801884    1.019406
         L2. |   -.648314   .0560016   -11.58   0.000    -.7587606   -.5378675
       _cons |   .0687627   .0779331     0.88   0.379    -.0849373    .2224628
------------------------------------------------------------------------------
```

Partial autocorrelation coefficients of an AR(2) process and OLS regression

Simulated AR(2) process:

```
. regress  sim_AR2  L2.sim_AR2

      Source |       SS       df       MS              Number of obs =     198
-------------+------------------------------           F(  1,   196) =    4.02
       Model |  11.4527874     1  11.4527874           Prob > F      =  0.0463
    Residual |  558.279637   196  2.84836549           R-squared     =  0.0201
-------------+------------------------------           Adj R-squared =  0.0151
       Total |  569.732424   197  2.89204276           Root MSE      =  1.6877

------------------------------------------------------------------------------
     sim_AR2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sim_AR2 |
         L2. |  -.1455914   .0726069    -2.01   0.046    -.2887824   -.0024003
       _cons |   .1034554   .1203402     0.86   0.391    -.1338725    .3407833
------------------------------------------------------------------------------
```

Partial autocorrelation coefficients of an AR(2) process and OLS regression

Simulated AR(2) process:

```
. regress  sim_AR2 L1.sim_AR2

      Source |       SS       df       MS              Number of obs =     199
-------------+------------------------------           F(  1,   197) =   88.51
       Model |  177.581013     1  177.581013           Prob > F      =  0.0000
    Residual |  395.262897   197  2.00641064           R-squared     =  0.3100
-------------+------------------------------           Adj R-squared =  0.3065
       Total |   572.84391   198  2.89315106           Root MSE      =  1.4165

------------------------------------------------------------------------------
     sim_AR2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sim_AR2 |
         L1. |   .5644949   .0600028     9.41   0.000     .4461646    .6828252
       _cons |   .0281964   .1006452     0.28   0.780    -.1702839    .2266766
------------------------------------------------------------------------------

. predict e12, residuals
(1 missing value generated)
```

Partial autocorrelation coefficients of an AR(2) process and OLS regression

Simulated AR(2) process:

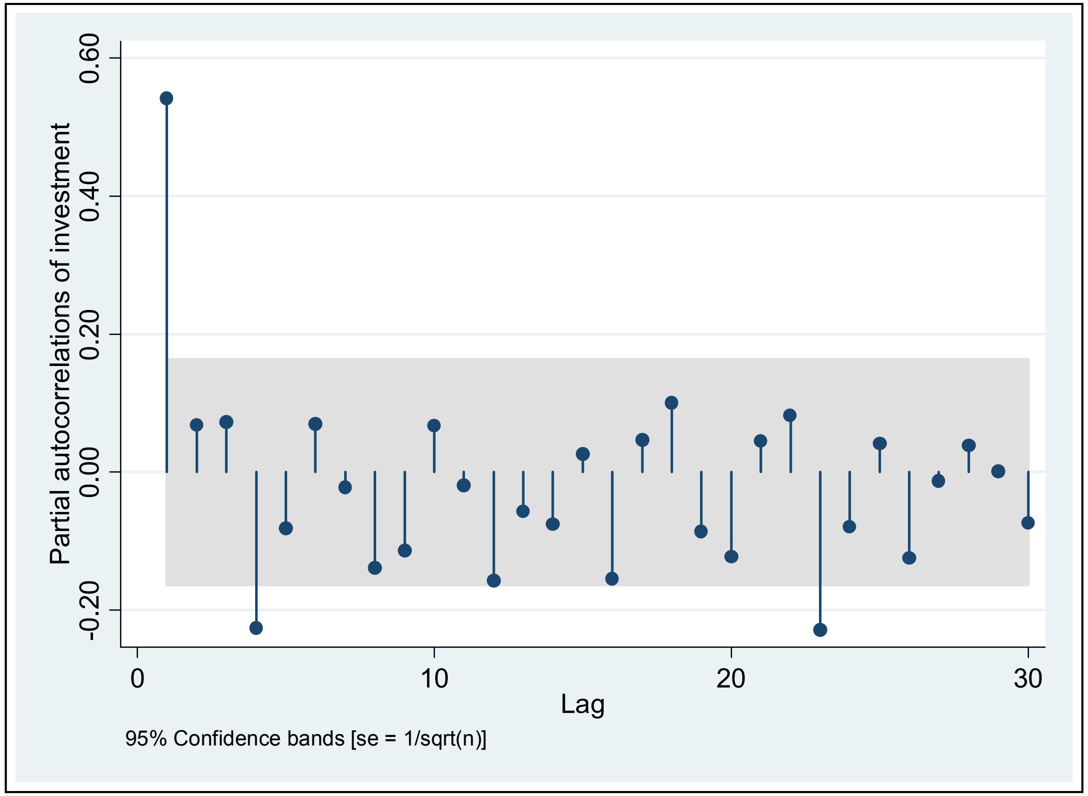
```
. regress  L2.sim_AR2 L1.sim_AR2

      Source |       SS       df       MS              Number of obs =     198
-------------+------------------------------           F(  1,   196) =   82.22
       Model |  159.677093     1  159.677093           Prob > F      =  0.0000
    Residual |  380.629201   196  1.94198572           R-squared     =  0.2955
-------------+------------------------------           Adj R-squared =  0.2919
       Total |  540.306294   197  2.74267154           Root MSE      =  1.3936

------------------------------------------------------------------------------
  L2.sim_AR2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sim_AR2 |
         L1. |   .5353322   .0590371     9.07   0.000     .4189027    .6517617
       _cons |   .0747103   .0992583     0.75   0.453    -.1210412    .2704618
------------------------------------------------------------------------------

. predict e32, residuals
(2 missing values generated)
```

Partial Autocorrelation Function for the inventory investment data

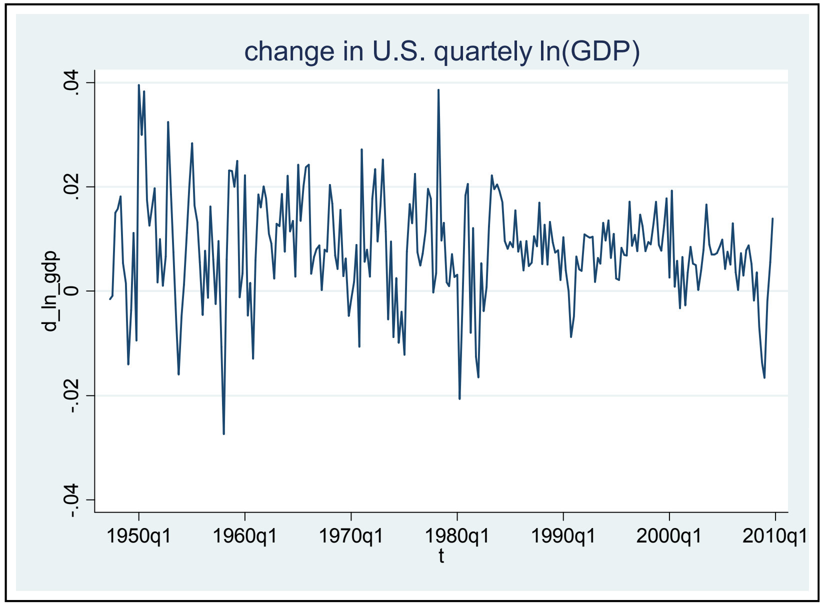


Note that the partial autocorrelations become close to zero after about four lags. There are indded almost no partial autocorrelations beyond four lags that fall outside the bounds ± 1.96 / $\sqrt{T}$.

Why is it "partial"?

Because OLS "partials out" effects of the intervening *y's* .

Example: PACF of order 2 for changes in U.S. real GDP



LAG	AC	PAC
1	0.3629	0.3635
2	0.2034	0.0833
3	-0.0080	-0.1262
4	-0.0922	-0.0898

Partial Autocorrelation

```
. reg   d_ln_gdp l.d_ln_gdp l2.d_ln_gdp

      Source |       SS       df       MS              Number of obs =     249
-------------+------------------------------           F(  2,   246) =   19.50
       Model |  .003368597     2  .001684298           Prob > F      =  0.0000
    Residual |  .021245317   246  .000086363           R-squared     =  0.1369
-------------+------------------------------           Adj R-squared =  0.1298
       Total |  .024613913   248   .00009925           Root MSE      =  .00929

------------------------------------------------------------------------------
    d_ln_gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    d_ln_gdp |
         L1. |   .3309458   .0635605     5.21   0.000     .2057535    .4561381
         L2. |   .0832515    .063452     1.31   0.191    -.0417271      .20823
             |
       _cons |   .0047485   .0008208     5.79   0.000     .0031319    .0063652
------------------------------------------------------------------------------
```

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(2)} y_{t-1} + \hat{\alpha}_2^{(2)} y_{t-2} + \hat{\varepsilon}_t$$

```
LAG       AC          PAC
--------------------------
1         0.3629      0.3635
2         0.2034      0.0833
3        -0.0080     -0.1262
4        -0.0922     -0.0898
```

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1^{(2)} y_{t-1} + \hat{\alpha}_2^{(2)} y_{t-2} + \hat{\varepsilon}_t$$

```
. reg   d_ln_gdp l.d_ln_gdp l2.d_ln_gdp
------------------------------------------------------------------------------
    d_ln_gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    d_ln_gdp |
         L1. |   .3309458   .0635605     5.21   0.000     .2057535    .4561381
         L2. |   .0832515    .063452     1.31   0.191    -.0417271      .20823
             |
       _cons |   .0047485   .0008208     5.79   0.000     .0031319    .0063652
------------------------------------------------------------------------------
```

Alternative way to get this result:

1. OLS regression of y_t on y_{t-1} and save the residuals
2. OLS regression of y_{t-2} on y_{t-1} and save the residuals
3. OLS regression of the residuals from step 1 on those from step2

Step 1

```
. reg   d_ln_gdp l.d_ln_gdp if insample==1

      Source |       SS       df       MS              Number of obs =     249
-------------+------------------------------           F(  1,   247) =   37.18
       Model |  .003219928     1  .003219928           Prob > F      =  0.0000
    Residual |  .021393986   247  .000086615           R-squared     =  0.1308
-------------+------------------------------           Adj R-squared =  0.1273
       Total |  .024613913   248   .00009925           Root MSE      =  .00931

------------------------------------------------------------------------------
    d_ln_gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    d_ln_gdp |
         L1. |    .361364   .0592679     6.10   0.000     .2446292    .4780988
             |
       _cons |   .0051689   .0007568     6.83   0.000     .0036783    .0066594
------------------------------------------------------------------------------

. predict res_d_ln_gdp_l_d_ln_gdp, res
```

Step 2

```
. reg   l2.d_ln_gdp l.d_ln_gdp

      Source |       SS       df       MS              Number of obs =     249
-------------+------------------------------           F(  1,    247) =   37.91
       Model |  .003291846     1  .003291846           Prob > F      =  0.0000
    Residual |  .021450485   247  .000086844           R-squared     =  0.1330
-------------+------------------------------           Adj R-squared =  0.1295
       Total |  .024742331   248  .000099767           Root MSE      =  .00932

------------------------------------------------------------------------------
 L2.d_ln_gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    d_ln_gdp |
        L1.  |   .3653773   .0593461     6.16   0.000     .2484885    .4822662
             |
       _cons |   .0050491   .0007578     6.66   0.000     .0035566    .0065416
------------------------------------------------------------------------------

. predict res_d_l2_ln_gdp_l_d_ln_gdp, res
```

Step 3

```
. reg  res_d_ln_gdp_l_d_ln_gdp res_d_l2_ln_gdp_l_d_ln_gdp

      Source |       SS       df       MS              Number of obs =     249
-------------+------------------------------           F(  1,   247) =    1.73
       Model |  .000148669     1  .000148669           Prob > F      =  0.1898
    Residual |  .021245317   247  .000086013           R-squared     =  0.0069
-------------+------------------------------           Adj R-squared =  0.0029
       Total |  .021393986   248  .000086266           Root MSE      =  .00927

------------------------------------------------------------------------------
res_d_ln_g.. |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
res_d_l2_l~p |   .0832515   .0633235     1.31   0.190    -.0414713    .2079743
       _cons |   8.73e-13   .0005877     0.00   1.00
------------------------------------------------------
```

