# Chapter 6

Univariate time series modelling and forecasting

# Univariate Time Series Models

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

**Some Notation and Concepts**

- A Strictly Stationary Process

A strictly stationary process is one where

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

- A Weakly Stationary Process

# Univariate Time Series Models (Cont'd)

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

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

- 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, $\gamma_s$, are known as autocovariances.

# Univariate Time Series Models (Cont'd)

- 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:

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

- If we plot $\tau_s$ against $s$=0,1,2,... then we obtain the autocorrelation function or correlogram.

# A White Noise Process

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

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

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

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

# A White Noise Process (Cont'd)

- For example, a 95 % confidence interval would be given by

$$\pm 1.96 \times \frac{1}{\sqrt{T}}$$

. 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 coefficient at lag $s$ is zero.

# Joint Hypothesis Tests

- We can also test the joint hypothesis that all $m$ of the $\tau_k$ 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 $\chi_m^2$.
- 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

- <u>Question:</u>

  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.

- <u>Solution:</u>

  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

# An ACF Example (Cont'd)

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

# Moving Average Processes

- Let $u_t$ $(t = 1, 2, 3, \ldots)$ be a sequence of independently and identically distributed (iid) random variables with $\mathrm{E}(u_t) = 0$ and $\mathrm{var}(u_t) = \sigma^2$, then

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

  is a $q$th order moving average model MA($q$).

- Its properties are

$$\mathrm{E}(y_t) = \mu$$
$$\mathrm{var}(y_t) = \gamma_0 = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma^2$$

  Covariances

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

# Example of an MA Problem

1. Consider the following MA(2) process:

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

where $u_t$ is a zero mean white noise process with variance $\sigma^2$.

   i. Calculate the mean and variance of $X_t$

   ii. Derive the autocorrelation function for this process (i.e. express the autocorrelations, $\tau_1$, $\tau_2$, ...as functions of the parameters $\theta_1$ and $\theta_2$).

   iii. If $\theta_1 = -0.5$ and $\theta_2 = 0.25$, sketch the acf of $X_t$.

# Solution

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

$$\begin{aligned}
\mathrm{E}(y_t) &= \mathrm{E}(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}) \\
&= \mathrm{E}(u_t) + \theta_1 \mathrm{E}(u_{t-1}) + \theta_2 \mathrm{E}(u_{t-2}) = 0 \\
\mathrm{var}(y_t) &= \mathrm{E}[y_t - \mathrm{E}(y_t)][y_t - \mathrm{E}(y_t)]
\end{aligned}$$

But $\mathrm{E}(y_t) = 0$, so

$$\begin{aligned}
\mathrm{var}(y_t) &= \mathrm{E}[(y_t)(y_t)] \\
\mathrm{var}(y_t) &= \mathrm{E}[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})] \\
\mathrm{var}(y_t) &= \mathrm{E}\left[u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2 + \textit{cross-products}\right]
\end{aligned}$$

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

# Solution (Cont'd)

So

$$\begin{aligned} \mathrm{var}(y_t) &= \gamma_0 = E\left[u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2\right] \\ &= \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2 \\ &= \left(1 + \theta_1^2 + \theta_2^2\right)\sigma^2 \end{aligned}$$

# Solution (Cont'd)

ii. The acf of $y_t$

$$\gamma_1 = \mathrm{E}[y_t - \mathsf{E}(y_t)][y_{t-1} - \mathsf{E}(y_{t-1})]$$

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

$$\gamma_1 = \mathrm{E}[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3})]$$

$$\gamma_1 = \mathrm{E}\left[\left(\theta_1 u_{t-1}^2 + \theta_1\theta_2 u_{t-2}^2\right)\right]$$

$$\gamma_1 = \theta_1\sigma^2 + \theta_1\theta_2\sigma^2$$

$$\gamma_1 = (\theta_1 + \theta_1\theta_2)\sigma^2$$

# **Solution** (Cont'd)

$$\begin{aligned}
\gamma_2 &= \mathsf{E}[y_t - \mathsf{E}(y_t)][y_{t-2} - \mathsf{E}(y_{t-2})] \\
\gamma_2 &= \mathrm{E}[y_t][y_{t-2}] \\
\gamma_2 &= \mathrm{E}[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-2} + \theta_1 u_{t-3} + \theta_2 u_{t-4})] \\
\gamma_2 &= \mathrm{E}\left[\left(\theta_2 u_{t-2}^2\right)\right] \\
\gamma_2 &= \theta_2 \sigma^2
\end{aligned}$$

