Chapter 6

Univariate time series modeling and forecasting

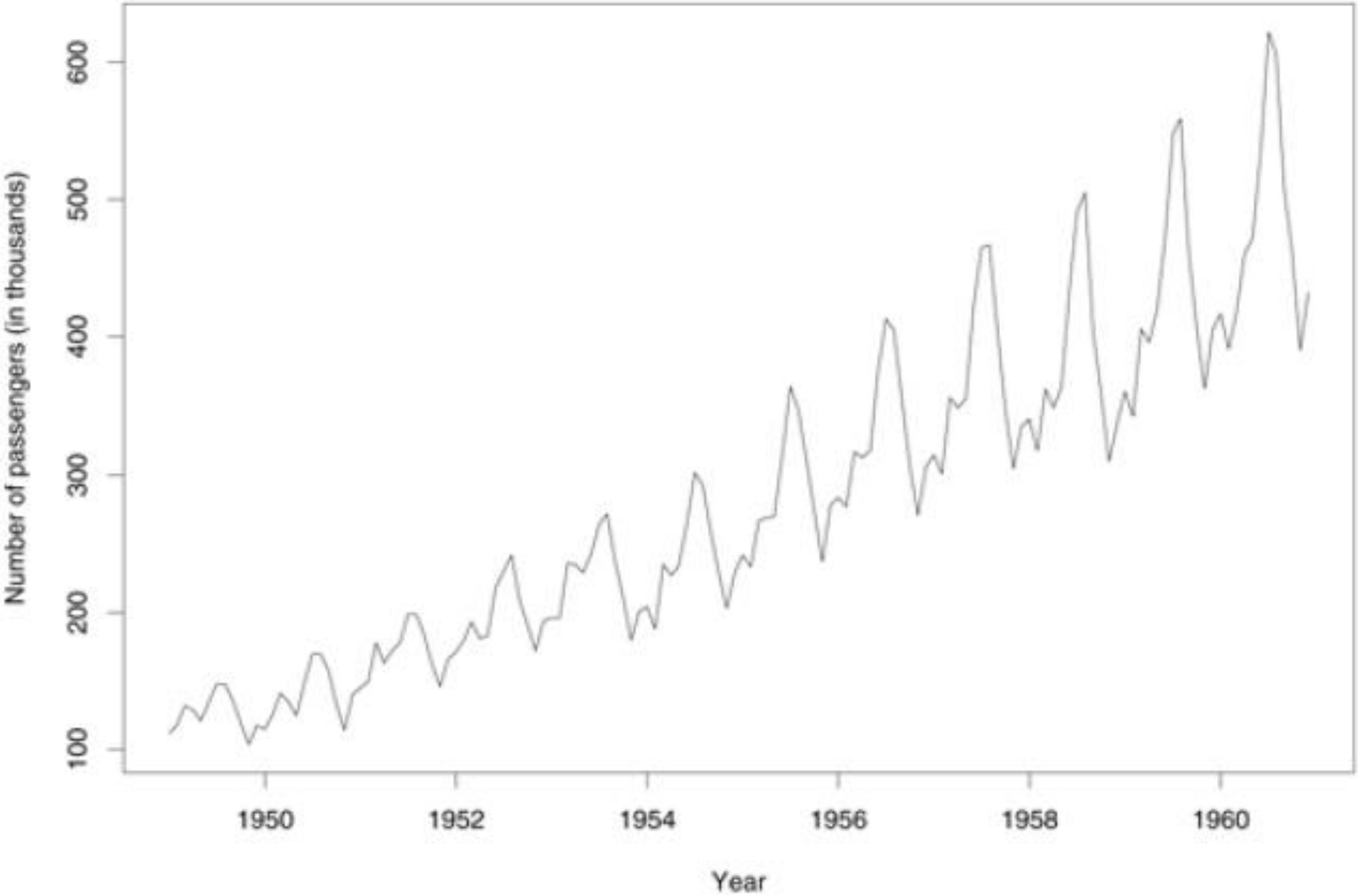


Figure 2.1. *The time plot of the Box-Jenkins airline data: Monthly totals, in thousands, of the numbers of international airline passengers from January 1949 to December 1960.*

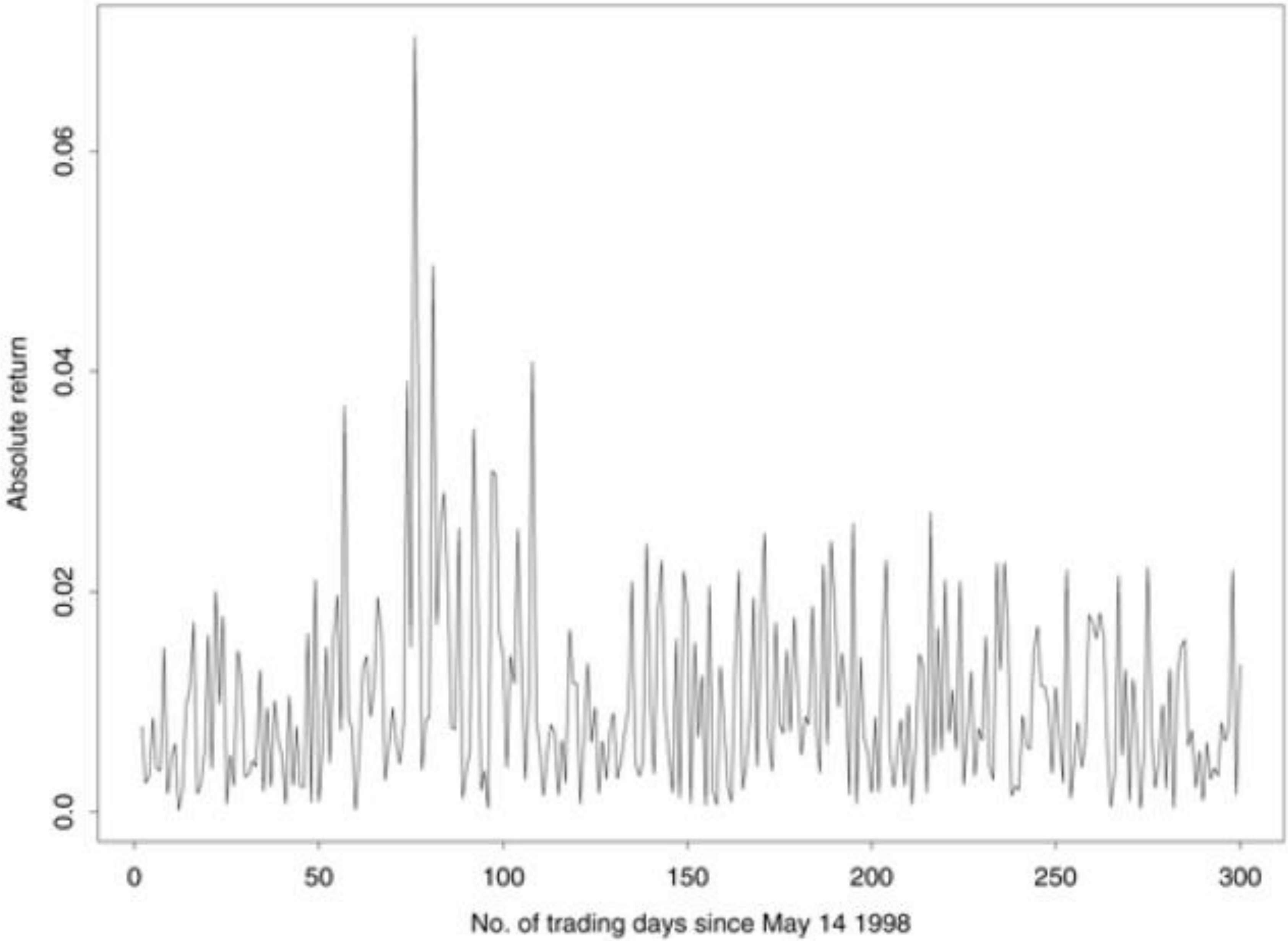


Figure 3.1. *A time plot of the absolute values of daily returns for the S&P 500 index on 300 trading days starting on May 14 1998.*

