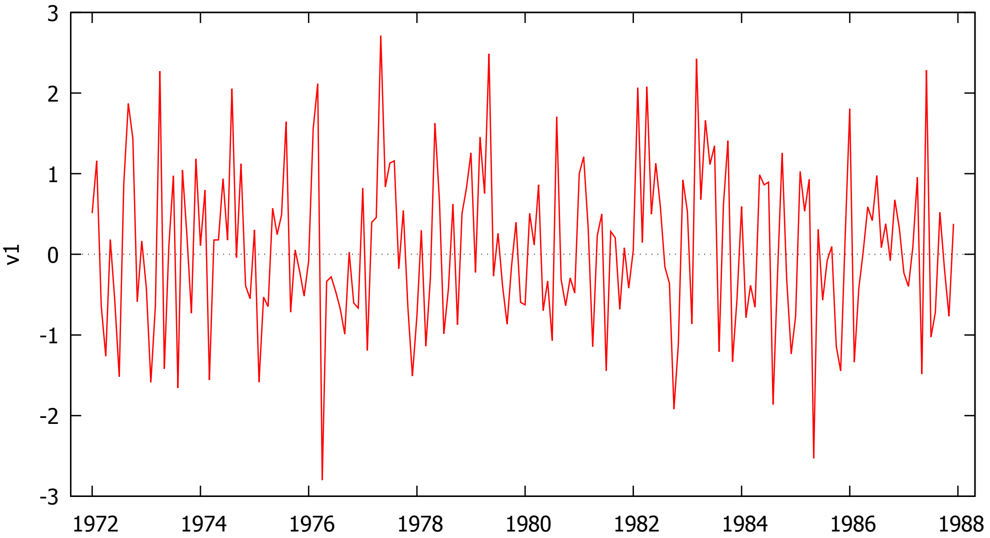


3
2
1
0
-1
-2
-3
v1
1972
1974
1976
1978
1980
1982
1984
1986
1988

From "C.W.J.Granger (1989): Forecasting in business and economics. Boston: Academic Press".

White Noise process (WN): purely random, no discernable structure or pattern.

Examples

You stand on the street corner and record the last digit of the license plates of passing cars.

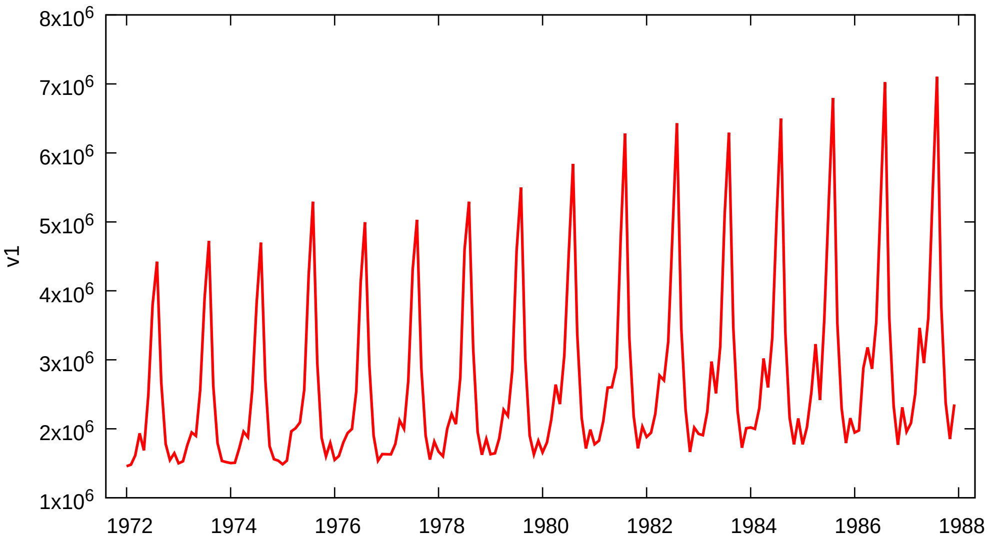
The last digit of the winning number at the "Lotteria di Capodanno".