# Solution (Cont'd)

$$\gamma_3 = \mathrm{E}[y_t - \mathrm{E}(y_t)][y_{t-3} - \mathrm{E}(y_{t-3})]$$

$$\gamma_3 = \mathrm{E}[y_t][y_{t-3}]$$

$$\gamma_3 = \mathrm{E}[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_{t-3} + \theta_1 u_{t-4} + \theta_2 u_{t-5})]$$

$$\gamma_3 = 0$$

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

# Solution (Cont'd)

We have the autocovariances, now calculate the autocorrelations:

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

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

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

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

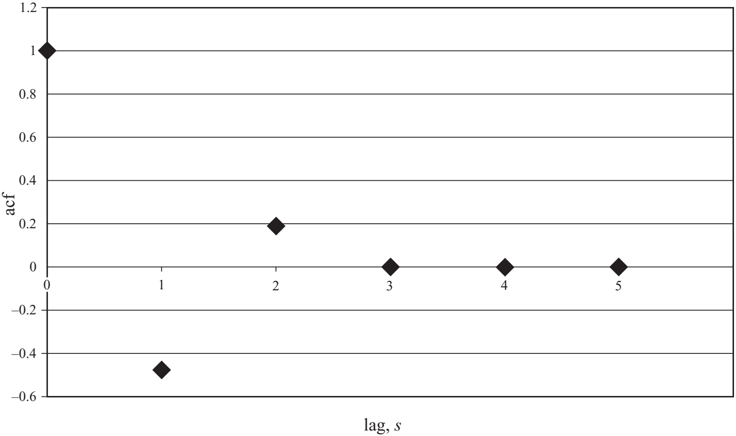
$$\tau_s = \frac{\gamma_s}{\gamma_0} = 0 \ \forall \ s > 2$$

# Solution (Cont'd)

iii. For $\theta_1 = -0.5$ and $\theta_2 = 0.25$, substituting these into the formulae above gives $\tau_1 = -0.476$, $\tau_2 = 0.190$.

# 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

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

- 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

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

# The Stationary Condition for an AR Model

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

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

- 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.

# 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($\infty$).
- 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,

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

# The Moments of an Autoregressive Process

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

$$E(y_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

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

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

- 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 $\mu = 0$ for simplicity.

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

iii. Derive the autocorrelation function for $y_t$.

# Solution

i. Unconditional mean:

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

But also

So

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

# Solution (Cont'd)

An infinite number of such substitutions would give

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

So long as the model is stationary, i.e. $|\phi_1| < 1$, then $\phi_1^{\infty} = 0$. So

$$\mathrm{E}(y_t) = \mu\big(1 + \phi_1 + \phi_1^2 + \cdots\big) = \frac{\mu}{1 - \phi_1}$$

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

## Solution (Cont'd)

From Wold's decomposition theorem:

$$\begin{aligned} y_t(1-\phi_1 L) &= u_t \\ y_t &= (1-\phi_1 L)^{-1} u_t \\ y_t &= \left(1+\phi_1 L+\phi_1^2 L^2+\cdots\right) u_t \end{aligned}$$

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

$$\text{var}(y_t) = \mathrm{E}[y_t - \mathrm{E}(y_t)][y_t - \mathrm{E}(y_t)]$$

# Solution (Cont'd)

but $\mathrm{E}(y_t) = 0$, since $\mu$ is set to zero.