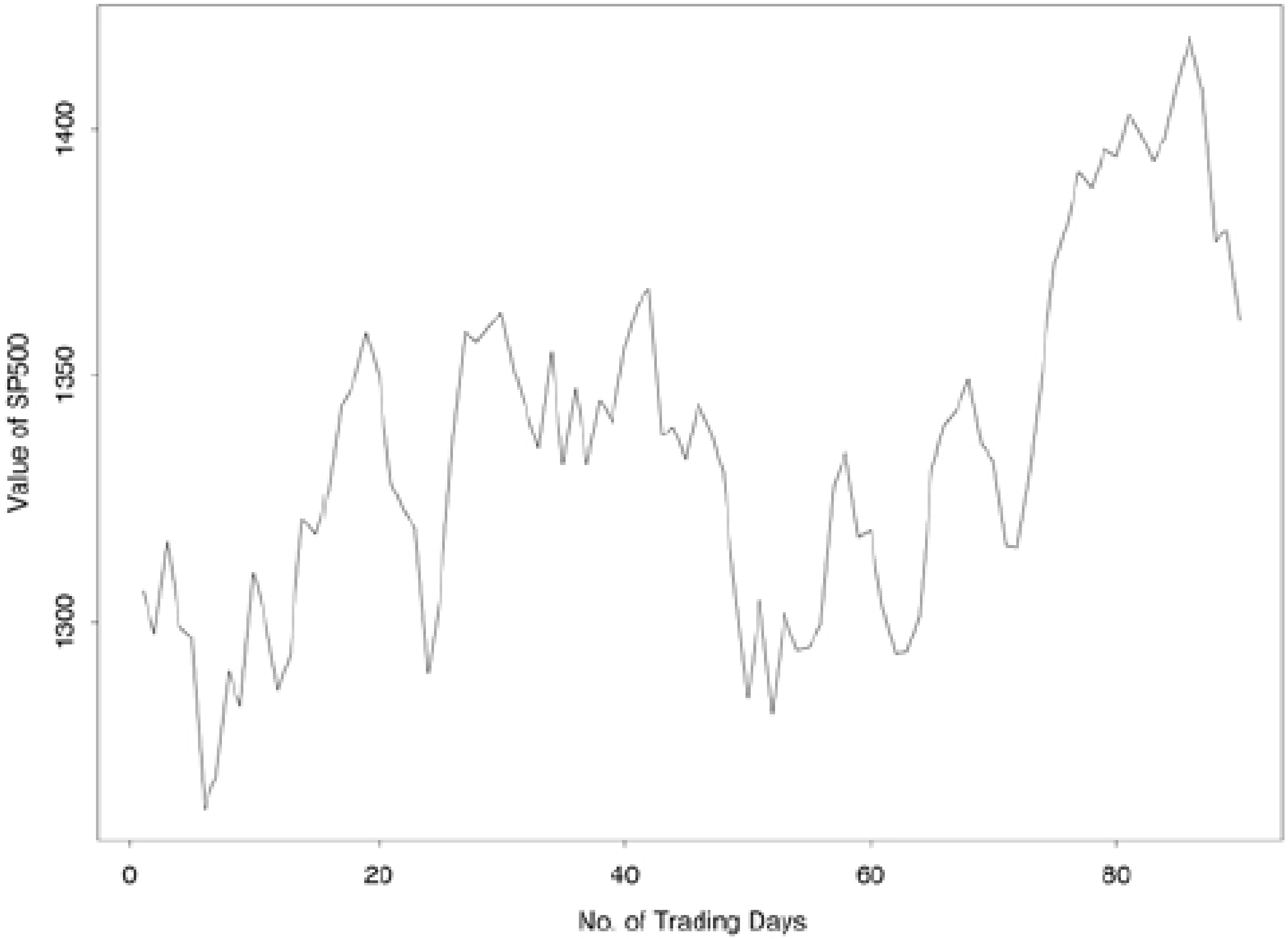


Figure 1.1. *A graph showing the Standard & Poor (S & P) 500 index for the U.S. stock market for 90 trading days starting on March 16 1999.*

I. Time Series: Introduction

Time Series: collection of observations made sequentially through time

Objectives of Time Series Analysis:

-Description: using summary statistics and/or graphical methods (Time Plot)

-Modeling: to find suitable statistical model to describe the Data-Generating-Process.

> ***-Univariate Model*** of a variable is based only on past values of that variable
>
> ***-Structural Model***: based on present and past values of other variables

-Forecasting (Predicting): to estimate future values of the series.

-Control: good forecasts enable the analyst to take action so as to control a given process (economic or industrial)

Time Series and Forecasting

“*Don’t never prophesy: if you prophesy right, nobody is going to remember and if you prophesy wrong, nobody is going to let you forget*”

Mark Twain

Applications of Time Series forecasting include:

- Economic Planning
- Sales forecasting
- Inventory control
- Production and Capacity Planning
- Budgeting
- Financial Risk Management

II. Univariate Time Series Models: Notations and Concepts

- Where we attempt to predict returns using only information contained in their past values.

- A **Strictly Stationary Process**

A strictly stationary process is one where

$$P\{y_{t_1} \le b_1, \ldots, y_{t_n} \le b_n\} = P\{y_{t_1+m} \le b_1, \ldots, y_{t_n+m} \le b_n\}$$

i.e. the probability measure for the sequence $\{y_t\}$ is the same as that for $\{y_{t+m}\}$ $\forall$ m.

- A **Weakly Stationary Process**

If a series satisfies the next three equations, it is said to be weakly or covariance stationary

1. $E(y_t) = \mu$, $\quad t = 1,2,\ldots,\infty$
2. $E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty$
3. $E(y_{t_1} - \mu)(y_{t_2} - \mu) = \gamma_{t_2 - t_1}$ $\quad \forall\ t_1, t_2$

Univariate Time Series Models (cont'd)

- So if the process is covariance stationary, all the variances are the same and all the covariances depend on the difference between t_1 and t_2. The moments

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

are known as **autocovariances**.

- However, the value of the autocovariances depend on the units of measurement of y_t.
- It is thus more convenient to use the **autocorrelations** which are the autocovariances normalised by dividing by the variance:

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

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

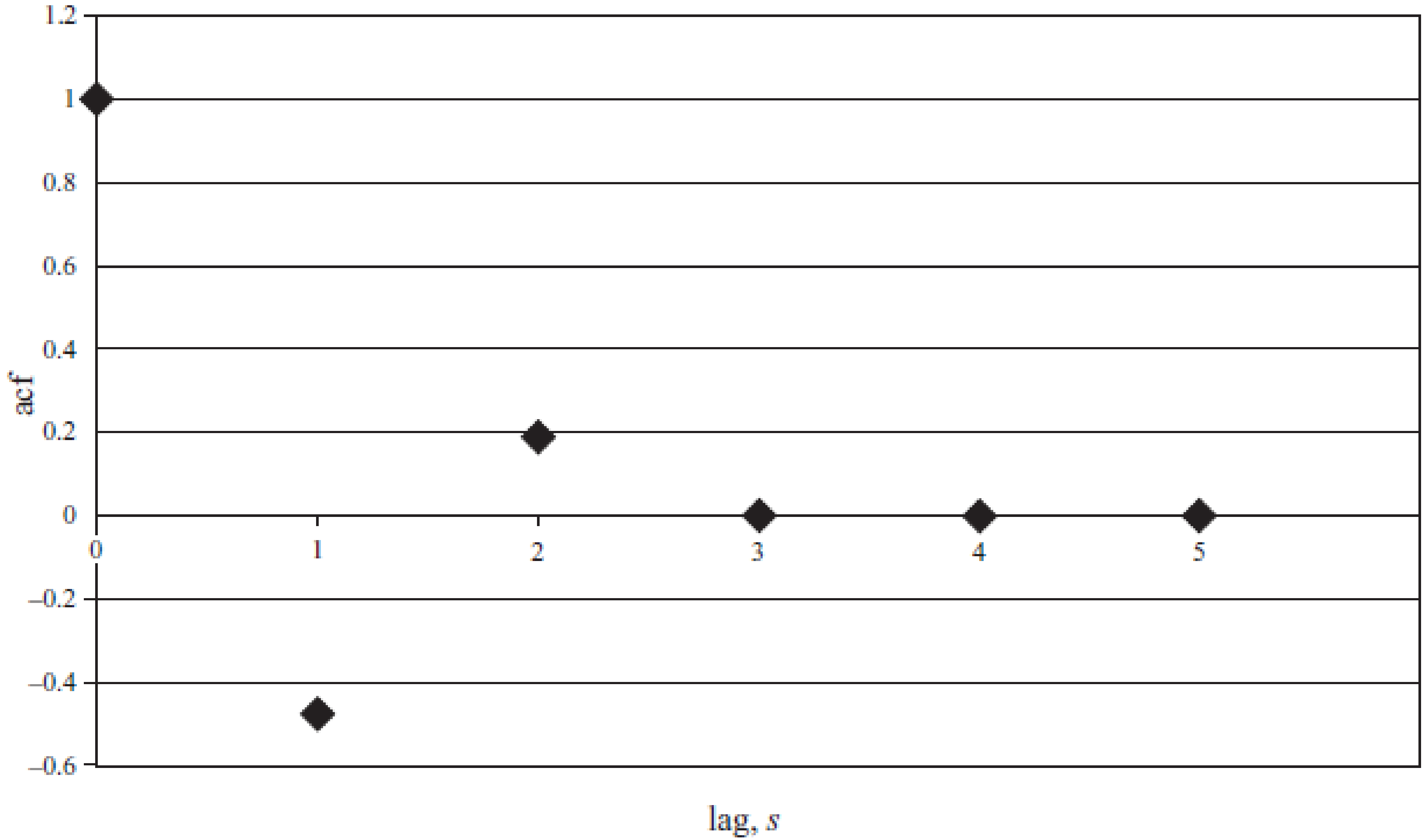


Figure 5.1 Autocorrelation function for sample MA(2) process

A White Noise Process (purely randomly process)

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

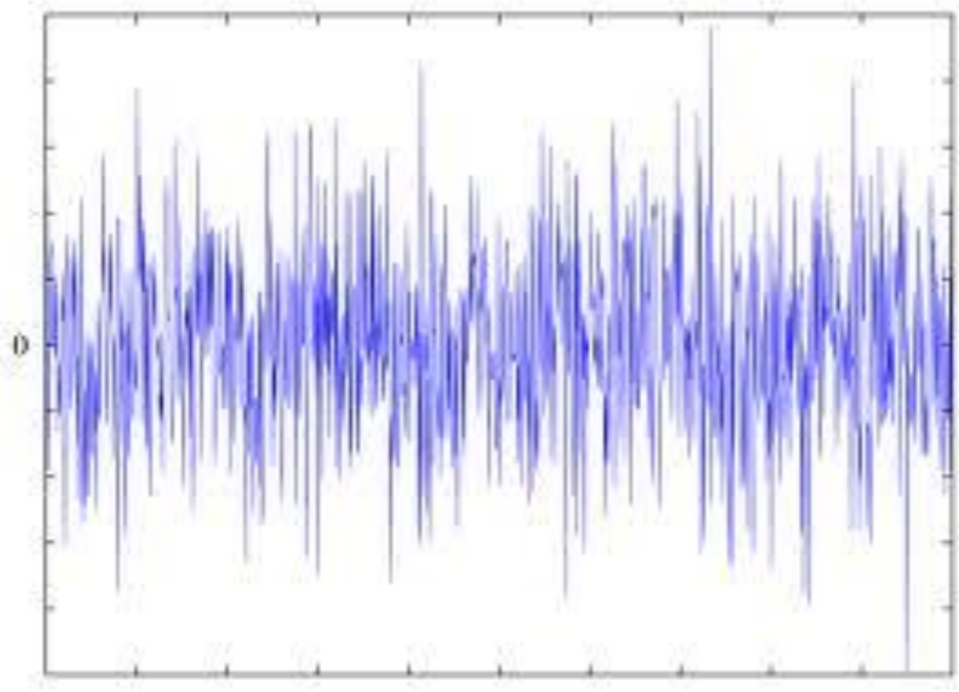
$$E(u_t) = \mu$$
$$Var(u_t) = \sigma^2$$
$$\gamma_s = \begin{cases} \sigma^2 & \text{if } s = 0 \\ 0 & \text{otherwise} \end{cases}$$

Autocorrelation function will be zero apart from a single peak of 1 at $s = 0$.