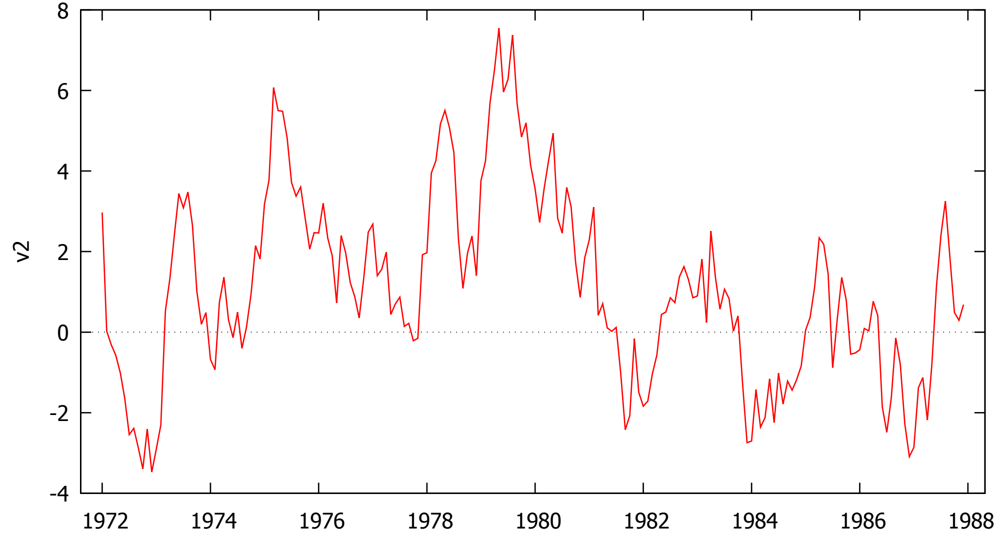
There is no correlation (covariance) between terms, and previous values do not help forecast future values.

Best forecast for future values: the mean (minim. expected loss function).

Zero mean white noise: $\{u_t\}$; best forecast: zero.

$E(u_t) = 0; \quad var(u_t) = \sigma^2; \quad cov(u_{t_1}, u_{t_2}) = 0; \quad$ best forecast $= 0$.


8x10^6
7x10^6
6x10^6
5x10^6
4x10^6
3x10^6
2x10^6
1x10^6
v1
1972
1974
1976
1978
1980
1982
1984
1986
1988



8
6
4
2
0
-2
-4
v2
1972
1974
1976
1978
1980
1982
1984
1986
1988

Zero mean White Noise (WN)

$u_t \quad u_{t-1} \quad u_{t-2} \quad u_{t-3} \quad u_{t-4} \quad u_{t-5} \quad u_{t-6} \quad u_{t-7} \quad u_{t-8} \ldots.$

There is no "relationship" (i.e. correlation) between any pair u_{t-i} and u_{t-j}. Therefore, if we combine two subsets of variables, there will be no relationship (i.e. correlation=0) if the two subsets are "disjoint". For example a combination of

u_t, u_{t-1} and u_{t-2} will be uncorrelated with a combination of

u_{t-5} and u_{t-6}

But if the two subsets have a common intersection, then there will be a correlation between the combinations; for example

$u_t + u_{t-1} + u_{t-2}$ will surely be correlated with

$$u_{t-2} + u_{t-3}$$

as u_{t-2} is common to both.

Moving average processes

MA(1): $y_t = \mu + u_t + \theta_1 u_{t-1}$ where u_t is zero mean WN

$u_t \quad u_{t-1} \quad u_{t-2} \quad u_{t-3} \quad u_{t-4} \quad u_{t-5} \quad u_{t-6} \quad u_{t-7} \quad u_{t-8}....$

$E(y_t) = \mu; \quad var(y_t) = (1+\theta_1^2)\sigma^2; \quad cov(y_t, y_{t-1}) = \theta_1\sigma^2; \quad cov(y_t, y_{t-2}) = 0;$

MA(2): $y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$

$E(y_t) = \mu; \quad var(y_t) = (1+\theta_1^2+\theta_2^2)\sigma^2; \quad cov(y_t, y_{t-1}) = (\theta_1 + \theta_1\theta_2)\sigma^2;$

$cov(y_t, y_{t-2}) = (\theta_1\theta_2)\sigma^2; \quad cov(y_t, y_{t-3}) = 0;$

MA(q): $y_t = \mu + u_t + \theta_1 u_{t-1} + ... + \theta_q u_{t-q}$

$E(y_t) = \mu; \quad var(y_t) = (1+\theta_1^2 + ... + \theta_q^2)\sigma^2; \quad cov(y_t, y_{t-1}) = ...;$

$cov(y_t, y_{t-2}) = ...;$ .... $cov(y_t, y_{t-q}) = ...;$ but $cov(y_t, y_{t-q-1}) = 0;$

Correlogram (autocorrelation function) “cuts off’ after lag $= q$.

