Sociedad Ecuatoriana de Estadística

Univariate time series modelling and forecasting

Andrés Peña M.

Univariate Time Series Models

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

Some Notation and Concepts

- <u>A Stationary Process</u>

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

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

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 the covariance function.
- The covariances, $\boldsymbol{\gamma_s}$, 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:

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

- 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

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

- Thus the autocorrelation function will be zero apart from a **single peak of 1** at s = 0. $\tau_s \sim$ 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 $\pm .196 \times \frac{1}{\sqrt{T}}$. 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 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:

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

  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.

  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

- 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)= σ_ε^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) = γ_0 = $(1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2)\sigma^2$

  Covariances

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

Example of an MA Problem

1. 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

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

$$\mathrm{Var}(X_t) = E[X_t - E(X_t)][X_t - E(X_t)]$$

but $E(X_t) = 0$, so

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

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

Solution (cont'd)

So $\mathrm{Var}(X_t) = \gamma_0 = \mathrm{E}\,[u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2]$

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

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

(ii) The acf of X_t.

$\gamma_1 \quad = \mathrm{E}[X_t - \mathrm{E}(X_t)][X_{t-1} - \mathrm{E}(X_{t-1})]$

$= \mathrm{E}[X_t][X_{t-1}]$

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

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

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

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

Solution (cont'd)

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

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

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

Solution (cont'd)

We have the autocovariances, now calculate the autocorrelations:

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

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

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

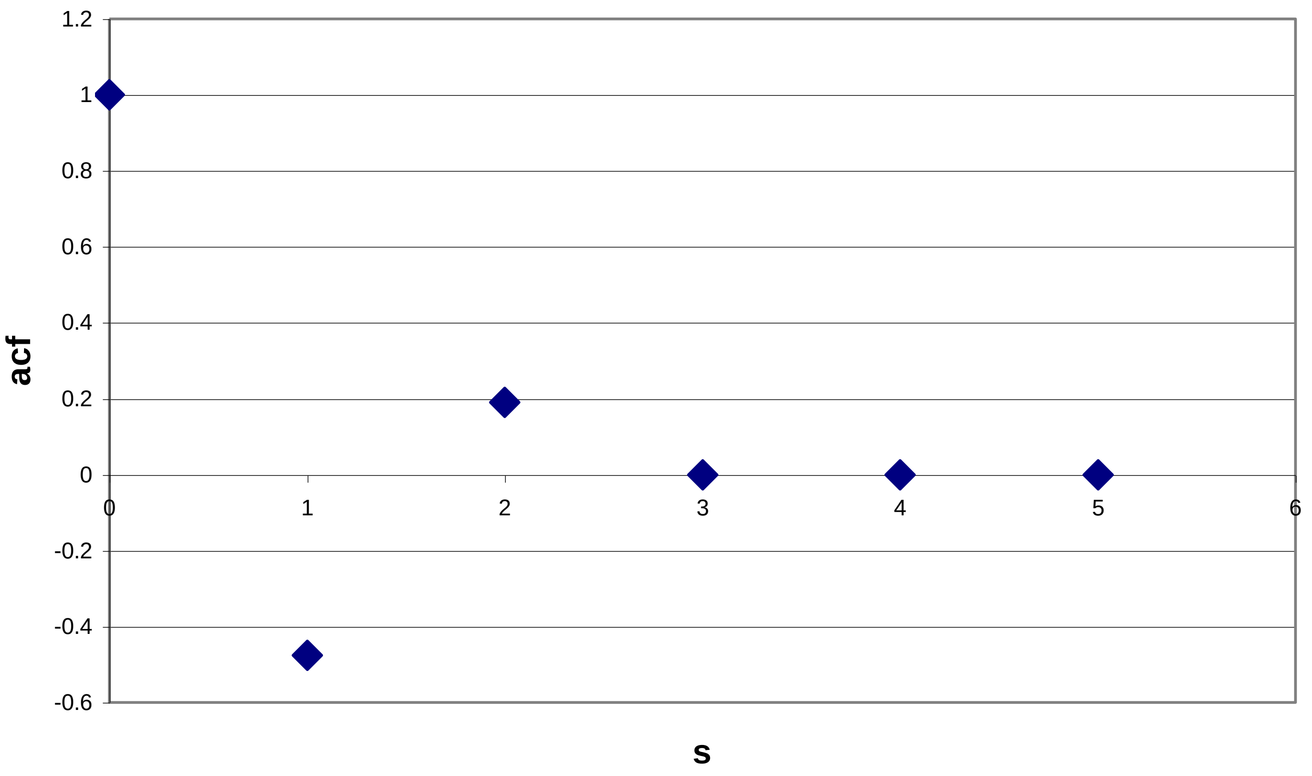
$$\tau_3 = \frac{\gamma_3}{\gamma_0} = 0$$

