

## 3.6: Model Selection

In this section, a rough guide for going about the data analysis will be provided. It consists of several parts, most of which have been discussed previously. The main focus is on the selection of  $p$  and  $q$  in the likely case that these parameters are unknown.

**Step 1.** Plot the data and check whether or not the variability remains reasonably stable throughout the observation period. If that is not the case, use preliminary transformations to stabilize the variance. One popular class is given by the *Box-Cox transformations* (Box and Cox, 1964)

$$f_{\lambda}(U_t) = \begin{cases} \lambda^{-1}(U_t^{\lambda} - 1), & U_t \geq 0, \lambda > 0. \\ \ln U_t & U_t > 0, \lambda = 0. \end{cases}$$

In practice  $f_0$  or  $f_{1/2}$  are often adequate choices. (Recall, for instance, the Australian wine sales data of Example 1.4.1.)

**Step 2.** Remove, if present, trend and seasonal components from the data. Chapter 1 introduced a number of tools to do so, based on the classical decomposition of a time series

$$Y_t = m_t + s_t + X_t$$

into a trend, a seasonality and a residual component. Note that differencing works also without the specific representation in the last display. If the data appears stationary, move on to the next step. Else apply, for example, another set of difference operations.

**Step 3.** Suppose now that Steps 1 and 2 have provided us with observations that are well described by a stationary sequence  $(X_t: t \in \mathbb{Z})$ . The goal is then to find the most appropriate ARMA( $p, q$ ) model to describe the process. In the unlikely case that  $p$  and  $q$  can be assumed known, utilize the estimation procedures of Section 3.5 directly. Otherwise, choose them according to one of the following criteria.

(a) The standard criterion that is typically implemented in software packages is a modification of *Akaike's information criterion*, see Akaike (1969), which was given by Hurvich and Tsai (1989). In this paper it is suggested that the ARMA model parameters be chosen to minimize the objective function

$$\text{AIC}_C(\phi, \theta, p, q) = -2 \ln L(\phi, \theta, S(\phi, \theta)/n) + \frac{2(p+q+1)n}{n-p-q-2}. \quad (3.6.1)$$

Here,  $L(\phi, \theta, \sigma^2)$  denotes the Gaussian likelihood defined in (3.5.4) and  $S(\phi, \theta)$  is the weighted sum of squares in (3.5.5). It can be seen from the definition that the  $\text{AIC}_C$  does not attempt to minimize the log-likelihood function directly. The introduction of the penalty term on the right-hand side of (3.6.1) reduces the risk of overfitting.

(b) For pure autoregressive processes, Akaike (1969) introduced criterion that is based on a minimization of the *final prediction error*. Here, the order  $p$  is chosen as the minimizer of the objective function

$$\text{FPE} = \hat{\sigma}^2 \frac{n+p}{n-p},$$

where  $\hat{\sigma}^2$  denotes the MLE of the unknown noise variance  $\sigma^2$ . For more on this topic and other procedures that help fit a model, we refer here to Section 9.3 of Brockwell and Davis (1991).

**Step 4.** The last step in the analysis is concerned with *diagnostic checking* by applying the goodness of fit tests of Section 1.5.

This page titled [3.6: Model Selection](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Alexander Aue](#).