```
LAG       AC         PAC
-------------------------
1        0.3629     0.3635
2        0.2034     0.0833
3       -0.0080    -0.1262
4       -0.0922    -0.0898
```

m^th^ population partial autocorrelation $\alpha_m^{(m)}$

Defined as the **last** coefficient in a **linear projection** of y on its *m* most recent values:

$$\hat{y}_{t+1|t} - \mu = \alpha_1^{(m)}(y_t - \mu) + \alpha_2^{(m)}(y_{t-1} - \mu) + \ldots + \alpha_m^{(m)}(y_{t-m+1} - \mu)$$

Linear Projection

$y^*_{t+1|t}$ is restricted to be a linear function of $\boldsymbol{y}_t$:

$$y^*_{t+1|t} = \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{y}_t = \alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1} + \ldots + \alpha_m y_{t-m+1}$$

Find the value for **α** such that it minimizes the MSE:

$$\min_{\boldsymbol{\alpha}} E\left[\left(y_{t+1} - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{y}_t\right)^2\right]$$

Then the forecast $\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{y}_t$ is called the linear projection of y_{t+1} on $\mathbf{y}_t$. It is the “population linear least squares regression” of y_{t+1} on $\mathbf{y}_t$

Sample PACF

Sample least quares regression

Example: for k = 2

$$\underset{\hat{\alpha}_0,\hat{\alpha}_1^{(2)},\hat{\alpha}_2^{(2)}}{\text{Min}} \sum_{t=3}^{T} \left[y_t - \left(\hat{\alpha}_0 + \hat{\alpha}_1^{(2)} y_{t-1} + \hat{\alpha}_2^{(2)} y_{t-2}\right)\right]^2$$

Population PACF

Population least quares regression (**Linear Projection)**

$$\min_{\alpha_0,\alpha_1,\alpha_2} E\left[y_t - \left(\alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2}\right)\right]^2$$

AR(*p*) process:

$$\alpha_m^{(m)} = 0 \quad \text{for } m = p+1, p+2, \ldots$$

Moreover, $\alpha_p^{(p)} = \varphi_p$

MA(q) process:

$\alpha_m^{(m)}$ asymptotically approaches zero

ACF and PACF of AR(1) Processes

$$\boxed{y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t}$$

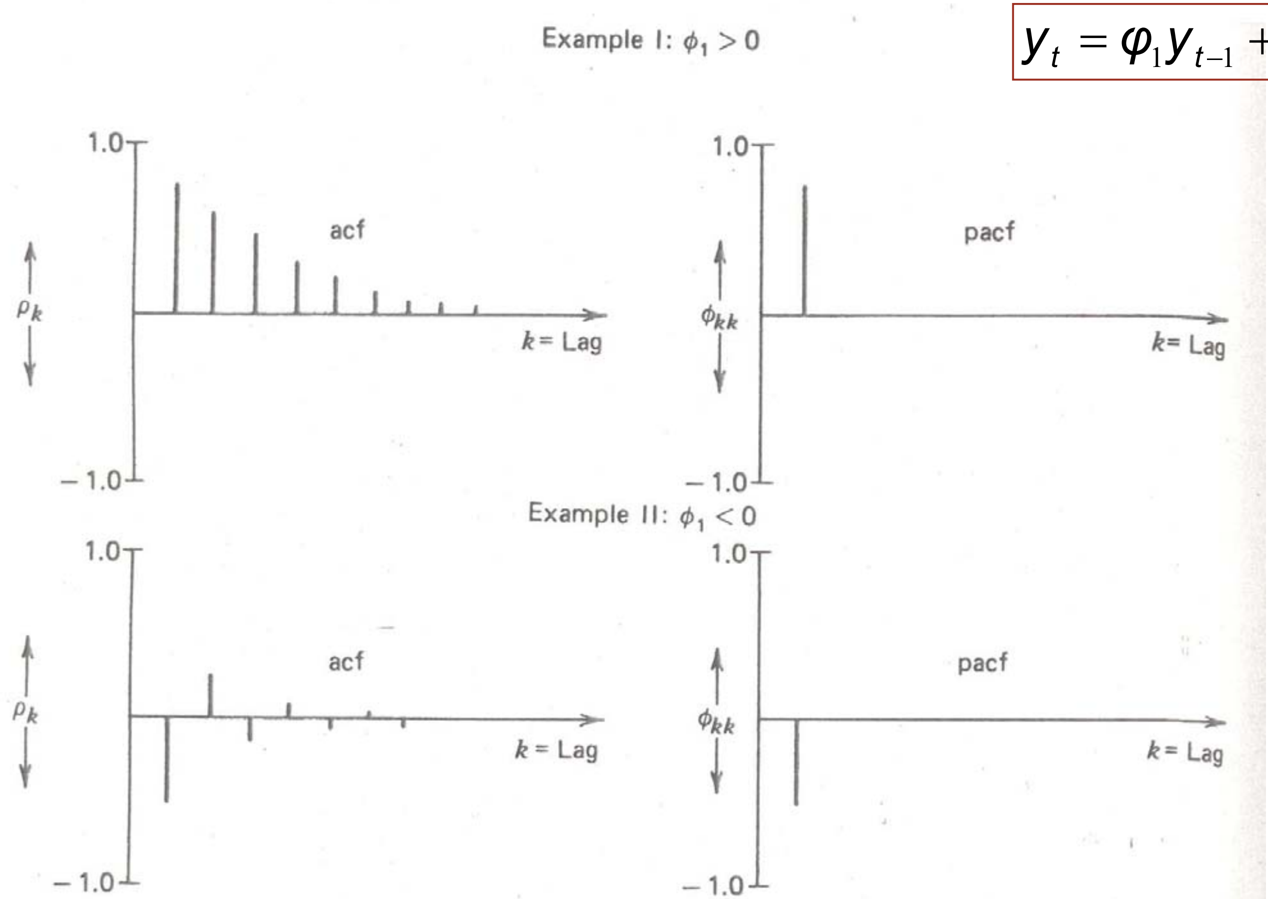


Figure 6.1 Examples of theoretical acf's and pacf's for two stationary AR(1) processes.

ACF and PACF of MA(1) Processes

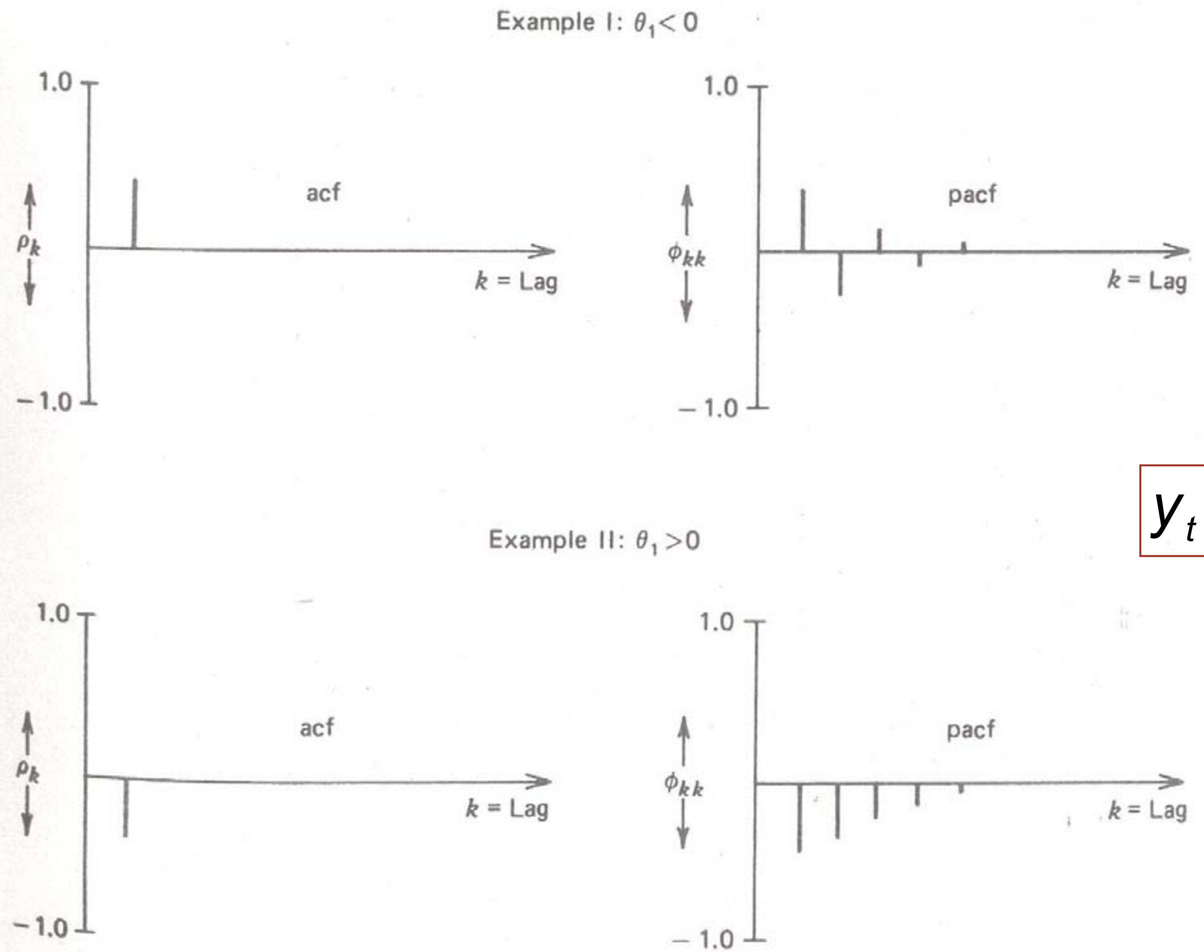


Figure 6.3 Examples of theoretical acf's and pacf's for two MA(1) processes.

$$\boxed{y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}}$$

ACF and PACF of AR(2) Processes

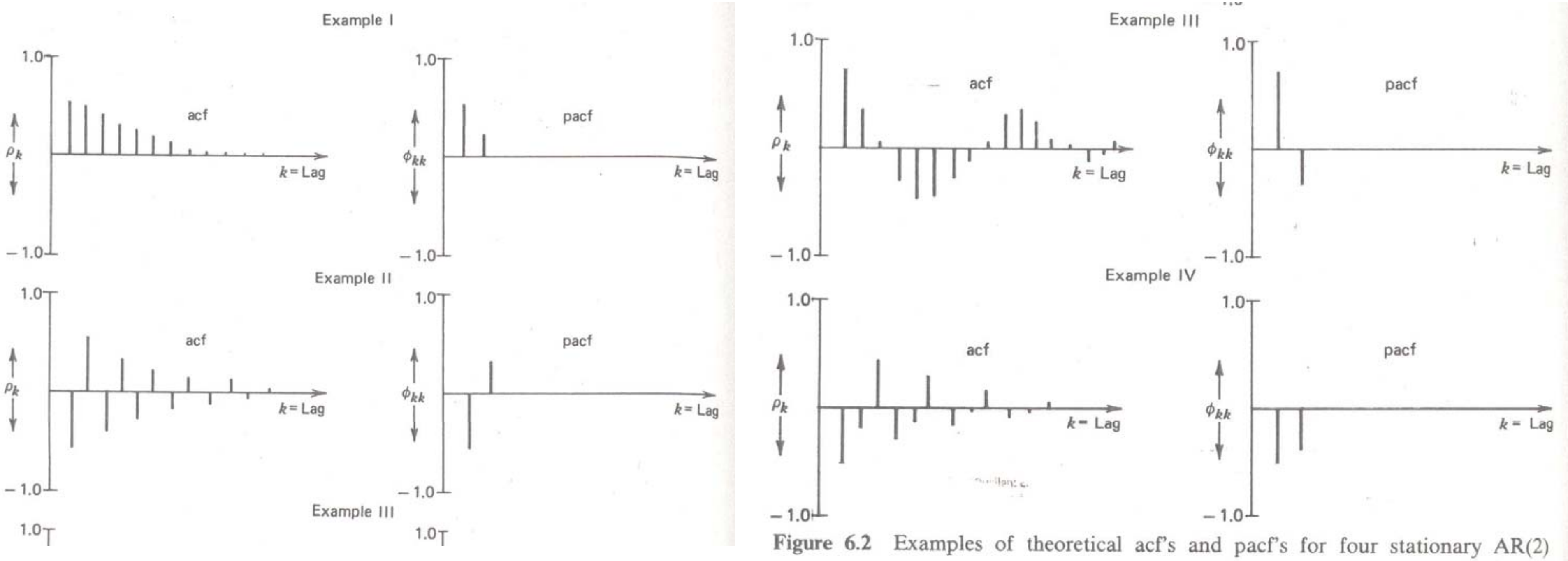


Figure 6.2 Examples of theoretical acf's and pacf's for four stationary AR(2) processes.

$$\boxed{y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \delta + \varepsilon_t}$$

ACF and PACF of MA(2) Processes

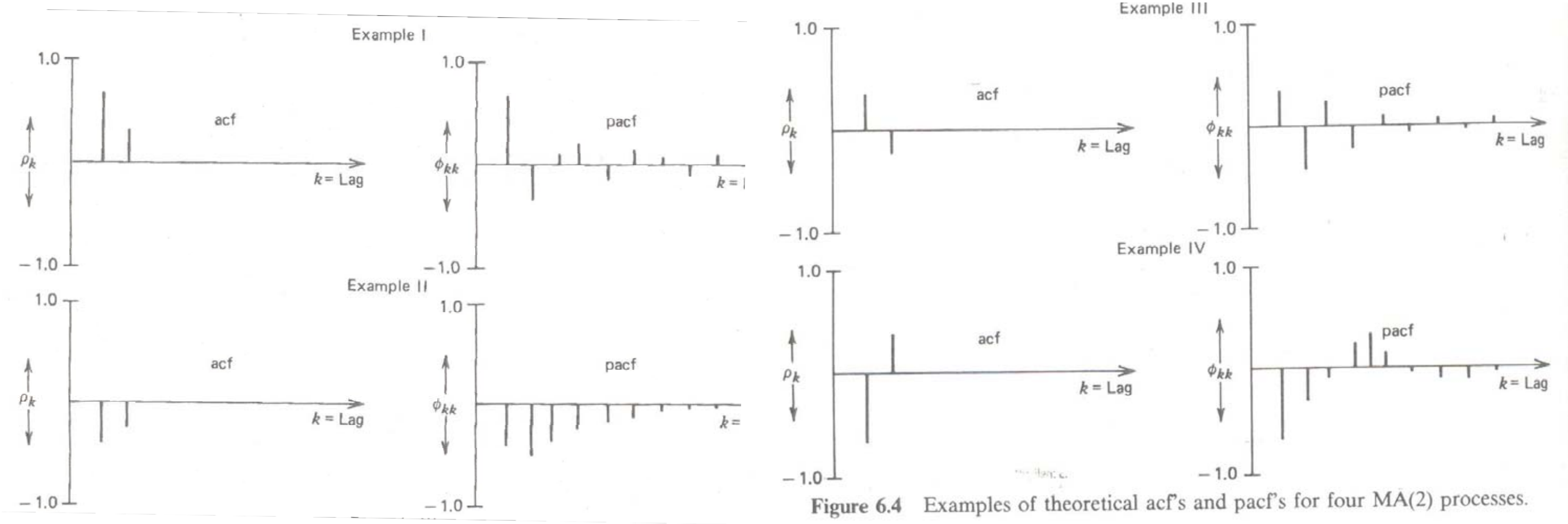


Figure 6.4 Examples of theoretical acf's and pacf's for four MA(2) processes.

$$\boxed{y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}}$$

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• **Identification**
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • **ARMA Models**
	• Estimation
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Mixed Autoregressive–Moving Average Models of order (*p*, *q*), ARMA (*p*, *q*):

Many stationary processes have the qualities of moving average as well as autoregressive processes.

$$y_t = \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

$$\text{Assumption : stationary process with mean } \mu$$

$$E(y_t) = \mu = \frac{\delta}{1 - \varphi_1 - \ldots - \varphi_p}$$

$$\text{Necessary condition : } \varphi_1 + \varphi_2 + \ldots + \varphi_p < 1$$

Why ARMA (*p*, *q*) may arise in practice:

Theorem: If x_t~ARMA(p,m) and y_t~ARMA(q,n) and x_t and y_t are independent then

$$z_t = x_t + y_t \sim \text{ARMA}(x,y)$$

where x ≤ p+q and y ≤ max(p+n,q+m)

Example: If two independent AR(1) series are summed

then x_t~ARMA(1,0) and y_t~ARMA(1,0)

then the resulting series will be ARMA(2,1)

ARMA (1, 1):

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Variances and Covariances (setting $\delta = 0$)

$$\gamma_0 = E\left[\left(\varphi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}\right)^2\right]$$

$$= \varphi_1^2 \gamma_0 - 2\varphi_1 \theta_1 E\left(y_{t-1}\varepsilon_{t-1}\right) + \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2$$

Since $E\left(y_{t-1}\varepsilon_{t-1}\right) = \sigma_\varepsilon^2$

$$\gamma_0\left(1 - \varphi_1^2\right) = \left(1 + \theta_1^2 + -2\varphi_1\theta_1\right)\sigma_\varepsilon^2$$

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\varphi_1\theta_1}{1 - \varphi_1^2}\sigma_\varepsilon^2$$

ARMA (1, 1):

$$\gamma_1 = E\left[y_{t-1}\left(\varphi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}\right)\right]$$

$$= \varphi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2$$

$$= \frac{\left(1-\varphi_1\theta_1\right)\left(\varphi_1 - \theta_1\right)}{1-\varphi_1^2}\sigma_\varepsilon^2$$

$$\gamma_2 = E\left[y_{t-2}\left(\varphi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}\right)\right]$$

$$= \varphi_1 \gamma_1$$

$$\gamma_k = \varphi_1 \gamma_{k-1} = \varphi_1^{k-1}\gamma_1 \quad k \geq 2$$

Autocorrelation Function of an ARMA (1, 1) process:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1-\varphi_1\theta_1)(\varphi_1-\theta_1)}{1+\theta_1^2-2\varphi_1\theta_1}\sigma_\varepsilon^2$$

$$\rho_k = \varphi_1\rho_{k-1} \qquad k \geq 2$$

The autocorrelation function begins at its starting value ρ_1 (which is a function of φ_1 and θ_1) and then decays geometrically from that starting value.

ACF and PACF of ARMA(1,1) Processes

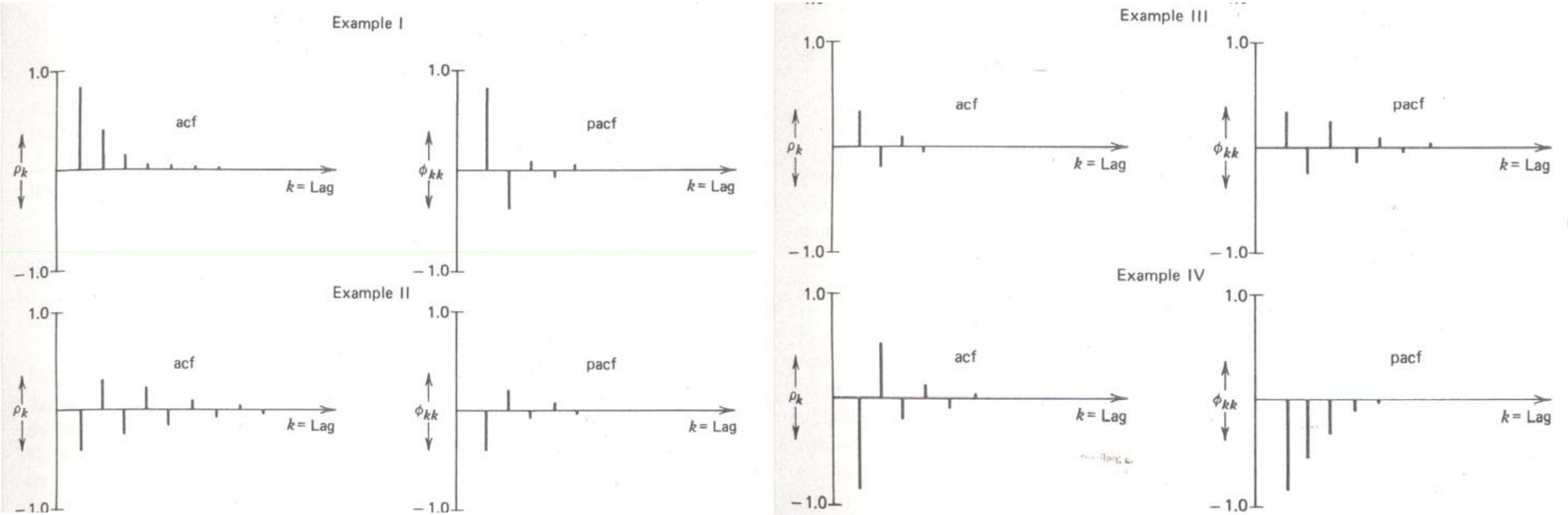


Figure 6.5 Examples of theoretical acf's and pacf's for six stationary ARMA(1, 1) processes.

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Primary distinguishing characteristics of theoretical acf's and pacf's for stationary processes

Process	acf	pacf
AR	Tails off toward zero (exponential decay or damped sine wave)	Cuts off to zero (after lag *p*)
MA	Cuts off to zero (after lag *q*)	Tails off toward zero (exponential decay or damped sine wave)
ARMA	Tails off toward zero	Tails off toward zero

Pankratz (1983), Forecasting with univariate Box-Jenkins models, p.122

Detailed characteristics of five common stationary processes

Process	acf	pacf
AR(1)	Exponential decay: (i) on the positive side if $\varphi_1 > 0$; (ii) alternating in sign starting on the negative side $\varphi_1 < 0$.	Spike at lag 1, then cuts off to zero: (i) spike is positive if $\varphi_1 > 0$; (ii) spike is negative if $\varphi_1 < 0$.
AR(2)	A mixture of exponential decays or a damped sine wave. The exact pattern depends on the signs and sizes of φ_1 and φ_2.	Spikes at lags 1 and 2, then cuts off to zero.
MA(1)	Spike at lag 1, then cuts off to zero: (i) spike is positive if $\theta_1 < 0$; (ii) spike is negative if $\theta_1 > 0$.	Damps out exponentially: (i) alternating in sign, starting on the positive side, if $\theta_1 < 0$; (ii) on the negative side, if $\theta_1 > 0$.
MA(2)	Spike at lag 1 and 2, then cuts off to zero.	A mixture of exponential decays or a damped sine wave. The exact pattern depends on the signs and sizes of θ_1 and θ_2.
ARMA(1,1)	Exponential decay from lag 1: (i) sign of ρ_1 = sign of $(\varphi_1 - \theta_1)$; (ii) all one sign if $\varphi_1 > 0$; (iii) alternating in sign if $\varphi_1 < 0$.	Exponential decay from lag 1: (i) $\varphi_{11} = \rho_1$; (ii) all one sign if $\theta_1 > 0$; (iii) alternate in sign if $\theta_1 < 0$.

Pankratz (1983), Forecasting with univariate Box-Jenkins models, p.123

Sample ACF/PACF vs. Population ACF/PACF

Test ρ_k = 0 for a particular k:

We can test whether individual values of the ACF and PACF are zero , using Bartlett's result:

For a white noise process, the sample (partial) autocorrelation coefficients (for k > 0) are distributed approximately according to a normal distribution with mean 0 and standard deviation $1/\sqrt{T}$

Test of H_0: $\rho_k = 0$ for a particular k

If a time series has been generated by a white noise process, the sample autocorrelation coefficients (for $k > 0$) are distributed approximately $N(0, 1/\sqrt{T})$.

Hence, with probability $1-\alpha$ a sample autocorrelation coefficient should fall within $[0 - z_{1-\alpha/2}\, 1/\sqrt{T};\ 0 + z_{1-\alpha/2}\, 1/\sqrt{T}]$

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • ARMA Models
	• **Estimation**
	• Diagnostic Checking
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Estimation

- (Yule-Walker-Equations)
- Least Squares Estimation
- Maximum Likelihood Estimation
- Conditional Maximum Likelihood Estimation

Least Squares Estimation of an AR(*p*) model

Choose the model parameters which minimizes the residual sum of squares.

AR(*p*): $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t$

$E\left[y_{t-j}\varepsilon_t\right] = 0 \quad \text{for } j = 1,2,3,\ldots,p$

=>OLS provides consistent estimators

ML Estimation

ARMA(p,q) model:

$$y_t = \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

with ε_t white noise

$E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_t \varepsilon_{t-k}) = 0$ for $k \neq 0$

Vector of population parameters:

$$\boldsymbol{\theta} = \begin{bmatrix} \varphi_1 & \cdots & \varphi_p & \delta & \theta_1 & \cdots & \theta_q & \sigma_\varepsilon^2 \end{bmatrix}^{\mathrm{T}}$$

ML Estimation

Calculate the probability density:

$$f_{Y_T, Y_{T-1}, \ldots, Y_1}\left(y_T, y_{T-1}, \ldots, y_1; \boldsymbol{\theta}\right)$$

This can loosely be seen as the probability of having observed the particular sample y_1, y_2, ..., y_T.

The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ is the value of $\boldsymbol{\theta}$ that maximizes this probability.

For this approach it is necessary to assume a distribution for the white noise process: $\varepsilon_t \sim \text{i.i.d. } N\left(0, \sigma_\varepsilon^2\right)$

Example: joint-normal pdf

$$f_{Y_T, Y_{T-1}, \ldots, Y_1}\left(y_T, y_{T-1}, \ldots, y_1; \boldsymbol{\theta}\right) = f\left(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) =$$

$$\frac{1}{(2\pi)^{k/2} \left|\boldsymbol{\Sigma}\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]$$

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_T) \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

For an i.i.d. random sample

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$$

$$|\boldsymbol{\Sigma}| = \sigma^{2n} \text{ and } \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\mathbf{I}$$

