III Box-Jenkins Methods

1. Pros and Cons of ARIMA Forecasting

    a) need for data

at least 50 and preferably 100 observations should be available to build a proper model

used most frequently for hourly or daily data, but with application to many high frequency cases and some useful quarterly applications

why? need a long series of data without structural change

particularly serious problem if dealing with seasonal data

in three years of monthly data, we have 36 observations, but,

only three observations on each of the monthly seasons

Therefore, the data costs are usually high

b) no automatic updating

unlike simple naive models or smoothing models there is no automatic updating feature

as new data become available the entire modelling procedure must be repeated

especially the diagnostic checking stage as the model may have broken down

the fact that these models tend to be unstable cause many economists to view them with suspicion

c) high cost

because of the large data requirements, the lack of convenient updating procedures, and the fact that they must be estimated using nonlinear estimation procedures, the B-J models tend to be high cost.

The amount of subjective input at the identification stage also make them somewhat more of an art than a science

d) unstable

the ARIMA model tend to be unstable, both with respect to changes in observations and changes in model specification.

many specification will yield no results, and as with most nonlinear estimation techniques, the results may not be unique

e) it works for short run

Main advantage -- for short-run forecasts with high frequency data the results may be hard to beat.

We will show in the final class of the term than these models are equivalent to others under special conditions (i.e., structural econometric, or exponential smoothing)

They also have the advantage of being less sensitive to the underlying assumptions of the nature of the data fluctuations than many other systems

However, while the general form will handle many functional forms, the specific form identified must match the actual data closely

2. Types of Time Series Processes

a) random walk

$$Y_t = \beta_o + Y_{t-1} + \varepsilon_t$$

b) white noise

$$Y_t = \beta_o + \varepsilon_t$$

c) stationary

a time series is said to be stationary if it is invariant with respect to time

thus it must have a constant mean over any time horizon

it must have a constant variance

the covariance between all lags of equal length must be the same

A white noise series is similar but requires in addition that the covariance between any pair of lags for a white noise series is zero, for stationarity all that is required is a constant value zero or not.

d) moving average

MA(1)

$$Y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

note difference to autoregressive form in econometrics

MA(q)

$$Y_t = \delta + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$$

e) autoregressive

AR(1)

$$Y_t = \delta + \varepsilon_t + \varphi_1 Y_{t-1}$$

AR(p)

$$Y_t = \delta + \varepsilon_t + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p}$$

f) mixed ARIMA

ARMA(p,q)

$$Y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p}$$

ARIMA(p,d,q)

is an ARMA model as given above, with d differences taken first to render the series stationary.

3. Autocorrelation and Partial Autocorrelation Functions

How does one decide on the appropriate functional form?

How can we tell if the series is stationary?

Key is to examine the autocorrelation (ACF) & partial autocorrelation (PACF) functions

define the autocorrelation function (ACF) with lag k as:

$COR(Y_t, Y_{t-k})$

measured by:

$$\rho_k = \frac{COV(Y_t, Y_{t+k})}{\sigma_{y_t} \sigma_{y_{t+k}}}$$

the theoretical autocorrelation function is generally unknown, but may be estimated from the sample autocorrelation function as follows:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{T} (Y_t - \overline{Y})^2}, \quad k = 1, 2, \ldots, K$$

where t is the length of the time series under study.

Normally one computes the first $K < N/4$ sample autocorrelations.

If we define the covariance between Y_t and Y_{t+k} as γ_k, then:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

thus for any stochastic process $\rho_0 = 1$.

It is also true that $\rho_k = \rho_{-k}$.

For a series to be white noise all $\rho_k = 0$, for $k > 0$. This may be tested using the Box-Pierce test statistic:

$$Q = T \sum_{k=1}^{K} \hat{\rho}_k^2$$

which is approximately distributed as chi square with K degrees of freedom
Thus a value below the critical value would lead to acceptance of white noise

If a series is stationary, the sample autocorrelations will tail off quickly as k increases.

The expression tails off means that the function decays in an exponential, sinusoidal, or geometric fashion, with a relatively large number of nonzero values.

A function is said to cut off if the function truncates abruptly with only a few nonzero values.

