Lecture Notes (2)

Univariate time series modelling and forecasting

1

1

Introduction: Univariate Time Series Models

- Univariate time series models are class of specifications where one attempts to model and to predict financial variables using only information contained in their own past values and possibly current and past values of an error term.
- This practice can be contrasted with structural models, which are multivariate in nature, and attempt to explain changes in a variable by reference to the movements in the current or past values of others (explanatory) variables. Time series models may be useful when a structural model is inappropriate.
- An important class of time series models is the family of AutoRegressive Integrated Moving Average (ARIMA) models, usually associated with Box and Jenkins (1976).
- In order to define, estimate and use ARIMA models, we first need to specify the notation and to define several important concepts.

2

Some Notation and Concepts

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

Some Notation and Concepts(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

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

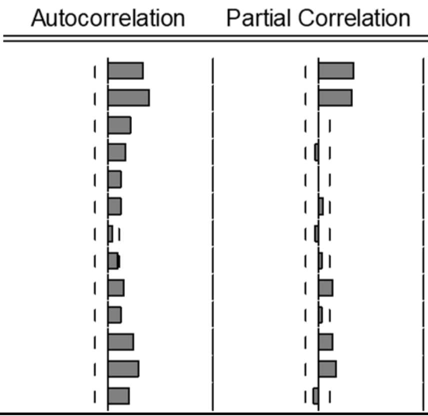
  are known as the covariance function.
- The covariances, γ_s, are known as **autocovariances** since they are the covariances of y with its own previous values.
- However, the value of the autocovariances depend on the units of measurement of y_t. It is thus more convenient to use the **autocorrelations** 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) or **correlogram.**

Correlogram for UK HP Data

Date: 10/26/20 Time: 14:26
Sample: 1991M01 2018M03
Included observations: 326



	AC	PAC	Q-Stat	Prob
1	0.358	0.358	42.056	0.000
2	0.418	0.333	99.750	0.000
3	0.231	0.016	117.44	0.000
4	0.182	-0.019	128.48	0.000
5	0.126	0.006	133.80	0.000
6	0.131	0.054	139.50	0.000
7	0.061	-0.030	140.74	0.000
8	0.096	0.036	143.87	0.000
9	0.160	0.144	152.56	0.000
10	0.137	0.039	158.87	0.000
11	0.255	0.148	180.86	0.000
12	0.299	0.179	211.38	0.000
13	0.209	-0.042	226.35	0.000

A White Noise Process

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

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

- Thus the autocorrelation function will be zero apart from a single peak of 1 at $s = 0$. $\hat{\tau}_s \sim$ approximately N(0, 1/T) where T = sample size.

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

- For example, a 95% confidence interval would be given by $\pm 1.96 \times (1/\sqrt{T})$. 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 coefficient at lag s is zero.

Joint Hypothesis Tests

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

$$Q = T\sum_{k=1}^{m} \hat{\tau}_k^2$$

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:

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

- 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.

  Test each of the individual coefficient for significance, and use both the Box-Pierce and Ljung-Box tests to establish whether they are jointly significant.

- Solution:

  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.

  Q=5.09 and *Q**=5.26

  Compared with a tabulated $\chi^2(5)$=11.1 at the 5% level, so the 5 coefficients are jointly insignificant.

Moving Average Processes

- The simplest class of time series model is that of the moving average process
- Let u_t (t=1,2,3,...) be a sequence of independently and identically distributed (iid) random variables with $E(u_t)=0$ and $Var(u_t)= \sigma^2$, then

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

  is a q^{th} order moving average model **MA(*q*).**

- Its properties are
  $E(y_t)=\mu$; $Var(y_t) = \gamma_0 = (1+\theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma^2$
  Covariances

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

Lag Operator Notation for MA(q)

- Using the lag operator (backshift operator) notation:
  $Ly_t = y_{t-1}$ $\qquad$ $L^i y_t = y_{t-i}$

- $y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q} = \mu + \theta(L)u_t$

  where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$

Example of an MA Problem

Consider the following MA(2) process:

$$X_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$$

where u_t is a zero mean white noise process with variance σ^2.