- If u_t is further normally distributed then $\hat{\tau}_s$ **~ approximately N(0,1/*T*)** where T = sample size and $\hat{\tau}_s$ denotes the autocorrelation at lag s

- We can use this to do significance tests for the autocorrelation coefficients by constructing a confidence interval.

White Noise Process: Example



Single Tests: τ_s =0

- $\hat{\tau}_s$ **~ approximately N(0,1/*T*)** where T = sample size and $\hat{\tau}_s$ denotes the autocorrelation at lag s
- We can use this to do significance tests for the autocorrelation coefficients by constructing a confidence interval.

- **H$_0$: true value of the autocorrelation coefficient at lag *s* is zero: τ_s =0**

- For example, a 95% confidence interval would be given by $\pm .196 \times \frac{1}{\sqrt{T}}$

- If the sample autocorrelation coefficient, $\hat{\tau}_s$, falls outside this region for any value of *s*, then we reject the null hypothesis that the true value of the autocorrelation coefficient at lag *s* is zero.

Joint Hypothesis Tests: $\tau_1 = 0$, $\tau_2 = 0, \ldots, \tau_m = 0$
(Data are independently distributed)

- We can also **test the joint hypothesis that all *m* of the τ_k autocorrelation coefficients are simultaneously equal to zero** using the ***Q*-statistic** developed by Box and Pierce:

$$Q = T\sum_{k=1}^{m}\tau_k^2$$

where T = sample size, m = maximum lag length
- The Q-statistic is asymptotically distributed as a χ_m^2.

- However, the Box Pierce test has poor small sample properties, so a variant has been developed, called the **Ljung-Box statistic**:

$$Q^* = T(T+2)\sum_{k=1}^{m}\frac{\tau_k^2}{T-k} \sim \chi_m^2$$

- This statistic is very useful as a **portmanteau (general) test of linear dependence** in time series.

An ACF Example

- **Question:**

  Suppose that a researcher had estimated the first 5 autocorrelation coefficients using a series of length 100 observations, and found them to be (from 1 to 5): 0.207, -0.013, 0.086, 0.005, -0.022.

  a) Test each of the individual coefficient for significance

  b) Use both the Box-Pierce and Ljung-Box tests to establish whether they are jointly significant.

- **Solution:**

  a) A coefficient would be significant if it lies outside (-0.196,+0.196) at the 5% level, so **only the first autocorrelation coefficient is significant**.

  b) Q=5.09 and Q^*=5.26

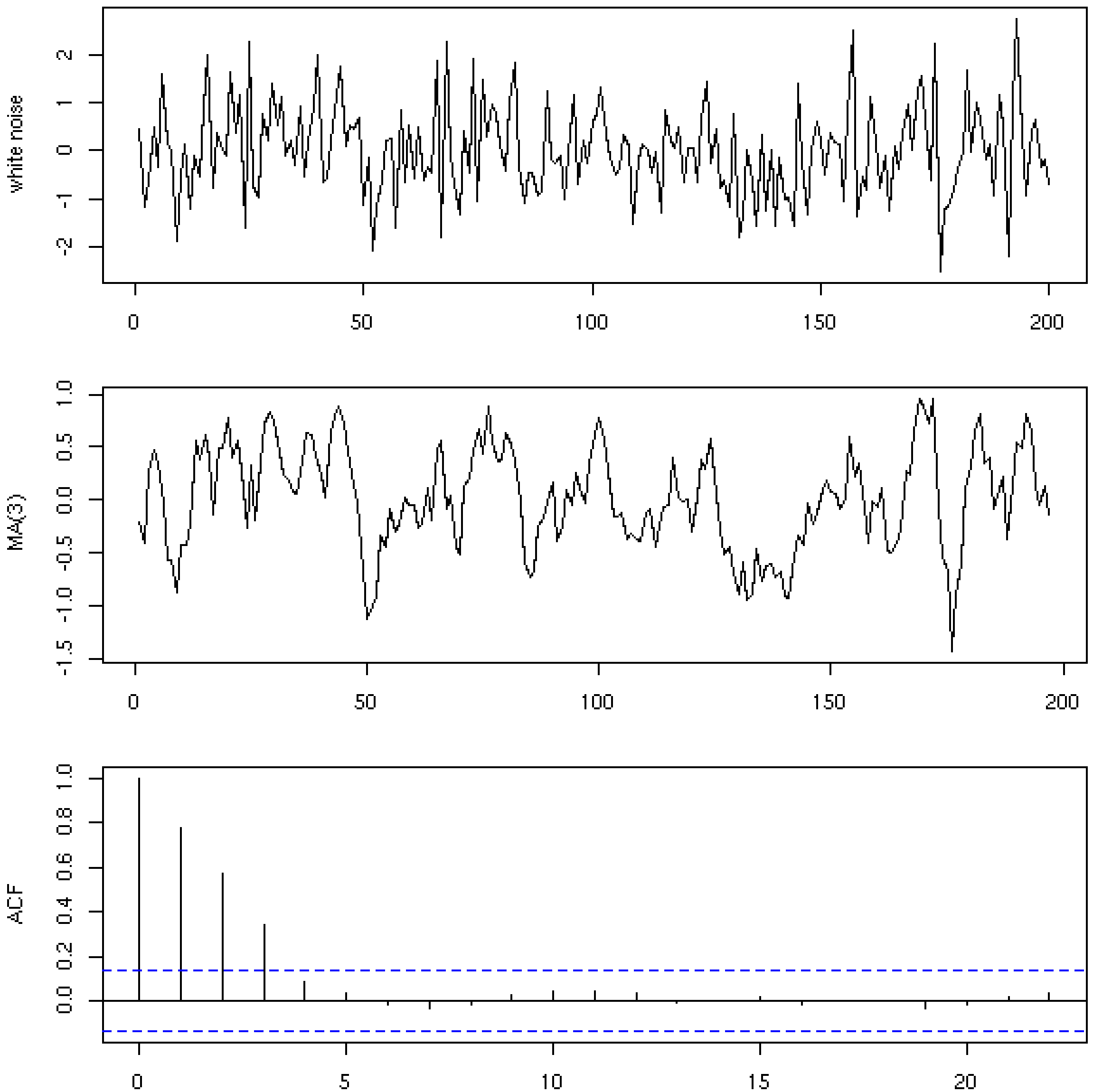
  Compared with a tabulated $\chi^2(5)$=11.07 at the 5% level, so the 5 coefficients are jointly insignificant (**low power of the joint test when 1 single coefficient is significant**).

III. Moving Average Processes

- Let u_t (t=1,2,3,...) be a white noise process with:
  E(u_t)=0 and $Var(u_t) = \sigma^2$, then
  $$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q}$$
  is a **q^{th} order moving average model MA(q).**

- Its properties are:
  E(y_t)=μ;
  Var(y_t) = $\gamma_0 = (1+ \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma^2$
  Covariances
  $$\gamma_s = \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \ldots + \theta_q\theta_{q-s})\sigma^2 & for \quad s = 1,2,\ldots,q \\ 0 \quad for \quad s > q \end{cases}$$

MA(q) is stationary



white noise
MA(3)
ACF

Example of an MA Problem

1. Consider the following **MA(2)** process:

where ε_t is a zero mean white noise process with variance .

(i) Calculate the **mean** and **variance** of X_t

(ii) Derive the **autocorrelation** function for this process (i.e. express the autocorrelations, τ_1, τ_2, ... as functions of the parameters θ_1 and θ_2).

(iii) If θ_1 = -0.5 and θ_2 = 0.25, sketch the ACF of X_t.

Solution

(i) If $E(u_t)=0$, then $E(u_{t-i})=0 \;\forall\; i$.

So

$$E(X_t) = E(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}) = E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2}) = 0$$

$$
\begin{aligned}
\text{Var}(X_t) &= E[X_t - E(X_t)][X_t - E(X_t)] \\
\text{but } E(X_t) &= 0, \text{ so} \\
\text{Var}(X_t) &= E[(X_t)(X_t)] \\
&= E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})] \\
&= E[\, u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2 \text{ +cross-products}]
\end{aligned}
$$

But E[cross-products]=0 since $\text{Cov}(u_t, u_{t-s})=0$ for $s \neq 0$.

Solution (cont'd)

So Var(X_t) = γ_0= E [ ]

=

= $(1+\theta_1^2+\theta_2^2)\sigma^2$

(ii) The ACF of X_t.

γ_1 = E[X_t-E(X_t)][X_{t-1}-E(X_{t-1})]

= E[X_t][X_{t-1}]

= E[($u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$)($u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3}$)]

= E[($\theta_1 u_{t-1}^2 + \theta_1\theta_2 u_{t-2}^2$)]

= $\theta_1\sigma^2 + \theta_1\theta_2\sigma^2$

= $(\theta_1 + \theta_1\theta_2)\sigma^2$

Solution (cont'd)

$$
\begin{aligned}
\gamma_2 \quad &= E[X_t - E(X_t)][X_{t-2} - E(X_{t-2})] \\
&= E[X_t][X_{t-2}] \\
&= E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-2} + \theta_1 u_{t-3} + \theta_2 u_{t-4})] \\
&= E[(\theta_2 u_{t-2}^2)] \\
&= \theta_2 \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
\gamma_3 \quad &= E[X_t - E(X_t)][X_{t-3} - E(X_{t-3})] \\
&= E[X_t][X_{t-3}] \\
&= E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-3} + \theta_1 u_{t-4} + \theta_2 u_{t-5})] \\
&= 0
\end{aligned}
$$

So $\gamma_s = 0$ for $s > 2$.