$$f(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}(\mathbf{y}-\boldsymbol{\mu})\right]$$

$$= \prod_{i=1}^{n} \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(Y_i - \mu)^2\right] \right\}$$

$$= f(Y_1)\cdot f(Y_2)\cdot \ldots \cdot f(Y_n)$$

Example: for an AR(1)

$$f_{Y_T,Y_{T-1},\dots,Y_1}\left(y_T,y_{T-1},\dots,y_1;\boldsymbol{\theta}\right)=f\left(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=$$

$$\frac{1}{(2\pi)^{k/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]$$

$$\boldsymbol{\mu}=\frac{\delta}{1-\varphi_1}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}\qquad \boldsymbol{\Sigma}=\frac{\sigma_\varepsilon^2}{1-\varphi_1^2}\begin{bmatrix}1 & \varphi_1 & \cdots & \varphi_1^{T-1}\\ \varphi_1 & 1 & \cdots & \varphi_1^{T-2}\\ \vdots & \vdots & \ddots & \vdots\\ \varphi_1^{T-1} & \varphi_1^{T-2} & \cdots & 1\end{bmatrix}$$

Alternative Approach

Joint density as a product of **conditional densities**!

Recall : $P(A \cap B) = P(A \mid B) \cdot P(B)$

Joint density of 1st and 2nd observations:

$$f_{Y_2,Y_1}(y_2, y_1; \boldsymbol{\theta}) = f_{Y_2|Y_1}(y_2 \mid y_1; \boldsymbol{\theta}) \cdot f_{Y_1}(y_1; \boldsymbol{\theta})$$

Joint density of first 3 observations:

$$f_{Y_3,Y_2,Y_1}(y_3, y_2, y_1; \boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3 \mid y_2, y_1; \boldsymbol{\theta}) \cdot f_{Y_2,Y_1}(y_2, y_1; \boldsymbol{\theta})$$

$$\Rightarrow \boxed{f_{Y_3,Y_2,Y_1}(y_3, y_2, y_1; \boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3 \mid y_2, y_1; \boldsymbol{\theta}) \cdot f_{Y_2|Y_1}(y_2 \mid y_1; \boldsymbol{\theta}) \cdot f_{Y_1}(y_1; \boldsymbol{\theta})}$$

This expression holds **in general** (i.e. not only for AR(1))!

Alternative Approach

Joint density as a product of **conditional densities**!

For the first 3 observations:

$$f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3 \mid y_2,y_1;\boldsymbol{\theta}) \cdot f_{Y_2|Y_1}(y_2 \mid y_1;\boldsymbol{\theta}) \cdot f_{Y_1}(y_1;\boldsymbol{\theta})$$

In general (for T observations)

$$f_{Y_T,Y_{T-1},\dots,Y_1,}(y_T,y_{T-1},\dots,y_1|\boldsymbol{\theta})$$

$$= f(y_T|y_{T-1},y_{T-2},\dots,y_1;\boldsymbol{\theta}) \cdot f(y_{T-1}|y_{T-2},y_{T-3},\dots,y_1;\boldsymbol{\theta}) \cdot \ldots \cdot f(y_2|y_1;\boldsymbol{\theta}) \cdot f(y_1|\boldsymbol{\theta})$$

$$= f_{Y_1}(y_1;\boldsymbol{\theta}) \cdot \prod_{t=2}^{T} f_{Y_t|Y_{t-1},Y_{t-2},\dots,Y_1}(y_t|y_{t-1},y_{t-2},\dots,y_1;\boldsymbol{\theta})$$

$$f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3 \mid y_2,y_1;\boldsymbol{\theta}) \cdot f_{Y_2|Y_1}(y_2 \mid y_1;\boldsymbol{\theta}) \cdot f_{Y_1}(y_1;\boldsymbol{\theta})$$

Example: ML Estimation of **AR(1) with normal errors**

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t \qquad \varepsilon_t \sim \text{i.i.d } N(0,\sigma_\varepsilon^2) \qquad \boldsymbol{\theta} = [\varphi_1 \quad \delta \quad \sigma_\varepsilon^2]^T$$

For Y_1: unconditional distribution of AR(1) model

$$E(Y_t) = \mu = \frac{\delta}{1-\varphi_1}$$

$$f_{Y_1}(y_1;\boldsymbol{\theta}) = f_{y_1}(y_1;\varphi_1,\delta,\sigma_\varepsilon^2) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2/(1-\varphi_1^2)}}\exp\left[\frac{-\{y_1-[\delta/1-\varphi_1]\}^2}{2\sigma_\varepsilon^2/(1-\varphi_1^2)}\right]$$

$$Var(Y_t) = E\left[(Y_t-\mu)^2\right] = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}$$

$$\boxed{f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3 \mid y_2,y_1;\boldsymbol{\theta}) \cdot f_{Y_2|Y_1}(y_2 \mid y_1;\boldsymbol{\theta}) \cdot f_{Y_1}(y_1;\boldsymbol{\theta})}$$

$$f_{Y_1}(y_1;\boldsymbol{\theta}) = f_{y_1}(y_1;\varphi_1, \delta, \sigma_\varepsilon^2) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2 / (1-\varphi_1^2)}} \exp\left[\frac{-\{y_1 - [\delta / 1-\varphi_1]\}^2}{2\sigma_\varepsilon^2 / (1-\varphi_1^2)}\right]$$

Conditional distribution of Y_2, given Y_1

$$\boxed{E(Y_2|Y_1 = y_1) = E(\varphi_1 y_1 + \delta + \varepsilon_2|Y_1 = y_1) = \varphi_1 y_1 + \delta}$$

$$f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[\frac{-\{y_2 - \delta - \varphi_1 y_1\}^2}{2\sigma_\varepsilon^2}\right]$$

$$\boxed{Var(Y_2|Y_1 = y_1) = Var(\varphi_1 y_1 + \delta + \varepsilon_2|Y_1 = y_1) = \sigma_\varepsilon^2}$$

$$\boxed{f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3 \mid y_2,y_1;\boldsymbol{\theta}) \cdot f_{Y_2|Y_1}(y_2 \mid y_1;\boldsymbol{\theta}) \cdot f_{Y_1}(y_1;\boldsymbol{\theta})}$$

$$f_{Y_1}(y_1;\boldsymbol{\theta}) = f_{y_1}(y_1;\varphi_1, \delta, \sigma_\varepsilon^2) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2 / (1-\varphi_1^2)}} \exp\left[\frac{-\{y_1 - [\delta / 1-\varphi_1]\}^2}{2\sigma_\varepsilon^2 / (1-\varphi_1^2)}\right]$$

$$f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[\frac{-\{y_2 - \delta - \varphi_1 y_1\}^2}{2\sigma_\varepsilon^2}\right]$$

Conditional distribution of Y_3, given Y_2 and Y_1

$$\boxed{E(Y_3|Y_2 = y_2, Y_1 = y_1) = E(\varphi_1 y_2 + \delta + \varepsilon_3 | y_2, y_1) = \varphi_1 y_2 + \delta = E(Y_3|Y_2 = y_2)}$$

$$f_{Y_3|Y_2Y_1}(y_3|y_2,y_1;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[\frac{-\{y_3 - \delta - \varphi_1 y_2\}^2}{2\sigma_\varepsilon^2}\right]$$

$$\boxed{Var(Y_3|Y_2 = y_2, Y_1 = y_1) = Var(\varphi_1 y_2 + \delta + \varepsilon_3 | y_2, y_1) = \sigma_\varepsilon^2 = Var(Y_3|Y_2 = y_2)}$$

$$\boxed{f_{Y_3,Y_2,Y_1}(y_3,y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2,Y_1}(y_3 \mid y_2,y_1;\boldsymbol{\theta})\cdot f_{Y_2|Y_1}(y_2 \mid y_1;\boldsymbol{\theta})\cdot f_{Y_1}(y_1;\boldsymbol{\theta})}$$

$$f_{Y_1}(y_1;\boldsymbol{\theta}) = f_{y_1}(y_1;\varphi_1,\delta,\sigma_\varepsilon^2) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2/(1-\varphi_1^2)}}\exp\left[\frac{-\{y_1-[\delta/1-\varphi_1]\}^2}{2\sigma_\varepsilon^2/(1-\varphi_1^2)}\right]$$

$$f_{Y_2|Y_1}(y_2|y_1;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left[\frac{-\{y_2-\delta-\varphi_1 y_1\}^2}{2\sigma_\varepsilon^2}\right]$$

$$f_{Y_3|Y_2Y_1}(y_3|y_2,y_1;\boldsymbol{\theta}) = f_{Y_3|Y_2}(y_3|y_2;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left[\frac{-\{y_3-\delta-\varphi_1 y_2\}^2}{2\sigma_\varepsilon^2}\right]$$

In general for an **AR(1)**

$$f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_1}(y_t|y_{t-1},y_{t-2},\ldots,y_1;\boldsymbol{\theta}) = f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\boldsymbol{\theta})$$

Exact ML Estimation of an AR(1) process

Likelihood of the complete sample

$$f_{Y_T, Y_{T-1}, \ldots, Y_1}(y_T, y_{T-1}, \ldots, y_1; \boldsymbol{\theta}) = f_{Y_1}(y_1; \boldsymbol{\theta}) \cdot \prod_{t=2}^{T} f_{Y_t | Y_{t-1}}(y_t \mid y_{t-1}; \boldsymbol{\theta})$$

Log likelihood function

$$L(\boldsymbol{\theta}) = \log\left[f_{Y_1}(y_1; \boldsymbol{\theta})\right] + \sum_{t=2}^{T} \log\left[f_{Y_t | Y_{t1}}(y_t \mid y_{t-1}; \boldsymbol{\theta})\right]$$

Exact ML Estimation of an AR(1) process

Log likelihood function

$$L(\boldsymbol{\theta}) = \log\left[f_{Y_1}(y_1;\boldsymbol{\theta})\right] + \sum_{t=2}^{T} \log\left[f_{Y_t|Y_{t1}}(y_t \mid y_{t-1};\boldsymbol{\theta})\right]$$

$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left[\sigma_\varepsilon^2 \big/ \left(1-\varphi_1^2\right)\right] - \frac{\{y_1 - [\delta/1-\varphi_1]\}^2}{2\sigma_\varepsilon^2 \big/ \left(1-\varphi_1^2\right)}$$

$$-\left[\frac{(T-1)}{2}\right]\cdot\log(2\pi) - \left[\frac{(T-1)}{2}\right]\cdot\log\left(\sigma_\varepsilon^2\right) - \sum_{t=2}^{T}\left[\frac{\{y_t - \delta - \varphi_1 y_{t-1}\}^2}{2\sigma_\varepsilon^2}\right]$$

No closed form solution. Optimal **θ** (i.e., the MLE) is found iteratively.

Maximization

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ that maximizes $L(\boldsymbol{\theta};\mathbf{y})$. Since the logarithm is a monotonic transformation, any value $\hat{\boldsymbol{\theta}}$ that maximizes the likelihood function also maximizes the log-likelihood function.

Fist-order condition:

$$s(\hat{\boldsymbol{\theta}};\mathbf{y}) = \left[\frac{\partial \log L(\boldsymbol{\theta};\mathbf{y})}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$$

To ensure that the found extrema is a maximum, we require that the Hessian matrix $H(\hat{\boldsymbol{\theta}})$ is negative definite.

Hessian Matrix

The Hessian Matrix $H(\boldsymbol{\theta};\mathbf{y})$ is the (*p* x *p*) matrix of the second-order partial derivatives.

$$H(\boldsymbol{\theta};\mathbf{y}) = \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$$

$$= \begin{pmatrix} \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{(\partial\theta_1)^2} & \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{\partial\theta_1\partial\theta_p} \\ \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{(\partial\theta_2)^2} & \cdots & \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{\partial\theta_2\partial\theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{\partial\theta_p\partial\theta_1} & \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{\partial\theta_p\partial\theta_2} & \cdots & \frac{\partial^2 \log L(\boldsymbol{\theta};\mathbf{y})}{(\partial\theta_p)^2} \end{pmatrix}$$

Hessian Matrix

The variance matrix of an ML estimator $\hat{\boldsymbol{\theta}}$ is

$$\text{Avar}\left(\hat{\boldsymbol{\theta}}\right) = \{-E[H(\boldsymbol{\theta};\mathbf{y})]\}^{-1} = I(\boldsymbol{\beta})^{-1} = \left\{E\left[s(\boldsymbol{\theta};\mathbf{y})s(\boldsymbol{\theta};\mathbf{y})'\right]\right\}^{-1}$$

The matrix $I(\boldsymbol{\beta})$ is also called the Fisher information matrix. Unfortunately, $\{-E[H(\boldsymbol{\theta};\mathbf{y})]\}^{-1}$ cannot be known in practice, since it depends on the unknown parameter.

Asymptotic Normality of the ML estimator

It can be shown that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(0, \left\{-E\left[H(\boldsymbol{\theta};\mathbf{y})\right]\right\}^{-1}\right)$$

As long as the model is correctly specified, the maximum likelihood estimator is consistent, asymptotically normally distributed around the true parameter value(s) and efficient.

Estimators of the variance matrix

Estimated Hessian matrix

$$\hat{\mathrm{Avar}}_1(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\left\{\frac{1}{n}\sum_{i=1}^{n} -\mathbf{H}_i(\hat{\boldsymbol{\theta}};\mathbf{y}_i)\right\}^{-1} = \left\{\sum_{i=1}^{n} -\mathbf{H}_i(\hat{\boldsymbol{\theta}};\mathbf{y}_i)\right\}^{-1}$$

