

## 7 Fitting an ARIMA model

### 7.1 The partial autocorrelation function

To provide a tool for helping to determine what model may be fitted to an observed time series, we introduce the concept of partial autocorrelation.

**Definition 7.1.1 (Partial Autocorrelation)**

Given a stationary process  $\{Y_t\}$ , for  $u \geq 1$ , the *partial autocorrelation at lag  $u$* ,  $\phi_{uu}$ , is defined as the correlation between  $Y_t$  and  $Y_{t-u}$  when the intervening variables,  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-u+1}$ , are held constant — or as the correlation between  $Y_t$  and  $Y_{t-u}$  adjusted for the intervening variables. The sequence  $\{\phi_{uu} : u \geq 1\}$  is known as the *partial autocorrelation function (pacf)*.  $\square$

Note that, because there are no intervening variables between  $Y_t$  and  $Y_{t-1}$ , the partial autocorrelation at lag 1 is equal to the autocorrelation at lag 1, i.e.,  $\phi_{11} = \rho_1$ .

### 7.2 The partial autocorrelation function for an $AR(p)$ process

The usefulness of the partial autocorrelation function as a tool for model selection is shown by considering its structure in the case of the  $AR(p)$  model,

$$Y_t = \sum_{k=1}^p \phi_k Y_{t-k} + \epsilon_t. \quad (1)$$

For  $u > p$ ,  $Y_t$  is not directly expressed in terms of  $Y_{t-u}$ . Hence we may deduce that for this model

$$\phi_{uu} = 0, \quad u > p.$$

Setting  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p+1}$  to be constant, we obtain

$$Y_t = c + \phi_p Y_{t-p} + \epsilon_t,$$

where  $c$  is a constant. Taking the unit of time to be  $p$ , this is the equation of an  $AR(1)$  model with parameter  $\phi_p$ , so that the correlation between  $Y_t$  and  $Y_{t-p}$  is equal to  $\phi_p$ . Thus for an  $AR(p)$  process

$$\phi_{pp} = \phi_p \neq 0.$$

Furthermore, it is usually the case that

$$\phi_{uu} \neq 0, \quad u < p.$$

### 7.3 Calculation of the pacf in the general case

It turns out that for an arbitrary stationary process the partial autocorrelations can be calculated from equations similar in form to the Yule-Walker equations.

Consider first the AR( $p$ ) model of equation (1) above. The corresponding autocorrelation function  $\{\rho_\tau\}$  satisfies the Yule-Walker Equations

$$\rho_\tau = \sum_{k=1}^p \phi_k \rho_{\tau-k}, \quad \tau \geq 1.$$

In particular, the first  $p$  Yule-Walker Equations are

$$\rho_\tau = \sum_{k=1}^p \phi_k \rho_{\tau-k}, \quad 1 \leq \tau \leq p. \quad (2)$$

Given the values  $\phi_k$ ,  $1 \leq k \leq p$  of the autoregressive coefficients, as shown in Section 5.4, the set of  $p$  simultaneous equations (2) can be solved to obtain  $\rho_\tau$ ,  $1 \leq \tau \leq p$ . Conversely, if we are not given the values  $\phi_k$ ,  $1 \leq k \leq p$  but are given the values  $\rho_\tau$ ,  $1 \leq \tau \leq p$ , we can solve the equations (2) to obtain  $\phi_k$ ,  $1 \leq k \leq p$ . In particular, we can obtain the value of  $\phi_{pp} \equiv \phi_p$ .

Now for any stationary process, with autocorrelation function  $\{\rho_\tau\}$ , it turns out that the partial autocorrelation at lag  $u$  may be found by solving the set of  $u$  simultaneous linear equations in  $\phi_{uk}$ ,  $1 \leq k \leq u$ ,

$$\rho_\tau = \sum_{k=1}^u \phi_{uk} \rho_{\tau-k}, \quad 1 \leq \tau \leq u, \quad (3)$$

when  $\phi_{uu}$  is the required partial autocorrelation.

Equations (3) are obtained as if we were matching the first  $u$  terms,  $\rho_\tau$ ,  $1 \leq \tau \leq u$ , of the autocorrelation function to the AR( $u$ ) model,  $Y_t = \sum_{k=1}^u \phi_{uk} Y_{t-k} + \epsilon_t$ . If the stationary process is an AR( $p$ ) process then for  $u = p$  the sets of equations (2) and (3) are actually identical, with

$$\phi_{pk} = \phi_k, \quad 1 \leq k \leq p.$$

We see again that for an AR( $p$ ) process  $\phi_{pp} = \phi_p$ . If for an AR( $p$ ) process we solve the set of equations (3) for  $u > p$  then we obtain  $\phi_{uk} = \phi_k$ ,  $1 \leq k \leq p$  and  $\phi_{uk} = 0$ ,  $p+1 \leq k \leq u$ , so we see again that for an AR( $p$ ) process  $\phi_{uu} = 0$ ,  $u > p$ .