Example: a maternity hospital

Every day the number of new patients that arrive is $100 + u_t$, where u_t is a zero mean WN.

Typically for 10% of the patients it is a day hospital, so they leave in the same day; 20% of the patients leave after one day, 30% after two days, 40% leave after three days.

Thus, the number of patients that leave the hospital in day t is

$N_t = 0.1(100 + u_t) + 0.2(100 + u_{t-1}) + 0.3(100 + u_{t-2}) + 0.4(100 + u_{t-3})$
$= 100 + 0.1u_t + 0.2u_{t-1} + 0.3u_{t-2} + 0.4u_{t-3}$ MA(3)

Today (t), very early in the morning, we want to predict how many patients wll leave the hospital.

Best forecast for today (t) is $= 100 + 0 + 0.2u_{t-1} + 0.3u_{t-2} + 0.4u_{t-3}$ as *today's* u_t is unpredictable, but the past values are known!

Best forecast for tomorrow ($t + 1$) is $= 100 + 0 + 0 + 0.3u_{t-1} + 0.4u_{t-2}$.

For the day after tomorrow ($t + 2$) is $= 100 + 0 + 0 + 0 + 0.4u_{t-1}$.

But, for all the following days the best forecast is $= 100$.

Autoregressive processes

AR(1): $y_t = \mu + \phi_1 y_{t-1} + u_t$

y_t depends on u_t and y_{t-1} which depends on u_{t-1} and y_{t-2} which depends on u_{t-2} and y_{t-3} which depends on u_{t-3} ....

thus, at any $lag = k$, y_t and y_{t-k} are "correlated".

Stationarity ($|\phi_1| < 1$):

$E(y_t) = E(y_{t-1}) = E(y_{t-2}) etc.$ and $var(y_t) = var(y_{t-1}) = var(y_{t-2}) etc.$

so that

$E(y_t) = \frac{\mu}{1-\phi_1}$; $var(y_t) = \frac{\sigma^2}{1-\phi_1^2}$;

$cov(y_t, y_{t-1}) = \phi_1 var(y_t)$; $cov(y_t, y_{t-2}) = \phi_1^2 var(y_t)$; $cov(y_t, y_{t-3}) = \phi_1^3 var(y_t)$;

AR(2): $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$

AR(p): $y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + u_t$

Autocovariances (autocorrelations) are "never" zero. Correlogram never "cuts off".

Diet.

I am slightly overweight. My dietician prescribes a diet. Following the diet prescriptions, the excess-weight should decrese 10% every month. In other words, every month the excess-weight should be 90% of that of the previous month.

The target-weight for me is 70 Kg. I suffer from an excess weight of 10 Kg: my weight is now 80 Kg.

With the diet, excess-weight next month should become 90% of $10 = 9$ Kg, so that my weight should become 79 Kg.

The following month the excess-weight should become 90% of $9 = 8.1$ Kg, so that my weight should become 78.1 Kg.

The following month the excess-weight should become 90% of $8.1 = 7.29$ Kg, so that my weight should become 77.29 Kg.

However.... Noise!

Excess-weight $y_t = 0.9\ y_{t-1} + u_t \qquad AR(1)$ process.

Partial autocorrelations

Partial autocorrelation of lag k measures correlation between y_t and y_{t-k}, after removing the effects of intermediate $y_{t-1}, y_{t-2}, ...$

For instance, an $AR(k)$ process

$$AR(k) : y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_k y_{t-k} + u_t$$

can be considered a linear regression model (autoregression) that shows the “direct connection” between y_t and y_{t-1}, between y_t and y_{t-2}, etc., till y_t and y_{t-k}.

But there is no “direct connection” between y_t and “lagged y” when the lag is more than k periods.

Partial autocorrelation of lag k is the coefficient ϕ_k. All partial autocorrelations of lags $> k$ are zero.

Partial correlogram (or partial autocorrelation function) “cuts off” after k lags.