The outer product of the score

$$\hat{\mathrm{Avar}}_2(\hat{\boldsymbol{\theta}}) = \left\{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{s}_{\mathrm{i}}(\hat{\boldsymbol{\theta}};\mathbf{y}_i)\,\mathrm{s}_{\mathrm{i}}(\hat{\boldsymbol{\theta}};\mathbf{y}_i)'\right\}^{-1}$$

Alternative Approaches

$$f_{Y_T,Y_{T-1},\dots,Y_1}\left(y_T,y_{T-1},\dots,y_1;\boldsymbol{\theta}\right)=f\left(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=$$

$$\frac{1}{(2\pi)^{k/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]$$

$$f_{Y_T,Y_{T-1},\dots,Y_1,}\left(y_T,y_{T-1},\dots,y_1|\boldsymbol{\theta}\right)$$

$$=f\left(y_T|y_{T-1},y_{T-2},\dots,y_1;\boldsymbol{\theta}\right)\cdot f\left(y_{T-1}|y_{T-2},y_{T-3},\dots,y_1;\boldsymbol{\theta}\right)\cdot\ldots\cdot f\left(y_2|y_1;\boldsymbol{\theta}\right)\cdot f\left(y_1|\boldsymbol{\theta}\right)$$

$$=f_{Y_1}\left(y_1;\boldsymbol{\theta}\right)\cdot\prod_{t=2}^{T}f_{Y_t|Y_{t-1},Y_{t-2},\dots,Y_1}\left(y_t|y_{t-1},y_{t-2},\dots,y_1;\boldsymbol{\theta}\right)$$

Hamilton (1994, pp. 120-122) shows for an AR(1) that the right hand sides are indeed equal!

Exact vs. conditional ML Estimation of an AR(1)

$$L(\boldsymbol{\theta}) = \xcancel{\log\left[f_{Y_1}(y_1;\boldsymbol{\theta})\right]} + \sum_{t=2}^{T} \log\left[f_{Y_t|Y_{t1}}(y_t \mid y_{t-1};\boldsymbol{\theta})\right]$$

$$= \xcancel{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left[\sigma_\varepsilon^2 / \left(1-\varphi_1^2\right)\right] - \frac{\left\{y_1 - \left[\delta / \left(1-\varphi_1\right)\right]\right\}^2}{2\sigma_\varepsilon^2 / \left(1-\varphi_1^2\right)}}$$

$$- \left[\frac{(T-1)}{2}\right] \cdot \log(2\pi) - \left[\frac{(T-1)}{2}\right] \cdot \log\left(\sigma_\varepsilon^2\right) - \sum_{t=2}^{T} \left[\frac{\left\{y_t - \delta - \varphi_1 y_{t-1}\right\}^2}{2\sigma_\varepsilon^2}\right]$$

Suppose we treat y_1 as known

$$\log\left[f_{Y_T, Y_{T-1}, \ldots, Y_2 | Y_1}\left(y_T, y_{T-1}, \ldots, y_2 \mid y_1;\boldsymbol{\theta}\right)\right]$$

$$= -\left[\frac{(T-1)}{2}\right] \cdot \log(2\pi) - \left[\frac{(T-1)}{2}\right] \cdot \log\left(\sigma_\varepsilon^2\right) - \sum_{t=2}^{T} \left[\frac{\left\{y_t - \delta - \varphi_1 y_{t-1}\right\}^2}{2\sigma_\varepsilon^2}\right]$$

Conditional ML Estimation of an AR(1) process

Treat the value of y_1 as deterministic and maximize the likelihood conditioned of the first observation:

$$f_{Y_T,Y_{T-1},\dots,Y_2|Y_1}(y_T, y_{T-1},\dots,y_2 \mid y_1;\boldsymbol{\theta}) = \prod_{t=2}^{T} f_{Y_t|Y_{t1}}(y_t \mid y_{t-1};\boldsymbol{\theta})$$

$$\log\left[f_{Y_T,Y_{T-1},\dots,Y_2|Y_1}(y_T, y_{T-1},\dots,y_2 \mid y_1;\boldsymbol{\theta})\right]$$
$$= -\left[\frac{(T-1)}{2}\right]\cdot\log(2\pi) - \left[\frac{(T-1)}{2}\right]\cdot\log(\sigma_\varepsilon^2) - \sum_{t=2}^{T}\left[\frac{\{y_t - \delta - \varphi_1 y_{t-1}\}^2}{2\sigma_\varepsilon^2}\right]$$

Conditional ML Estimation of an AR(1) process

Maximization with respect to δ and φ_1 is equivalent to minimizing:

$$\sum_{t=2}^{T}\{y_t-\delta-\varphi_1 y_{t-1}\}^2$$

This can be achieved by an OLS regression of y_t on a constant and it's own lagged values:

$$\begin{bmatrix}\hat{\delta}\\ \hat{\varphi}_1\end{bmatrix}=\begin{bmatrix}T-1 & \sum_{t=2}^{T} y_{t-1}\\ \sum_{t=2}^{T} y_{t-1} & \sum_{t=2}^{T} y_{t-1}^2\end{bmatrix}^{-1}\begin{bmatrix}\sum_{t=2}^{T} y_t\\ \sum_{t=2}^{T} y_{t-1}y_t\end{bmatrix}$$

Conditional ML Estimation of an AR(1) process

Conditional MLE of the variance is found by differentiating the log likelihood with respect to σ_ε^2:

$$\frac{\partial \log(\boldsymbol{\theta})}{\partial \sigma_\varepsilon^2} = -\frac{(T-1)}{2\sigma_\varepsilon^2} + \sum_{t=2}^{T}\left[\frac{\{y_t - \delta - \varphi_1 y_{t-1}\}^2}{2\sigma_\varepsilon^4}\right] = 0$$

$$\Leftrightarrow \hat{\sigma}_\varepsilon^2 = \sum_{t=2}^{T}\left[\frac{\{y_t - \hat{\delta} - \hat{\varphi}_1 y_{t-1}\}^2}{T-1}\right]$$

Conditional ML Estimation of an AR(*p*) process

$$\log\left[f_{Y_T,Y_{T-1},\dots,Y_{p+1}|Y_p,\dots,Y_1}\left(y_T,y_{T-1},\dots,y_{p+1}\mid y_p,\dots,y_1;\boldsymbol{\theta}\right)\right]$$

$$=-\left[\frac{(T-p)}{2}\right]\cdot\log(2\pi)-\left[\frac{(T-p)}{2}\right]\cdot\log\left(\sigma_\varepsilon^2\right)$$

$$-\sum_{t=2}^{T}\left[\frac{\left\{y_t-\delta-\varphi_1 y_{t-1}-\varphi_2 y_{t-2}-\dots-\varphi_p y_{t-p}\right\}^2}{2\sigma_\varepsilon^2}\right]$$

$$\min\sum_{t=p+1}^{T}\left(y_t-\hat{\delta}-\hat{\varphi}_1 y_{t-1}-\hat{\varphi}_2 y_{t-2}-\dots-\hat{\varphi}_p y_{t-p}\right)^2$$

$$\hat{\sigma}_\varepsilon^2=\frac{1}{T-p}\sum_{t=p+1}^{T}\left(y_t-\hat{\delta}-\hat{\varphi}_1 y_{t-1}-\hat{\varphi}_2 y_{t-2}-\dots-\hat{\varphi}_p y_{t-p}\right)^2$$

Conditional ML Estimation of a MA(1) process

MA(1): $y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$

Gaussian white noise assumed: $\varepsilon_t \sim$ i.i.d. $N\left(0, \sigma_\varepsilon^2\right)$

Vector of population parameters: $\boldsymbol{\theta} = \left[\theta_1 \quad \mu \quad \sigma_\varepsilon^2\right]^\mathrm{T}$

If the value of ε_{t-1} is known:

$$Y_t | \varepsilon_{t-1} \sim N\left(\left(\mu - \theta_1 \varepsilon_{t-1}\right), \sigma_\varepsilon^2\right)$$

$$f_{Y_t|\varepsilon_{t-1}} \sim \left(y_t | \varepsilon_{t-1}; \boldsymbol{\theta}\right) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[\frac{-\left\{y_t - \mu + \theta_1 \varepsilon_{t-1}\right\}^2}{2\sigma_\varepsilon^2}\right]$$

Conditional ML Estimation of a MA(1) process

If $\varepsilon_0 = 0$:

$$(Y_1 | \varepsilon_0 = 0) \sim N(\mu, \sigma_\varepsilon^2)$$

$$\varepsilon_1 = y_1 - \mu$$

$$f_{Y_2|Y_1,\varepsilon_0=0}(y_2 | y_1, \varepsilon_0 = 0; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[\frac{-\{y_2 - \mu + \theta_1\varepsilon_1\}^2}{2\sigma_\varepsilon^2}\right]$$

Since ε_0 is known, the full sequence $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$ can be calculated from $\{y_1, y_2, \ldots, y_T\}$:

$$\varepsilon_t = y_t - \mu + \theta_1\varepsilon_{t-1}$$

Conditional ML Estimation of a MA(1) process

Conditional density of the t^{th} observation:

$$f_{Y_t|Y_{t-1},Y_{t-2},\ldots,Y_1,\varepsilon_0=0}\left(y_t|y_{t-1},y_{t-2},\ldots,y_1,\varepsilon_0=0;\boldsymbol{\theta}\right)$$

$$=f_{Y_t|\varepsilon_{t-1}}\left(y_t|\varepsilon_{t-1},\varepsilon_0=0;\boldsymbol{\theta}\right)$$

$$=\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left[\frac{-\left\{y_t-\mu+\theta_1\varepsilon_{t-1}\right\}^2}{2\sigma_\varepsilon^2}\right]$$

$$=\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left[\frac{-\left\{\varepsilon_t\right\}^2}{2\sigma_\varepsilon^2}\right]$$

Conditional ML Estimation of a MA(1) process

Conditional log likelihood function:

$$L(\boldsymbol{\theta}) = \log\left[f_{Y_T, Y_{T-1}, \ldots, Y_1, \varepsilon_0 = 0}\left(y_T, y_{T-1}, \ldots, y_1 | \varepsilon_0 = 0; \theta\right)\right]$$

$$= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\sigma_\varepsilon^2\right) - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}$$

Conditional ML Estimation of a MA(1) process

$$\varepsilon_t = (y_t - \mu) + \theta_1 \cdot (y_{t-1} - \mu) + \theta_1^2 (y_{t-2} - \mu) + \ldots + \theta_1^{t-1}(y_1 - \mu) + \theta_1^t \varepsilon_0$$

“If $|\theta_1|$ is substantially less than unity, the effect of imposing $\varepsilon_0 = 0$ will quickly die out and the conditional likelihood will give a good approximation to the unconditional likelihood for a reasonable large sample size.“

Conditional ML Estimation of a MA(q) process

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

$$\varepsilon_0 = \varepsilon_1 = \ldots = \varepsilon_{-q+1} = 0 \rightarrow \boldsymbol{\varepsilon_0} = \begin{bmatrix} \varepsilon_0 & \varepsilon_{-1} & \cdots & \varepsilon_{-q+1} \end{bmatrix}^{\mathrm{T}}$$

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

$$L(\boldsymbol{\theta}) = \log\left[f_{Y_T, Y_{T-1}, \ldots, Y_1, \boldsymbol{\varepsilon}_0 = \mathbf{0}} \left(y_T, y_{T-1}, \ldots, y_1 \middle| \boldsymbol{\varepsilon_0} = \mathbf{0}; \boldsymbol{\theta} \right) \right]$$

$$= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log\left(\sigma_\varepsilon^2\right) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}$$

Conditional ML Estimation of an ARMA(*p*,*q*) process

ARMA(*p*,*q*):

$$y_t = \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

Vector of population parameters:

$$\boldsymbol{\theta} = \begin{bmatrix} \varphi_1 & \cdots & \varphi_p & \delta & \theta_1 & \cdots & \theta_q & \sigma_\varepsilon^2 \end{bmatrix}^\mathrm{T}$$

Conditioning on **y**'s and **ε**'s, **taking initial values**:

$$\mathbf{y_0} = \begin{bmatrix} y_1 & y_2 & \cdots & y_p \end{bmatrix}^\mathrm{T}$$

$$\boldsymbol{\varepsilon}_\mathbf{0} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_q \end{bmatrix}^\mathrm{T}$$

Conditional ML Estimation of an ARMA(*p*,*q*) process

Given $\boldsymbol{y}_0$ and $\boldsymbol{\varepsilon}_0$ the sequence $\{\varepsilon_{q+1}, \varepsilon_{q+2}, \ldots, \varepsilon_T\}$ can be calculated from $\{y_{p+1}, y_{p+2}, \ldots, y_{p+T}\}$:

$$\varepsilon_t = y_t - \delta - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \ldots - \varphi_p y_{t-p}$$
$$+ \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

Conditional log likelihood function

$$L(\boldsymbol{\theta}) = \log\left[f_{Y_T, Y_{T-1}, \ldots, Y_{p+1} | \mathbf{Y_0}, \boldsymbol{\varepsilon}_0}\left(y_T, y_{T-1}, \ldots, y_{p+1} | \mathbf{y}_0, \boldsymbol{\varepsilon}_0; \boldsymbol{\theta}\right)\right]$$

$$= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\sigma_\varepsilon^2\right) - \sum_{t=p+1}^{T} \frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}$$

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	    • Autocorrelation Function
	    • Moving Average and Autoregressive Models
	    • Partial Autocorrelation Function
	    • ARMA Models
	• Estimation
	• **Diagnostic Checking**
	• Forecasting

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Example: AR (1)

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$$

Estimated AR(1) model $\quad E(y_t) = \mu = \dfrac{\delta}{1-\varphi_1}, \quad Var(y_t) = \dfrac{\sigma_\varepsilon^2}{1-\varphi_1^2}$