(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

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

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

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

(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)

γ_2 = 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[($\theta_2 u_{t-2}^2$)]
$= \theta_2 \sigma^2$

γ_3 = 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

So γ_s = 0 for s > 2.

Solution (cont'd)

We have the autocovariances, now calculate the autocorrelations:

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

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

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

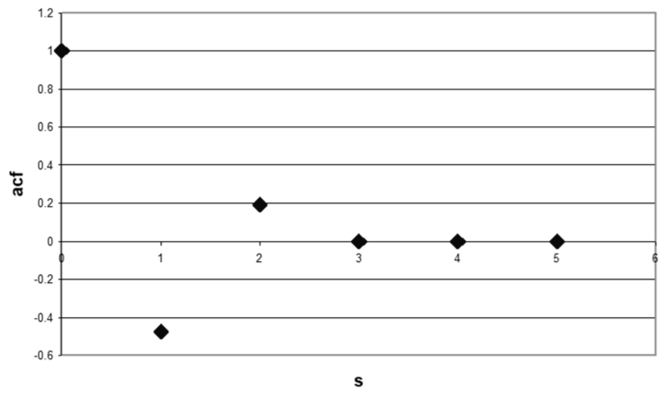
$$\tau_3 = \frac{\gamma_3}{\gamma_0} = 0$$

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

(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 plot will appear as follows:



Autoregressive Processes

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

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

- Or using the lag operator notation:

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

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

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

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 (refer to Box 6.1 on page 260).

- **Example 1:** 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 $y_t = 3y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + u_t$ stationary?
  The characteristic roots are 1, 2/3, and 2. Since only one of these lies outside the unit circle, the process is non-stationary.

  (Refer to Page 261)

Wold's Decomposition Theorem

- States that any stationary series can be decomposed into the sum of two unrelated processes, a purely deterministic part and a purely stochastic part, which will be an MA(∞).
- For the AR(p) model, $\phi(L)y_t = u_t$, ignoring the intercept, the Wold decomposition is

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

where,

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

- Refer to pages 262 ~ 263

The Moments of an Autoregressive Process

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

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

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

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

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

Sample AR Problem

- Consider the following simple AR(1) model

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

(i) 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.

Solution

(i) Unconditional mean:

$E(y_t) = E(\mu + \phi_1 y_{t-1})$
$= \mu + \phi_1 E(y_{t-1})$

But also

So $E(y_t) = \mu + \phi_1 (\mu + \phi_1 E(y_{t-2}))$
$= \mu + \phi_1 \mu + \phi_1^2 E(y_{t-2}))$

$E(y_t) = \mu + \phi_1 \mu + \phi_1^2 E(y_{t-2}))$
$= \mu + \phi_1 \mu + \phi_1^2 (\mu + \phi_1 E(y_{t-3}))$
$= \mu + \phi_1 \mu + \phi_1^2 \mu + \phi_1^3 E(y_{t-3})$

Solution (cont'd)

An infinite number of such substitutions would give

$E(y_t) = \mu(1+\phi_1+\phi_1^2+...) + \phi_1^{\infty} y_0$

So long as the model is stationary, i.e. , then $\phi_1^{\infty} = 0$.

So $E(y_t) = \mu(1+\phi_1+\phi_1^2+...) = \dfrac{\mu}{1-\phi_1}$

(ii) Calculating the variance of y_t: $y_t = \phi_1 y_{t-1} + u_t$

From Wold's decomposition theorem:

$$y_t(1-\phi_1 L) = u_t$$
$$y_t = (1-\phi_1 L)^{-1} u_t$$
$$y_t = (1+\phi_1 L+\phi_1^2 L^2+\ldots)u_t$$

Solution (cont'd)

So long as $|\phi_1| < 1$, this will converge.

$$y_t = u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \ldots$$

$Var(y_t) = E[y_t - E(y_t)][y_t - E(y_t)]$

but $E(y_t) = 0$, since we are setting $\mu = 0$.

