

ARIMA Modelling and Forecasting

Introduction

- Describe the stationarity of the AR process
- Determine the mean and variance of the AR process
- Assess the importance of Box-Jenkins methodology
- Describe the various types of forecasts
- Evaluate the measures of forecast importance

Stationarity of the AR process

- If an AR model is not stationary, this implies that previous values of the error term will have a non-declining effect on the current value of the dependent variable.
- This implies that the coefficients on the MA process would not converge to zero as the lag length increases.
- For an AR model to be stationary, the coefficients on the corresponding MA process decline with lag length, converging on 0.

AR Process

- The test for stationarity in an AR model (with p lags) is that the roots of the characteristic equation lie outside the unit circle (i.e. > 1), where the characteristic equation is:

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

Unit Root

- When testing for stationarity for any variable, we describe it as testing for a 'unit root', this is based on this same idea.
- The most basic AR model is the AR(1) model, on which most tests for stationarity are based, such as the Dickey-Fuller test (covered later)

Box-Jenkins Methodology

- This is a method for estimating ARIMA models, based on the ACF and PACF as a means of determining the stationarity of the variable in question and the lag lengths of the ARIMA model.
- Although the ACF and PACF methods for determining the lag length in an ARIMA model are commonly used, there are other methods termed information criteria which can also be used (covered later)

Box-Jenkins

- The Box-Jenkins approach typically comprises four parts:
 - Identification of the model
 - Estimation, usually OLS
 - Diagnostic checking (mostly for autocorrelation)
 - Forecasting

ACF and PACF

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *.	. *.	1	0.0930	0.0930	7.580	0.384
. * .	. * .	2	-0.073	-0.082	1.2307	0.540
. .	. .	3	-0.023	-0.008	1.2772	0.735
. * .	. * .	4	-0.106	-0.111	2.3010	0.681
. .	. .	5	-0.003	0.0162	3.017	0.806
. *.	. *.	6	0.1400	0.1244	4.1334	0.659
. *.	. *.	7	0.1620	0.1416	6.6260	0.469
. * .	. * .	8	-0.103	-0.128	7.6356	0.470
. .	. .	9	-0.047	-0.003	7.8513	0.549
. * .	. * .	10	-0.124	-0.117	9.3624	0.498
. .	. *.	11	0.0280	0.0889	4.434	0.581
. **	. *.	12	0.2170	0.1661	4.231	0.286

Identification

- Identification of the most appropriate model is the most important part of the process, where it becomes as much 'art' as 'science'.
- The first step is to determine if the variable is stationary, this can be done with the correlogram. If it is not stationary it needs to be first-differenced. (it may need to be differenced again to induce stationarity)
- The next stage is to determine the p and q in the ARIMA (p, I, q) model (the I refers to how many times the data needs to be differenced to produce a stationary series)

Identification

- To determine the appropriate lag structure in the AR part of the model, the PACF or Partial correlogram is used, where the number of non-zero points of the PACF determine where the AR lags need to be included.
- To determine the MA lag structure, the ACF or correlogram is used, again the non-zero points suggest where the lags should be included.
- Seasonal dummy variables may also need to be included if the data exhibits seasonal effects.

Diagnostic Checks

- With this approach we only test for autocorrelation usually, using the Q or Ljung-Box statistic.
- If there is evidence of autocorrelation, we need to go back to the identification stage and respecify the model, by adding more lags.
- A criticism of this approach is that it fails to identify if the model is too big or over-parameterised, it only tells us if it is too small.

ARIMA Example

- Following the Box-Jenkins methodology, the following ARIMA(2,1,1) model was produced:

$$\Delta \hat{y}_t = 0.7 + 0.6\Delta y_{t-1} + 0.3\Delta y_{t-2} + 0.1u_{t-1}$$

(0.1) (0.2) (0.1) (0.01)

$$\overline{R}^2 = 0.12, LB(2) = 3.21$$

Forecasting

- One of the most important tests of how well a model performs is how well it forecasts.
- One of the most useful models for forecasting is the ARIMA model.
- To produce dynamic forecasts the model needs to include lags of either the variables or error terms.

Types of Forecast

- Forecasts can be either in-sample or out-of-sample forecasts.
- In general the out-of sample forecasts are a better test of how well the model works, as the forecast uses data not included in the estimation of the model.
- To conduct out-of-sample forecasts, we need to leave some observations at the end of our sample for this purpose

Types of Forecasts

- A one-step-ahead is a forecast for the next observation only.
- A multi-step-ahead forecast is for $1, 2, 3, \dots, s$ steps ahead.
- A recursive window for the forecast means that the initial estimation date is fixed but the additional observations are added one by one to the estimation time span.
- A rolling window is where the estimation time period is fixed but the start and end dates successively increase by 1.

Conditional Forecasting

- A conditional expectation is one that is taken for time $t + 1$, conditional upon or given all information available up to and including time t (this is important later). It is written as:

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

Ω_t – all information at time t

Measuring Forecast Accuracy

- To determine how accurate a forecast is, the simplest method is to plot the forecast against the actual values as a direct comparison
- In addition it may be worthwhile to compare the turning points, this is particularly important in finance.
- There are a number of methods to determine accuracy of the forecast, often more than one is included in a set of results.

Tests of Forecast Accuracy

- Tests of forecast accuracy are based on the difference between the forecast of the variables value at time t and the actual value at time t . The closer the two are together and the smaller the forecast error, the better the forecast.
- There are a variety of statistics measuring this accuracy, mostly based on an average of the errors between the actual and forecast values at time t .
- However these statistics provide little information in isolation, as they are unbounded from above and their value depends on the units of the variables being forecast.

Conclusion

- When using AR models, whether the series is stationary or not determine how stable it is.
- Box-Jenkins methodology is part art and part science.
- Forecasting of time series is an important measure of how well a model works
- There are many measures of how accurate a forecast is, usually a number are calculated to determine if the forecast is acceptable, although all have faults.