$$
\begin{aligned}
\mathrm{var}(y_t) &= \mathrm{E}[(y_t)(y_t)] \\
&= \mathrm{E}\Big[\big(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots\big)\big(u_t + \phi_1 u_{t-1} \\
&\quad +\phi_1^2 u_{t-2} + \cdots\big)\Big] \\
&= \mathrm{E}\left[u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \cdots + \textit{cross-products}\right] \\
&= \mathrm{E}\left[u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \cdots\right] \\
&= \sigma_u^2 + \phi_1^2 \sigma_u^2 + \phi_1^4 \sigma_u^2 + \cdots \\
&= \sigma_u^2 \left(1 + \phi_1^2 + \phi_1^4 + \cdots\right) \\
&= \frac{\sigma_u^2}{(1 - \sigma_u^2)}
\end{aligned}
$$

# Solution (Cont'd)

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

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

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

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

## **Solution** (Cont'd)

under the result above that $E(y_t) = E(y_{t-1}) = 0$. Thus

$$\begin{aligned}\gamma_1 &= E\big[\big(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots\big)\big(u_{t-1} + \phi_1 u_{t-2} \\ &\quad + \phi_1^2 u_{t-3} + \cdots\big)\big] \\ \gamma_1 &= E\big[\phi_1 u_{t-1}^2 + \phi_1^3 u_{t-2}^2 + \cdots + cross - products\big] \\ \gamma_1 &= \phi_1\sigma^2 + \phi_1^3\sigma^2 + \phi_1^5\sigma^2 + \cdots \\ \gamma_1 &= \frac{\phi_1\sigma^2}{(1-\phi_1^2)}\end{aligned}$$

For the second autocorrelation coefficient,

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

# Solution (Cont'd)

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

$$
\begin{aligned}
\gamma_2 &= \mathrm{E}[y_t y_{t-2}] \\
\gamma_2 &= \mathrm{E}\left[\left(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots\right)\left(u_{t-2} + \phi_1 u_{t-3} \right.\right. \\
&\quad \left.\left. + \phi_1^2 u_{t-4} + \cdots\right)\right] \\
\gamma_2 &= \mathrm{E}\left[\phi_1^2 u_{t-2}^2 + \phi_1^4 u_{t-3}^2 + \cdots + \textit{cross-products}\right] \\
\gamma_2 &= \phi_1^2\sigma^2 + \phi_1^4\sigma^2 + \cdots \\
\gamma_2 &= \phi_1^2\sigma^2\left(1 + \phi_1^2 + \phi_1^4 + \cdots\right) \\
\gamma_2 &= \frac{\phi_1^2\sigma^2}{\left(1 - \phi_1^2\right)}
\end{aligned}
$$

## **Solution** (Cont'd)

If these steps were repeated for $\gamma_3$, the following expression would be obtained

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

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

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

# Solution (Cont'd)

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

$$
\begin{aligned}
\tau_0 &= \frac{\gamma_0}{\gamma_0} = 1 \\
\tau_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\left(\dfrac{\phi_1 \sigma^2}{(1-\phi_1^2)}\right)}{\left(\dfrac{\sigma^2}{(1-\phi_1^2)}\right)} = \phi_1 \\
\tau_2 &= \frac{\gamma_2}{\gamma_0} = \frac{\left(\dfrac{\phi_1^2 \sigma^2}{(1-\phi_1^2)}\right)}{\left(\dfrac{\sigma^2}{(1-\phi_1^2)}\right)} = \phi_1^2 \\
\tau_3 &= \phi_1^3
\end{aligned}
$$

$$\tau_s = \phi_1^s$$

# The Partial Autocorrelation Function (denoted $\tau_{kk}$)

- 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 $\tau_{kk}$ measures the correlation between $y_t$ and $y_{t-k}$ after removing the effects of $y_{t-k+1}, y_{t-k+2}, \ldots, y_{t-1}$