```
ARIMA regression

Sample:  1955q1 to 1979q4                       Number of obs      =       100
                                                Wald chi2(1)       =    169.73
Log likelihood = -106.0871                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
  savingrate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
savingrate   |
       _cons |   6.013607   .4074939    14.76   0.000     5.214934    6.812281
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .8117232   .0623055    13.03   0.000     .6896067    .9338398
-------------+----------------------------------------------------------------
      /sigma |   .6952772   .0296021    23.49   0.000     .6372581    .7532963
------------------------------------------------------------------------------
```

Estimated coefficients are highly significant.

Pankratz (1983) “Forecasting with univariate Box-Jenkins models”

Autoregressive Process of order 1, AR(1):

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$$

Estimation:

$$\hat{y}_t = \hat{\varphi}_1 y_{t-1} + \hat{\delta}$$

$$\hat{\varphi}_1 = .8117232$$

$$\hat{\delta} = 6.013607 \cdot (1 - .8117232) = 1.1322227$$

$$\hat{y}_t = .8117232 \cdot y_{t-1} + 1.1322227$$

$$\hat{\sigma}_\varepsilon = 0.6952772$$

Residuals:

$$\hat{\varepsilon}_t = y_t - \hat{y}_t$$

Pankratz (1983) “Forecasting with univariate Box-Jenkins models”

How should the residuals "behave"?

➔they are estimates of the "true" residuals

Example: AR(1)

estimated residuals

$$\begin{aligned}\hat{\varepsilon}_t &= y_t - \hat{y}_t \\ &= y_t - \hat{\varphi}_1 y_{t-1} + \hat{\delta}\end{aligned}$$

true residuals

$$\varepsilon_t = y_t - \varphi_1 y_{t-1} + \delta$$

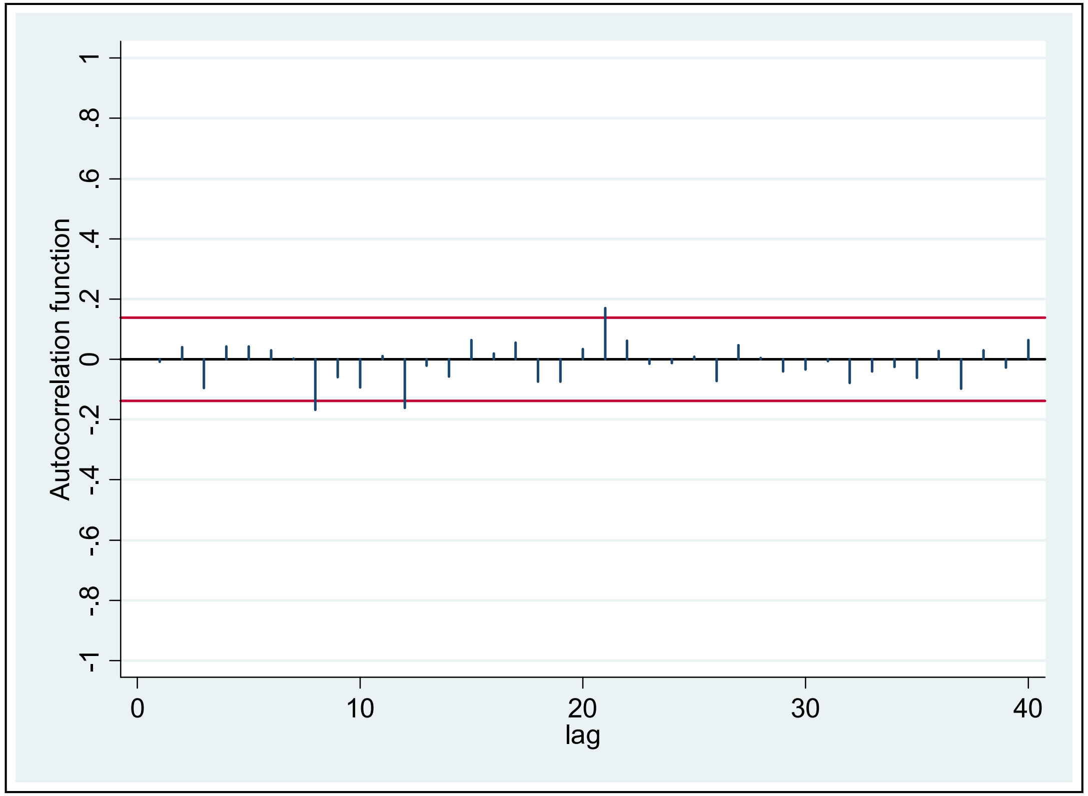
The "true" residuals ε_t of an ARMA model are white-noise, i.e. their ACF and PACF are zero for all lags. Hence, if we estimated the correct model for y_t, the ACF and PACF of the estimated residuals should be zero.

Test $\rho_k = 0$ for a particular k:

We can test whether individual values of the residual ACF and PACF are zero , using Bartlett's result:

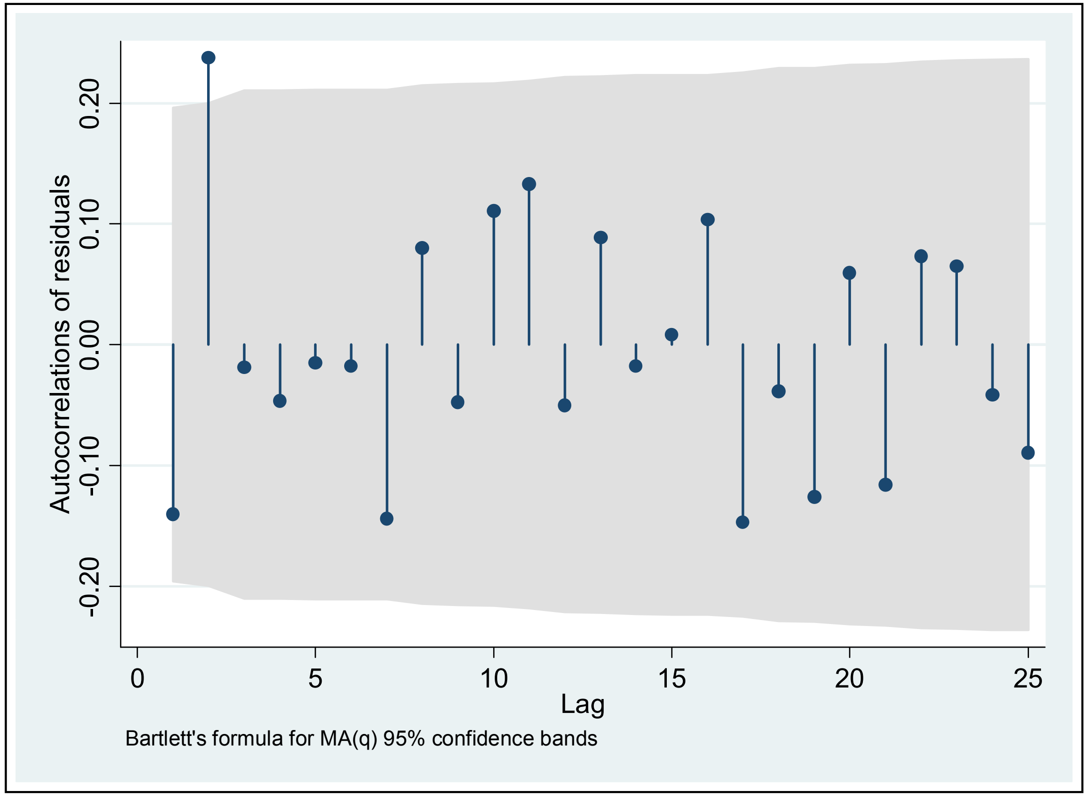
For a white noise process, the sample (partial) autocorrelation coefficients (for k > 0) are distributed approximately according to a normal distribution with mean 0 and standard deviation $1/\sqrt{T}$

Autocorrelation Function for i.i.d. N(0, 1) noise:



Since approximately 95% of the sample autocorrelation should fall between the bounds $\pm\, 1.96 / \sqrt{T}$, one would expect roughly 40(.05) = 2 values to fall outside the bounds.

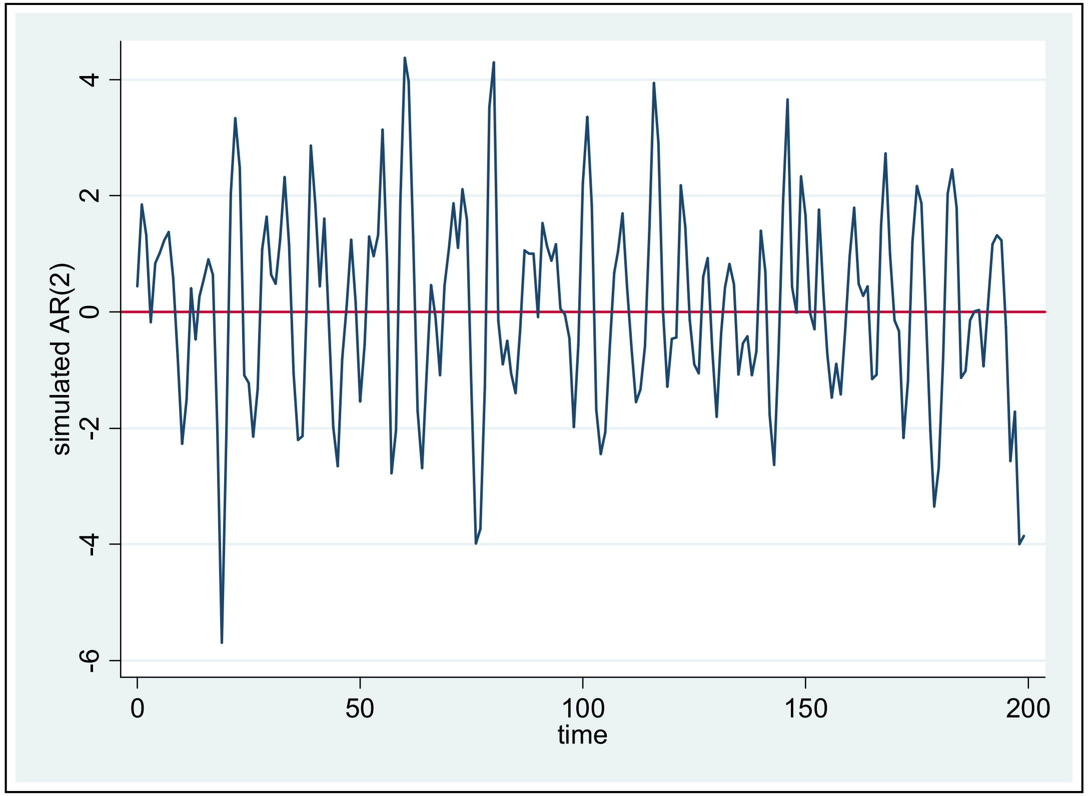
Example: ACF of the residuals



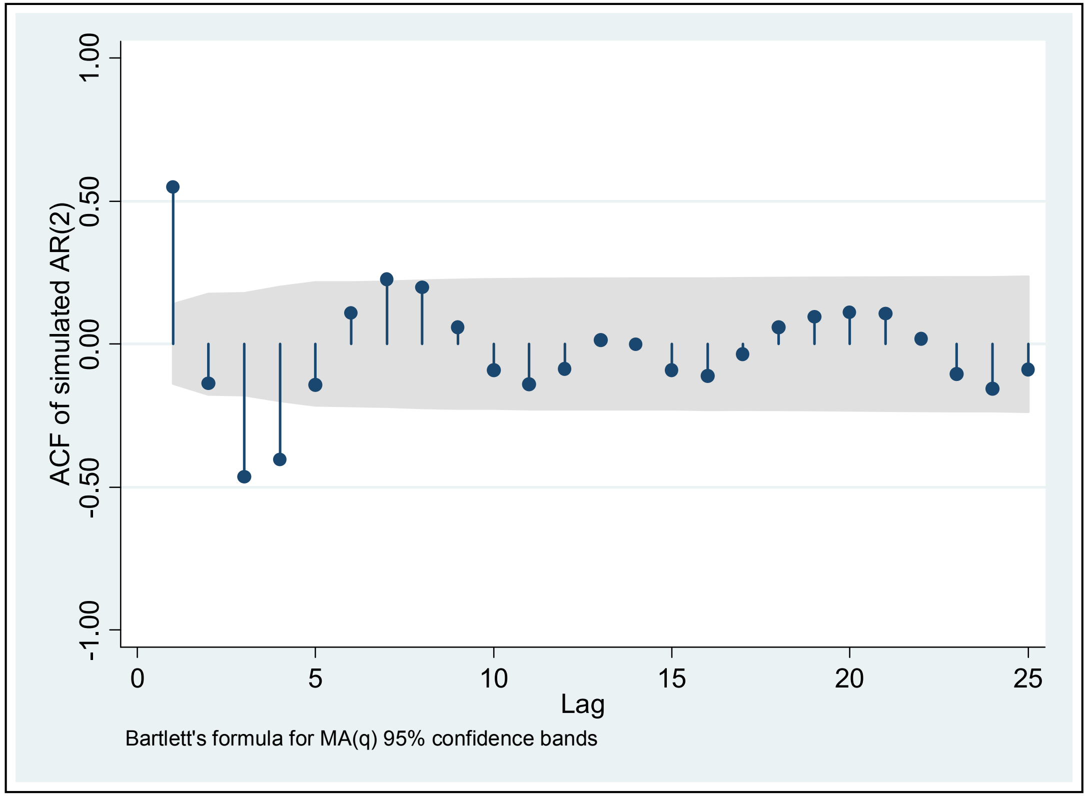
The **ACF** of the residuals suggests that the AR(1) model is not adequate because of the significant spike at lag 2.

Pankratz (1983) “Forecasting with univariate Box-Jenkins models”

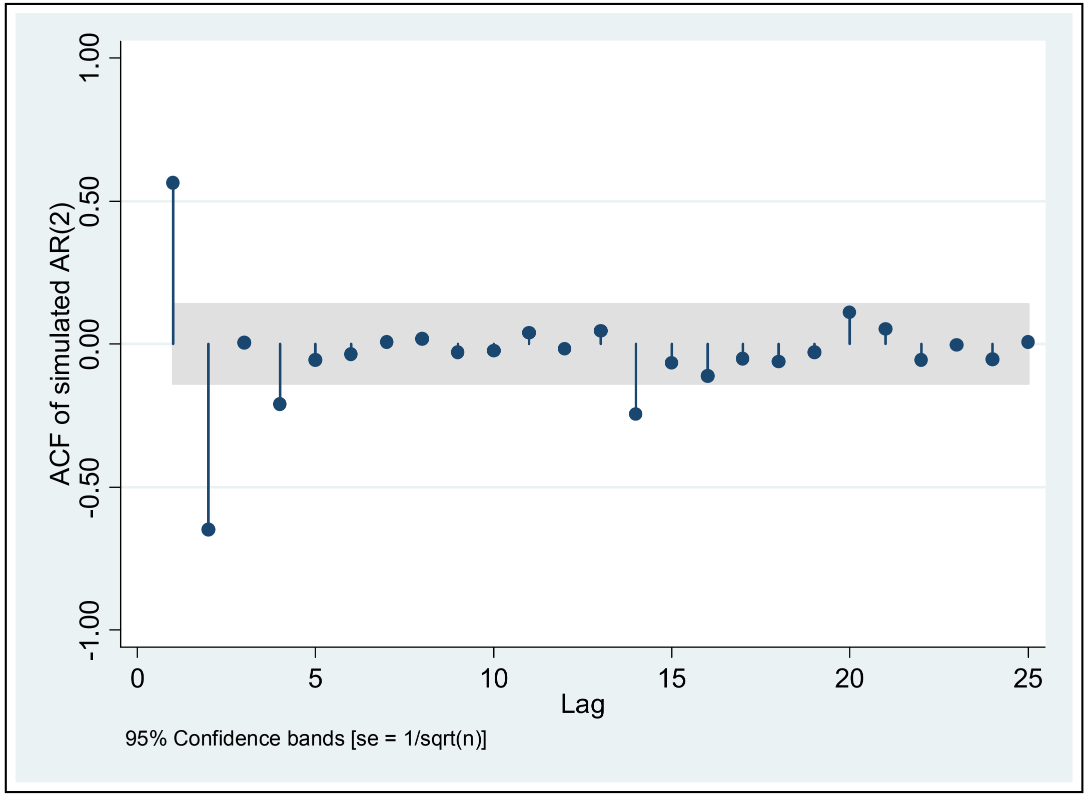
Simulated AR(2)



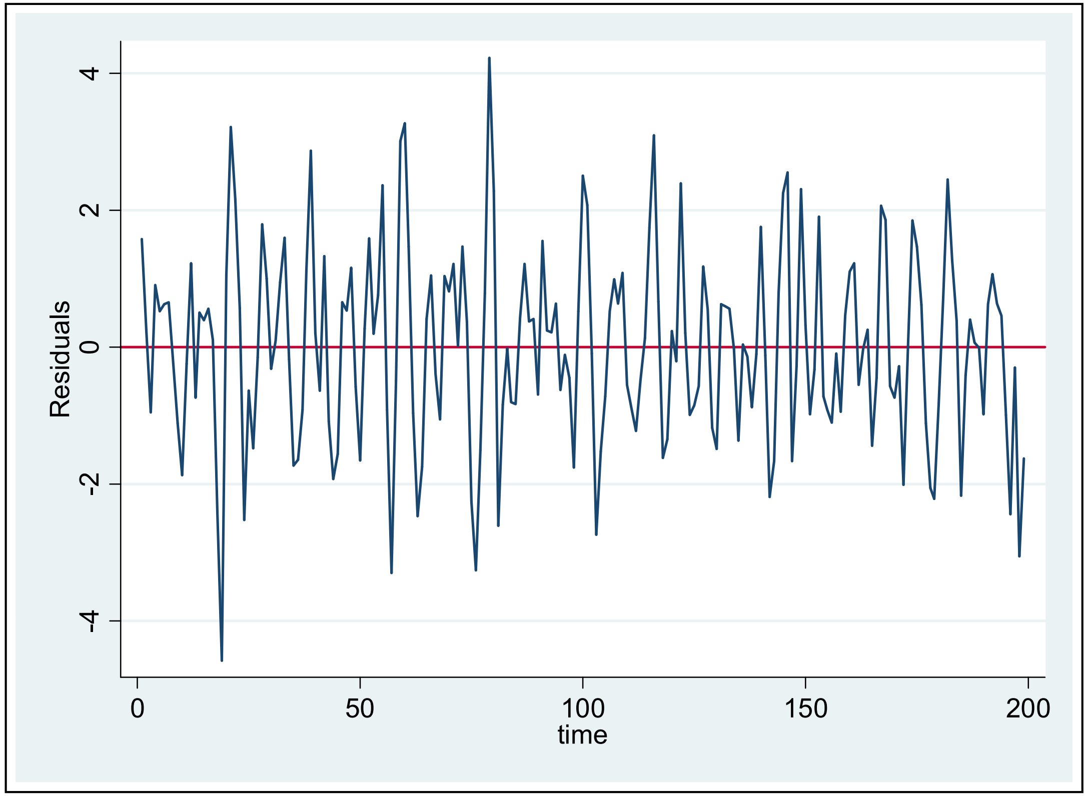
ACF simulated AR(2)



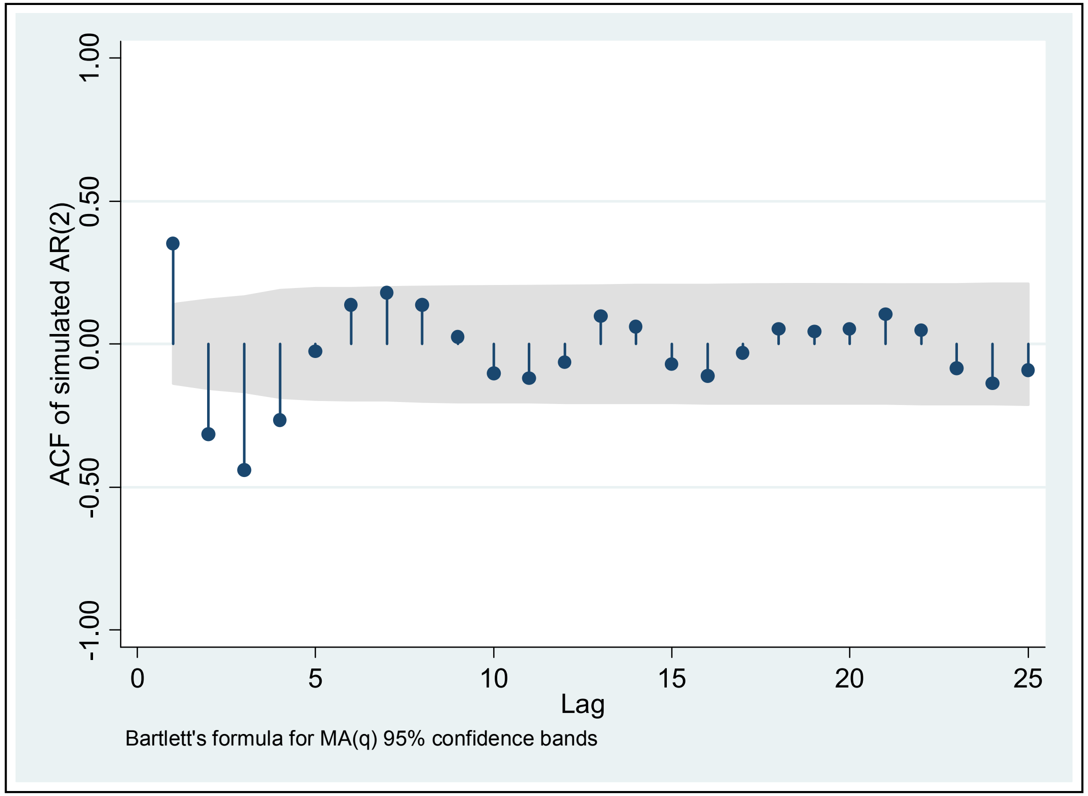
PACF simulated AR(2)



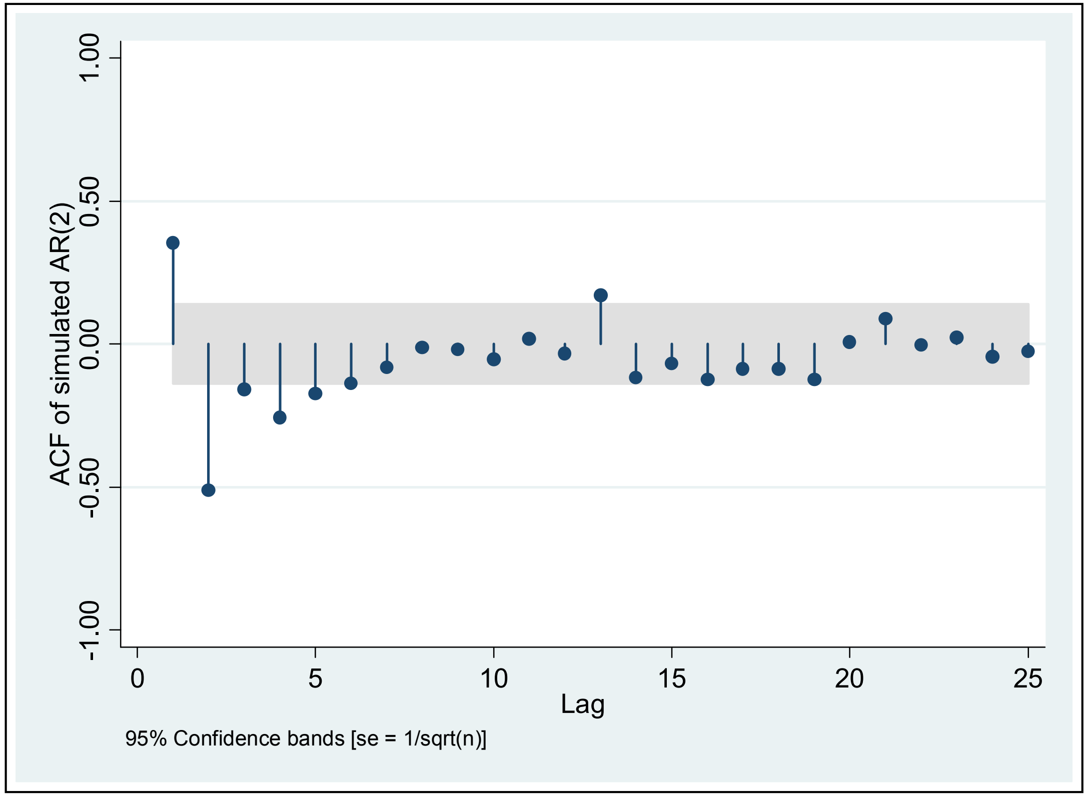
Residuals of an AR(1) model of a simulated AR(2) process



ACF of the residuals of an AR(1) model of a simulated AR(2) process



PACF of the residuals of an AR(1) model of a simulated AR(2) process



Testing that the residuals are a White Noise process:

Joint Hypothesis Test (Box and Pierce)

H_0: All autocorrelation coefficients are zero

$$Q = T\sum_{k=1}^{K}\hat{\rho}_k^2 \quad \sim\chi^2 \quad \text{with } K-p-q \text{ degrees of freedom}$$

Refined Test (Box and Ljung)

$$Q = T(T+2)\sum_{k=1}^{K}\frac{1}{T-k}\hat{\rho}_k^2 \quad \sim\chi^2 \quad \text{with } K-p-q \text{ degrees of freedom}$$

Comparison of different candidate models:

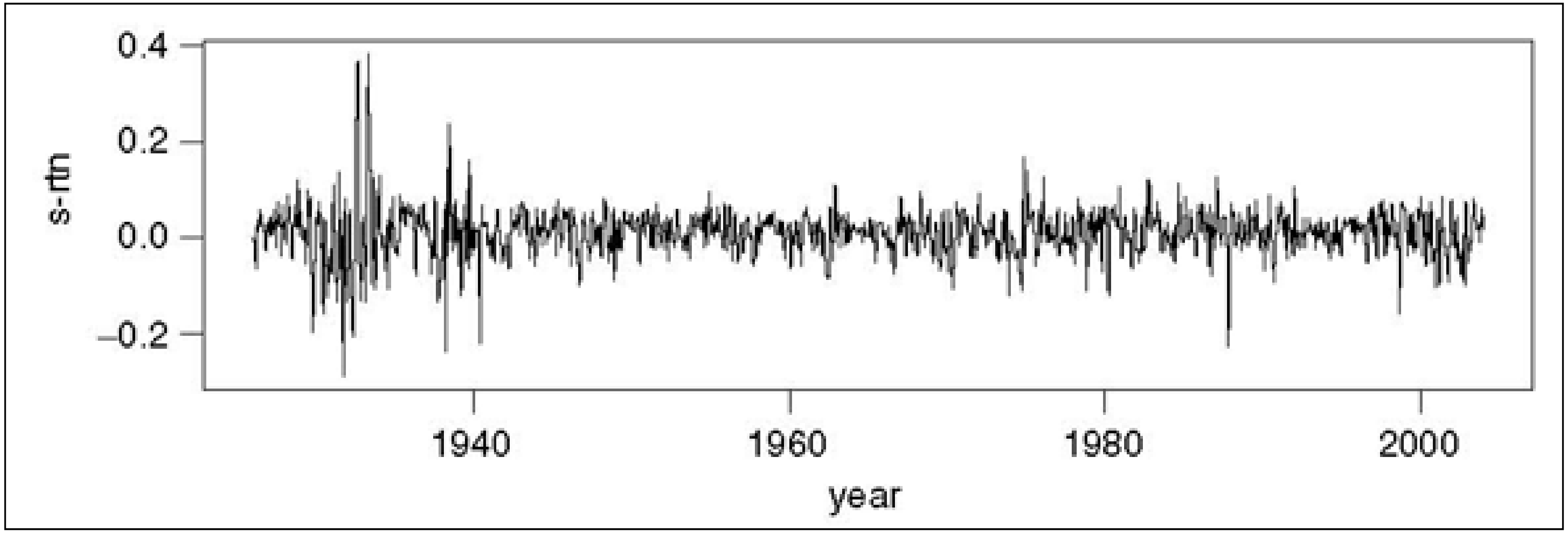
Akaike's Information Criterion (AIC)

$$AIC = \log \hat{\sigma}^2 + 2\frac{p+q}{T}$$

Bayesian Information Criterion (Schwarz)

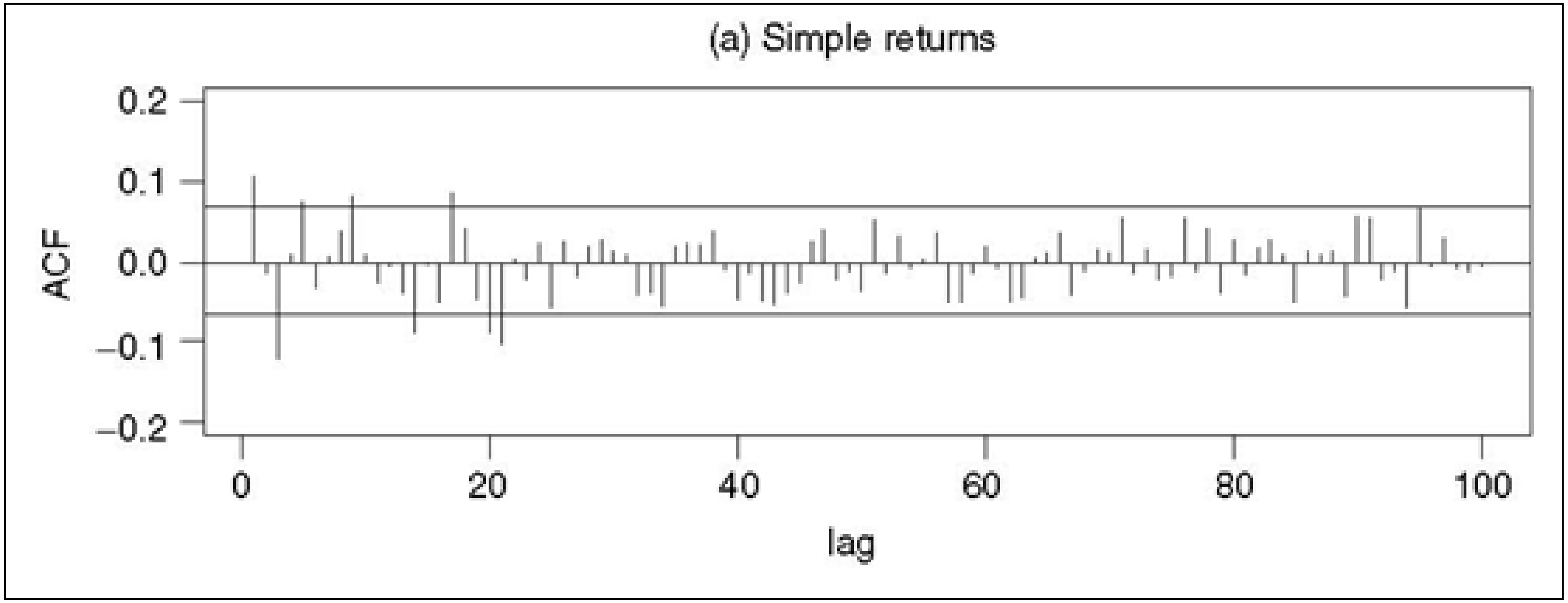
$$BIC = \log \hat{\sigma}^2 + \frac{p+q}{T}\log T$$

Example: monthly simple returns of CRSP index



Monthly simple returns of the value-weighted CRSP index from January 1926 to **December 2003**.

Tsay (2005) “Analysis of Financial Time Series ”, p. 18

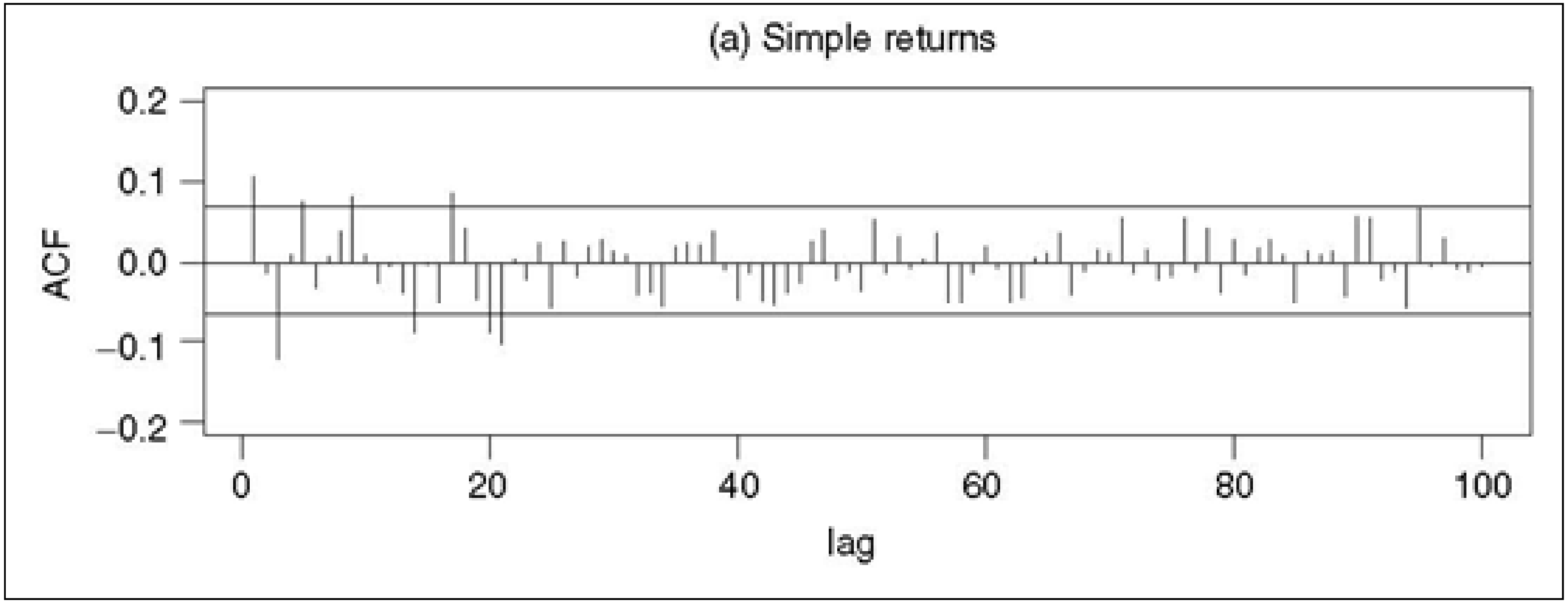


Sample ACF of monthly simple returns of the value-weighted CRSP index from January 1926 to **December 1997**. Horizontal lines denote two standard-error intervals of the sample ACF.

$$T = 72 \cdot 12 = 864 \Rightarrow 2 \Big/ \sqrt{T} = 2 \Big/ \sqrt{864} \approx 0.06$$

Tsay (2005) "Analysis of Financial Time Series ", p. 29

CRSP Example: Identification (ACF)



There are some significant serial correlations at the 5% level. The Ljung–Box statistics give $Q(5) = 27.8$ and $Q(10) = 36.0$ for the simple returns .The p-values of these .. test statistics are all less than 0.0003, suggesting that monthly returns of the value-weighted index are serially correlated.

Tsay (2005) "Analysis of Financial Time Series ", p. 29

CRSP Example: Identification (PACF)

Table 2.1. Sample Partial Autocorrelation Function and Akaike Information Criterion for the Monthly Simple Returns of CRSP Value-Weighted Index from January 1926 to December 1997

p	1	2	3	4	5
PACF	0.11	−0.02	−0.12	0.04	0.07
AIC	−5.807	−5.805	−5.817	−5.816	−5.819
p	6	7	8	9	10
PACF	−0.06	0.02	0.06	0.06	−0.01
AIC	−5.821	−5.819	−5.820	−5.821	−5.818

As an example, consider the monthly simple returns of CRSP value-weighted index from January 1926 to December 1997. Table 2.1 gives the first 10 lags of a sample PACF of the series. With $T = 864$, the asymptotic standard error of the sample PACF is approximately 0.03. Therefore, using the 5% significance level, we identify an AR(3) or AR(5) model for the data (i.e., $p = 3$ or 5).

CRSP Example: Identification (AIC)

Table 2.1. Sample Partial Autocorrelation Function and Akaike Information Criterion for the Monthly Simple Returns of CRSP Value-Weighted Index from January 1926 to December 1997

p	1	2	3	4	5
PACF	0.11	−0.02	−0.12	0.04	0.07
AIC	−5.807	−5.805	−5.817	−5.816	−5.819
p	6	7	8	9	10
PACF	−0.06	0.02	0.06	0.06	−0.01
AIC	−5.821	−5.819	−5.820	−5.821	−5.818

Table 2.1 also gives the AIC for $p = 1, \ldots, 10$. The AIC values are close to each other with minimum −5.821 occurring at $p = 6$ and 9, suggesting that an AR(6) model is preferred by the criterion. This example shows that different approaches for order determination may result in different choices of p. There is no evidence to suggest that one approach outperforms the other in a real application. Substantive information of the problem under study and simplicity are two factors that also play an important role in choosing an AR model for a given time series.

index in Table 2.1. The fitted model is

$$r_t = 0.0103 + 0.104 r_{t-1} - 0.010 r_{t-2} - 0.120 r_{t-3} + \hat{a}_t, \quad \hat{\sigma}_a = 0.054.$$

The standard errors of the coefficients are 0.002, 0.034, 0.034, and 0.034, respectively. Except for the lag-2 coefficient, all parameters are statistically significant at the 1% level.

For this example, the AR coefficients of the fitted model are small, indicating that the serial dependence of the series is weak, even though it is statistically significant at the 1% level. The significance of $\hat{\phi}_0$ of the entertained model implies that the expected mean return of the series is positive. In fact, $\hat{\mu} = 0.0103/(1 - 0.104 + 0.010 + 0.120) = 0.01$, which is small, but has an important long-term implication.

Tsay (2005) "Analysis of Financial Time Series ", p. 18

CRSP Example: Diagnostic Checking

$$r_t = 0.0103 + 0.104r_{t-1} - 0.010r_{t-2} - 0.120r_{t-3} + \hat{a}_t, \quad \hat{\sigma}_a = 0.054.$$

The standard errors of the coefficients are 0.002, 0.034, 0.034, and 0.034, respectively. Except for the lag-2 coefficient, all parameters are statistically significant at the 1% level.

