Chapter 3, Part V: More on Model Identification; Examples

Automatic Model Identification Through AIC

As mentioned earlier, there is a clear need for automatic, objective methods of identifying the best ARMA model for the data at hand. Objective methods become particularly crucial when trained experts in model building are not available. Furthermore, even for experts, objective methods provide a very useful additional tool, since the correlogram and partial correlogram do not always point clearly to a single best model.

Perhaps the most widely used objective method is AIC (Akaike Information Criterion, see Akaike (1974) [1]). To use AIC, we evaluate a statistic, AIC(p,q), for each ARMA(p,q) model under consideration. We then select the model which gives the smallest value of AIC.

The AIC criterion is defined as

$$AIC(p,q) = -2\log\,[\text{likelihood}(p,q)] + 2(p+q) \quad .$$

Here, likelihood(p,q) denotes the maximized value of the likelihood function for an ARMA(p,q) model. Roughly speaking, the larger likelihood(p,q), the better the ARMA(p,q) model fits the data. Thus, the best fitting models are the ones which give the *smallest* values of $-2\log\,[\text{likelihood}(p,q)]$. It follows that this first term of AIC evaluates the *fidelity* of the model. The second term of AIC, $2(p+q)$, constitutes a *penalty function* to guard against overfitting (i.e., using too many parameters). Thus, AIC provides an automatic implementation of Box-Jenkins' advice that we should use the most parsimonious model which adequately fits the data.

AIC has been widely tested and found to provide good model choices. However, it is by no means foolproof, and should be used *in conjunction* with the correlogram and partial correlogram, if possible. In particular, if these diagrams point to a more parsimonious model than that selected by AIC, the more parsimonious model should be used. The correlogram is also a crucial tool for diagnosing non-stationarity, a task for which AIC is of no use whatsoever. If an ARMA model is to be identified on

1 Akaike, H. (1974), "A New Look at the Statistical Model Identification," IEEE Trans. Autom. Control, AC-19, 716-723.

differenced data, then AIC (as well as the correlogram and partial correlogram) must of course be based on the differenced (not raw) data.

When Are Ordinary and Partial Correlations Significantly Different From Zero?

If the ordinary and partial correlogram are to be used for model identification, we must have a practical way of deciding where these diagrams cut off. In an MA(q) model, for example, the *theoretical* correlogram cuts off after lag q, so that all subsequent values are exactly zero. In practice, though, the theoretical correlation function is unknown and the best we can do is to plot the *sample* correlation function. Beyond lag q, these sample correlations will not be exactly zero, but they will be small. The key question is, *how* small will they be likely to be? A simple guideline is the following:

> A sample correlation or partial correlation is judged to be significantly different from zero if it lies outside the band $\pm 2/\sqrt{n}$, where n is the sample size.

The justification for this rule is that if the theoretical correlation (or partial correlation) is zero, then the sample correlation (or partial correlation) will be a random variable with a standard error of *approximately* $1/\sqrt{n}$. Accordingly, it is unlikely (probability .05) that the sample correlation will be outside the band $\pm 2/\sqrt{n}$ if the true correlation is indeed zero. Consequently, if the sample value *does* lie outside the band, we may conclude (with a high degree of confidence) that the true theoretical value is not zero.

Keep in mind that, although the method just described provides a useful rule of thumb, it should not be used blindly. Its theoretical underpinnings are only approximate, since the theoretical standard error is only approximately $1/\sqrt{n}$. Furthermore, if we examine a whole sequence of sample correlations, then we must remember that the probability .05 quoted above refers only to a given prespecified *individual* correlation. It does not apply *simultaneously* to a whole sequence of lags. The probability that *some* sample correlation will lie outside the band $\pm 2/\sqrt{n}$ is substantially larger than .05, even if all the corresponding theoretical values are zero.

Numerical Examples

Model identification is carried out for six series (with lengths ranging from 200 to 322) in Granger, pp. 68-74. Five of the six series were computer-generated. Granger identified all models from the correlogram and partial correlogram. In addition to these diagrams, we will use AIC for some series. Note that here, AIC is put at a considerable (and artificial) disadvantage since it is based only on the last 50 terms of the series, given by Granger in Table 3.2. Meanwhile, the correlogram and partial correlogram (Table 3.3) are based on the full series. Since interpretation of these diagrams is based on subjective judgements, some of the descriptions given here will differ from Granger's.