- At lag 1, the acf = pacf always

- At lag 2,

$$\tau_{22} = \left(\tau_2 - \tau_1^2\right) / \left(1 - \tau_1^2\right)$$

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

# The Partial Autocorrelation Function (denoted $\tau_{kk}$) (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($\infty$), 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

$$\begin{aligned} \phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \quad \text{and} \\ \theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \end{aligned}$$

or

$$\begin{aligned} y_t = {} & \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 u_{t-1} \\ & + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} + u_t \end{aligned}$$

with

$$\mathrm{E}(u_t) = 0; E\left(u_t^2\right) = \sigma^2; E(u_t u_s) = 0, t \neq s$$

# 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

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

- The autocorrelation function for an ARMA process will display combinations of behaviour derived from the AR and MA parts, but for lags beyond $q$, the acf will simply be identical to the individual AR($p$) model.

# Summary of the Behaviour of the acf for AR and MA Processes

An autoregressive process has

- a geometrically decaying acf
- number of spikes of pacf = AR order

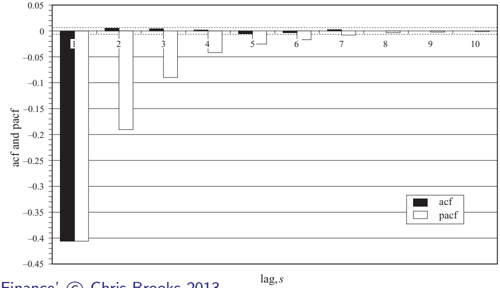
A moving average process has

- Number of spikes of acf = MA order
- a geometrically decaying pacf

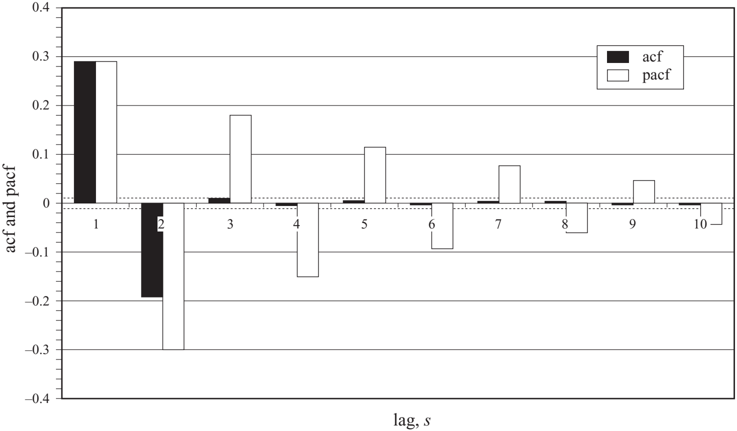
# Some sample acf and pacf plots for standard processes

- 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.

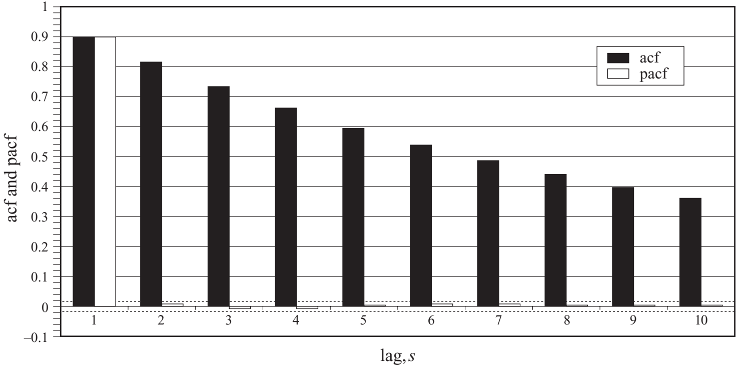
Figure: Sample autocorrelation and partial autocorrelation functions for an MA(1) model: $y_t = -0.5u_{t-1} + u_t$