Consider the residual series of the fitted AR(3) model for the monthly value-weighted simple returns. We have $Q(12) = 16.9$ with p-value 0.050 based on its asymptotic chi-squared distribution with 9 degrees of freedom. Thus, the null hypothesis of no residual serial correlation in the first 12 lags is barely not rejected at the 5% level. However, since the lag-2 AR coefficient is not significant at the 5% level, one can refine the model as

$$r_t = 0.0102 + 0.103r_{t-1} - 0.122r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0542,$$

where all the estimates are significant at the 5% level. The residual series gives $Q(12) = 17.2$ with p-value 0.070 (based on χ^2_{10}). The model is adequate in modeling the dynamic linear dependence of data.

Schedule

Deter-ministic Models	• Components of a Time Series • Additive and Multiplicative Models • Simple Trend Models • Smoothing Techniques • Seasonal Adjustment

Stationary Stochastic Processes	• Introduction
	• Identification
	  • Autocorrelation Function
	  • Moving Average and Autoregressive Models
	  • Partial Autocorrelation Function
	  • ARMA Models
	• Estimation
	• Diagnostic Checking
	• **Forecasting**

Non-stationary Stochastic Processes	• Introduction • Nonstationarity and Trends • ARIMA Models • Unit Root Tests • Seasonal ARIMA

Optimal Forecasts

Forecast error $$e_{T+l} = y_{T+l} - \underbrace{\tilde{y}_{T+l} \mid \Omega_T}$$

Forecast horizon

l-period ahead forecast
conditional on information available at *T*

How should we form $\tilde{y}_{T+1} \mid \Omega_T$? Minimize forecast MSE

$$\min_{\tilde{y}_{T+l}} MSE(\tilde{y}_{T+l}) = E\left[e_{T+l}^2 | \Omega_T\right] = E\left[(y_{T+l} - \tilde{y}_{T+l})^2 | \Omega_T\right]$$

Solution : $\tilde{y}_{T+1} \mid \Omega_T = E[Y_{T+1} \mid \Omega_T]$

Assumptions (about Ω_T)

1. the ARMA parameters are known

   (i.e., we assume that we know the true model; in practice, we do not; estimation error is usually a smaller source of prediction errors if T is not too small)

2. for MA and mixed processes: all present and past disturbances are known

   (this is effectively the same as assuming an infinite realisation of observations; in practice, we only have a finite number of past observations, but if T is not too small this hardly matters)

3. the disturbance term ε_t is normally distributed

   (to get prediction intervals; already assumed that for estimation)

Harvey (1981) “Time Series Models”, p.157-164

Forecasting an **AR(1) Process**

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t$$

One-period forecast

$$y_{T+1} = \varphi_1 y_T + \delta + \varepsilon_{T+1}$$

$$\tilde{y}_{T+1|T} = E\left(y_{T+1} | y_T, \ldots, y_1\right) = \varphi_1 y_T + \delta$$

Two-period forecast

$$y_{T+2} = \varphi_1 y_{T+1} + \delta + \varepsilon_{T+2} = \varphi_1\left(\varphi_1 y_T + \delta + \varepsilon_{T+1}\right) + \delta + \varepsilon_{T+2}$$

$$= \varphi_1^2 y_T + (\varphi_1 + 1)\delta + \varphi_1 \varepsilon_{T+1} + \varepsilon_{T+2}$$

$$\tilde{y}_{T+2|T} = \varphi_1 \tilde{y}_{T+1|T} + \delta = \varphi_1^2 y_T + (\varphi_1 + 1)\delta$$

Forecasting an **AR(1) Process**

***l*-period** forecast

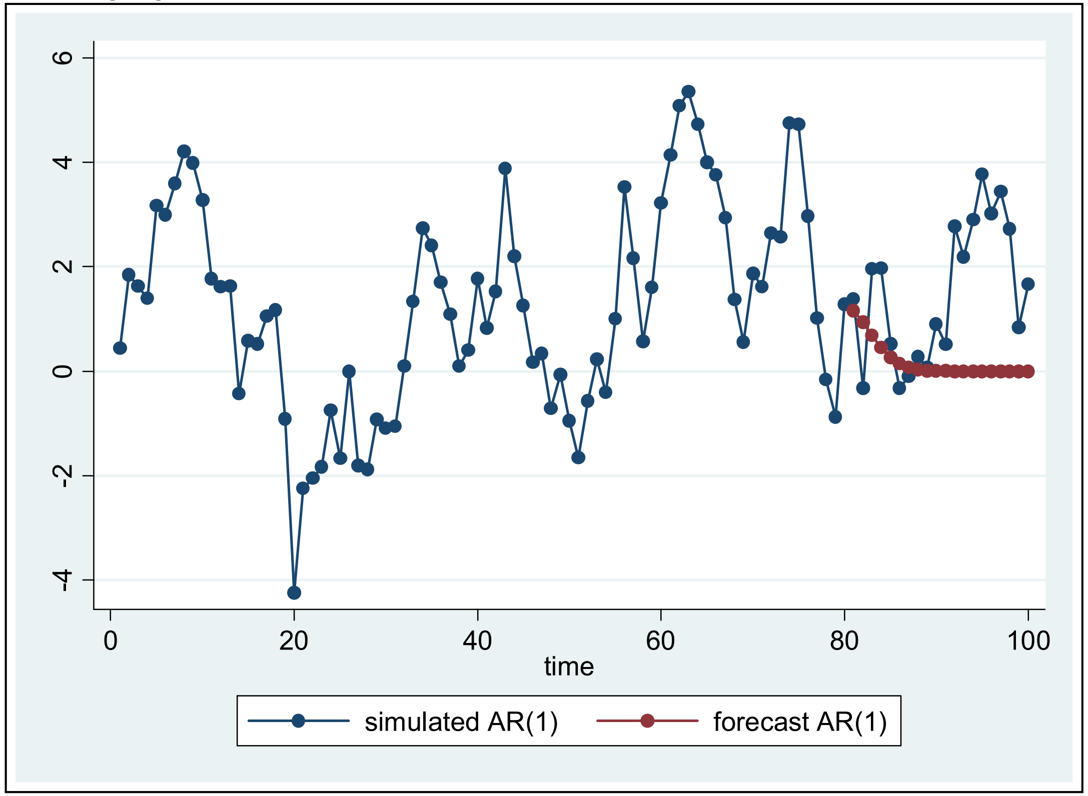
$$\tilde{y}_{T+l|T} = \varphi_1^l y_T + \left(\varphi_1^{l-1} + \varphi_1^{l-2} + \ldots + \varphi_1 + 1\right)\delta$$

Note: $\lim_{l \to \infty} \tilde{y}_{T+l|T} = \delta \sum_{j=0}^{\infty} \varphi_1^j = \dfrac{\delta}{1-\varphi_1} = \mu$

The forecast decays geometrically toward μ as the forecast horizon l increases.

Example:

AR(1) process $y_t = 0.9 y_{t-1} + \varepsilon_t$



Forecast Error **AR(1) Process**

One-period forecast error

$$
\begin{aligned}
e_{T+1} &= y_{T+1} - \tilde{y}_{T+1|T} \\
&= \varphi_1 y_T + \delta + \varepsilon_{T+1} - \varphi_1 y_T - \delta \\
&= \varepsilon_{T+1}
\end{aligned}
$$

MSE

$$E\left[e_{T+1}^2\right] = Var\left(e_{T+1}\right) = E\left(\varepsilon_{T+1}^2\right) = Var\left(\varepsilon_{T+1}\right) = \sigma_\varepsilon^2$$

Forecast Error **AR(1) Process**

One-period forecast error $E\left[e_{T+1}^2\right] = \sigma_\varepsilon^2$

Two-period forecast error

$$
\begin{aligned}
e_{T+2} &= y_{T+2} - \tilde{y}_{T+2|T} \\
&= \varphi_1^2 y_T + (\varphi_1 + 1)\delta + \varphi_1 \varepsilon_{T+1} + \varepsilon_{T+2} - \left(\varphi_1^2 y_T + (\varphi_1 + 1)\delta\right) \\
&= \varphi_1 \varepsilon_{T+1} + \varepsilon_{T+2}
\end{aligned}
$$

$$MSE \quad E\left[e_{T+2}^2\right] = \varphi_1^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 = \left(1 + \varphi_1^2\right)\sigma_\varepsilon^2$$

Forecast Error **AR(1) Process**

One-period forecast error $E\left[e_{T+1}^2\right] = \sigma_\varepsilon^2$

Two-period forecast error $E\left[e_{T+2}^2\right] = \left(1 + \varphi_1^2\right)\sigma_\varepsilon^2$

Similarly,

Three-period forecast error $E\left[e_{T+3}^2\right] = \left(1 + \varphi_1^2 + \varphi_1^4\right)\sigma_\varepsilon^2$

and for $l \to \infty$

$$E\left[e_{T+l}^2\right] = \left(1 + \varphi_1^2 + \varphi_1^4 + \ldots\right)\sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \varphi_1^2}$$

which is the **unconditional variance of y_t**

Forecasting an **MA(1) Process**

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

One-period forecast

$$y_{T+1} = \mu + \varepsilon_{T+1} - \theta_1 \varepsilon_T$$

$$\tilde{y}_{T+1|T} = E\left(y_{T+1} \middle| \underbrace{y_T, \ldots, y_1, y_0, y_{-1}, \ldots}\right) = \mu - \theta_1 \varepsilon_T$$

Recall our assumption:

2. for MA and mixed processes: all present and past disturbances are known

   (this is effectively the same as assuming an infinite realisation of observations; in practice, we only have a finite number of past observations, but if T is not too small this hardly matters)

Forecasting an **MA(1) Process**

One-period forecast

$$\tilde{y}_{T+1|T} = E\left(y_{T+1} \middle| y_T,\ldots,y_1,y_0,y_{-1},\ldots\right) = \mu - \theta_1 \varepsilon_T$$

Under this assumption (all present and past disturbances are known), ε_T is known (i.e., it is in the information set) and computing the conditional mean (the forecast) is very easy.

But this is unrealistic. In practice, we could calculate ε_T recursively, assuming that ε_0 =0.

Recursion:

$$\varepsilon_1 = y_1 + \theta_1 \varepsilon_0$$
$$\varepsilon_2 = y_2 + \theta_1 \varepsilon_1$$
$$\varepsilon_3 = y_3 + \theta_1 \varepsilon_2$$
$$\vdots$$
$$\varepsilon_T = y_T + \theta_1 \varepsilon_{T-1}$$

if $\varepsilon_0 = 0 \Rightarrow$

$$\varepsilon_1^* = y_1$$
$$\varepsilon_2^* = y_2 + \theta_1 y_1$$
$$\varepsilon_3^* = y_3 + \theta_1 y_2 + \theta_1^2 y_1$$
$$\vdots$$
$$\varepsilon_T^* = y_T + \theta_1 y_{T-1} + \ldots + \theta_1^{T-1} y_1$$

How important is the error from using ε_T^* instead of ε_T ?

$$\varepsilon_T = y_T + \theta_1 y_{T-1} + \ldots + \theta_1^{T-1} y_1 + \theta_1^T \varepsilon_0$$
$$\varepsilon_T^* = y_T + \theta_1 y_{T-1} + \ldots + \theta_1^{T-1} y_1$$

→ Small error if T is large and $|\theta_1|<1$ (process is “invertible”; see below)

Finite sample Prediction and Invertibility

$$y_{T+1} = \mu + \varepsilon_{T+1} - \theta_1 \varepsilon_T$$

$$\tilde{y}^*_{T+1} = \mu + \underbrace{\varepsilon_{T+1}}_{=0} - \theta_1 \varepsilon^*_T$$

$$e_{T+1} = \varepsilon_{T+1} - \theta_1 \left(\varepsilon_T - \varepsilon^*_T \right)$$

$$= \varepsilon_{T+1} - \theta_1 \left(\theta_1^T \varepsilon_0 \right) = \varepsilon_{T+1} - \theta_1^{T+1} \varepsilon_0$$

$$E\left[e^2_{T+1}\right] = E\left[\left(\varepsilon_{T+1} - \theta_1^{T+1}\varepsilon_0\right)^2\right]$$

$$= E\left[\varepsilon^2_{T+1} - 2\theta_1^{T+1}\varepsilon_{T+1}\varepsilon_0 + \theta_1^{2(T+1)}\varepsilon_0^2\right]$$

$$= \sigma_\varepsilon^2 + \theta_1^{2(T+1)}\sigma_\varepsilon^2 = \left(1 + \theta_1^{2(T+1)}\right)\sigma_\varepsilon^2$$

Finite sample Prediction and Invertibility