$Var(y_t) = E[(y_t)(y_t)]$

$= E[\left(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + ..\right)\left(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + ..\right)]$

$= E[(u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \ldots + cross - products)]$

$= E[(u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \ldots)]$

$= \sigma_u^2 + \phi_1^2 \sigma_u^2 + \phi_1^4 \sigma_u^2 + \ldots$

$= \sigma_u^2 (1 + \phi_1^2 + \phi_1^4 + \ldots)$

$= \dfrac{\sigma_u^2}{(1-\phi_1^2)}$

Solution (cont'd)

(iii) Turning now to calculating the acf, first calculate the autocovariances:

$\gamma_1 = \text{Cov}(y_t, y_{t-1}) = \text{E}[y_t\text{-E}(y_t)][y_{t-1}\text{-E}(y_{t-1})]$

Since a_0 has been set to zero, $\text{E}(y_t) = 0$ and $\text{E}(y_{t-1}) = 0$, so

$\gamma_1 = \text{E}[y_t y_{t-1}]$

$$\gamma_1 = \text{E}[(u_t + \phi_1 u_{t-1} + \phi_1{}^2 u_{t-2} + \ldots)(u_{t-1} + \phi_1 u_{t-2} + \phi_1{}^2 u_{t-3} + \ldots)]$$

$$= \text{E}[\ \phi_1\ u_{t-1}{}^2 + \phi_1{}^3 u_{t-2}{}^2 + \ldots + cross - products]$$

$$= \phi_1 \sigma^2 + \phi_1^3 \sigma^2 + \phi_1^5 \sigma^2 + \ldots$$

$$= \frac{\phi_1 \sigma^2}{(1-\phi_1^2)}$$

Solution (cont'd)

For the second autocorrelation coefficient,

$\gamma_2 = \text{Cov}(y_t, y_{t-2}) = \text{E}[y_t\text{-E}(y_t)][y_{t-2}\text{-E}(y_{t-2})]$

Using the same rules as applied above for the lag 1 covariance

$\gamma_2 = \text{E}[y_t y_{t-2}]$

$$= \text{E}[(u_t + \phi_1 u_{t-1} + \phi_1{}^2 u_{t-2} + \ldots)(u_{t-2} + \phi_1 u_{t-3} + \phi_1{}^2 u_{t-4} + \ldots)]$$

$$= \text{E}[\ \phi_1{}^2 u_{t-2}{}^2 + \phi_1{}^4 u_{t-3}{}^2 + \ldots + cross - products]$$

$$= \phi_1^2 \sigma^2 + \phi_1^4 \sigma^2 + \ldots$$

$$= \phi_1^2 \sigma^2 (1 + \phi_1^2 + \phi_1^4 + \ldots)$$

$$= \frac{\phi_1^2 \sigma^2}{(1-\phi_1^2)}$$

Solution (cont'd)

- If these steps were repeated for γ_3, the following expression would be obtained

$$\gamma_3 = \frac{\phi_1^3 \sigma^2}{(1-\phi_1^2)}$$

and for any lag *s*, the autocovariance would be given by

$$\gamma_s = \frac{\phi_1^s \sigma^2}{(1-\phi_1^2)}$$

The acf can now be obtained by dividing the covariances by the variance:

Solution (cont'd)

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

$$\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{\left(\frac{\phi_1 \sigma^2}{(1-\phi_1^2)}\right)}{\left(\frac{\sigma^2}{(1-\phi_1^2)}\right)} = \phi_1 \qquad \tau_2 = \frac{\gamma_2}{\gamma_0} = \frac{\left(\frac{\phi_1^2 \sigma^2}{(1-\phi_1^2)}\right)}{\left(\frac{\sigma^2}{(1-\phi_1^2)}\right)} = \phi_1^2$$

$$\tau_3 = \phi_1^3$$

$$\ldots$$

$$\tau_s = \phi_1^s$$

- ***Note that use of the Yule-Walker Equations would have given the same answer.***

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 always
- 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*), this can be written as an AR(∞), if MA(q) process is invertible (refer to page 267 "The invertibility condition"). So there are direct connections between y_t and all its previous values.
- For an MA(*q*), the theoretical pacf will be geometrically declining.