Formal tests of stationarity i.e., "unit root" tests were covered in the topic on co-integration in the econometric forecasting section of the course.

The partial autocorrelation function can be thought of as the simple autocorrelation between two random variables in a conditional distribution.

It is also possible to show that the partial autocorrelation coefficient φ_{kk} is the kth coefficient in an autoregressive process of order k

$$\hat{\rho}_j = \varphi_{k1}\hat{\rho}_{j-1} + \varphi_{k2}\hat{\rho}_{j-2} + \ldots + \varphi_{kk}\hat{\rho}_{j-k}, \; j = 1, 2, \ldots, k$$

The estimated values of the φ_{kk}'s is the partial autocorrelation function

Alternatively the values of the partial autocorrelation function may be estimated from the equation:

$$Y_t = b_o + b_1 Y_{t-1} + b_2 Y_{t-2} + \ldots + b_k Y_{t-k}$$

Here the partial autocorrelation coefficient φ_{kk} is estimated by b_k.

4. Four Key Steps in Forecasting Using the B-J Method

a) identification

Identification involves looking at the ACF and the PACF to decide what functional form best fits the data.

If the ACF does not tail off quickly, the series is likely not stationary and some transformation is required.

Take repeated differences, if the problems is a trend

try log transformation or other, if the problem is related to changing variance

The following table gives general rule for identification of the likely model

Model	ACF	PACF
AR(p)	Tails off	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off
ARMA(p,q)	Tails off	Tails off

For the ARMA(p,q) process the tail off will show a significant "cut" in the PACF after lag p and in the ACF after lag q

Frequently the estimated functions are not clear cut and several models may seem appropriate.

Thus we may estimate several, rejecting those that fail the checking stage and finally picking the acceptable model with the fewest parameters (parsimony)

b) estimation

The functional form of the ARIMA model is highly nonlinear due to the presence of the moving average terms.

Like any nonlinear estimation routine, starting values must be given.

In SHAZAM starting value of zero with the restrict option forces these coefficients to be constrained to zero.

As with any nonlinear estimator, the results may not be unique.

Estimation problems such as a failure to get convergence frequently results is the data series is not stationary.

c) diagnostic checking

models must be check for validity before the forecasts are generated

if the model passes then we proceed, if not

must return to the identification stage and pick the next most probable model

What is required?

check coefficients for significance

the coefficients of order p and q should be significant (interior need not be)

residuals should (must) be white noise

A test of whether the residuals form a white process is given by a modified version of the Box-Pierce Q statistic discussed above in the form:

$$Q = (T - d) \sum_{k=1}^{K} \hat{\rho}_k^2$$

where rho hat is the autocorrelations of the residuals, d is the order of differencing to obtain a stationary series, T is the length of the series, and K is the number of autocorrelations being checked.

Here if Q is larger than the critical value for the chi squared distribution with K-p-q degrees of freedom, the model should considered inadequate.

d) forecasting

Forecasting with this system is straight forward, the forecast is the expected value, evaluated at a particular point in time.

Confidence intervals may also be easily derived from the standard errors of the residuals.

Details given below with examples for specific functional forms.