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

(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 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:

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

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

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

or $\quad \phi(L) y_t = \mu + u_t \qquad$ 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**.
- 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.

The Moments of an Autoregressive Process

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

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

- 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 autocorrelation function **will decay exponentially to zero**.

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(∞), 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

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

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

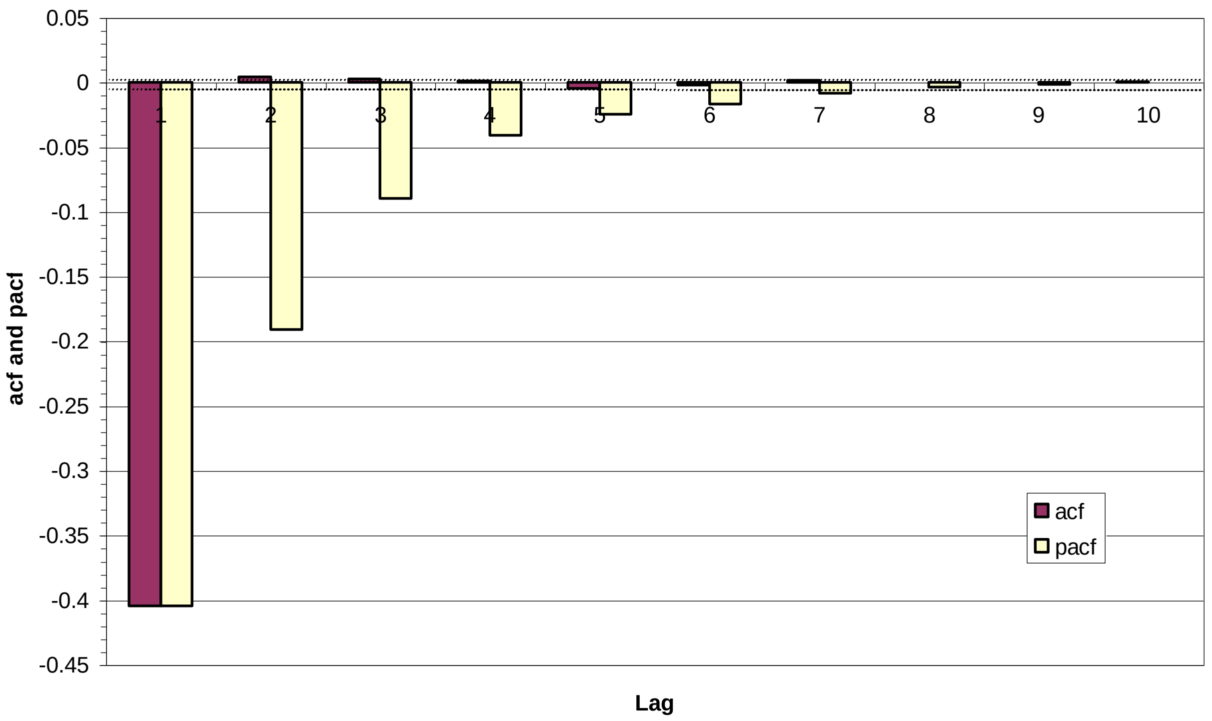
A **moving average process** has

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

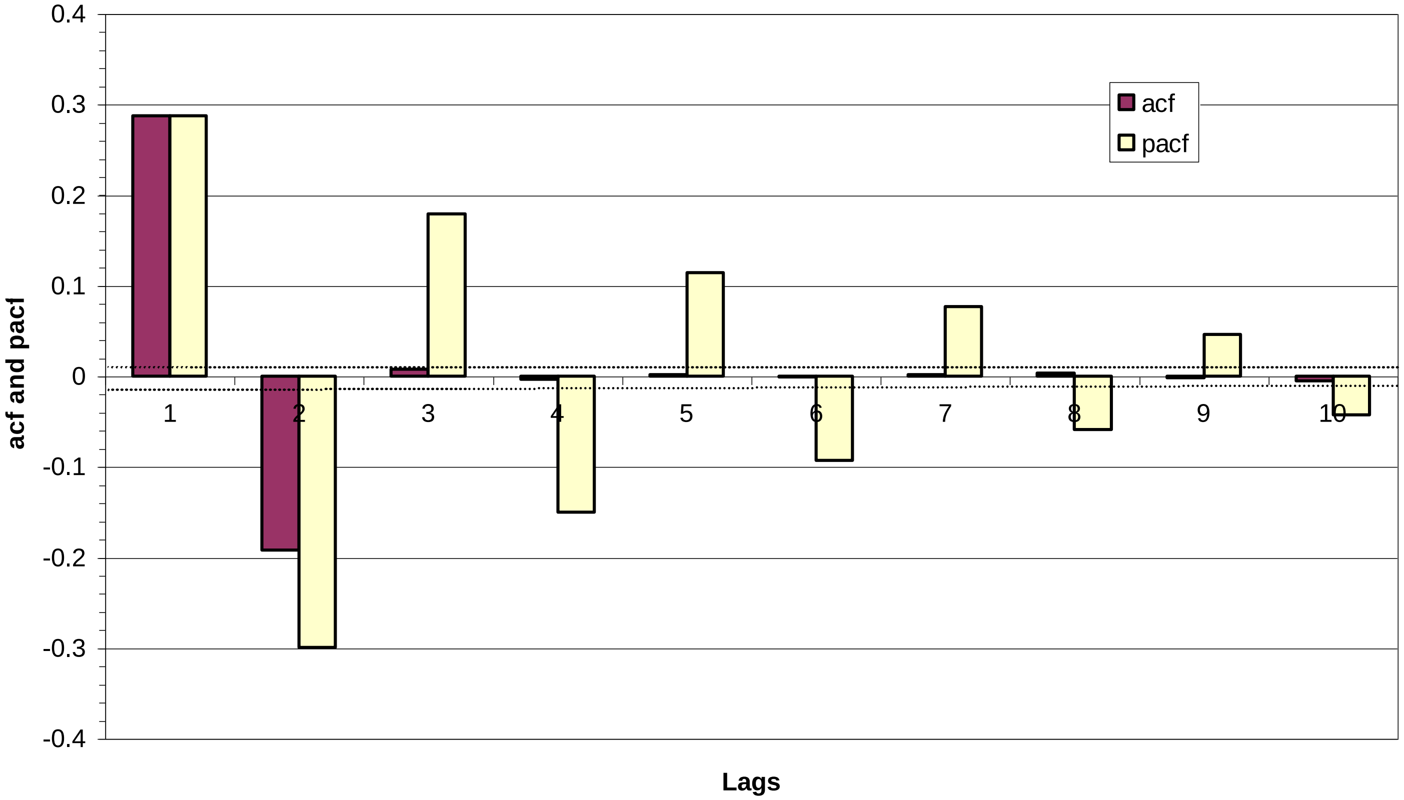
Some sample acf and pacf plots for standard processes

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.