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

31

Summary of the Behaviour of the acf and pacf for AR, MA and ARMA Processes

- The mean of an ARMA(p, q) series is given by

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

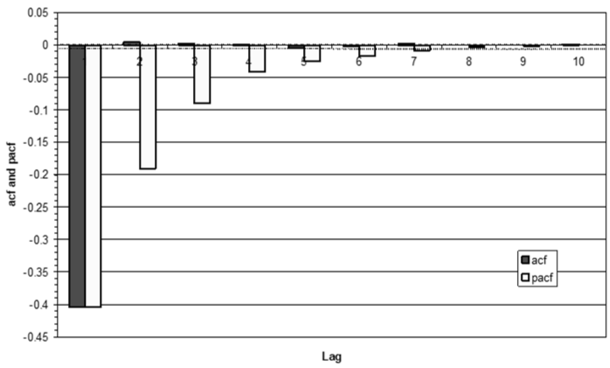
- An autoregressive process AR(p) has
  - a geometrically decaying acf
  - number of spikes of pacf = p =AR order
- A moving average process MA(q) has
  - Number of spikes of acf = q = MA order
  - a geometrically decaying pacf
- A combination autoregressive moving average process ARMA(p, q) has
  - a geometrically decaying acf
  - a geometrically decaying pacf

32

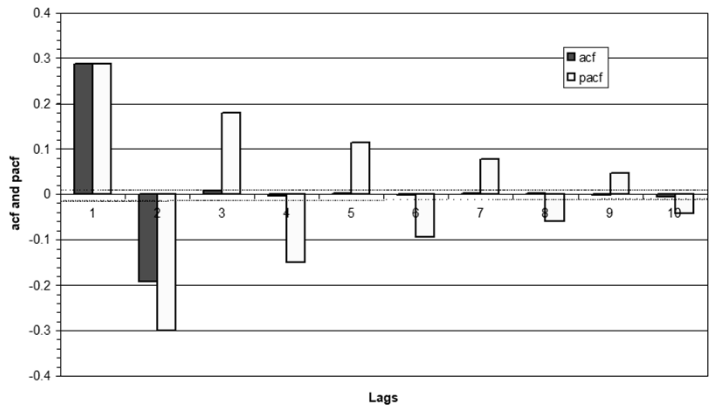
Some sample acf and pacf plots for standard processes

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.

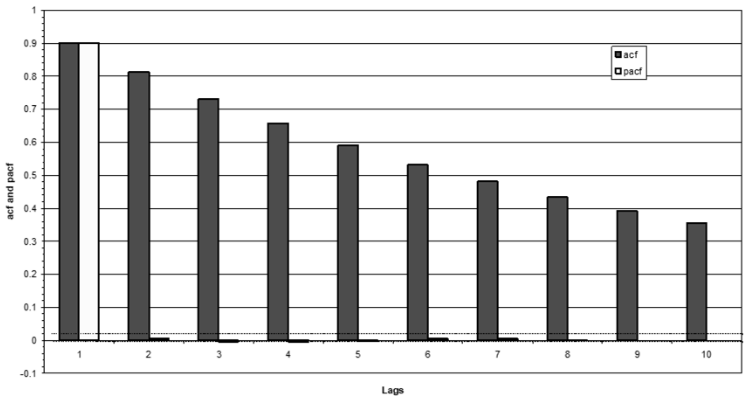
ACF and PACF for an MA(1) Model: $y_t = -0.5u_{t-1} + u_t$



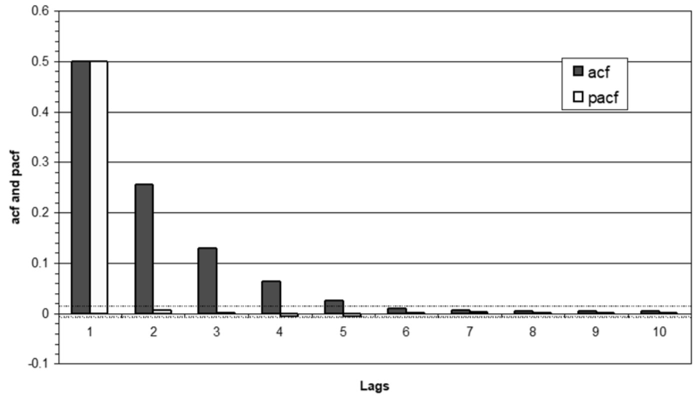
ACF and PACF for an MA(2) Model: $y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$