Connections between AR and MA.

y_t are observed at any t, while the zero mean WN u_t are not observed

AR(1): $y_t = \mu + u_t + \phi_1 y_{t-1}$

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

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

$= \mu + u_t + \phi_1\{\mu + u_{t-1} + \phi_1[\mu + u_{t-2} + \phi_1(\mu + u_{t-3} + \phi_1 y_{t-4})]\}$

$= \mu + \phi_1\mu + \phi_1^2\mu + \phi_1^3\mu + ... + u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \phi_1^3 u_{t-3} + \phi_1^4 u_{t-4} + ...$

AR(1) can be represented as MA(∞) if $|\phi_1| < 1$ (stationary)

(particular case of Wold's decomposition theorem on stationary processes)

MA(1): $y_t = \mu + u_t + \theta_1 u_{t-1}$ thus $u_t = -\mu + y_t - \theta_1 u_{t-1}$

$= -\mu + y_t - \theta_1\{-\mu + y_{t-1} - \theta_1 u_{t-2}\}$

$= -\mu + y_t - \theta_1\{-\mu + y_{t-1} - \theta_1[-\mu + y_{t-2} - \theta_1 u_{t-3}]\}$

$= -\mu + y_t - \theta_1\{-\mu + y_{t-1} - \theta_1[-\mu + y_{t-2} - \theta_1(-\mu + y_{t-3} - \theta_1 u_{t-4})]\}$

$= -\mu + \theta_1\mu - \theta_1^2\mu + \theta_1^3\mu + ... + y_t - \theta_1 y_{t-1} + \theta_1^2 y_{t-2} - \theta_1^3 y_{t-3} + \theta_1^4 y_{t-4} + ...$

MA(1) can be represented as AR(∞) if $|\theta_1| < 1$ (invertible)

Mixed Autoregressive - Moving Average processes: $ARMA$

$ARMA(1,1)$: $y_t = \mu + \phi_1 y_{t-1} + u_t + \theta_1 u_{t-1}$

$ARMA(p,q)$: $y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + ... + \theta_q u_{t-q}$

Correlations and partial correlations are never zero.

Why use complex models like $ARMA$?

1) If neither correlogram nor partial correlogram "cut off" after some lags, "maybe" $ARMA$ could provide a "parsimonious" representation of the process.

2) Theorem: The sum of two independent processes $x_t \sim ARMA(p_1, q_1)$ and $w_t \sim ARMA(p_2, q_2)$ is $ARMA(r, s)$, where $r = p_1 + p_2$ and $s =$ "the larger" between $p_1 + q_2$ and $p_2 + q_1$.

For example, in the very simple case $x_t \sim AR(1) \sim ARMA(1, 0)$, but the variables x_t are not directly observable, because observed values are "contaminated" by some measurement error $w_t \sim WN \sim ARMA(0, 0)$, then the "observed" process $y_t = x_t + w_t$ is $ARMA(1, 1)$.

$x_t = \phi x_{t-1} + u_t$ and $w_t \sim WN$ then

$y_t = x_t + w_t$

$= \phi x_{t-1} + u_t + w_t$

$= \phi x_{t-1} + \phi w_{t-1} + u_t - \phi w_{t-1} + w_t$

$= \phi y_{t-1} + u_t \quad + \quad w_t - \phi w_{t-1}$

$AR(1) \qquad\qquad MA(1)$

Three problems.

1) Identification: given "observed" series (y_t) how can we realise if the underlying process is MA or AR (of some order) or none of them (Box-Jenkins 1970)?

2) Estimation: if the process is MA or AR (of some order), how can we estimate the coefficients (θ and ϕ), the variance of u_t, the variances and covariances of all coefficients for tests and diagnostic?

3) Forecast: If the process is MA or AR (of some order), what would be the optimum forecast one-step ahead, two-steps ahead, etc.?

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, γ_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 τ_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

$$E(y_t) = \mu$$

$$\mathrm{var}(y_t) = \sigma^2$$

$$\gamma_{t-r} = \begin{cases} \sigma^2 & \text{if} \quad t = r \\ 0 & \textit{otherwise} \end{cases}$$

- Thus the autocorrelation function will be zero apart from a single peak of 1 at s=0. $\hat{\tau}_s \sim \textit{approx.}\ N(0, 1/T)$ 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 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^2_m.
- However, the Box Pierce test has poor small sample properties, so a variant has been developed, called the Ljung-Box statistic:

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

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

Moving Average Processes

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

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

  is a qth 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 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).