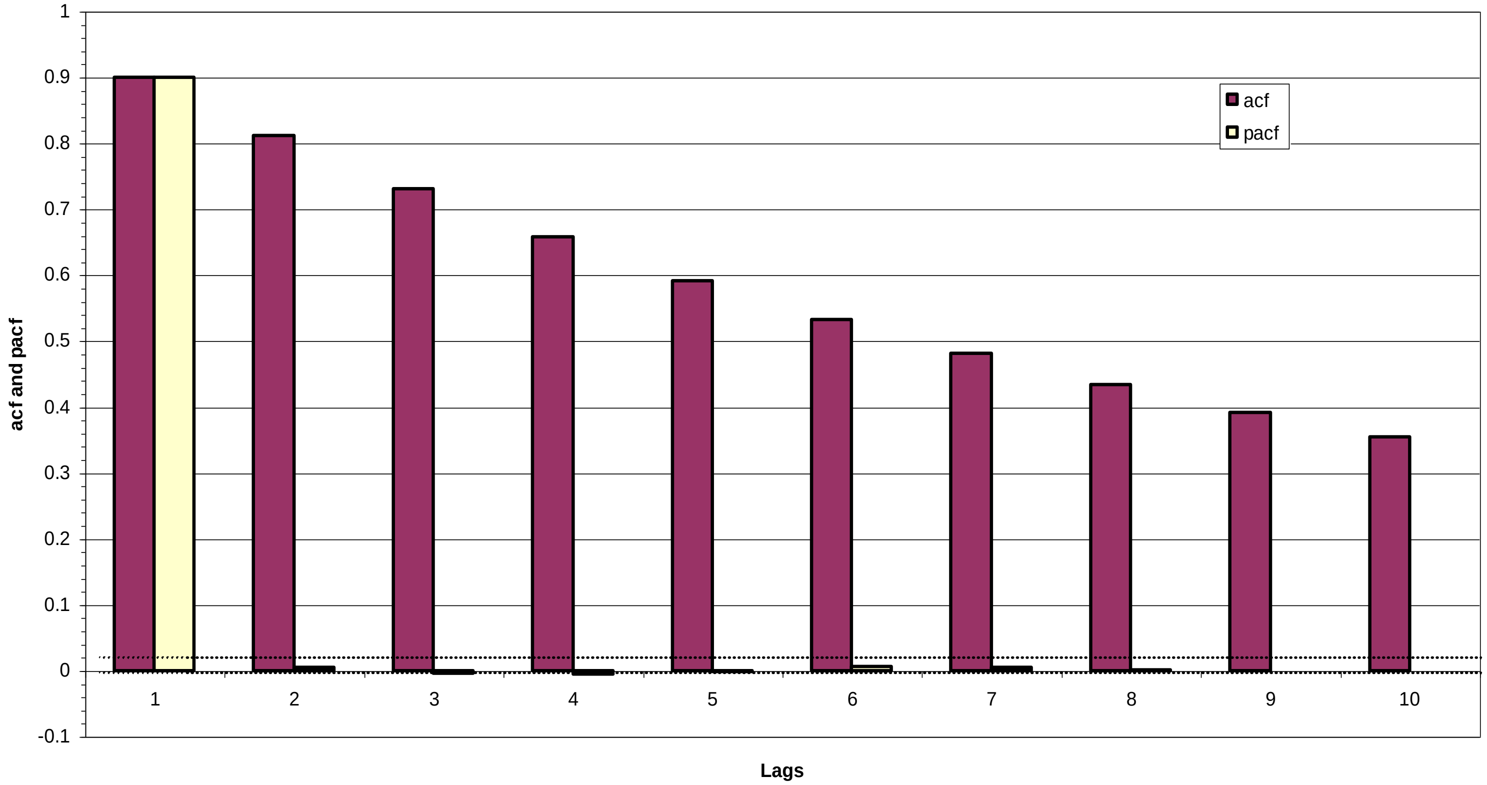
ACF and PACF for an MA(1) Model: $y_t = -0.5u_{t-1} + u_t$