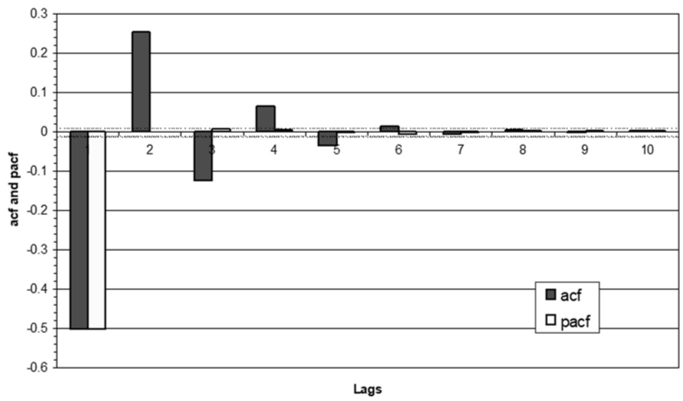
ACF and PACF for a slowly decaying AR(1) Model: $y_t = 0.9y_{t-1} + u_t$



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

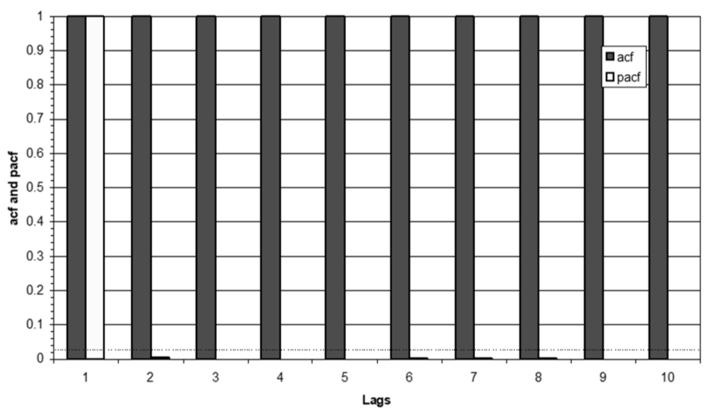


ACF and PACF for a more rapidly decaying AR(1) Model with Negative Coefficient: $y_t = -0.5y_{t-1} + u_t$



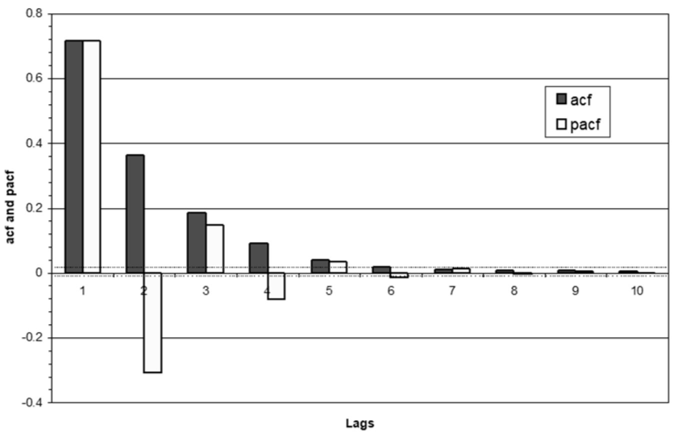
37

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



38

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



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. Their approach was a practical and pragmatic one, involving 3 steps:

  Step 1. Identification

  Step 2. Estimation

  Step 3. Model diagnostic checking

<u>Step 1:</u>

- Involves determining the order of the model.
- Use of graphical procedures (e.g . plotting the acf and pacf)

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

Step 2:

- Estimation of the parameters of the model specified in step 1.
- Can be done using least squares or maximum likelihood depending on the model.

Step 3:

- Model checking – i.e. determining whether the model specified and estimated is adequate.