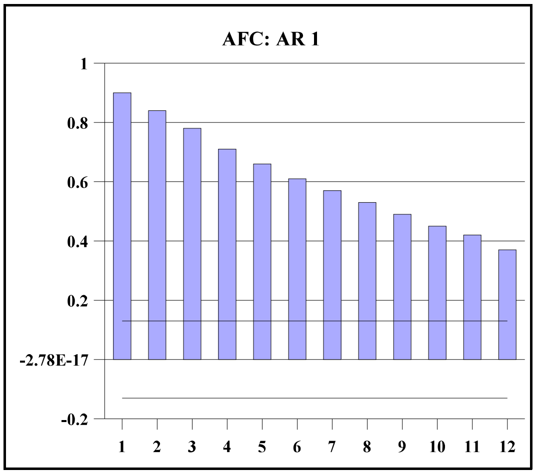
5. Examples of the Full Process

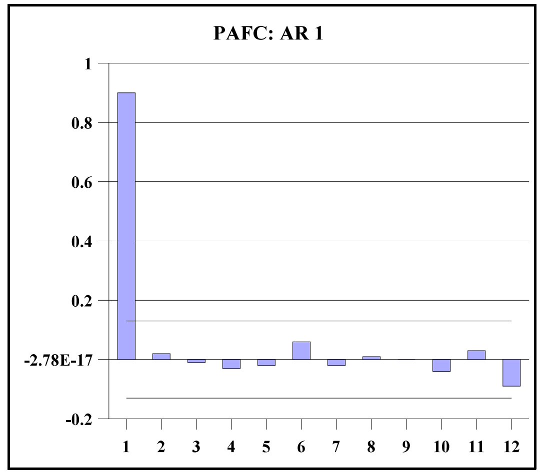
a) Autoregressive models

Autoregressive models are identified with an ACF that tails off and a PACF that cuts off after p lags.

The estimated values of ρ_k may be used as efficient starting values for the estimation process.

Consider a simple AR(1) process:





The basic model is given by:

$$Y_t = \varphi_1 Y_{t-1} + \delta + \varepsilon_t$$

a one-period ahead forecast here is given by:

$$\hat{Y}_{t+1} = E[\,Y_{t+1} \mid Y_t, \ldots, Y_1\,]$$
$$= \varphi_1 Y_t + \delta$$

a two-period ahead forecast is given by:

$$\hat{Y}_{t+2} = \varphi_1 \hat{Y}_{t+1} + \delta$$
$$= \varphi_1 (\varphi_1 Y_t + \delta) + \delta$$
$$= \varphi_1^2 Y_t + (\varphi_1 + 1)\delta$$

a h-period ahead forecast would be:

$$\hat{Y}_{t+h} = \varphi_1^h Y_t + (\varphi_1^{h-1} + \varphi_1^{h-2} + \ldots + \varphi_1 + 1)\delta$$

$$\lim_{h \to \infty} \hat{Y}_{t+h} = \frac{\delta}{1 - \varphi_1}$$

This limiting forecast is the mean of Y_t. Therefore, if Y_t is stationary, $\varphi_1 < 1$ is a necessary condition. Values of φ_1 outside of the unit circle will result in an unstable result.

For higher order AR processes, e.g., order 2, the mean of Y is:

$$\overline{Y_t} = \frac{\delta}{1 - \varphi_1 - \varphi_2}$$

and thus the necessary condition for stationarity is:

$$\varphi_1 + \varphi_2 < 1$$

Confidence intervals

What is the forecast error in this case? For AR(1):

$$e_{t+h} = Y_{t+h} - \hat{Y}_{t+h}$$

$$= \varepsilon_{t+h} + \varphi_1 \varepsilon_{t+h-1} + \ldots + \varphi_1^{h-1} \varepsilon_{t+1}$$

$$E[e_{t+h}] = 0$$

the forecast error variance is:

$$E[e_t^2] = (1 + \varphi_1^2 + \varphi_1^4 + \ldots + \varphi_1^{2h-2})\sigma_\varepsilon^2$$

Thus the confidence interval expands non-linearly as t increase.

The actual 95% confidence interval is centred on the expected value plus or minus 1.96 times the square root of the forecast error variance.

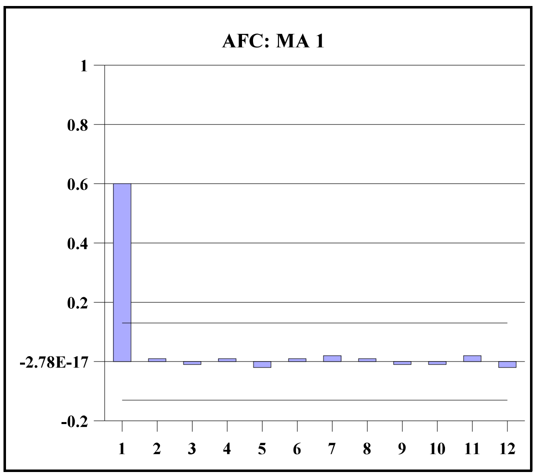
In conclusion we find than for AR(1) models the forecast decays towards the mean of the series as the length of the forecast horizon increases, and our confidence in the forecasts decline rapidly.