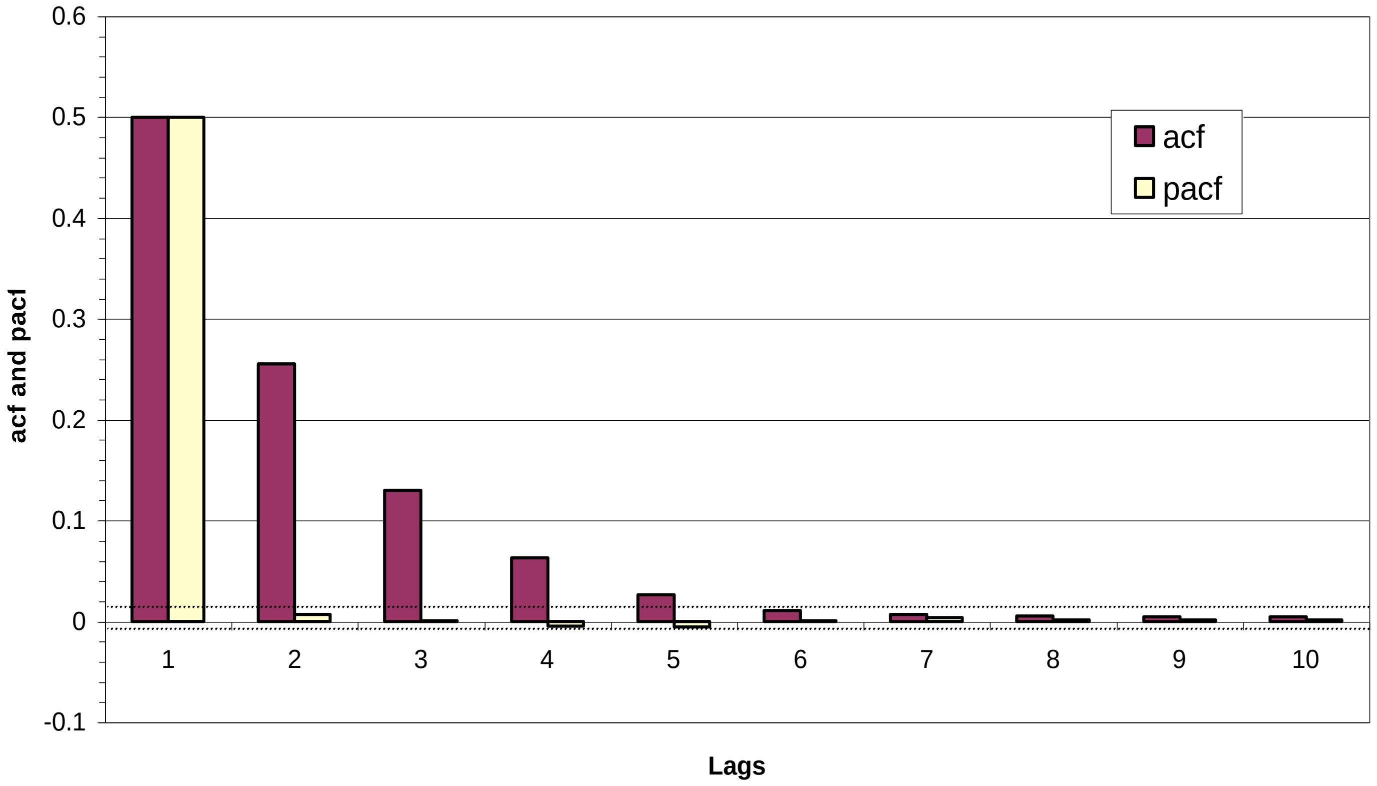
ACF and PACF for an MA(2) Model: $y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$