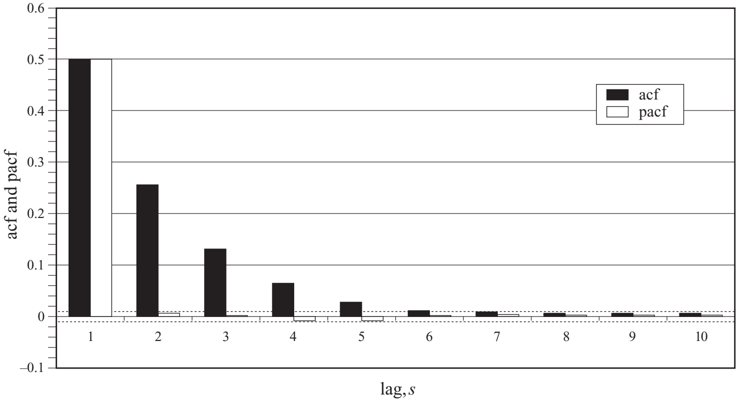
# ACF and PACF for an MA(2) Model: $y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$



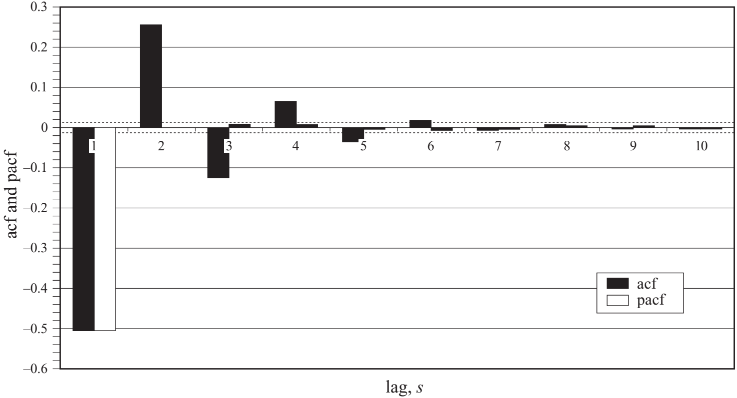
# ACF and PACF for a slowly decaying AR(1) Model: $y_t = 0.9y_{t-1} + u_t$



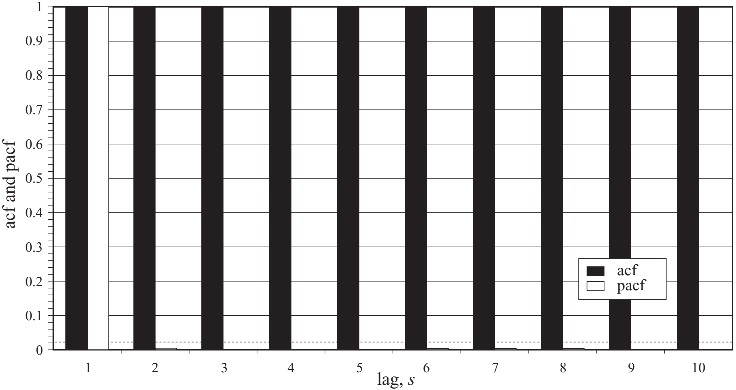
# ACF and PACF for a more rapidly decaying AR(1) Model: $y_t = 0.5y_{t-1} + u_t$



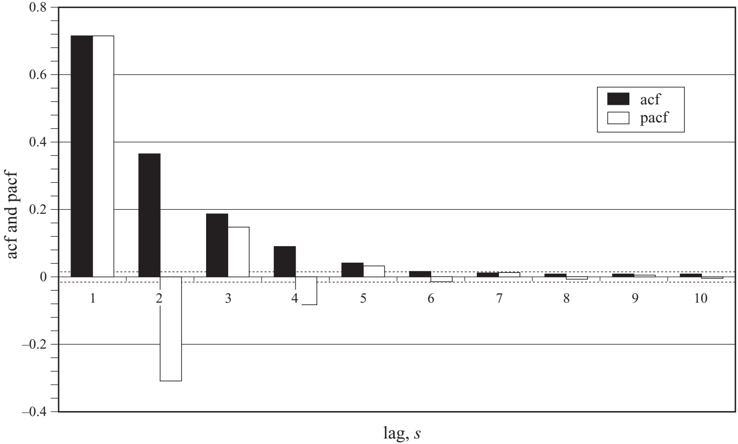
# ACF and PACF for a more rapidly decaying AR(1) Model with Negative Coefficient: $y_t = -0.5y_{t-1} + u_t$