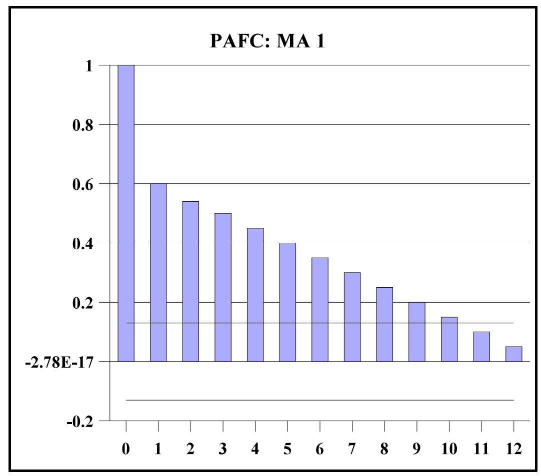
b) moving average models

Moving average models are identified with an ACF that cuts off after q periods and a PACF that tails off.

The estimated values of ρ_{kk} may not be efficient starting values for the estimation process. This results from the fact the moving average component is highly nonlinear and thus ad hoc guesses are used. Values close to zero are generally preferred.

Consider a simple MA(1) process:





The basic model is given by:

$$Y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

again a one-period ahead forecast is given by:

$$\hat{Y}_{t+1} = E[Y_{t+1} \mid Y_t, \ldots, Y_1]$$
$$= \delta - \theta_1 \varepsilon_t$$

For a MA(1) process a h-period ahead forecast, for h > 1 is:

$$\hat{Y}_{t+h} = \delta$$

Thus for more than q periods ahead, the forecast always equals the mean of the series.

Confidence bands

$$E[e_{t+h}^2] = E[(Y_{t+h} - \hat{Y}_{t+h})^2]$$
$$= E[(\varepsilon_{t+h} - \theta_1 \varepsilon_{t+h-1})^2] = (1 + \theta_1^2)\sigma_\varepsilon^2$$

Thus the confidence bands widen for q periods and then remain constant there after.

c) The ARMA(1,1) process

This is a combination of the AR(1) and the MA(1) models

The combined impact lasts for one period and then the AR process continues

Again the forecasts decay to the mean of the series as h increases and the confidence interval increases nonlinearly.

d) The ARIMA(1,1,0) process

This is the same as the AR(1) model, expect that the modelling is done on the first differences of the original series.

Here as h increases, the forecast becomes the mean of the trend of the series and the confidence intervals increase even more rapidly than is the case for the simple AR(1) model.

e) The ARIMA(0,1,1) process

This process is the same as the exponential smoothing model for $\alpha = 1 - \theta$

General Conclusion:

A MA(q) process has a memory of only q periods

An AR(1) process has an infinite memory, but, only recent observations have a large impact.

Therefore, ARIMA models are best used for short-term forecasting where:

$$h \leq p + q$$

and since simple processes are preferred, that usually means one or two periods ahead.

Seasonal ARIMA Models

ARIMA processes can be extended to handle seasonal variations.

Seasonal differencing: $Y_t - Y_{t-4}$, may be used to remove nonstationarity caused by moving seasonality in the data.

The ACF and PACF may be used to identify the order of the process in the standard manner. Look for spikes at seasonal lags.

Notation in the form: ARIMA(p,d,q)(P,D,Q), where p, d, and q are the order of the AR, level of differencing and MA from the regular process and P, D, Q are the corresponding values from the seasonal process.

Example of specification of ARIMA(1,0,2)(1,0,0) for quarterly data:

$$\hat{Y}_{t+1} = \delta + \varphi_1 Y_t + \varphi_{s1} Y_{t-3} - \theta_1 \varepsilon_t - \theta_2 \varepsilon_{t-1} + \varepsilon_{t+1}$$

Advanced Topics in Time Series Analysis

Invertibility and Stationarity

It can be shown that if the AR(1) process is stationary, it is equivalent to a moving average process of infinite order (and thus with infinite memory).

In fact, for any stationary autoregressive process of any order there exists an equivalent moving average process of infinite order and thus the autoregressive process is invertible into a moving average process.

Similarly, if certain invertibility conditions are met, any finite order moving average process has an equivalent autoregressive process of infinite order.