$$MSE\left(\tilde{y}^*_{T+1|T}\right) = E\left\{\left[\varepsilon_{T+1} - \theta_1\left(\varepsilon_T - \varepsilon^*_T\right)\right]^2\right\} = \left[1 + \theta_1^{\ 2(T+1)}\right]\sigma^2_\varepsilon$$

If T is large, the difference between ε_T and ε^*_T will be negligible, provided that $|\theta_1| < 1$.

The problem of obtaining the exact *MMSE* of a future observation for an MA or mixed process is sometimes referred to as the **finite sample prediction** problem.

Harvey (1981), Time Series Models, p.157-164

Forecasting an **MA(1) Process**

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

One-period forecast $y_{T+1} = \mu + \varepsilon_{T+1} - \theta_1 \varepsilon_T$

$$\tilde{y}_{T+1|T} = E\left(y_{T+1} | y_T, \ldots, y_1, y_0, y_{-1}, \ldots\right) = \mu - \theta_1 \varepsilon_T$$

Two-period forecast $y_{T+2} = \mu + \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1}$

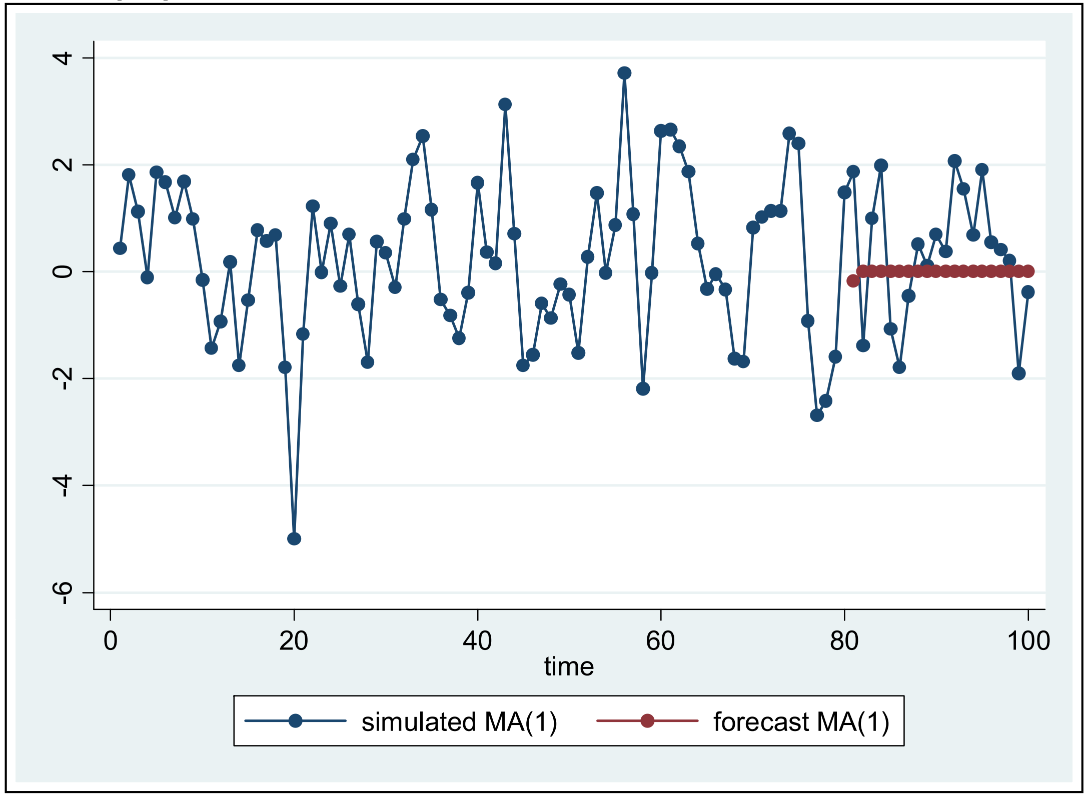
$$\tilde{y}_{T+2|T} = \mu$$

***l*-period** forecast

$$\tilde{y}_{T+l|T} = E\left(y_{T+l} | y_T, \ldots, y_1\right) = E\left(\mu + \varepsilon_{T+l} - \theta_1 \varepsilon_{T+l-1}\right) = \mu$$

Example:

MA(1) process $y_t = \varepsilon_t + 0.8\varepsilon_{t-1}$



Forecast Error **MA(1) Process**

One-period forecast error

$$
\begin{aligned}
e_{T+1} &= y_{T+1} - \tilde{y}_{T+1|T} \\
&= (\mu + \varepsilon_{T+1} - \theta_1 \varepsilon_T) - (\mu - \theta_1 \varepsilon_T) \\
&= \varepsilon_{T+1}
\end{aligned}
$$

MSE

$$
E\left[e_{T+1}^2\right] = \sigma_\varepsilon^2
$$

Forecast Error **MA(1) Process**

Two-period forecast error

$$e_{T+2} = \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1}$$

MSE

$$E\left[e_{T+2}^2\right] = \left(1 + \theta_1^2\right)\sigma_\varepsilon^2$$

***l*-period** forecast for *l* > 1

$$\tilde{y}_{T+l|T} = 0$$

MSE

$$MSE\left(\tilde{y}_{T+1|T}\right) = Var\left(y_{T+l}\right) = \left(1 + \theta_1^2\right)\sigma_\varepsilon^2$$

Forecasting an **ARMA(1,1) Process**

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

One-period forecasts

$$\begin{aligned}\tilde{y}_{T+1|T} &= E(\varphi_1 y_T + \delta + \varepsilon_{T+1} - \theta_1 \varepsilon_T | y_T, \ldots, y_1, y_0, y_{-1}, \ldots) \\ &= \varphi_1 y_T + \delta - \theta_1 \varepsilon_T\end{aligned}$$

Two-period forecasts

$$\begin{aligned}\tilde{y}_{T+2|T} &= E(\varphi_1 y_{T+1} + \delta + \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1} | y_T, \ldots, y_1, y_0, y_{-1}, \ldots) \\ &= \varphi_1 \tilde{y}_{T+1|T} + \delta \\ &= \varphi_1^2 y_T + (\varphi_1 + 1)\delta - \varphi_1 \theta_1 \varepsilon_T\end{aligned}$$

Forecasting an **ARMA(1,1) Process**

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

***l*-period** forecast

$$\begin{aligned} \tilde{y}_{T+l|T} &= \varphi_1 \tilde{y}_{T+(l-1)|T} + \delta \\ &= \varphi_1^l y_T + \left(\varphi_1^{l-1} + \ldots + \varphi_1 + 1\right)\delta - \varphi_1^{l-1}\theta_1 \varepsilon_T \end{aligned}$$

Note: $\lim_{l \to \infty} \tilde{y}_T(l) = \dfrac{\delta}{1-\varphi_1} = \mu$

Forecasting an **ARMA(p,q) Process** (I)

ARMA(p,q) process **at time $T + l$**

$$y_{T+l} = \varphi_1 y_{T+l-1} + \ldots + \varphi_p y_{T+l-p} + \varepsilon_{T+l} - \theta_1 \varepsilon_{T+l-1} - \ldots - \theta_q \varepsilon_{T+l-q}$$

Future values of ε_t are unknown, and cannot be predicted. If this equation is used to predict y_{T+l}, the ε_t's must be set equal to their expected values of zero for $t > T$.

Harvey (1981), Time Series Models, p.157-164

Forecasting an **ARMA(*p*,*q*) Process** (II)

ARMA(p,q) process **at time $T + l$**

$$y_{T+l} = \varphi_1 y_{T+l-1} + \ldots + \varphi_p y_{T+l-p} + \varepsilon_{T+l} - \theta_1 \varepsilon_{T+l-1} - \ldots - \theta_q \varepsilon_{T+l-q}$$

Future values of y_t are also unknown, but predictions may be computed from recursion:

$$\tilde{y}_{T+l|T} = \varphi_1 \tilde{y}_{T+l-1|T} + \ldots + \varphi_p \tilde{y}_{T+l-p|T} + \tilde{\varepsilon}_{T+l|T} - \ldots - \theta_q \tilde{\varepsilon}_{T+l-q|T}$$

where $\tilde{y}_{T+j|T} = y_{T+j}$ for $j \le 0$ and $\tilde{\varepsilon}_{T+j|T} = \begin{cases} 0 & \text{for } j > 0 \\ \varepsilon_{T+j} & \text{for } j \le 0 \end{cases}$

Harvey (1981), Time Series Models, p.157-164

Example 1: AR(1) process

$$\tilde{y}_{T+l|T} = \varphi_1 \tilde{y}_{T+l-1|T} \qquad l = 1,2,\ldots$$

The starting value is given by: $\tilde{y}_{T|T} = y_T$

$$\tilde{y}_{T+l|T} = \varphi_1^l y_T$$

“Thus the predicted value decline exponentially towards zero, and the forecast function has exactly the same form as the autocovariance function.”

Harvey (1981), Time Series Models, p.157-164

Example 2: MA(1) process

MA(1) process **at time *T* + 1**:

$$y_{T+1} = \varepsilon_{T+1} - \theta_1 \varepsilon_T$$

Since ε_{T+1} is unknown, it is set equal to zero in the corresponding prediction equation which is:

$$\tilde{y}_{T+1|T} = -\theta_1 \varepsilon_T$$

For $l > 1$, $\tilde{y}_{T+l|T} = 0$, and so knowledge of the data generating process is of no help in predicting more than one period ahead.

Harvey (1981), Time Series Models, p.157-164

Example 3: ARMA(2,2) process

$$y_t = 0.6y_{t-1} + 0.2y_{t-2} + \varepsilon_t + 0.3\varepsilon_{t-1} - 0.4\varepsilon_{t-2}$$

$$y_T = 4, \quad y_{T-1} = 5, \quad \varepsilon_T = 1, \quad \varepsilon_{T-1} = 0.5$$

Then

$$\tilde{y}_{T+1|T} = 0.6y_T + 0.2y_{T-1} + 0.3\varepsilon_T - 0.4\varepsilon_{T-1} = 3.5$$

$$\tilde{y}_{T+2|T} = 0.6\tilde{y}_{T+1|T} + 0.2y_T - 0.4\varepsilon_T = 2.5$$

Thereafter forecasts are generated by the difference equation

$$\tilde{y}_{T+l|T} = 0.6\tilde{y}_{T+l-1|T} + 0.2\tilde{y}_{T+l-2|T} \qquad l = 3,4,\ldots$$

Harvey (1981), Time Series Models, p.157-164

ARMA(*p*,*q*) process at time *T* + *l*:

$$\tilde{y}_{T+l} = \varphi_1 \tilde{y}_{T+l-1} + \ldots + \varphi_p y_{T+l-p} + \tilde{\varepsilon}_{T+l} - \theta_1 \tilde{\varepsilon}_{T+l-1} - \ldots - \theta_q \tilde{\varepsilon}_{T+l-q|T}$$

Recursive forecasting recipe:

1. replace unknown y_{T+l} by their forecasts for $l > 0$;
2. "forecasts" of y_{T+l} , $l \leq 0$, are simply the known values y_{T+l}
3. since ε_t is white noise, the optimal forecast of ε_{T+l}, $l > 0$, is simply zero
4. "forecasts" of ε_{T+l}, $l \leq 0$, are just the known values ε_{T+l}

Example: $p = 4$ and $q = 3$

for $l = 1$:

$$y_{T+1} = \varphi_1 y_T + \varphi_2 y_{T-1} + \varphi_3 y_{T-2} + \varphi_4 y_{T-3} + \varepsilon_{T+1} - \theta_1 \varepsilon_T - \theta_2 \varepsilon_{T-1} - \theta_3 \varepsilon_{T-2}$$

$$\tilde{y}_{T+1|T} = \varphi_1 y_T + \varphi_2 y_{T-1} + \varphi_3 y_{T-2} + \varphi_4 y_{T-3} - \theta_1 \varepsilon_T - \theta_2 \varepsilon_{T-1} - \theta_3 \varepsilon_{T-2}$$

for $l = 2$:

$$y_{T+2} = \varphi_1 y_{T+1} + \varphi_2 y_T + \varphi_3 y_{T-1} + \varphi_4 y_{T-2} + \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1} - \theta_2 \varepsilon_T - \theta_3 \varepsilon_{T-1}$$

$$\tilde{y}_{T+2|T} = \varphi_1 \tilde{y}_{T+1|T} + \varphi_2 y_T + \varphi_3 y_{T-1} + \varphi_4 y_{T-2} - \theta_2 \varepsilon_T - \theta_3 \varepsilon_{T-1}$$

Forecast Intervals

Example: 1-step ahead for AR(1)

$$y_t = \varphi_1 y_{t-1} + \delta + \varepsilon_t \qquad \tilde{y}_T(1) = \tilde{y}_{T+1} = E(y_{T+1} | \Omega_T) = \varphi_1 y_T + \delta$$

$$e_{T+1} = y_{T+1} - \tilde{y}_{T+1} = \varepsilon_{T+1} \qquad \text{if } \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \Rightarrow e_{T+1} \sim N(0, \sigma_\varepsilon^2)$$

$$P(-z_{1-\alpha/2} \leq \frac{e_{T+1}}{\sigma_\varepsilon} \leq z_{1-\alpha/2}) = 1 - \alpha$$

$$\Rightarrow P(\tilde{y}_{T+1} - z_{1-\alpha/2}\sigma_\varepsilon \leq y_{T+1} \leq \tilde{y}_{T+1} + z_{1-\alpha/2}\sigma_\varepsilon) = 1 - \alpha$$

$$95\% \text{ interval} \quad y_{T+1} = \tilde{y}_{T+1} \pm 1.96\sigma_\varepsilon$$

$$\text{Note : we need} \quad E\left[e_{T+1}^2\right] = MSE\left(\tilde{y}_{T+1|T}\right) = E\left[\left(y_{T+1} - \tilde{y}_{T+1|T}\right)^2\right]$$