# 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): $y_t = 0.5y_{t-1} + 0.5u_{t-1} + u_t$



# 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:

- Involves determining the order of the model.
- Use of graphical procedures
- A better procedure is now available

Step 2:

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

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

Step 3:

- Model checking

Box and Jenkins suggest 2 methods:

- deliberate overfitting
- residual diagnostics

# Some More Recent Developments in ARMA Modelling

- Identification would typically not be done using acf's.
- 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).

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

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

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

# Information Criteria for Model Selection (Cont'd)

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, 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.

# Exponential Smoothing

- Another modelling and forecasting technique
- How much weight do we attach to previous observations?
- Expect recent observations to have the most power in helping to forecast future values of a series.
- The equation for the model

$$S_t = \alpha y_t + (1 - \alpha) S_{t-1} \tag{1}$$

Where
$\alpha$ is the smoothing constant, with $0 \leq \alpha \leq 1$,
$y_t$ is the current realised value,
$S_t$ is the current smoothed value.

## Exponential Smoothing (Cont'd)

$$S_{t-1} = \alpha y_{t-1} + (1-\alpha)S_{t-2} \qquad (2)$$

- and lagging again

$$S_{t-2} = \alpha y_{t-2} + (1-\alpha)S_{t-3} \qquad (3)$$

- Substituting into (1) for $S_{t-1}$ from (2)

$$S_t = \alpha y_t + (1-\alpha)(\alpha y_{t-1} + (1-\alpha)S_{t-2})$$

$$S_t = \alpha y_t + (1-\alpha)\alpha y_{t-1} + (1-\alpha)^2 S_{t-2} \qquad (4)$$

# Exponential Smoothing (Cont'd)

- Substituting into (4) for $S_{t-2}$ from (3)

$$\begin{aligned} S_t &= \alpha y_t + (1-\alpha)\alpha y_{t-1} + (1-\alpha)^2 S_{t-2} \\ &= \alpha y_t + (1-\alpha)\alpha y_{t-1} + (1-\alpha)^2(\alpha y_{t-2} + (1-\alpha)S_{t-3}) \\ &= \alpha y_t + (1-\alpha)\alpha y_{t-1} + (1-\alpha)^2 \alpha y_{t-2} + (1-\alpha)^3 S_{t-3} \end{aligned}$$

- $T$ successive substitutions of this kind would lead to

$$S_t = \left( \sum_{i=0}^{T} \alpha(1-\alpha)^i y_{t-i} \right) + (1-\alpha)^T S_0$$

Since $\alpha \geq 0$, the effect of each observation declines geometrically as the variable moves another observation forward in time.

# Exponential Smoothing (Cont'd)

- Forecasts are generated by

$$f_{t,s} = S_t$$

  for all steps into the future $s = 1, 2, \ldots$

- This technique is called single (or simple) exponential smoothing.

- It doesn't work well for financial data because
  - there is little structure to smooth
  - it cannot allow for seasonality
  - it is an ARIMA(0,1,1) with MA coefficient (1-$\alpha$) - (See Granger & Newbold, p174)
  - forecasts do not converge on long term mean as $s \to \infty$

- Can modify single exponential smoothing

# Exponential Smoothing (Cont'd)

- to allow for trends (Holt's method)
- or to allow for seasonality (Winter's method).

- Advantages of Exponential Smoothing
  - Very simple to use
  - Easy to update the model if a new realisation becomes available.