Solution (cont'd)

We have the autocovariances, now calculate the **autocorrelations**:

$$\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\theta_1 + \theta_1\theta_2)\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{(\theta_1 + \theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)}$$

$$\tau_2 = \frac{\gamma_2}{\gamma_0} = \frac{(\theta_2)\sigma^2}{(1+\theta_1^2+\theta_2^2)\sigma^2} = \frac{\theta_2}{(1+\theta_1^2+\theta_2^2)}$$

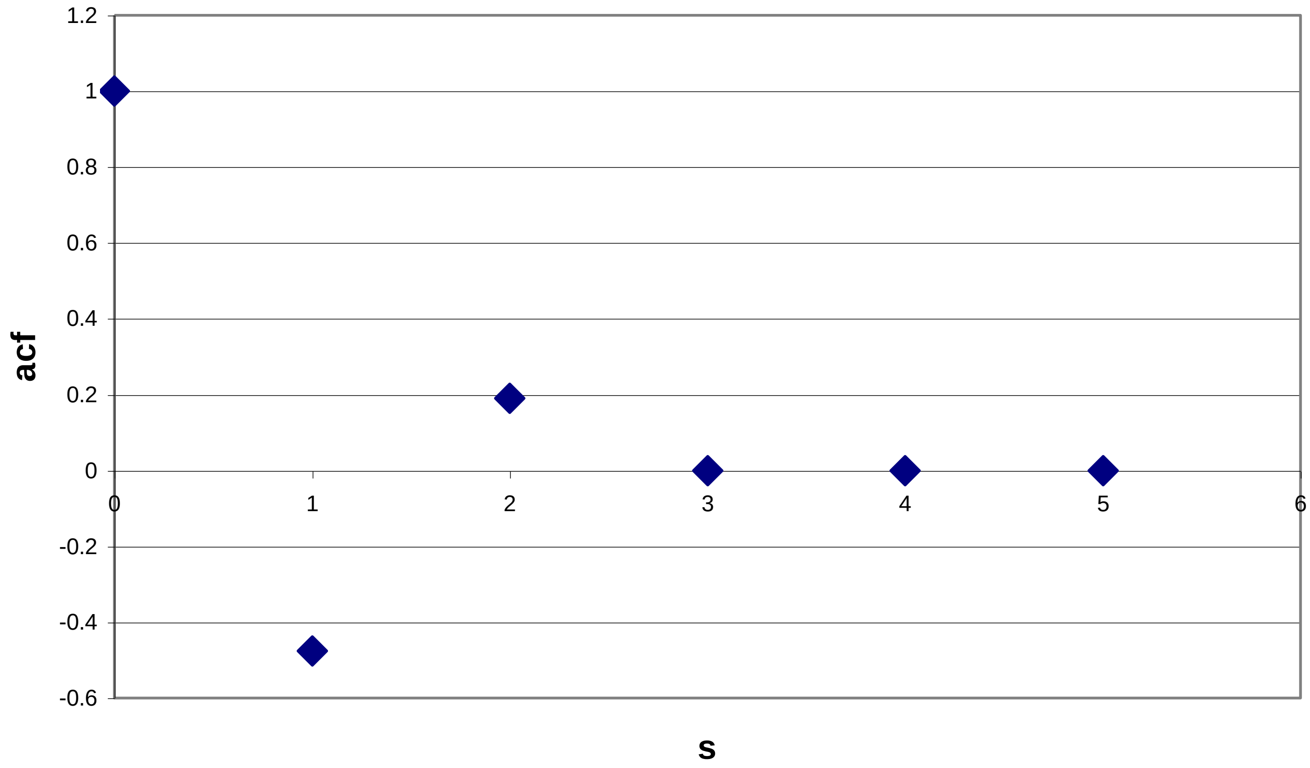
$$\tau_3 = \frac{\gamma_3}{\gamma_0} = 0$$

$$\tau_s = \frac{\gamma_s}{\gamma_0} = 0 \;\forall\; s > 2$$

(iii) For θ_1 = -0.5 and θ_2 = 0.25, substituting these into the formulae above gives τ_1 = -0.476, τ_2 = 0.190.

ACF Plot

Thus the ACF (autocorrelation) plot will appear as follows:



IV. Autoregressive Processes

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

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

where u_t is a white noise process with zero mean.

- Or using the **lag operator notation**:

$Ly_t = y_{t-1}$ $\qquad\qquad$ $L^i y_t = y_{t-i}$

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

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

or $\phi(L) y_t = \mu + u_t$ $\qquad$ where $\qquad$ $\phi(L) = 1 - (\phi_1 L + \phi_2 L^2 + \ldots \phi_p L^p)$.

φ = 0.3
φ = 0.9
φ1 = 0.3
φ2 = 0.3
φ1 = 0.9
φ2 = -0.8
AR(0)
AR(1)
AR(2)

The Stationary Condition for an AR Model

- **The condition for stationarity of a general AR(p) model is that the roots of** $1-\phi_1 z-\phi_2 z^2-\ldots-\phi_p z^p=0$ **all lie outside the unit circle i.e. |z|>1**

- A stationary AR(p) model is required for it to have an MA(∞) representation.

- **Example 1 (Random walk):** Is $y_t = y_{t-1} + u_t$ stationary?
  The characteristic root is 1, so it is a unit root process (so non-stationary)

- **Example 2:** Is $y_t = 2y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + u_t$ stationary?

$$y_t = 3y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + u_t \quad (5.63)$$

Again, the first stage is to express this equation using the lag operator notation, and then taking all the terms in y over to the LHS

$$y_t = 3Ly_t - 2.75L^2y_t + 0.75L^3y_t + u_t \quad (5.64)$$

$$(1 - 3L + 2.75L^2 - 0.75L^3)y_t = u_t \quad (5.65)$$

The characteristic equation is

$$1 - 3z + 2.75z^2 - 0.75z^3 = 0 \quad (5.66)$$

which fortunately factorises to

$$(1 - z)(1 - 1.5z)(1 - 0.5z) = 0 \quad (5.67)$$

so that the roots are $z = 1$, $z = 2/3$, and $z = 2$. Only one of these lies outside the unit circle and hence the process for y_t described by (5.63) is not stationary.

Wold's Decomposition Theorem

- **An AR(p) stationary series with no constant and no other terms can be expressed as MA(∞)**

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

$$y_t = \psi(L)u_t$$

where,

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1}$$

Characteristics of an Autoregressive Process

- The moments of an autoregressive process are as follows. The **mean** is given by

$$E(y_t) = \frac{\mu}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$$

- The **autocovariances** and **autocorrelation functions** can be obtained by solving what are known as the **Yule-Walker equations**:

$$\begin{aligned}
\tau_1 &= \phi_1 + \tau_1\phi_2 + \ldots + \tau_{p-1}\phi_p \\
\tau_2 &= \tau_1\phi_1 + \phi_2 + \ldots + \tau_{p-2}\phi_p \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\tau_p &= \tau_{p-1}\phi_1 + \tau_{p-2}\phi_2 + \ldots + \phi_p
\end{aligned}$$

- **If the AR model is stationary, the ACF will decay geometrically to zero.**

Example 5.4

Consider the following simple AR(1) model

$$y_t = \mu + \phi_1 y_{t-1} + u_t$$

(i) Given that the series is stationary, calculate the (unconditional) mean of y_t

For the remainder of the question, set μ=0 for simplicity.

(ii) Calculate the (unconditional) variance of y_t.

(iii) Derive the autocorrelation function for y_t.

V. The Partial Autocorrelation Function (denoted τ_{kk})

- Measures the correlation between an observation k periods ago and the current observation, after controlling for observations at intermediate lags (i.e. all lags < k).

- So τ_{kk} measures the correlation between y_t and y_{t-k} after removing the effects of y_{t-k+1} , y_{t-k+2} , …, y_{t-1} .

- At lag 1, the ACF = PACF: $\tau_{11} = \tau_1$

- At lag 2, $\tau_{22} = (\tau_2 - \tau_1^2) / (1 - \tau_1^2)$

- For lags 3+, the formulae are more complex.

The Partial Autocorrelation Function (denoted τ_{kk}) (cont'd)

- **The PACF is useful for telling the difference between an AR process and an ARMA process**.
- In the case of an **AR(*p*),** there are direct connections between y_t and y_{t-s} only for $s \leq p$.
- So for an AR(p), the theoretical **PACF will be zero after lag *p***.
- In the case of an **MA(*q*),** if it is invertible (*roots of characteristic equations* $\theta(z)=0$ *lie outside the unit circle*), it can be written as an AR(∞), so there are direct connections between y_t and **all** its previous values.
- For an MA(q), the theoretical **PACF will be geometrically declining**.