Box and Jenkins suggest 2 methods:

- deliberate overfitting
- residual diagnostics

(Refer to page 274)

Some More Recent Developments in ARMA Modelling

- Identification would typically not be done using graphical plots of the acf and pacf.
- We want to form a parsimonious model.
- Reasons:
  - variance of estimators is inversely proportional to the number of degrees of freedom.
  - models which are profligate might be inclined to fit to data specific features
- This gives motivation for using information criteria, which embody 2 factors
  - a term which is a function of the RSS
  - some penalty for adding extra parameters
- The object is to choose the number of parameters which minimises 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. $p \le \overline{p}, q \le \overline{q}$
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, and will typically pick "bigger" models.

Constructing ARIMA Models in EViews

- Refer to pages 276~281.
- Using the monthly UK house price series.
- Objective of this exercise is to build an ARMA model for the house price changes. We use Box-Jenkins approach, i.e., the follow the three steps:
  - Step 1: Plot acf and pacf - possible ARMA (p, q)
  - Step 2: Estimate parameters for all models – ARMA(0, 0) ~ ARMA(5, 5)
  - Step 3: Model checking/selecting - AIC or SBIC (Summary table on page 280) ARMA(4, 5) by AIC and ARMA(2, 0) = AR(2) by SBIC.

Forecasting/ Prediction

- Why forecast?

  Forecasts are made essentially because they are useful! Financial decisions often involve a long-term commitment of resources, the returns to which will depend upon what happens in the future. In this context, the decisions made today will reflect forecasts of the future state of the world, and the more accurate those forecasts are, the more utility (or money!) is likely to be gained from acting on them.

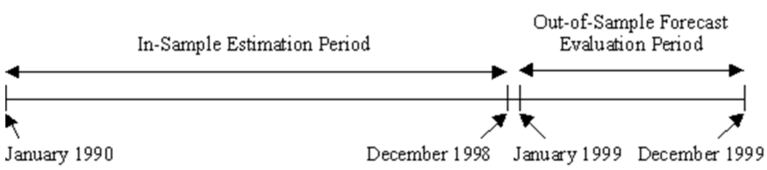
- An important test of the adequacy of a model. e.g.
  - Forecasting tomorrow's return on a particular share
  - Forecasting the price of a house given its characteristics
  - Forecasting the riskiness of a portfolio over the next year
  - Forecasting the volatility of bond returns

- We can distinguish two approaches:
  - Econometric (structural) forecasting
  - Time series forecasting

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

- Multi-step ahead versus single-step ahead forecasts
- Recursive versus rolling windows
- To understand how to construct forecasts, we need the idea of conditional expectations: $E(y_{t+s} \mid \Omega_t)$
- We cannot forecast a white noise process: $E(u_{t+s} \mid \Omega_t) = 0 \ \forall \ s > 0$.
- The two simplest naïve forecasting "methods" (Box 6.3, on page 288)
  1. Assume no change : $f_{t.s} = y_t$
  2. Forecasts are the long term average $f_{t.s} = \bar{y}$

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:

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

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

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

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

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

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

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

Forecasting with MA Models

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

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

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

- 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)

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

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

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

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

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

Forecasting with AR Models

- Say we have estimated an AR(2)

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

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

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

Forecasting with AR Models (cont'd)

$$\begin{aligned} f_{t,3} = E(y_{t+3|t}) &= E(\mu + \phi_1 y_{t+2} + \phi_2 y_{t+1} + u_{t+3}) \\ &= \mu + \phi_1 E(y_{t+2}) + \phi_2 E(y_{t+1}) \\ &= \mu + \phi_1 f_{t,2} + \phi_2 f_{t,1} \end{aligned}$$

- 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.

Forecasting in EViews

Forecasting using ARMA models in EViews

- Suppose that the AR(2) model selected for the house price percentage changes series were estimated using observations Feb 1991 ~ Dec 2010, leaving 29 remaining observations to construct forecasts for and to test forecast accuracy (i.e. for the period Jan 2011 ~ May 2013)

  (Details refer to pages 296 ~ 299)