# Forecasting in Econometrics

- Forecasting = prediction.
- 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

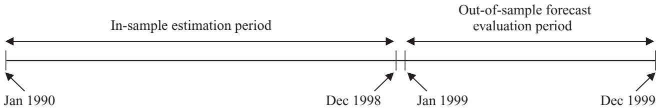
We can distinguish two approaches:

- Econometric (structural) forecasting
- Time series forecasting

- 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

- 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:

$$E(y_{t+1} \mid \Omega_t)$$

- We cannot forecast a white noise process:

$$\mathrm{E}(u_{t+s}|\Omega_t) = 0 \, \forall, s > 0$$

- The two simplest forecasting "methods"
  1. Assume no change: $f(y_{t+s}) = y_t$
  2. Forecasts are the long term average $f(y_{t+s}) = \bar{y}$

# Models for Forecasting

- Structural models

$$\text{e.g.} \quad \begin{aligned} y &= X\beta + u \\ y &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t \end{aligned}$$

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

$$\begin{aligned} \mathrm{E}(y_t \,|\Omega_{t-1}\,) &= \mathrm{E}(\beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t) \\ &= \beta_1 + \beta_2 \mathrm{E}(x_{2t}) + \beta_3 \mathrm{E}(x_{3t}) + \cdots + \beta_k \mathrm{E}(x_{kt}) \end{aligned}$$

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

$$\begin{aligned} \mathrm{E}(y_t) &= \beta_1 + \beta_2 \bar{x}_2 + \beta_3 \bar{x}_3 + \cdots + \beta_k \bar{x}_k \\ &= \bar{y}\,!! \end{aligned}$$

# Models for Forecasting (Cont'd)

- <u>Time Series Models</u>

  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:

$$f_{t,s} = \sum_{i=1}^{p} a_i f_{t,s-i} + \sum_{j=1}^{q} b_j u_{t+s-j}$$

where

$$\begin{aligned} f_{t,s} &= y_{t+s},\ s \leq 0 \\ u_{t+s} &= 0,\ s > 0 \\ &= u_{t+s},\ s \leq 0 \end{aligned}$$

# 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.

# Forecasting with MA Models (Cont'd)

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

$$
\begin{aligned}
f_{t,1} &= E(y_{t+1|t}) = \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} \\
f_{t,2} &= E(y_{t+2|t}) = E(\mu + \theta_1 u_{t+1} + \theta_2 u_t + \theta_3 u_{t-1} + u_{t+2} \mid \Omega_t) \\
&= E(y_{t+2|t}) = \mu + \theta_2 u_t + \theta_3 u_{t-1} \\
f_{t,3} &= E(y_{t+3|t}) = E(\mu + \theta_1 u_{t+2} + \theta_2 u_{t+1} + \theta_3 u_t + u_{t+3} \mid \Omega_t) \\
&= E(y_{t+3|t}) = \mu + \theta_3 u_t \\
f_{t,4} &= E(y_{t+4|t}) = \mu \\
f_{t,s} &= E(y_{t+s|t}) = \mu \ \forall\ s \geq 4
\end{aligned}
$$

# Forecasting with AR Models

- Say we have estimated an AR(2)

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

# Forecasting with AR Models (Cont'd)

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

# Forecasting with AR Models (Cont'd)

- We can see immediately that

$$
\begin{aligned}
f_{t,4} &= \mu + \phi_1 f_{t,3} + \phi_2 f_{t,2} \quad \text{etc, so} \\
f_{t,s} &= \mu + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2}
\end{aligned}
$$

- 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:

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

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

- *MAE* is given by

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

- Mean absolute percentage error:

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

- 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.

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

- A measure more closely correlated with profitability:

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

$$\text{where} \qquad \begin{aligned} z_{t+s} &= 1 \quad \text{if } (y_{t+s} f_{t,s}) > 0 \\ z_{t+s} &= 0 \quad \textit{otherwise} \end{aligned}$$

# 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"
- "data mining" issues
- simple versus complex models
- financial or economic theory

# Statistical Versus Economic or Financial loss functions

- Statistical evaluation metrics may not be appropriate.
- How well does the forecast perform in doing the job we wanted it for?

**Limits of forecasting: What can and cannot be forecast?**

- All statistical forecasting models are essentially extrapolative
- 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
  - anchoring
  - illusory patterns
  - "group-think".
- The Usually Optimal Approach

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