If forecasts are to be useful, the processes must be both stationary and invertible.

Example: MA(1)

$$\begin{aligned} Y_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &= (1 - \theta_1 B)\varepsilon_t \end{aligned}$$

$$\textit{or solving for } \varepsilon_t \textit{ yields:}$$

$$\varepsilon_t = (1 - \theta_1 B)^{-1} Y_t$$

Now if $|\theta_1| < 1$, we may write this as:

$$\varepsilon_t = \left(\sum_{j=0}^{\infty} \theta_1^j B^j\right) Y_t$$

$$or$$

$$\varepsilon_t = (1 + \theta_1 B^1 + \theta_1^2 B^2 + \ldots) Y_t$$

which is an infinite order autoregressive process with weights $\varphi_j = -\theta_1^j$.

Here we have inverted the MA(1) process to an AR(∞) process.

For an MA(2) process the invertibility conditions are:

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$|\theta_2| < 1$$

Note the similarity to the stationarity conditions for an AR(2) process:

$$\varphi_1 + \varphi_2 < 1$$
$$\varphi_2 - \varphi_1 < 1$$
$$|\varphi_2| < 1$$

The AR(p) process is stationary only if the roots of $\Phi_p(B) = 0$ lie outside the unit circle but is invertible for all values of the weights $\{\varphi_i\}$.

The MA(q) process is stationary regardless of the values of the weights $\{\theta_i\}$ but is invertible only if the roots of $\Theta_q(B) = 0$ lie outside the unit circle.

ARIMA Process for Sum of Series (Aggregation models)

Why are mixed processes likely to be found in business and economic applications?

Consider the following theorem:

If

$$x_t \sim ARMA(p_1, q_1),\; y_t \sim ARMA(p_2, q_2),$$

x_t, y_t are independent series, and $z_t = x_t + y_t$ then typically

$$z_t \sim ARMA(P, Q),$$

where $P = p_1 + p_2$, and $Q = p_1 + q_2$ or $p_2 + q_1$, whichever is larger.

If both x_t and y_t are AR(1), then z_t would be ARMA(2,1).

Only if all components are MA would aggregates not be mixed processes.

If the true economic series is AR(p) but is measured with white noise error, MA(0), then the observed process will appear to be ARMA(p,p).

Transfer Functions

A transfer function model relates a dependent variable to lagged values of itself, current and lagged values of one or more independent variables, and an error term that is partially explained by a time-series model. It is of the form:

$$Y_t = \nu^{-1}(B)\,\omega(B)\,X_t + \varphi^{-1}(B)\,\theta(B)\,\eta_t$$

All parameters should be estimated simultaneously, but this is computationally difficult an is frequently done in stages. i.e., causal model first, then ARIMA model of the residuals.

Vector Autocorrelations (VAR) and Vector ARIMA Models

Most of the ARIMA models considered have been univariate, involving only one variable.

For a univariate autoregression of order p, one would regress a variable on p lags of itself.

A multi-variate autoregression, or a vector autoregression, or a VAR, could involve N variables.

In an N-variable vector autoregression of order p, one would regress the relevant left-hand-side variable on p lags of itself, and p lags of every other variable.

The key point is that, in contrast to the univariate case, vector autoregressions allow for cross-variable dynamics.

Consider a 2-variable VAR(1):

$$Y_{1,t} = \varphi_{11} Y_{1,t-1} + \varphi_{12} Y_{2,t-1} + \varepsilon_{1,t}$$
$$Y_{2,t} = \varphi_{21} Y_{1,t-1} + \varphi_{22} Y_{2,t-1} + \varepsilon_{2,t}$$

The disturbance variance-covariance structure is:

$$\varepsilon_{1,t} \sim WN(0, \sigma_1^2)$$
$$\varepsilon_{2,t} \sim WN(0, \sigma_2^2)$$
$$cov(\varepsilon_{1,t}, \varepsilon_{2,t} = \sigma_{12}$$

These systems are easy to estimate by OLS or seemingly unrelated regression, SHAZAM (system command)

Vector ARIMA models would be highly non-linear and a real mess to estimate.

Thus even though theoretically better Vector arma models are seldom used.