VI. ARMA Processes

- By combining the AR(p) and MA(q) models, we can obtain an **ARMA(p,q)** model: $\phi(L)y_t = \mu + \theta(L)u_t$

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

and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$

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

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

The Invertibility Condition

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

- The **mean** of an ARMA series is given by

$$E(y_t) = \frac{\mu}{1-\phi_1-\phi_2-...-\phi_p}$$

- **Remark**: **The ACF for an ARMA process will display combinations of behaviour derived from the AR and MA parts**

Summary of the Behaviour of the ACF for AR and MA Processes

An **autoregressive process AR(p)** has

- a geometrically decaying ACF
- number of non-zero points of PACF = AR order =p

A **moving average process MA(q)** has

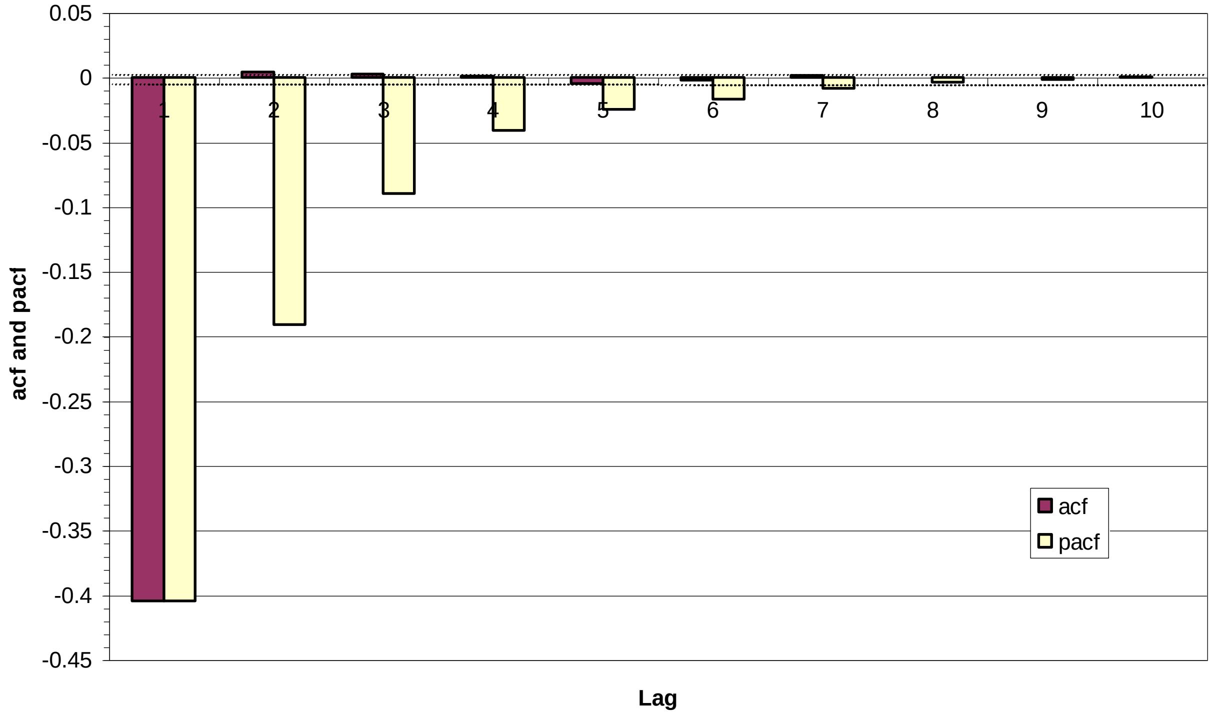
- Number of non-zero points of ACF = MA order =q
- a geometrically decaying PACF

ARMA models have both ACF and PACF geometrically decaying to 0.

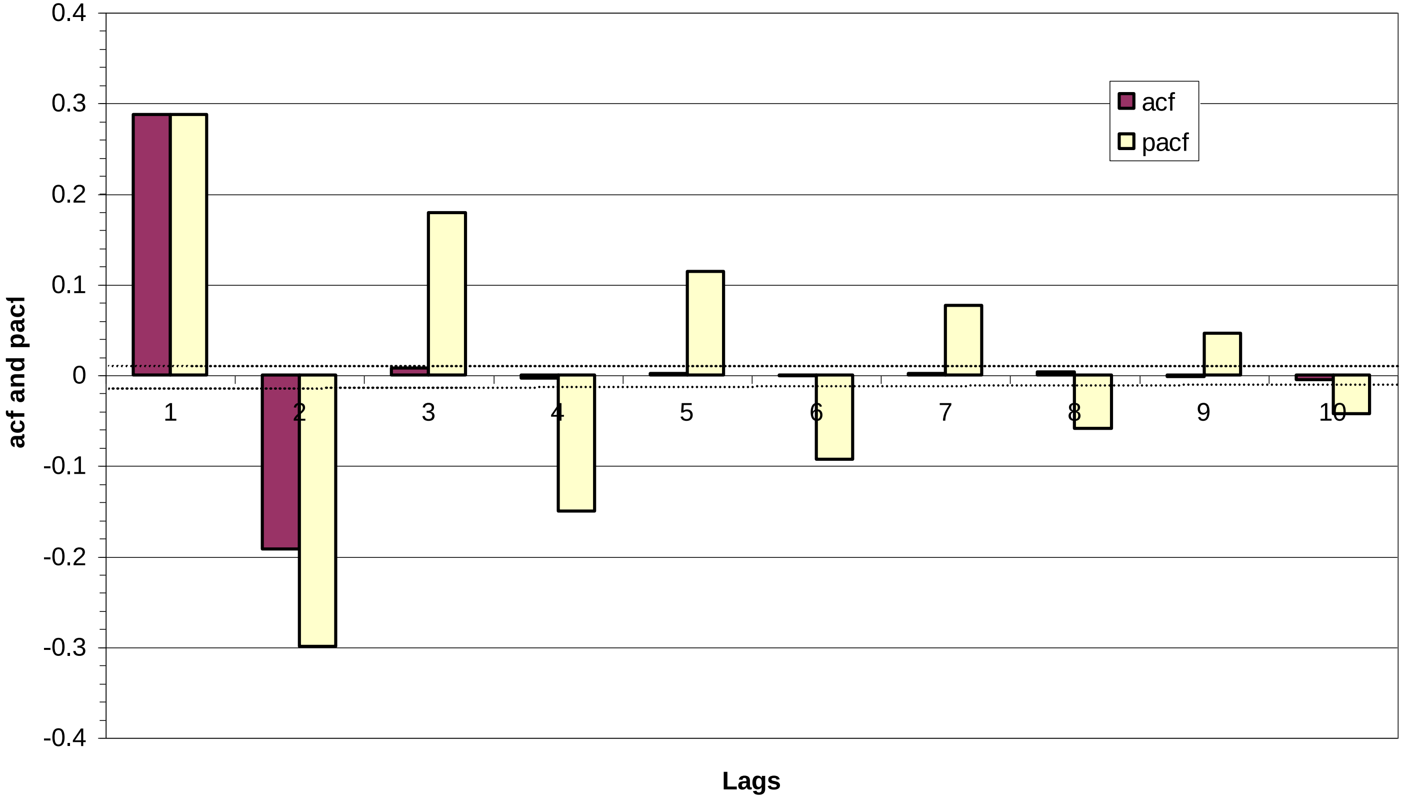
Some sample acf and pacf plots for standard processes

The ACF and PACF are not produced analytically from the relevant formulae for a model of that type, but rather are estimated using 100,000 simulated observations with disturbances drawn from a normal distribution.

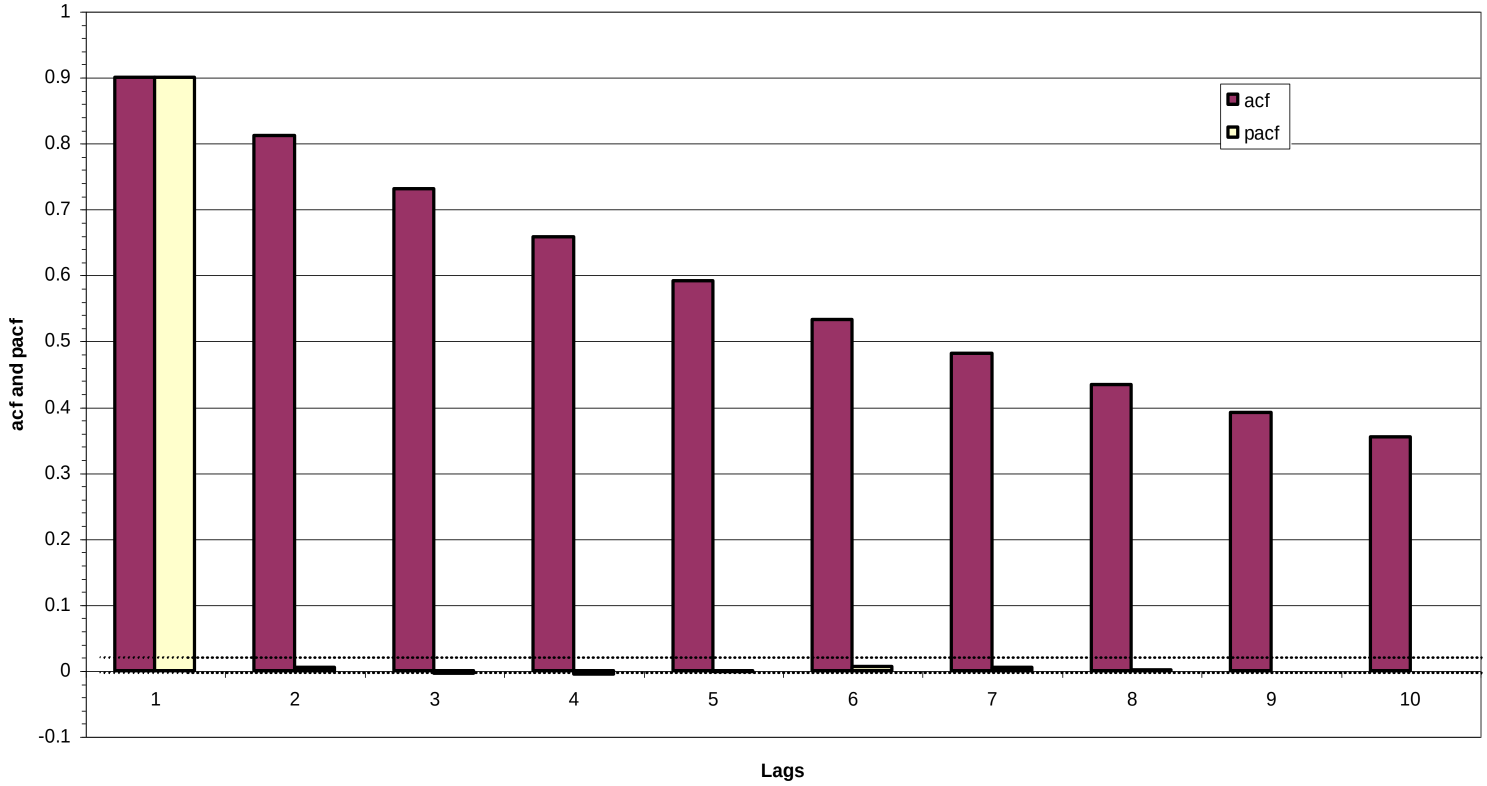
ACF and PACF for an MA(1) Model: $y_t = -0.5u_{t-1} + u_t$