For Series E, the correlations (given for lags 1 to 12 in Table 3.3) are all within the ± 2 Standard Error bound, $(\pm .14)$. Thus, none are significantly different from zero, and we therefore identify the model as white noise. The partial correlations are also essentially zero for all lags shown, indicating that the correct model is AR(0), which is just another name for the white noise model. In truth, series E *is* white noise, so the identification was correct.

For series F, neither the correlation function nor the partial correlation function seems to cut off at some small value. Indeed, the lag 5 correlation is -.15, and the lag 4 partial correlation is -.14. (Meanwhile, Granger chooses to ignore the lag 5 correlation and hence identifies the model as MA(1).) For additional guidance, we turn to AIC, whose values are given here:

$AIC(p\ ,q\ )$ **For Series F**

	$q=0$	$q=1$	$q=2$
$p=0$		210.9	212.9
$p=1$	212.0	212.9	220.5
$p=2$	213.2	214.2	221.5

The smallest value of AIC is 210.9, for (p,q)=(0,1). This suggests an MA(1) as the appropriate model. Since the correlation function is essentially consistent with this conclusion (except at lag 5), we can identify the model as MA(1). Again, this turns out to be the correct identification.

For Series G, the correlation function clearly dies down, while the partial correlation function clearly cuts off beyond lag 1. Thus, we can identify the model as AR(1). Examination of AIC gives contradictory results (but remember that here, AIC was based only on the final 50 terms of the series):

$AIC(p\ ,q\ )$ **For Series G**

	$q=0$	$q=1$	$q=2$
$p=0$		203.5	205.3
$p=1$	207.9	205.3	216.1
$p=2$	207.3	215.7	228.0

According to AIC, the best model is MA(1). Since this conclusion is clearly contradicted by the correlation function, we stick with the AR(1). This is in fact the correct choice. Furthermore if AIC were evaluated on the *full* series, it would have been very likely to identify the correct model as well.

For series H, the correlation function dies down slowly, while the partial correlation function cuts off beyond lag 2. Thus, we may identify the model as AR(2). Note, though, that we might suspect non-stationarity since the correlations die down so slowly. Thus, we might have tried differencing the series. The correlation and partial correlation functions both seem to cut off beyond lag 1, giving the contrary indications that that the differences are MA(1) or AR(1). AIC values for the differences are given next:

$AIC(p\ ,q)$ **For Differences Of Series H**

	$q=0$	$q=1$	$q=2$
$p=0$		197.5	199.3
$p=1$	198.6	222.6	231.5
$p=2$	213.5	223.3	237.3

According to AIC, the two best models are MA(1) and AR(1). Thus, we are left with three reasonable models for the original series: AR(2), ARIMA(0,1,1) (if the first difference is MA(1)), and ARIMA(1,1,0) (if the first difference is AR(1)). In fact, the correct model was AR(2), but Granger shows that the ARIMA(1,1,0) choice leads to the fitted model

$$x_t - x_{t-1} = -.23(x_{t-1} - x_{t-2}) + \varepsilon_t \quad ,$$

or

$$x_t = .77x_{t-1} + .23x_{t-2} + \varepsilon_t \quad ,$$

which is almost identical to the fitted AR(2) model

$$x_t = .73x_{t-1} + .19x_{t-2} + \varepsilon_t \quad .$$

For series J, the correlation function dies down *extremely* slowly, staying quite close to 1, and a difference is clearly required. For the differences, the correlations are fairly small beyond lag 1, except for a large value at lag 9. The partial correlations for the differences seem to cut off beyond lag 1, so the differences seem to be AR(1). Thus, we identify the original series as ARIMA(1,1,0). On the other hand, Granger identifies the difference series as MA(1) (ignoring the lag-9 correlation) and thus identifies the original series as ARIMA(0,1,1). This is, in fact, the correct model.

Series K is the U.S. Index of Industrial Production, monthly data from Jan. 1948 to Oct. 1974. The data are seasonally adjusted (see next Chapter for details). The correlation function shows that a difference is required. For the differences, the correlation function dies down while the partial correlation function cuts off beyond lag 1 (except for a large value at lag 12, which is probably an artifact of the seasonal adjustment). We thus identify the differences as AR(1), and the original series as

ARIMA(1,1,0). Since Series J consists of real data, we cannot say whether this model is correct.