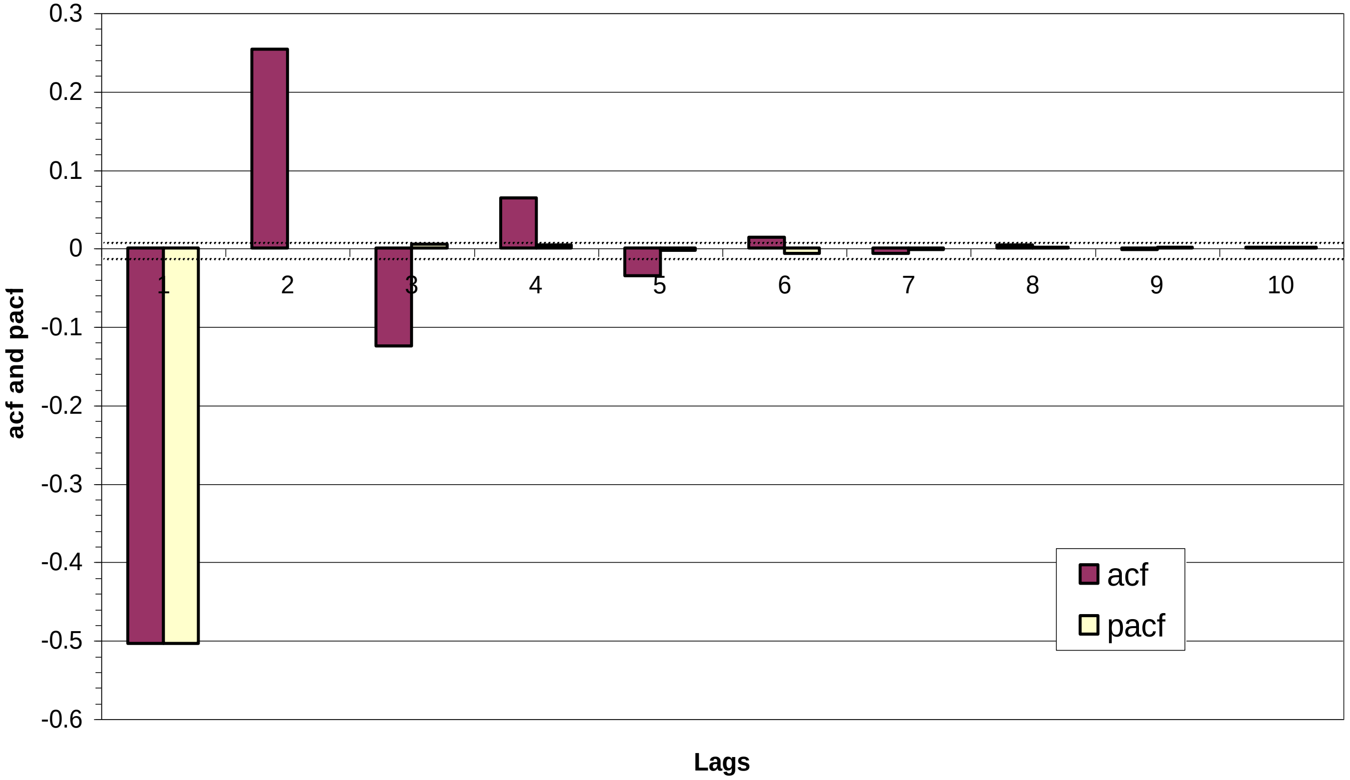
ACF and PACF for a slowly decaying AR(1) Model: $y_t = 0.9y_{t-1} + u_t$