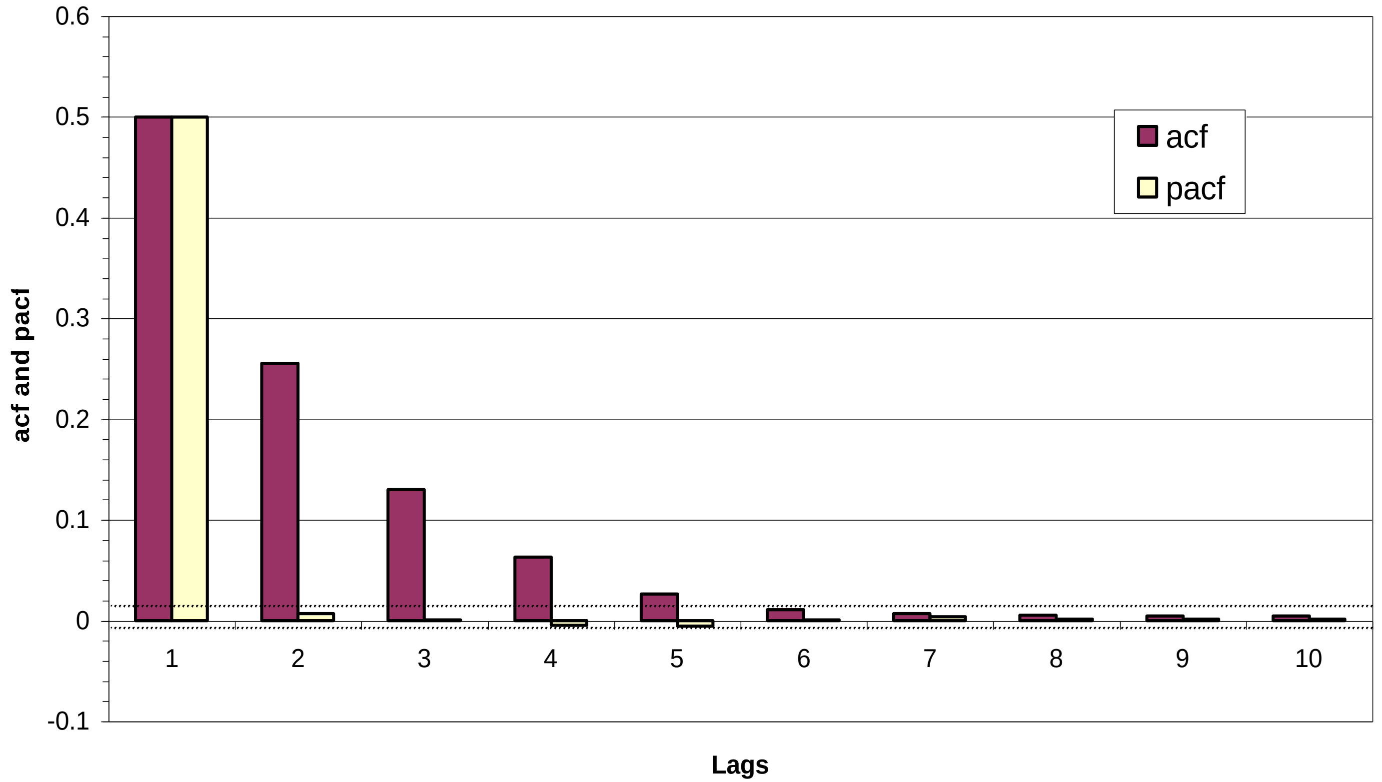
ACF and PACF for an MA(2) Model: $y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$



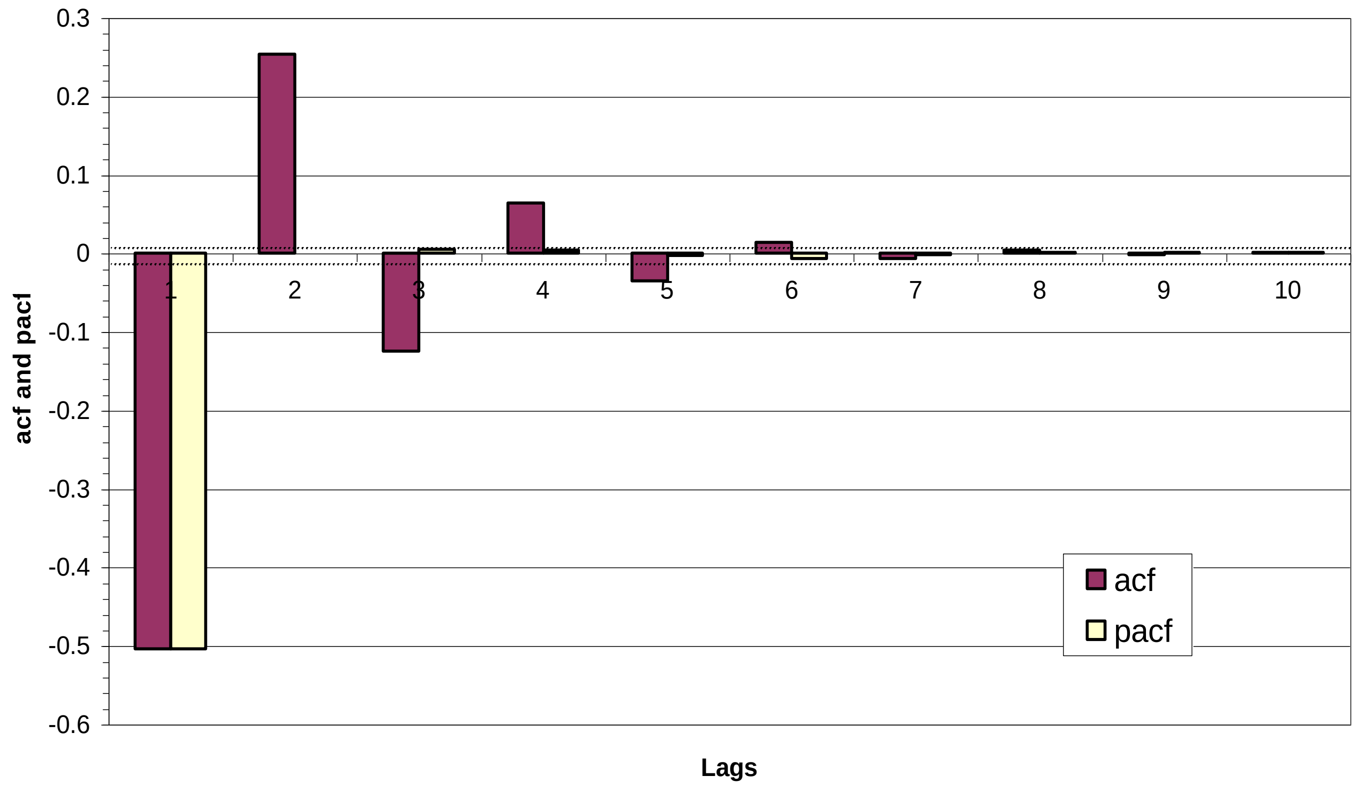
ACF and PACF for a slowly decaying AR(1) Model: $y_t = 0.9y_{t-1} + u_t$



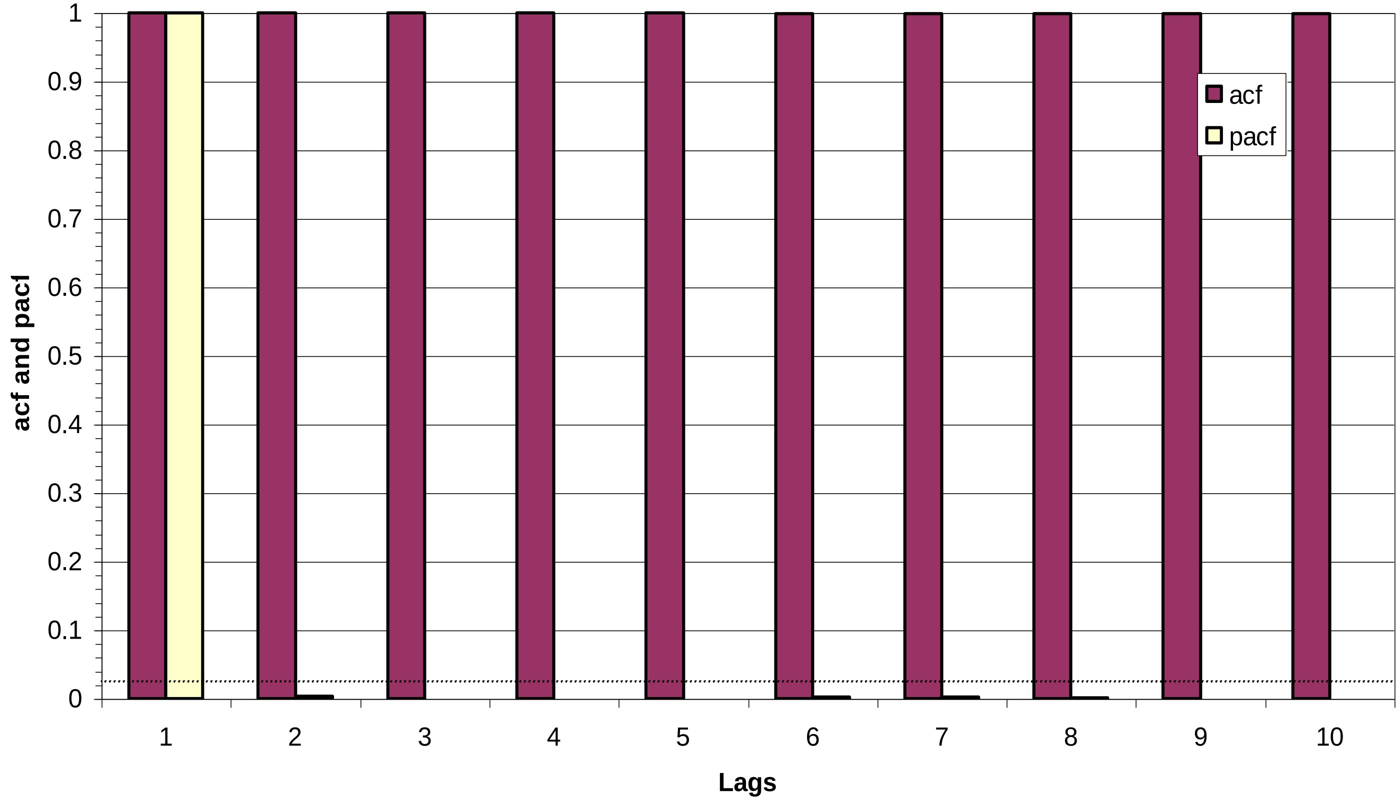
ACF and PACF for a more rapidly decaying AR(1) Model: $y_t = 0.5y_{t-1} + u_t$