In practice, for an observed time series  $y_1, y_2, \dots, y_T$ , the autocorrelations  $\rho_\tau$  are unknown but are estimated by the sample autocorrelations  $r_\tau$ ,  $1 \leq \tau \leq T-1$ . Correspondingly, the  $\rho_\tau$  in Equations (3) can be replaced by the  $r_\tau$  to find the *method of moments* estimates  $\hat{\phi}_{uu}$  of  $\phi_{uu}$ . Thus for each  $u$  in the range  $1 \leq u \leq T-1$  we solve the set of equations

$$r_\tau = \sum_{k=1}^u \hat{\phi}_{uk} r_{\tau-k}, \quad 1 \leq \tau \leq u \quad (4)$$

to obtain  $\hat{\phi}_{uk}$ ,  $1 \leq k \leq u$ , and in particular  $\hat{\phi}_{uu}$ , and hence the *sample partial autocorrelation function (pacf)*,  $\{\hat{\phi}_{uu} : 1 \leq u \leq T-1\}$ .

- Note that Equation (3) with  $u = 1$  yields  $\phi_{11} = \rho_1$  and that, correspondingly, Equation (4) with  $u = 1$  yields  $\hat{\phi}_{11} = r_1$ .

## 7.4 Model selection

The process of fitting an ARIMA model to an observed time series may be carried out as a three stage iterative procedure.

1. *Model selection (or identification)* – choose one or possibly more ARIMA models, which appear to provide a suitable model for the data.
2. *Parameter estimation* – estimate the parameters of the chosen model.
3. *Diagnostic checking* – check whether the fitted model really does adequately fit the data. If it does not, return to Stage 1 and attempt to find a better model.

It should be noted that there may turn out to be more than one plausible model – there is not necessarily a clear-cut best model. In this section, we examine Stage 1.

The first step in analyzing an observed time series is to plot the data. Features such as trend, seasonality and discontinuities, if present, will usually be apparent from simple inspection of the plot. If there are clear-cut discontinuities then it may be advisable to analyze the series by first breaking it up into homogeneous segments.

In some circumstances it may be desirable to transform the data, most commonly by taking logarithms. For example, if there is a trend present and from inspection of the plot it appears that the standard deviation of the data is proportional to the local mean, the logarithmic transformation has the effect of making the variance homogeneous. Also, if the trend in the observed series is exponential, the logarithmically transformed series will have a linear trend, which may be dealt with by taking differences or, equivalently, fitting some  $\text{ARIMA}(p, d, q)$  model with  $d \geq 1$ . It is always useful to plot the data again after any transformation or differencing to see if the resulting series appears to be stationary.

Transformations and differencing of the data may be useful, but they should not be carried out unless there is a genuine reason to do so. There is a danger in over-transforming or over-differencing in that you are distancing yourself from the original data and, in addition, may be causing undesirable side-effects to arise. For instance, over-differencing gives rise to an increase in the variability of the series.

We shall for the present restrict attention to non-seasonal models. In searching for an appropriate  $\text{ARIMA}(p, d, q)$  model for the observed series (possibly after transformation), the smallest possible number of parameters for adequate representation of the data should be employed. This is sometimes referred to as the *principle of parsimony*.

This principle should be adopted partly because, in a model with a large number of parameters, the estimates of the parameters will be subject to large errors. Furthermore, a model with a smaller number of parameters is easier to interpret.

In most situations that are amenable to ARIMA modelling, a suitable model may be found with each of  $p, d$  and  $q$  less than or equal to 2. Often, one of  $p, d$  and  $q$  is zero.

Inspection of the sample acf and pacf may help to decide whether a stationary model is appropriate for an observed time series and, if so, which of the ARMA models is a suitable

one to try to fit to the data. The following table summarizes the relevant properties of the (theoretical) acfs and pacfs of the ARMA models.

model	acf	pacf
AR( $p$ )	dies away geometrically	zero after lag $p$
MA( $q$ )	zero after lag $q$	dies away geometrically
ARMA( $p, q$ )	dies away geometrically	dies away geometrically

If the data are to be represented by an ARMA( $p, q$ ) model with  $p \geq 1$  and  $q \geq 1$  then both the acf and the pacf should eventually die away geometrically (in general as a mixture of exponentials or damped sine waves) with neither exhibiting an obvious cutoff point.