The ***MSE*** is easy to derive **for MA processes**

Example: MA(1)

$$
\begin{aligned}
e_{T+1} &= y_{T+1} - \tilde{y}_{T+1|T} \\
&= \underbrace{\left(\mu + \varepsilon_{T+1} - \theta_1 \varepsilon_T\right) - \left(\mu - \theta_1 \varepsilon_T\right)}_{\text{Forecast error is just a function of shocks !!}} \\
&= \varepsilon_{T+1}
\end{aligned}
$$

$$
\begin{aligned}
MSE\left(\tilde{y}_{T+1|T}\right) &= E\left[\left(y_{T+1} - \tilde{y}_{T+1|T}\right)^2\right] \\
&= E\left[e_{T+1}^2\right] = E\left[\varepsilon_{T+1}^2\right] = Var\left[\varepsilon_{T+1}\right] = \sigma_\varepsilon^2
\end{aligned}
$$

Harvey (1981), Time Series Models, p.157-164

General Solution "MA representation"

Any stationary ARMA process may be expressed as an infinite MA:

$$y_{T+l} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{T+l-j}$$

$$= \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \ldots + \psi_{l-1} \varepsilon_{T+1} + \psi_l \varepsilon_T + \psi_{l+1} \varepsilon_{T-1} + \ldots$$

The **conditional expectation** of y_{t+l} is

$$\tilde{y}_{T+l|T} = \sum_{j=0}^{\infty} \psi_{l+j} \varepsilon_{T-j}$$

$$= \psi_l \varepsilon_T + \psi_{l+1} \varepsilon_{T-1} + \ldots$$

Harvey (1981), Time Series Models, p.157-164

General Solution "MA representation"

Forecast Error

$$e_{T+l} = \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \ldots \psi_{l-1} \varepsilon_{T+1}$$

MSE

$$E[e_{T+l}^2 | \Omega_T] = (\varepsilon_{T+1} + \psi_1 \varepsilon_{T+l-1} + \ldots + \psi_{l-1} \varepsilon_{T+1})(\varepsilon_{T+1} + \psi_1 \varepsilon_{T+l-1} + \ldots + \psi_{l-1} \varepsilon_{T+1})$$

Since $E[\varepsilon_i \varepsilon_j] = 0, i \neq j$

$$E[e_{T+l}^2 | \Omega_T] = (1 + \psi_1^2 + \ldots + \psi_{l-1}^2) \sigma_\varepsilon^2$$

Harvey (1981), Time Series Models, p.157-164

General Solution "MA representation"

Forecast Error

$$MSE\left(\tilde{y}_{T+l|T}\right) = E\left[\left(y_{T+l} - \tilde{y}_{T+l|T}\right)^2\right] = \left(1 + \psi_1^2 + \ldots + \psi_{l-1}^2\right)\sigma_\varepsilon^2$$

Prediction Interval (the disturbance term ε_t is normally distributed)

$$y_{T+l} = \tilde{y}_{T+l|T} \pm 1.96\left(1 + \psi_1^2 + \ldots + \psi_{l-1}^2\right)^{\frac{1}{2}}\sigma_\varepsilon$$

$$= \tilde{y}_{T+l|T} \pm 1.96\left(1 + \sum_{j=1}^{l-1}\psi_j^2\right)^{\frac{1}{2}}\sigma_\varepsilon$$

Harvey (1981), Time Series Models, p.157-164

How do we find $\psi_1, \ldots, \psi_{l-1}$?

$$y_t = \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

$$\left(1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p\right) y_t = \left(1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q\right) \varepsilon_t$$

$$a(L) y_t = b(L) \varepsilon_t$$

$$y_t = c(L) \varepsilon_t$$

$$= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{with} \quad \psi_0 = 1$$

So the ψ_1, ψ_2, ... coefficients in $c(L)$, can be obtained by equating coefficients of L^j, j = 1, 2, ... in $a(L)c(L) = b(L)$.

How do we find $\psi_1, \ldots, \psi_{l-1}$?

Example ARMA(1,1): $y_t = 0.26 y_{t-1} + \varepsilon_t - 0.28\varepsilon_{t-1}$

$$\underbrace{(1 - 0.26L)}_{a(L)} y_t = \underbrace{(1 - 0.28L)}_{b(L)} \varepsilon_t$$

$$a(L)c(L) = b(L)$$

$$(1 - 0.26L)(1 + \psi_1 L + \psi_2 L^2 + \ldots) = 1 - 0.28L$$

$$1 - 0.26L + \psi_1 L - 0.26\psi_1 L^2 + \psi_2 L^2 - 0.26\psi_2 L^3 + \ldots = 1 - 0.28L$$

$$-0.26 + \psi_1 = -0.28$$

$$\psi_1 = -0.28 + 0.26 = -0.02$$

How do we find ψ_1, ..., ψ_{l-1}?

Example ARMA(1,1): $y_t = 0.26 y_{t-1} + \varepsilon_t - 0.28\varepsilon_{t-1}$

$$(1-0.26L)\left(1+\psi_1 L+\psi_2 L^2+...\right)=1-0.28L$$

$$1-0.26L+\psi_1 L-0.26\psi_1 L^2+\psi_2 L^2-0.26\psi_2 L^3+...=1-0.28L$$

$$-0.26\psi_1+\psi_2=0$$

$$\psi_2=0.26\cdot\psi_1=0.26\cdot(-0.02)=\text{-}0.0052$$

Prediction Interval

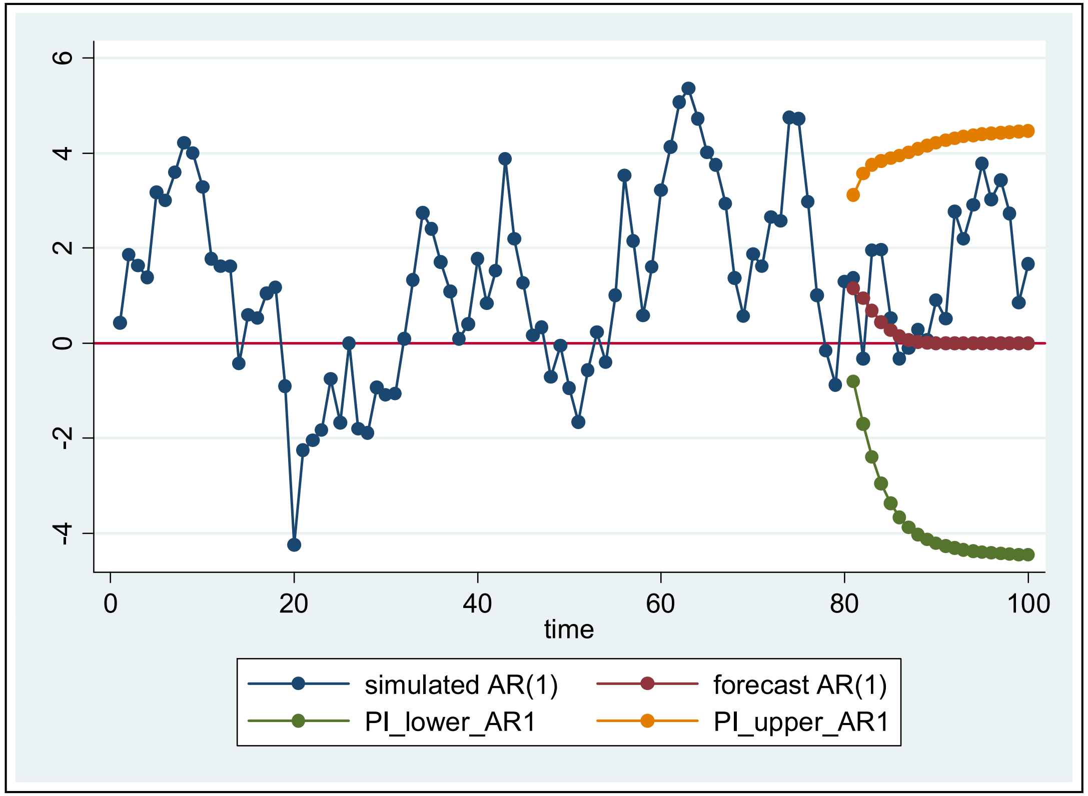
"Since the *MMSE* of y_{T+l} is a linear function of the disturbances, it will be normally distributed when the disturbances are normal. A 95% prediction interval is given by:"

$$y_{T+l} = \tilde{y}_{T+l|T} \pm 1.96\left(1+\psi_1^2+\ldots+\psi_{l-1}^2\right)^{\frac{1}{2}}\sigma_\varepsilon$$

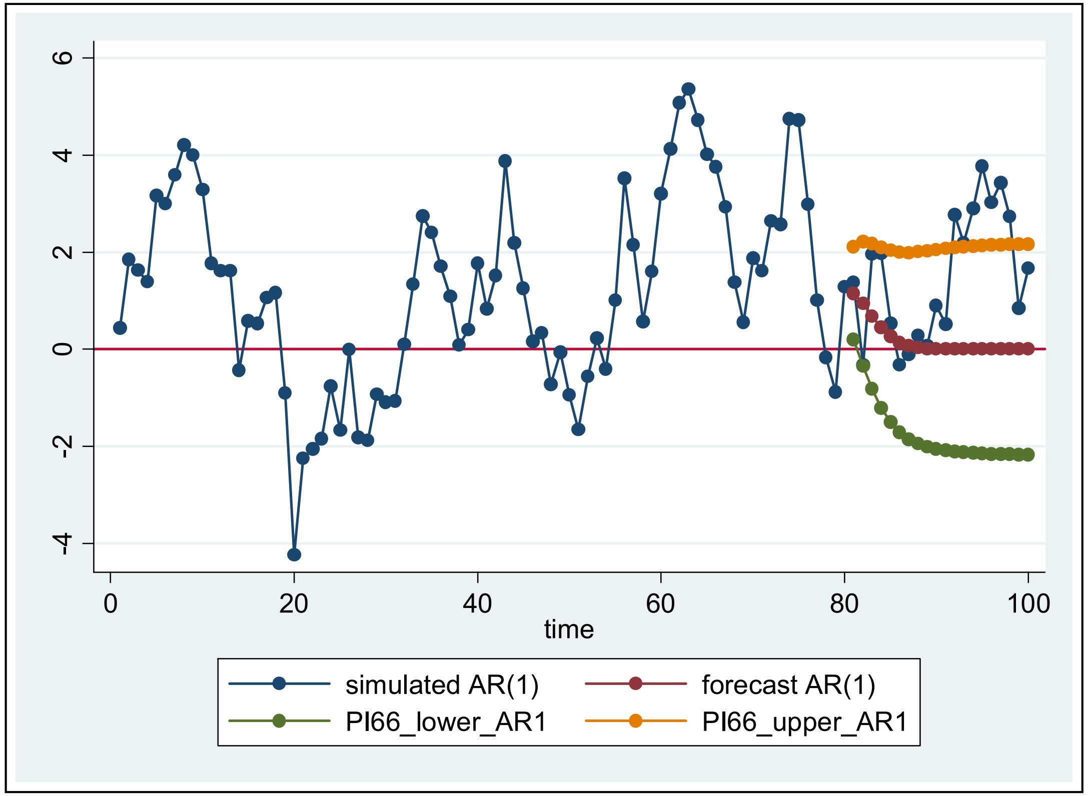
$$= \tilde{y}_{T+l|T} \pm 1.96\left(1+\sum_{j=1}^{l-1}\psi_j^2\right)^{\frac{1}{2}}\sigma_\varepsilon$$

Harvey (1981), Time Series Models, p.157-164

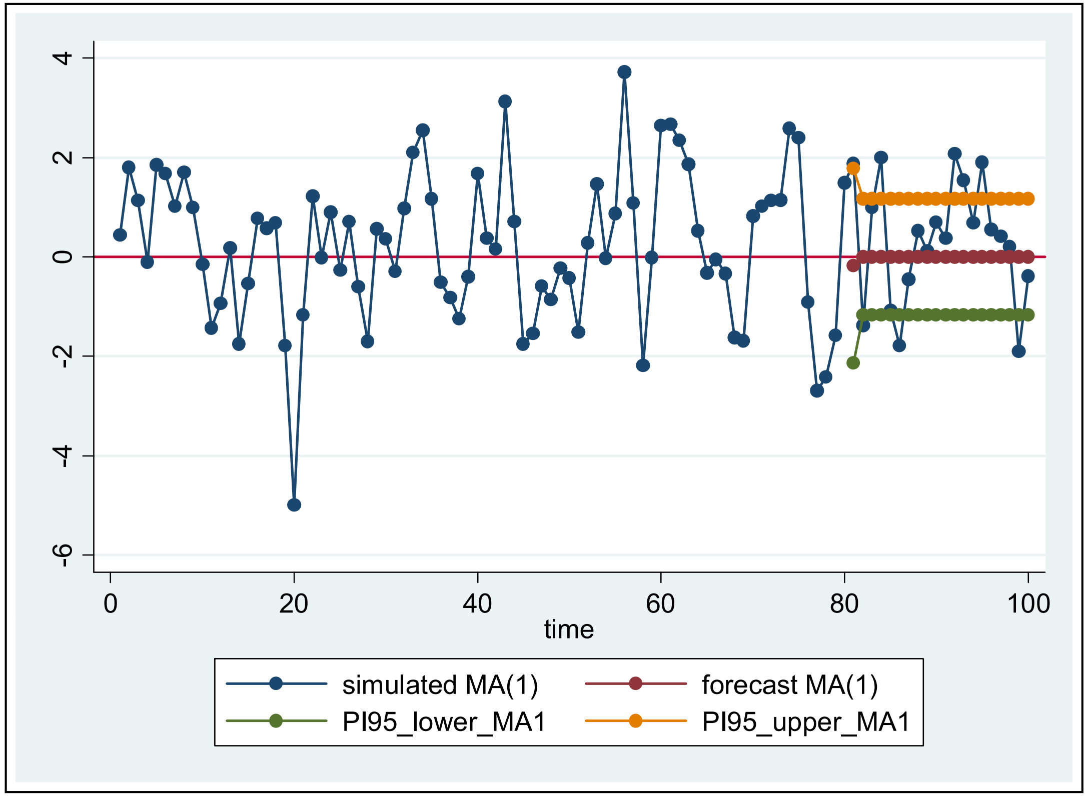
Prediction Interval 95% – AR(1)



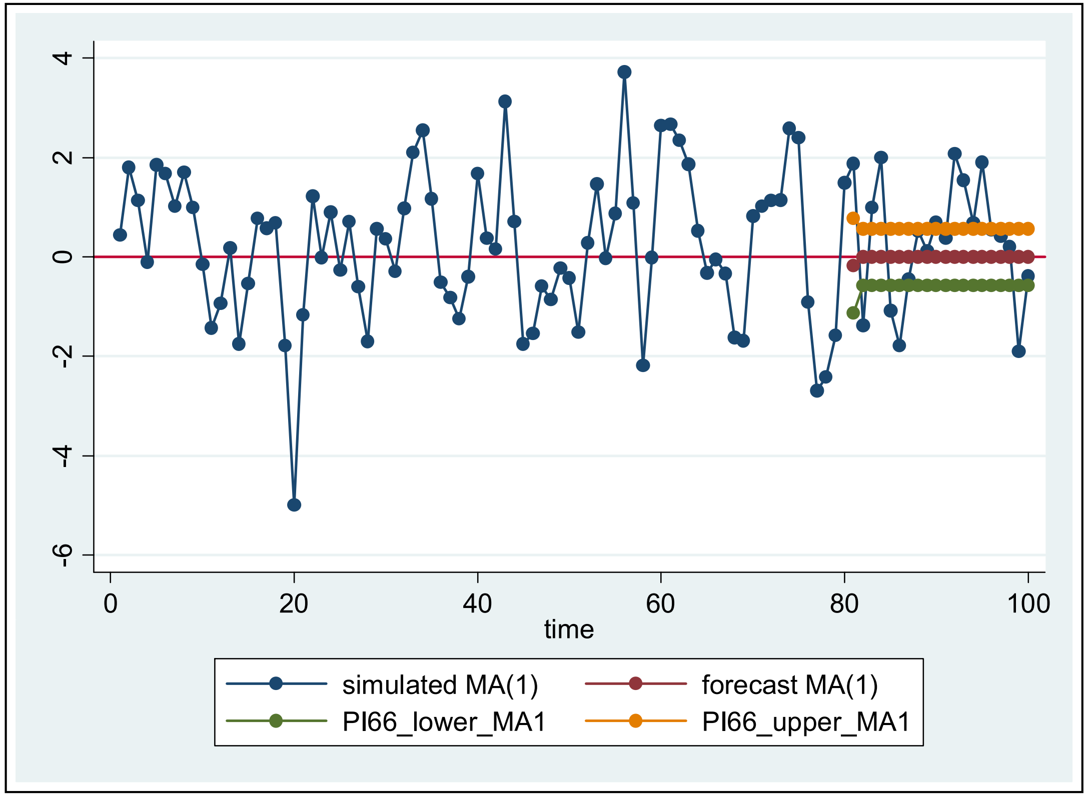
Prediction Interval 66% – AR(1)



Prediction Interval 95% – MA(1)



Prediction Interval 66% – MA(1)



Prediction with Estimated Parameters

Predictions made with θ and φ replaced by their estimates creates an additional source of variability.

Example: AR(1) φ_1 replaced by its ML estimator

$$\hat{y}_{T+l|T} = \hat{\varphi}_1^l y_T$$

$$y_{T+l} - \hat{y}_{T+l|T} = \underbrace{\left(y_{T+l} - \tilde{y}_{T+l|T}\right)}_{\text{prediction error when } \varphi \text{ is known}} + \underbrace{\left(\tilde{y}_{T+l|T} - \hat{y}_{T+l|T}\right)}_{\text{error arising from the estimation of } \varphi}$$

$$y_{T+l} - \hat{y}_{T+l|T} = \left(y_{T+l} - \tilde{y}_{T+l|T}\right) + \left(\varphi_1^l - \hat{\varphi}_1^l\right) y_T$$

Harvey (1981), Time Series Models, p.157-164

Prediction with Estimated Parameters

AR(1) φ_1 replaced by its ML estimator

$$y_{T+l} - \hat{y}_{T+l|T} = \left(y_{T+l} - \tilde{y}_{T+l|T}\right) + \left(\varphi_1^l - \hat{\varphi}_1^l\right) y_T$$

$$MSE\left(\hat{y}_{T+l|T}\right) = MSE\left(\tilde{y}_{T+l|T}\right) + y_T^2 E\left[\left(\hat{\varphi}_1 - \varphi_1\right)^2\right]$$

y_t is taken as fixed, whereas $\hat{\varphi}_1$ is a random variable

"Replacing $E\left[\left(\hat{\varphi}_1 - \varphi_1\right)^2\right]$ by its asymptotic variance gives an approximation to the mean square error, i.e."

$$MSE\left(\hat{y}_{T+l|T}\right) \cong \sigma_\varepsilon^2 + y_T^2 \frac{\left(1 - \varphi_1^2\right)}{T}$$

Harvey (1981), Time Series Models, p.157-164

Example: monthly log returns of (value-weighted) CRSP index

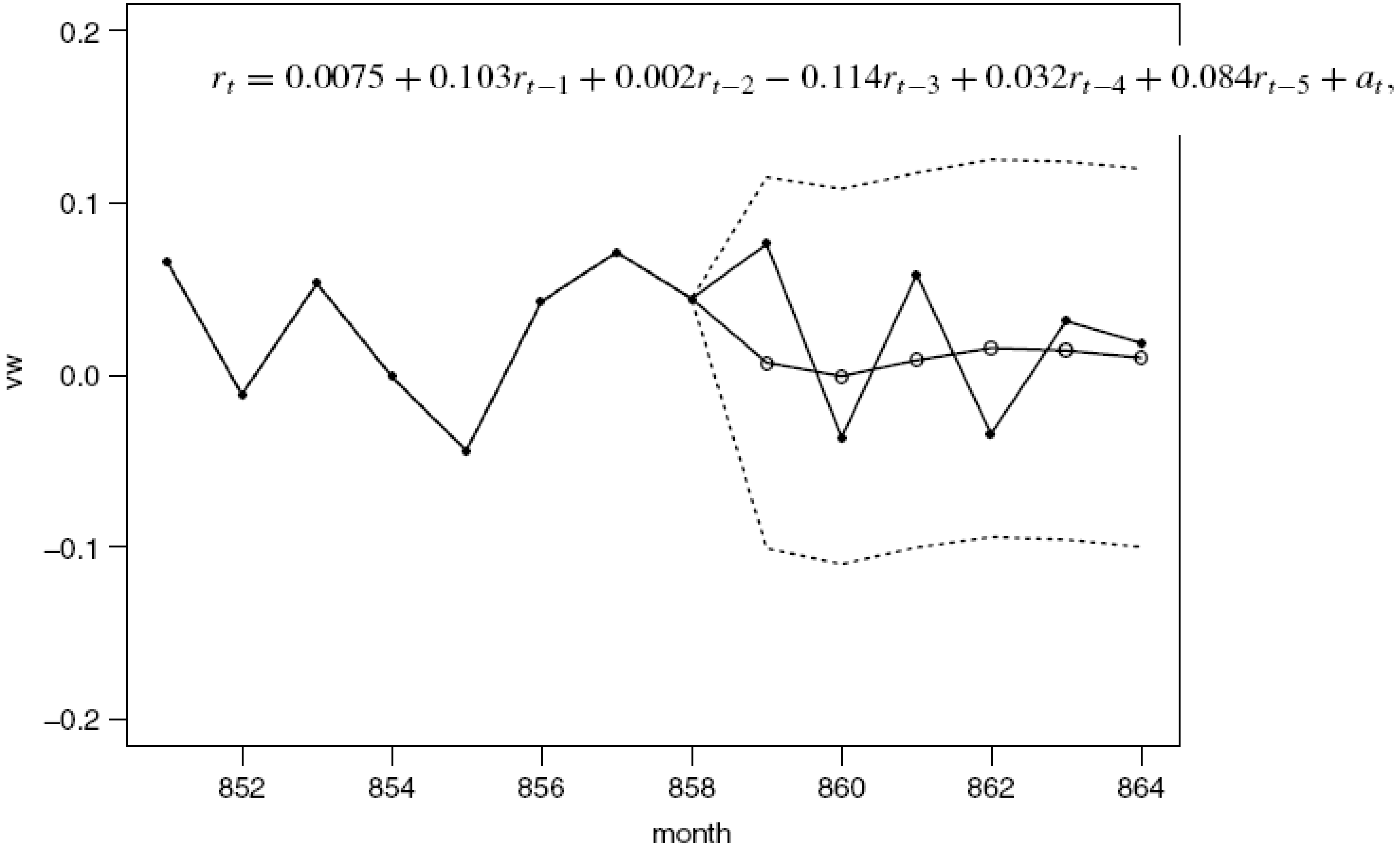


Figure 2.7. Plot of 1-step to 6-step ahead out-of-sample forecasts for the monthly log returns of the CRSP value-weighted index. The forecast origin is $t = 858$. The forecasts are denoted by ○ and the actual observations by black dots. The two dashed lines denote two standard-error limits of the forecasts.

CRSP Example: Forecasting

$$r_t = 0.0102 + 0.103 r_{t-1} - 0.122 r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0542,$$

Table 2.2. Multistep Ahead Forecasts of an AR(3) Model for the Monthly Simple Returns of CRSP Value-Weighted Index with Forecast Origin of 858

Step	1	2	3	4	5	6
Forecast	0.0088	0.0020	0.0050	0.0097	0.0109	0.0106
Standard error	0.0542	0.0546	0.0546	0.0550	0.0550	0.0550
Actual	0.0762	−0.0365	0.0580	−0.0341	0.0311	0.0183

Table 2.2 contains the 1-step to 6-step ahead forecasts and the standard errors of the associated forecast errors at the forecast origin 858 for the monthly simple return of the value-weighted index using an AR(3) model that was reestimated using the first 858 observations.

Tsay (2005) “Analysis of Financial Time Series ”, p. 18

CRSP Example: Forecasting

$$r_t = 0.0102 + 0.103 r_{t-1} - 0.122 r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0542,$$

Table 2.2. Multistep Ahead Forecasts of an AR(3) Model for the Monthly Simple Returns of CRSP Value-Weighted Index with Forecast Origin of 858

Step	1	2	3	4	5	6
Forecast	0.0088	0.0020	0.0050	0.0097	0.0109	0.0106
Standard error	0.0542	0.0546	0.0546	0.0550	0.0550	0.0550
Actual	0.0762	−0.0365	0.0580	−0.0341	0.0311	0.0183

The actual returns are also given. Because of the weak serial dependence in the series, the forecasts and standard deviations of forecast errors converge to the sample mean and standard deviation of the data quickly. For the first 858 observations, the sample mean and standard error are 0.0098 and 0.0550, respectively.