

# Box-Jenkins modelling

Rob J Hyndman

25 May 2001

---

The Box-Jenkins approach to modelling ARIMA processes was described in a highly influential book by statisticians George Box and Gwilym Jenkins in 1970. An ARIMA process is a mathematical model used for forecasting. Box-Jenkins modelling involves identifying an appropriate ARIMA process, fitting it to the data, and then using the fitted model for forecasting. One of the attractive features of the Box-Jenkins approach to forecasting is that ARIMA processes are a very rich class of possible models and it is usually possible to find a process which provides an adequate description to the data.

The original Box-Jenkins modelling procedure involved an iterative three-stage process of model selection, parameter estimation and model checking. Recent explanations of the process (e.g., Makridakis, Wheelwright and Hyndman, 1998) often add a preliminary stage of data preparation and a final stage of model application (or forecasting).

- 1 **Data preparation** involves transformations and differencing. Transformations of the data (such as square roots or logarithms) can help stabilize the variance in a series where the variation changes with the level. This often happens with business and economic data. Then the data are *differenced* until there are no obvious patterns such as trend or seasonality left in the data. "Differencing" means taking the difference between consecutive observations, or between observations a year apart. The differenced data are often easier to model than the original data.
- 2 **Model selection** in the Box-Jenkins framework uses various graphs based on the transformed and differenced data to try to identify potential ARIMA processes which might provide a good fit to the data. Later developments have led to other model selection tools such as Akaike's Information Criterion.
- 3 **Parameter estimation** means finding the values of the model coefficients which provide the best fit to the data. There are sophisticated computational algorithms designed to do this.
- 4 **Model checking** involves testing the assumptions of the model to identify any areas where the model is inadequate. If the model is found to be inadequate, it is necessary to go back to Step 2 and try to identify a better model.
- 5 **Forecasting** is what the whole procedure is designed to accomplish. Once the model has been selected, estimated and checked, it is usually a straight forward task to compute forecasts. Of course, this is done by computer.

Although originally designed for modelling time series with ARIMA processes, the underlying strategy of Box and Jenkins is applicable to a wide variety of statistical modelling situations. It provides a convenient framework which allows an analyst to think about the data, and to find an appropriate statistical model which can be used to help answer relevant questions about the data.

See entry for ARIMA processes.

**References**

- BOX, G.E.P. and G.M. JENKINS (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.
- MAKRIDAKIS, S., S.C. WHEELWRIGHT, and R.J. HYNDMAN (1998) *Forecasting: methods and applications*, New York: John Wiley & Sons.
- PANKRATZ, A. (1983) *Forecasting with univariate Box-Jenkins models: concepts and cases*, New York: John Wiley & Sons.