If the acf dies away slowly, more or less linearly, and not geometrically then the data are not stationary and differencing is necessary. An ARIMA( $p, d, q$ ) model with  $d \geq 1$  will be required. The data may be differenced once or more and the resulting acf and pacf examined to see if the differenced data appear to be stationary and amenable to being modelled as an ARMA process.

## 7.5 Properties of sample acfs and pacfs

Because the acf and pacf that we see are sample rather than theoretical functions, their values are subject to sampling errors. For example, an underlying theoretical cutoff point will not result in sample values that are exactly zero beyond it but, hopefully, in sample values that do not differ significantly from zero.

If a realization  $y_1, y_2, \dots, y_T$  of length  $T$  is observed from a stationary process  $\{Y_t\}$  and the sample acf  $\{r_\tau\}$  calculated, it should be noted that in general for small lag differences the  $r_\tau$  can be highly correlated with each other, perhaps leading to a very distorted picture of the underlying acf. In the general case, there are rather complex approximate expressions for the covariances and variances of the  $r_\tau$ . For some special cases, however, these expressions are very simple.

An important example occurs when the underlying process is taken to be a white noise process. If  $T$  is large then the correlations among the  $r_\tau$  are negligible and the  $r_\tau$ ,  $\tau > 0$  are approximately NID(0,  $1/T$ ). Hence, to investigate whether an observed time series may be modelled as a white noise process, we may consider the approximate 95% probability limits  $\pm 2/\sqrt{T}$ . If not more than about one in twenty of the calculated  $r_\tau$  fall outside these limits then this may be taken to support the hypothesis of white noise.

More generally, if the underlying process is an MA( $q$ ) process then

$$\text{var}(r_\tau) \approx \frac{1 + 2(\rho_1^2 + \rho_2^2 + \dots + \rho_q^2)}{T},$$

and, to investigate the hypothesis that the MA( $q$ ) model is correct, we may consider approximate 95% probability limits at

$$\pm 2 \sqrt{\frac{1 + 2(r_1^2 + r_2^2 + \dots + r_q^2)}{T}}, \quad \tau > q.$$

If not more than about one in twenty of the calculated  $r_\tau$ ,  $\tau > q$  fall outside the limits then this may be taken to support the hypothesis.

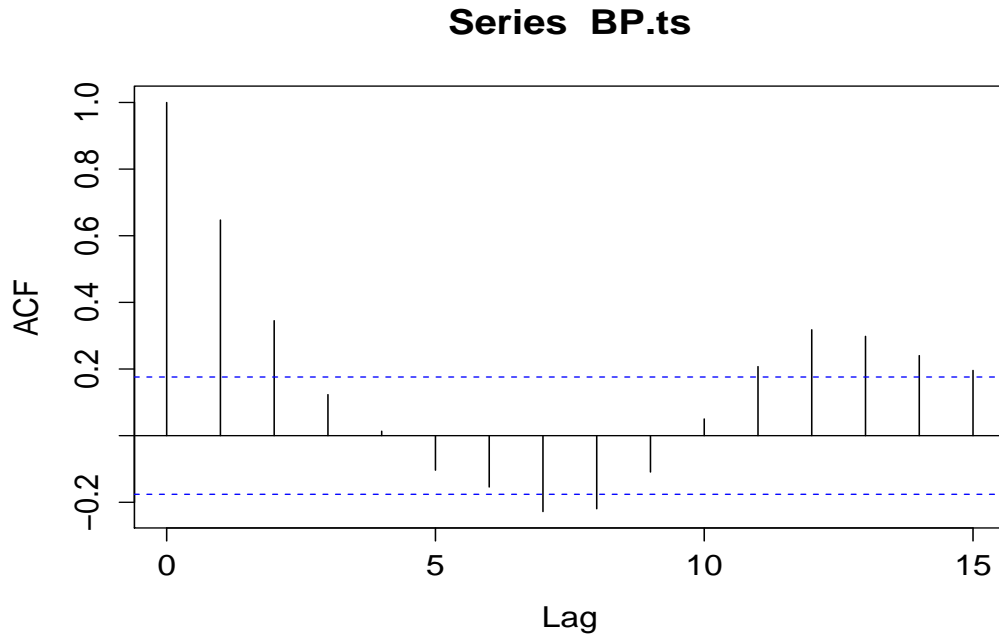
There also exist approximate expressions for the variances of the sample partial autocorrelations  $\hat{\phi}_{uu}$ . If the underlying process is an AR( $p$ ) process then for  $u > p$  the  $\hat{\phi}_{uu}$  are approximately NID(0, 1/ $T$ ). Hence, to investigate the hypothesis that the AR( $p$ ) model is correct, we may consider 95% probability limits at  $\pm 2/\sqrt{T}$ ,  $u > p$ . Again, if not more than about one in twenty of the calculated  $\hat{\phi}_{uu}$ ,  $u > p$  fall outside the limits then this may be taken to support the hypothesis.