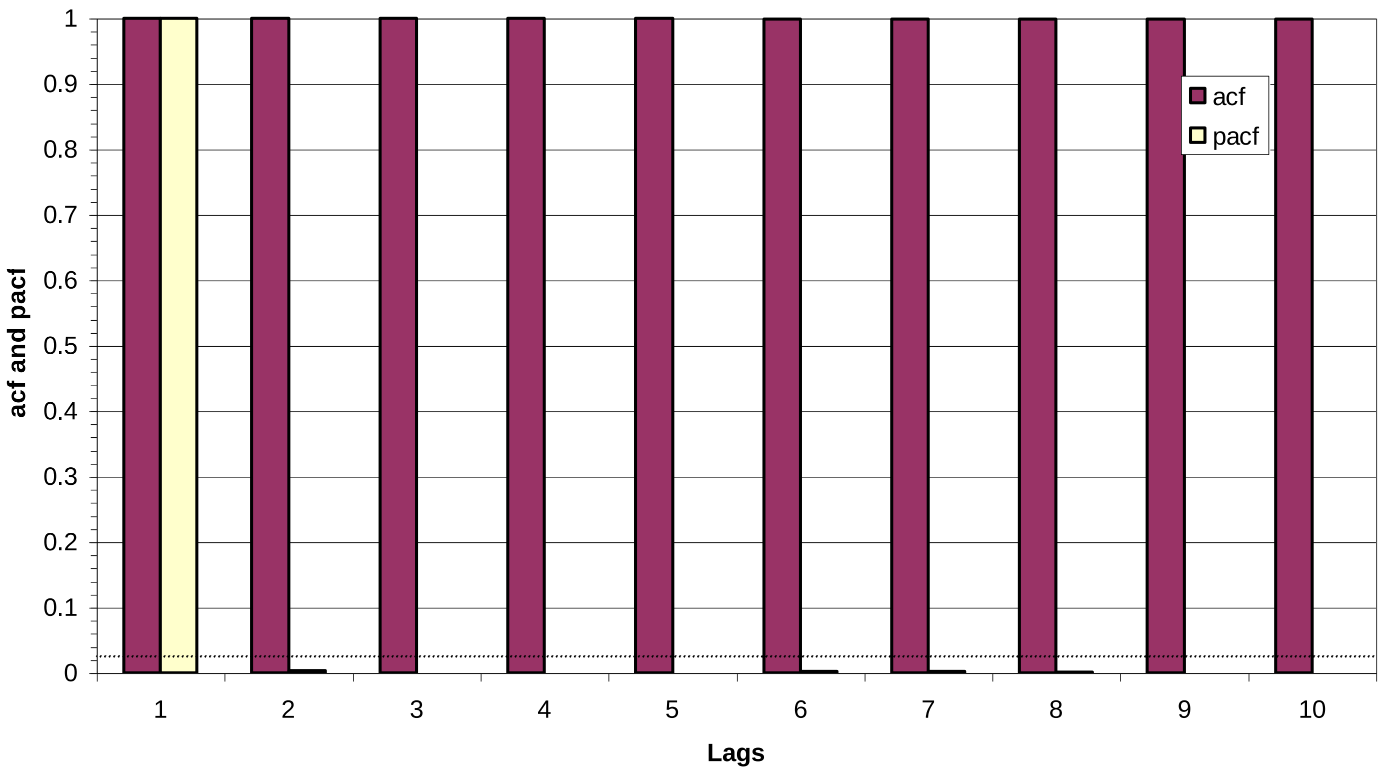
ACF and PACF for a more rapidly decaying AR(1) Model: $y_t = 0.5y_{t-1} + u_t$