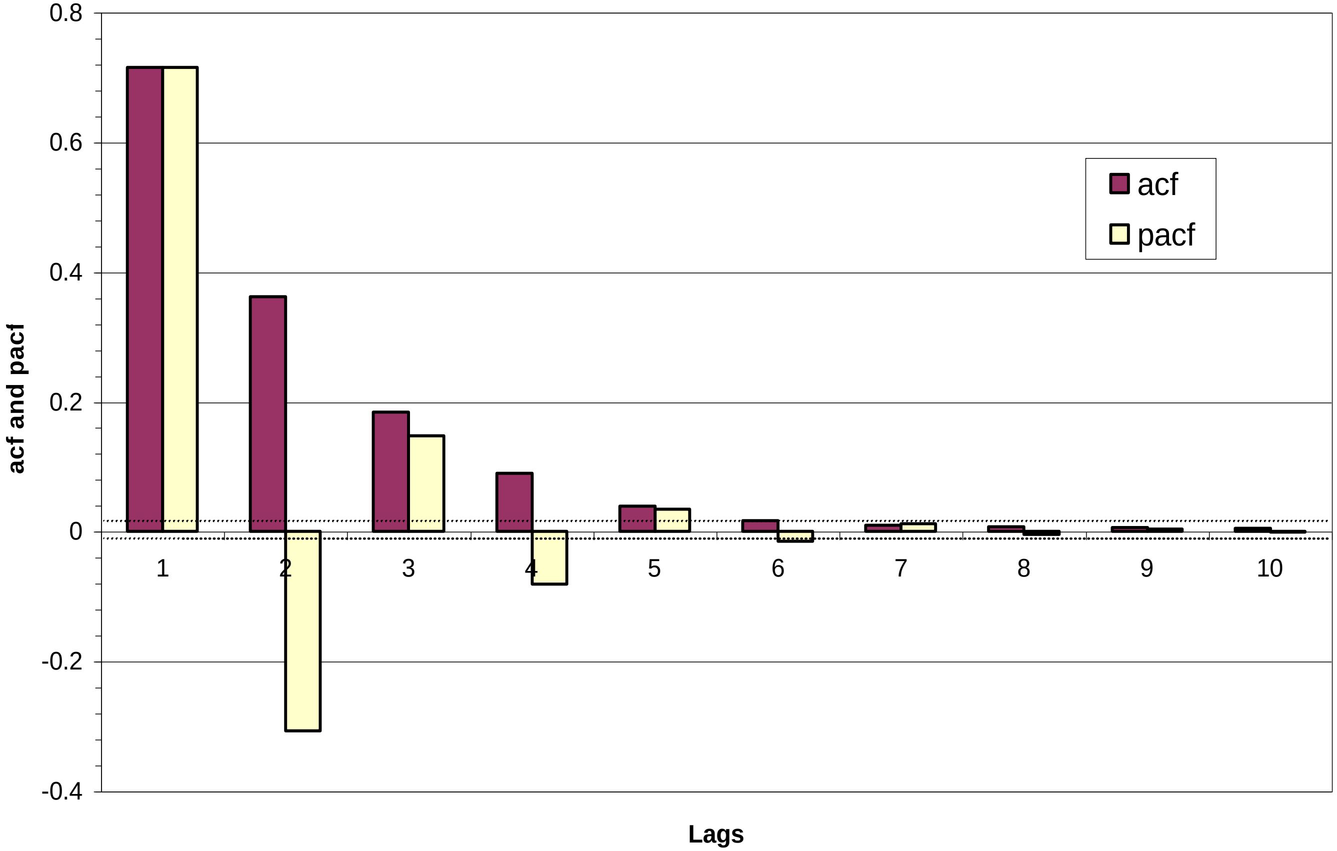
ACF and PACF for a more rapidly decaying AR(1) Model with Negative Coefficient: $y_t = -0.5y_{t-1} + u_t$



ACF and PACF for a Non-stationary Model (i.e. a unit coefficient): $y_t = y_{t-1} + u_t$



ACF and PACF for an ARMA(1,1): $y_t = 0.5y_{t-1} + 0.5u_{t-1} + u_t$



VII. Building ARMA Models
- The Box Jenkins Approach

- **Box and Jenkins** (1970) were the first to approach the task of *estimating an ARMA model* in a systematic manner. There are 3 steps to their approach:

  1. Identification
  2. Estimation
  3. Model diagnostic checking

Step 1: Identification

- Involves determining the order of the model.
- Use of graphical procedures (Time plot, ACF, PACF)

Building ARMA Models
- The Box Jenkins Approach (cont'd)

Step 2: Estimation

- Estimation of the parameters
- Can be done using *least squares* or *maximum likelihood* depending on the model.

Step 3: Model diagnostic checking

Box and Jenkins suggest 2 methods:

- **overfitting** (fitting a larger model than required to capture the dynamics of the data, extra terms would be insignificant)
- **residual diagnostics** (checking the residuals for evidence of linear dependence: ACF, PACF, Ljung-Box Test. If linear dependence is present, the model is not appropriate). More common.

Some More Recent Developments in ARMA Modelling

- We want to form a **parsimonious model**: one with all features of data using as few parameters as possible

- This gives motivation for using **information criteria**, which embody 2 factors:

  - a term which is a function of the RSS (fitness of the model)

  - some penalty for adding extra parameters (parsimonious model)

- **The objective is to choose the parameters which minimise the information criterion.**

Information Criteria for Model Selection

- The information criteria vary according to how stiff the penalty term is.
- The three most popular criteria are **Akaike's** (1974) information criterion (AIC), **Schwarz's** (1978) Bayesian information criterion (SBIC), and the **Hannan-Quinn criterion** (HQIC).

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

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

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

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

SBIC embodies a stiffer penalty term than *AIC*.

- Which IC should be preferred if they suggest different model orders?
  - ***SBIC* is strongly consistent but inefficient.**
  - ***AIC* is not consistent, but more efficient**, and will typically pick "bigger" models.

ARIMA Models

- As distinct from ARMA models, the I stands for **integrated**.
- An **integrated autoregressive process** is one with a characteristic root on the unit circle.
- Typically researchers difference the variable as necessary and then build an ARMA model on those differenced variables.
- An **ARMA(*p*,*q*) model in the variable differenced *d* times** is equivalent to an **ARIMA(*p*,*d*,*q*) model on the original data**.

XI. Forecasting in Econometrics

- **Forecasting = prediction**: determine the values that a series is likely to take

- **Examples:**
- Forecasting tomorrow's return on a particular share
- Forecasting the price of a house given its characteristics
- Forecasting the riskiness of a portfolio over the next year
- Forecasting the volatility of bond returns

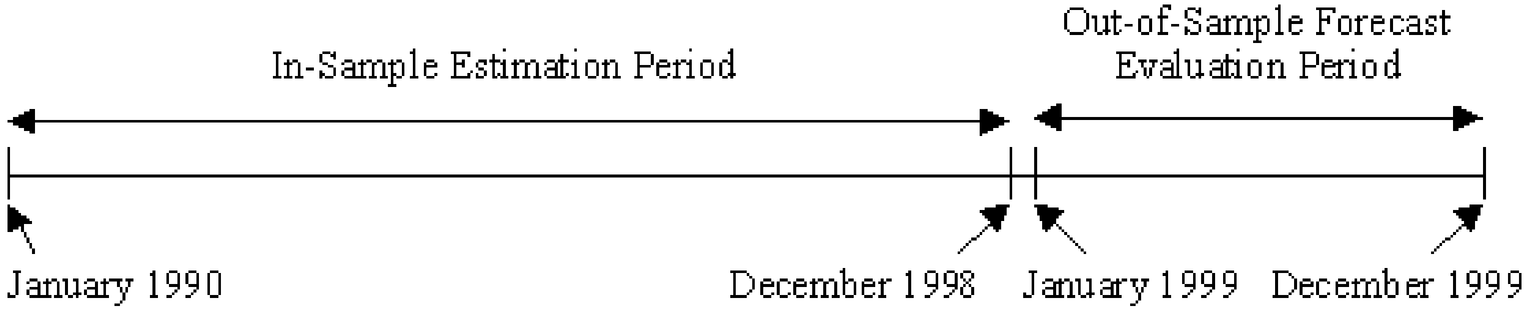
- We can distinguish **two approaches**:
- **Econometric (structural) forecasting:** relates a dependent variable to one or more independent variables
- **Time series forecasting:** involves trying to forecast future values given previous values and/or previous values of errors

- The distinction between the two types is somewhat blurred (e.g, VARs).