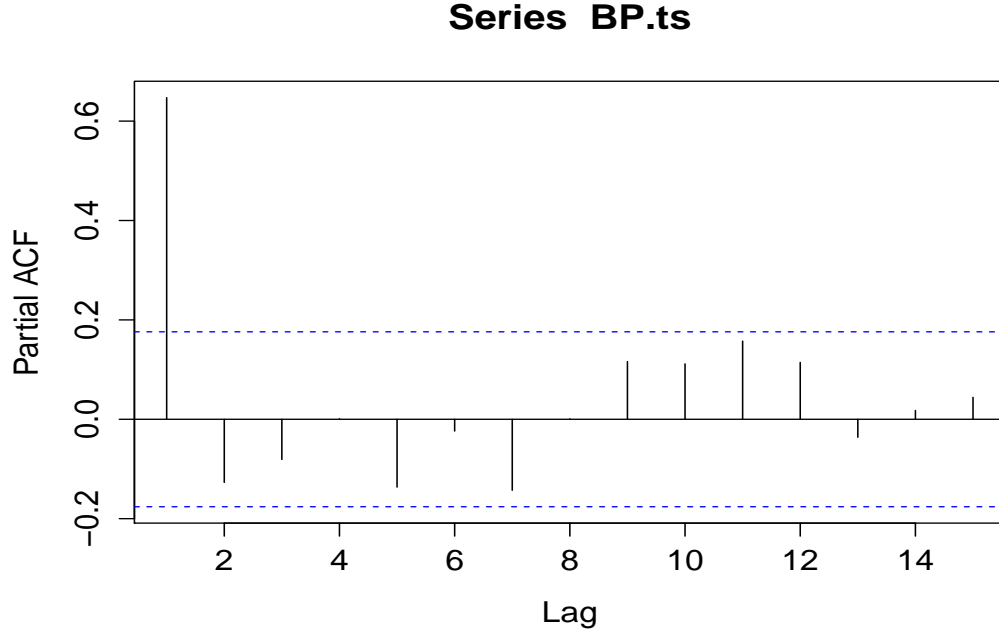
Identification of  $p$  and  $q$  for a mixed ARMA model is more problematic, and several iterations of the fitting procedure may be needed.

## 7.6 Bread price example

The following output gives a part of the acf seen before in Section 3.5 and, in addition, part of the pacf.

```
> BP.acf<-acf(BP.ts, 15)
> BP.pacf<-acf(BP.ts,15,type="partial")
```





The white noise 95% probability limits for the acf are at

$$\pm \frac{2}{\sqrt{T}} = \pm \frac{2}{\sqrt{124}} = \pm 0.180.$$

The first two and several other autocorrelations lie outside these limits. Hence the data cannot be modelled as white noise. From the fuller listing in Chapter 3, it looks as if the acf may die away exponentially. Hence the data may plausibly be represented by an ARMA model. There is no obvious cutoff point in the acf, so a MA model does not seem appropriate. The oscillation in the acf possibly suggests an AR(2) model, although an AR(1) model may be sufficient.

The pacf appears to have a definite cutoff point at lag 1, which suggests an AR(1) model. To test the hypothesis that the AR(1) model is correct, we may consider the 95% probability limits for the pacf at  $\pm 2/\sqrt{T} = \pm 0.180$ ,  $u > 1$ . All the listed partial autocorrelations beyond the first fall within these limits, which supports the hypothesis of an AR(1) model.

For an AR(1) model,  $\rho_1 = \phi$ , where  $\phi$  is the autoregressive parameter. Hence a first estimate of  $\phi$  is given by the method of moments estimator  $\hat{\phi} = r_1 = 0.647$ .

## 7.7 The Akaike information criterion

R uses the *Akaike information criterion* (AIC) to compare the goodness of fit of different models, assuming that the data are normally distributed, i.e., that the  $\epsilon_t$  are NID(0,  $\sigma^2$ ). The AIC is defined by

$$\begin{aligned} \text{AIC} &= -2 \text{ maximized log-likelihood} + 2n \\ &\approx T \ln \hat{\sigma}^2 + 2n + \text{const}, \end{aligned}$$

where  $T$  is the length of the observed series after any differencing,  $n$  is the number of fitted parameters and  $\hat{\sigma}^2$  is the estimated white noise variance, which in R is referred to as `sigma^2`. The model with the smallest value of the AIC is judged to be the most appropriate.