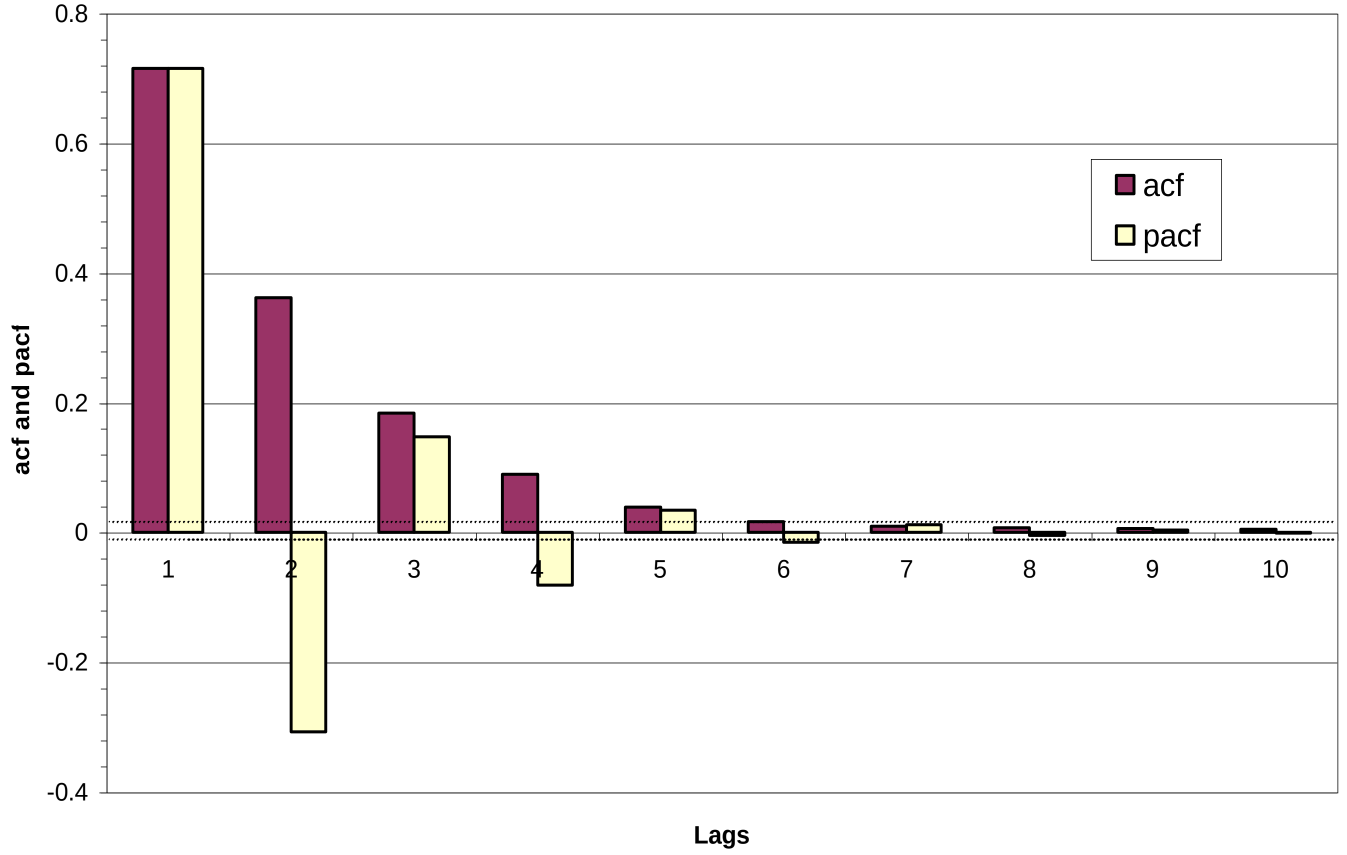
ACF and PACF for a more rapidly decaying AR(1) Model with Negative Coefficient: $y_t = -0.5y_{t-1} + u_t$



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



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

<u>Step 1:</u>

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

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

Step 2:

- 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).

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

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

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

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

  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.

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

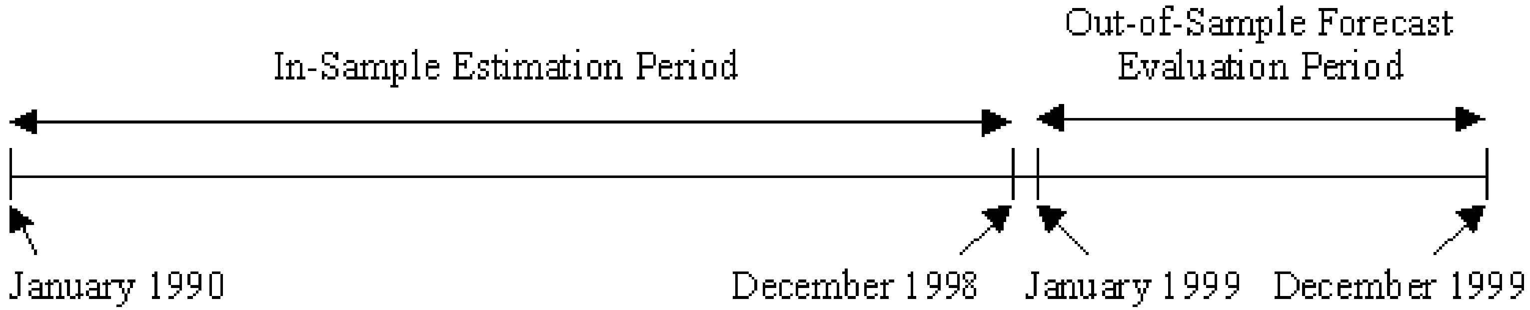
- We can distinguish two approaches:
  - Econometric (structural) forecasting
  - Time series forecasting

- 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.

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.

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:

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

MAE is given by

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

Mean absolute percentage error:

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

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.