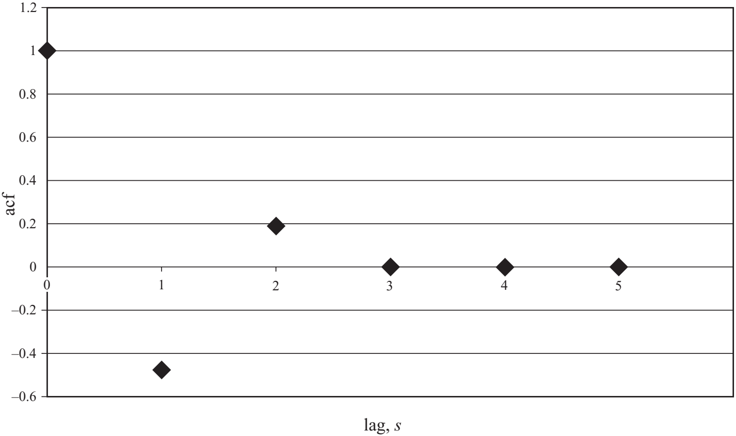
   iii. If $\theta_1 = -0.5$ and $\theta_2 = 0.25$, sketch the acf of X_t.

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(∞) 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(∞).
- 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.

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 $y_{t-k+1}, y_{t-k+2}, \ldots, y_{t-1}$
- 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 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(∞), 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

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

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

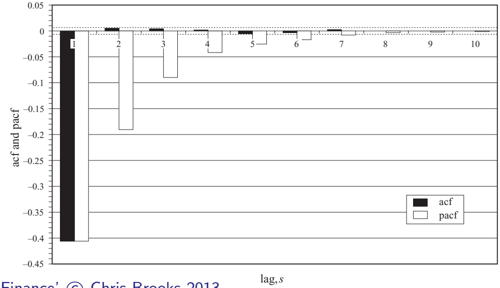
A moving average process has

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

Some sample acf and pacf plots for standard processes

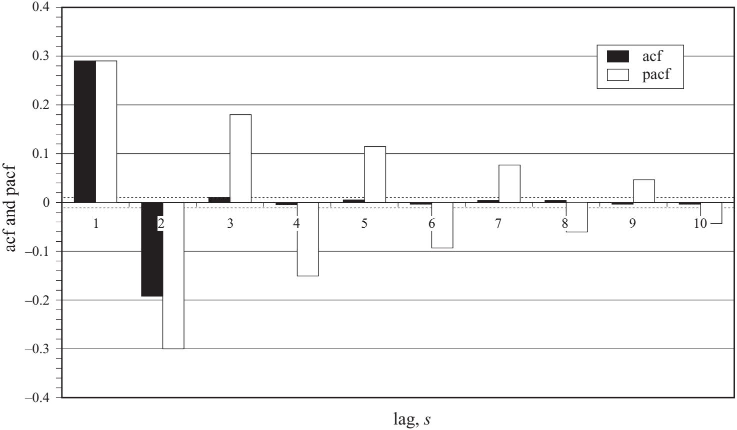
- The acf and pacf are not produced analytically from the relevant formulae for a model of that type, but rather are estimated using 100,000 simulated observations with disturbances drawn from a normal distribution.

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

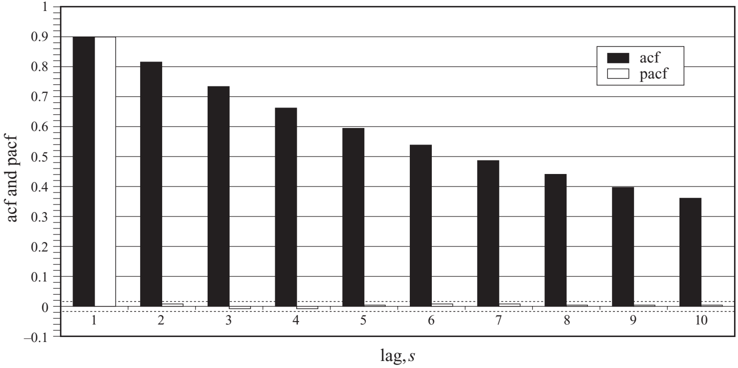


ACF and PACF for an MA(2) Model:

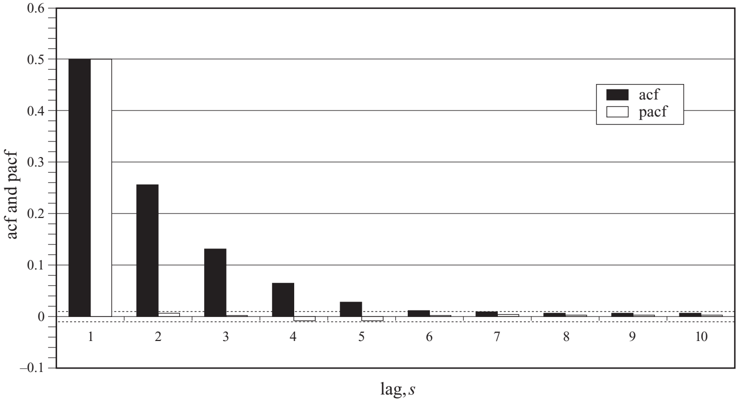
$$y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$$



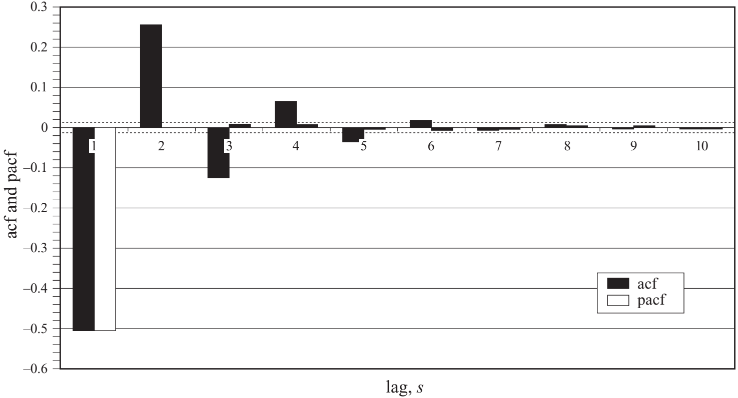
ACF and PACF for a slowly decaying AR(1) Model: $y_t = 0.9y_{t-1} + u_t$



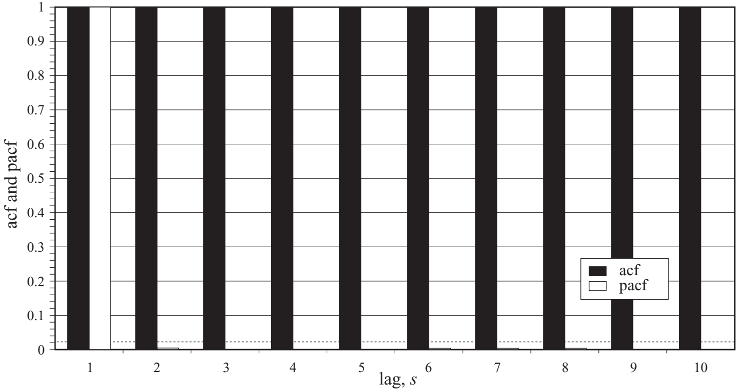
ACF and PACF for a more rapidly decaying AR(1) Model: $y_t = 0.5y_{t-1} + u_t$



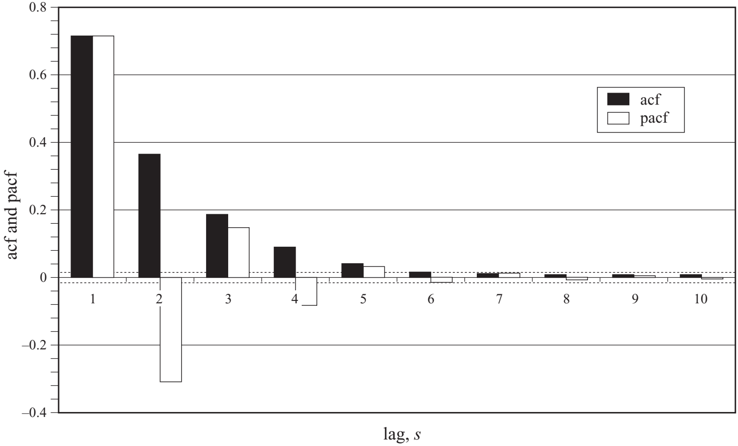
ACF and PACF for a more rapidly decaying AR(1) Model with Negative Coefficient: $y_t = -0.5y_{t-1} + u_t$



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).

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

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

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

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

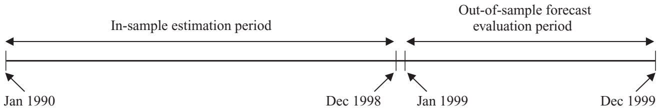
We can distinguish two approaches:

- 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 $f(y_{t+s}) = \bar{y}$

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.