In the special case where competing models each have the same number of parameters, according to the AIC we choose the model with the smallest value of the estimated prediction variance,  $\hat{\sigma}^2$ , as the one that provides the best fit and is likely to provide the most accurate predictions. More generally, where the number of parameters used differs among the competing models, the AIC weighs the benefit of any decrease in  $\hat{\sigma}^2$  against the disadvantage of a larger number of parameters, the estimates of which are less reliable. In other words, the AIC adjusts the value of the estimated prediction variance by including a penalty term for the number of parameters in the model.

- As we are only interested in comparing the AICs of different models, it is sufficient to evaluate the AIC up to an arbitrary additive constant.
- In using the Akaike criterion, all the models being compared should be applied to the same transformation, if any, of the data and should all involve the same degree of differencing.
- The AIC tends to select rather high order models. For example, it will tend to overestimate the true order of an autoregressive model.

## 7.8 Bread price example: parameter estimation

The function `arima` is used to fit ARIMA models using a maximum likelihood method.

Except when specifically fitting an ARMA model (i.e. no differencing present), the `arima` function does not include a process mean in the model but assumes a zero-mean series. If an ARMA model is being fitted, then `arima` will estimate the process mean. The `order` parameter of the `arima` function is used to specify the model to be fitted.

```
> BP.ar1m<-arima(BP.ts, order=c(1,0,0))
> BP.ar1m
```

Call:

```
arima(x = BP.ts, order = c(1, 0, 0))
```

Coefficients:

	ar1	intercept
	0.6429	5.6608
s.e.	0.0678	0.2307

```
sigma^2 estimated as 0.8655: log likelihood = -167.26, aic = 340.52
```

```
>
```

```
> BP.ar1m$coef
```

	ar1	intercept
	0.6429364	5.6608179

```

> BP.ar1m$var.coef
              ar1      intercept
ar1          0.0045933960 0.0001634083
intercept 0.0001634083 0.0532077436
> BP.ar1m$sigma2
[1] 0.8655124
> BP.ar1m$aic
[1] 340.5203

```

The object `BP.ar1m` contains the results of fitting an AR(1) model. The estimated value of the autoregressive parameter is  $\hat{\phi} = 0.6429364$  with standard error  $\sqrt{0.004593396} = 0.067775$ . The estimated white noise variance is  $\hat{\sigma}^2 = 0.8655124$ .

A 95% confidence interval for  $\phi$  is given by

$$\begin{aligned}
0.6429364 \pm t_{121,0.025} \times 0.067775 &= 0.6429364 \pm (1.9798)(0.067775) \\
&= (0.509, 0.777).
\end{aligned}$$

```

> BP.ar2m<-arima(BP.ts, order=c(2,0,0))
> BP.ar2m

```

```

Call:
arima(x = BP.ts, order = c(2, 0, 0))

```

Coefficients:

```

              ar1      ar2  intercept
              0.7231 -0.1235      5.6546
s.e.    0.0888    0.0892    0.2051

```

```

sigma^2 estimated as 0.8521:  log likelihood = -166.31,  aic = 340.62
>

```

```

> BP.ar2m$coef
              ar1      ar2  intercept
0.7231081 -0.1235387  5.6546006
> BP.ar2m$var.coef
              ar1      ar2      intercept
ar1          0.0078771805 -0.0051498171 -0.0001560912
ar2          -0.0051498171  0.0079488694  0.0003603017
intercept -0.0001560912  0.0003603017  0.0420676123
> BP.ar2m$sigma2
[1] 0.8520988
> BP.ar2m$aic
[1] 340.6158

```

The object `BP.ar2m` contains the results of fitting an AR(2) model. The estimated values of the autoregressive parameters are  $\hat{\phi}_1 = 0.7231081$  and  $\hat{\phi}_2 = -0.1235387$ , with standard errors  $\sqrt{0.0078771805} = 0.088753$  and  $\sqrt{0.0079488694} = 0.089156$ , respectively. The estimated white noise variance is  $\hat{\sigma}^2 = 0.8520988$ .



The corresponding fitted autoregressive model is

$$Y_t - 5.6546 = 0.7231(Y_{t-1} - 5.6546) - 0.1235(Y_{t-2} - 5.6546) + \epsilon_t,$$

i.e.,

$$Y_t = 2.264 + 0.723Y_{t-1} - 0.124Y_{t-2} + \epsilon_t.$$