In-Sample Versus Out-of-Sample

- Expect the “forecast” of the model to be good **in-sample**.

- Say we have some data - e.g. monthly FTSE returns for 120 months: 1990M1 – 1999M12. We could use all of it to build the model, or keep some observations back:



- A good test of the model since we have not used the information from 1999M1 onwards when we estimated the model parameters.

How to produce forecasts

- **One-step-ahead forecast:** forecast generated for the next observation only
- **Multi-step ahead:** forecasts generated for *1, 2, 3… s* steps ahead
- **Recursive** forecasting for *s* steps ahead: initial date is fixed, add one observation at a time to the estimation period
- **Rolling windows** for *s* steps ahead: length of in-sample period is fixed, start and end dates increase successively by 1

- To understand how to construct forecasts, we need the idea of **conditional expectations**: $E(y_{t+1} \mid \Omega_t)$
- Optimal forecast for a white noise process: $E(u_{t+s} \mid \Omega_t) = 0 \ \forall \ s > 0$.

- The two **simplest forecasting "methods"**

  1. Assume no change : $E(y_{t+s} \mid \Omega_t) = y_t$
  2. Forecasts are the long term average $f(y_{t+s}) = \bar{y}$

Models for Forecasting

- **Structural models**

e.g. $y = X\beta + u$

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

To forecast *y*, we require the conditional expectation of its future value:

$$E\left(y_t \middle| \Omega_{t-1}\right) = E\left(\beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t\right)$$

$$= \beta_1 + \beta_2 E\left(x_{2t}\right) + \ldots + \beta_k E\left(x_{kt}\right)$$

But what are $\mathrm{E}(x_{2t})$ etc.? We could use $\bar{x}_2$, so

$$E\left(y_t\right) = \beta_1 + \beta_2 \bar{x}_2 + \ldots + \beta_k \bar{x}_k$$

$$= \bar{y}\,!!!$$

→ problem!!!

Models for Forecasting (cont'd)

- **Time Series Models**

  The current value of a series, y_t, is modelled as a function only of its previous values and the current value of an error term (and possibly previous values of the error term).

- Models include:
  - simple unweighted averages
  - exponentially weighted averages
  - ARIMA models
  - Non-linear models – e.g. threshold models, GARCH, bilinear models, etc.

Forecasting with ARMA Models

The forecasting model typically used is of the form:

$$f_{t,s} = \mu + \sum_{i=1}^{p} \phi_i f_{t,s-i} + \sum_{j=1}^{q} \theta_j u_{t+s-j}$$

where $f_{t,s} = y_{t+s}$, $s \leq 0$; $u_{t+s} = 0$, $s > 0$

$\qquad\qquad\qquad\qquad\qquad\qquad = u_{t+s}$, $s \leq 0$

Forecasting with MA Models

- An MA(q) only has memory of q.

  e.g. say we have estimated an MA(3) model:

$$
\begin{aligned}
y_t &= \mu + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \theta_3 u_{t-3} + u_t \\
y_{t+1} &= \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + u_{t+1} \\
y_{t+2} &= \mu + \theta_1 u_{t+1} + \theta_2 u_t + \theta_3 u_{t-1} + u_{t+2} \\
y_{t+3} &= \mu + \theta_1 u_{t+2} + \theta_2 u_{t+1} + \theta_3 u_t + u_{t+3}
\end{aligned}
$$

- We are at time t and we want to forecast 1,2,..., s steps ahead.

- We know y_t , y_{t-1}, ..., and u_t , u_{t-1}

Forecasting with MA Models (cont'd)

$$f_{t,1} = E(y_{t+1\,|\,t}) = E(\mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + u_{t+1})$$
$$= \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2}$$

$$f_{t,2} = E(y_{t+2\,|\,t}) = E(\mu + \theta_1 u_{t+1} + \theta_2 u_t + \theta_3 u_{t-1} + u_{t+2})$$
$$= \mu + \theta_2 u_t + \theta_3 u_{t-1}$$

$$f_{t,3} = E(y_{t+3\,|\,t}) = E(\mu + \theta_1 u_{t+2} + \theta_2 u_{t+1} + \theta_3 u_t + u_{t+3})$$
$$= \mu + \theta_3 u_t$$

$$f_{t,4} = E(y_{t+4\,|\,t}) = \mu$$

$$f_{t,s} = E(y_{t+s\,|\,t}) = \mu \qquad \forall\, s \geq 4$$

Forecasting with AR Models

- Say we have estimated an AR(2)

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$
$$y_{t+1} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}$$
$$y_{t+2} = \mu + \phi_1 y_{t+1} + \phi_2 y_t + u_{t+2}$$
$$y_{t+3} = \mu + \phi_1 y_{t+2} + \phi_2 y_{t+1} + u_{t+3}$$

$$\begin{aligned} f_{t,1} = \mathrm{E}(y_{t+1 \mid t}) &= \mathrm{E}(\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}) \\ &= \mu + \phi_1 \mathrm{E}(y_t) + \phi_2 \mathrm{E}(y_{t-1}) \\ &= \mu + \phi_1 y_t + \phi_2 y_{t-1} \end{aligned}$$

$$\begin{aligned} f_{t,2} = \mathrm{E}(y_{t+2 \mid t}) &= \mathrm{E}(\mu + \phi_1 y_{t+1} + \phi_2 y_t + u_{t+2}) \\ &= \mu + \phi_1 \mathrm{E}(y_{t+1}) + \phi_2 \mathrm{E}(y_t) \\ &= \mu + \phi_1 f_{t,1} + \phi_2 y_t \end{aligned}$$

Forecasting with AR Models (cont'd)

$$f_{t,3} = \mathrm{E}(y_{t+3\mid t}) = \mathrm{E}(\mu + \phi_1 y_{t+2} + \phi_2 y_{t+1} + u_{t+3})$$
$$= \mu + \phi_1 \mathrm{E}(y_{t+2}) + \phi_2 \mathrm{E}(y_{t+1})$$
$$= \mu + \phi_1 f_{t,2} + \phi_2 f_{t,1}$$

- We can see immediately that

$f_{t,4} = \mu + \phi_1 f_{t,3} + \phi_2 f_{t,2}$ etc., so

$$f_{t,s} = \mu + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2}$$

- Can easily generate ARMA(p,q) forecasts in the same way.

How can we test whether a forecast is accurate or not?

- For example, say we predict that tomorrow's return on the FTSE will be 0.2, but the outcome is actually -0.4. Is this accurate? **Define $f_{t,s}$ as the forecast made at time t for s steps ahead** (i.e. the forecast made for time $t+s$), and **y_{t+s} as the realised value of y at time $t+s$**.

- Some of the most popular criteria for assessing the accuracy of time series forecasting techniques are:

Mean Squared Error:

$$MSE = \frac{1}{N}\sum_{t=1}^{N} (y_{t+s} - f_{t,s})^2$$

Mean Absolute Error:

$$MAE = \frac{1}{N}\sum_{t=1}^{N} \left| y_{t+s} - f_{t,s} \right|$$

Mean Absolute Percentage Error:

$$MAPE = 100 \times \frac{1}{N}\sum_{t=1}^{N} \left| \frac{y_{t+s} - f_{t,s}}{y_{t+s}} \right|$$

Example

Forecast error aggregation

Steps ahead	Forecast	Actual	Squared error	Absolute error
1	0.20	−0.40	$(0.20 - -0.40)^2 = 0.360$	$\lvert 0.20 - -0.40\rvert = 0.600$
2	0.15	0.20	$(0.15-0.20)^2 = 0.002$	$\lvert 0.15-0.20\rvert = 0.050$
3	0.10	0.10	$(0.10-0.10)^2 = 0.000$	$\lvert 0.10-0.10\rvert = 0.000$
4	0.06	−0.10	$(0.06 - -0.10)^2 = 0.026$	$\lvert 0.06 - -0.10\rvert = 0.160$
5	0.04	−0.05	$(0.04 - -0.05)^2 = 0.008$	$\lvert 0.04 - -0.05\rvert = 0.090$

$$MSE = (0.360 + 0.002 + 0.000 + 0.026 + 0.008)/5 = 0.079$$

$$MAE = (0.600 + 0.050 + 0.000 + 0.160 + 0.090)/5 = 0.180$$

How can we test whether a forecast is accurate or not? (cont'd)

- It has, however, also recently been shown (Gerlow *et al.*, 1993) that the accuracy of forecasts according to traditional statistical criteria are not related to trading profitability.

- A measure more closely correlated with profitability:

% correct sign predictions $= \frac{1}{N}\sum_{t=1}^{N} z_{t+s}$

where $z_{t+s} = 1$ if $(y_{t+s} \,.\, f_{t,s}) > 0$

$z_{t+s} = 0$ otherwise

Forecast Evaluation Example

- Given the following forecast and actual values, calculate the MSE, MAE and percentage of correct sign predictions:

Steps Ahead	Forecast	Actual
1	0.20	-0.40
2	0.15	0.20
3	0.10	0.10
4	0.06	-0.10
5	0.04	-0.05

- MSE = 0.079, MAE = 0.180, % of correct sign predictions = 40

What factors are likely to lead to a good forecasting model?

- “signal” versus “noise”
- simple versus complex models
- financial or economic theory

Statistical Versus Economic or Financial loss functions

Limits of forecasting: What can and cannot be forecast?

- Forecasting models are prone to break down around turning points
- Series subject to structural changes or regime shifts cannot be forecast
- Predictive accuracy usually declines with forecasting horizon
- Forecasting is not a substitute for judgement

Back to the original question: why forecast?

- **Why not use “experts” to make judgemental forecasts?**
- Judgemental forecasts bring a different set of problems:

  e.g., psychologists have found that expert judgements are prone to the following biases:

  - over-confidence
  - inconsistency
  - recency
  - illusory patterns
  - “group-think”…

- **Optimal Approach**

  To use a **statistical forecasting model** built on solid theoretical foundations **supplemented by expert judgements and interpretation**.