University of Oxford

Statistical Methods
Autocorrelation

Identification and Estimation

Dr. Órlaith Burke

Michaelmas Term, 2011

Department of Statistics, 1 South Parks Road,
Oxford OX1 3TG
# Contents

1 Model Identification ................................................. 1
   1.1 The Sample ACF .............................................. 1
       1.1.1 Using the ACF to determine a model ................. 10
   1.2 The PACF ...................................................... 10
       1.2.1 AR(1) Process ........................................ 12
       1.2.2 AR(p) Process ....................................... 12
       1.2.3 MA(1) Process ........................................ 13
       1.2.4 MA(q) process ...................................... 13
   1.3 The Sample PACF ............................................. 14

2 Estimating the Model Parameters .................................. 17
   2.1 Estimating Drift and Trend terms .......................... 17
       2.1.1 Estimating Drift ...................................... 18
       2.1.2 Estimating a Linear Trend ............................ 19
   2.2 Method of Moments Estimators for AR, MA and ARMA models . 20
       2.2.1 AR Models ............................................ 20
       2.2.2 MA Models ............................................ 21
       2.2.3 ARMA(1,1) Model .................................... 21
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>Alternative Estimation Techniques</td>
<td>22</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Least Squares Estimation</td>
<td>22</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Maximum Likelihood Estimation</td>
<td>24</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Unconditional Least Squares</td>
<td>25</td>
</tr>
</tbody>
</table>
Model Identification

We have been introduced to ARIMA models and have examined in detail the behaviour of various models. But all of our investigations so far have looked at theoretical models for time series. Our next task is to consider how we might fit these models to real time series data.

To begin, in this chapter we examine how to establish what are appropriate values for $p, d$ and $q$ in fitting an ARIMA($p, d, q$) model to a set of data.

1.1 The Sample ACF

The first technique we use to determine values for $p, d$ and $q$ is to compute the sample ACF from the data and compare this to known properties of the ACF for ARIMA models.

The sample autocorrelation function is defined as

$$ r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}. $$

1.1.1 The Sample ACF

The first technique we use to determine values for $p, d$ and $q$ is to compute the sample ACF from the data and compare this to known properties of the ACF for ARIMA models.

The sample autocorrelation function is defined as

$$ r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}. $$

The sample autocorrelation $r_k$ provides us with an estimate for the true autocorrelation $\rho_k$. In order to make proper statistical determinations about $\rho_k$ we need to
investigate the sampling distribution behaviour. Unfortunately the sampling behaviour is too complicated for us to go into detail in these notes. Instead, we consider the following asymptotic result:

Let $\rho_k$ and $r_k$ be the autocorrelation function and sample autocorrelation function respectively of stationary ARMA process. In the limit as $n \to \infty$ the joint distribution of

\[ (\sqrt{n}(r_1 - \rho_1), \sqrt{n}(r_2 - \rho_2), \ldots, \sqrt{n}(r_m - \rho_m)) \] (1.2)

approaches a joint normal distribution with zero means and with variance-covariance matrix $c_{ij}$ where

\[
c_{ij} = \sum_{k=-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+i} + 2\rho_i\rho_j\rho_k^2).
\] (1.3)

For large $n$ we could therefore say that $r_k$ is approximately normally distributed with mean $\rho_k$ and variance $c_{kk}/n$. We will not examine $r_k$ in any more detail however we take note that the $r_k$ should not be expected to match $\rho_k$ in great detail. We may see ripples or trends in $r_k$ which are not present in $\rho_k$ but are artifacts of sampling.
AR1: $Y_t = 0.9Y_{t-1} + \epsilon_t$
1. MODEL IDENTIFICATION

AR1: \( Y_t = 0.4Y_{t-1} + \epsilon_t \)
1. MODEL IDENTIFICATION

$AR1: Y_t = -0.7Y_{t-1} + \epsilon_t$
AR1: $Y_t = 1.0Y_{t-1} + \epsilon_t$
1. MODEL IDENTIFICATION

MA1: \( Y_t = 1.0 \epsilon_{t-1} + \epsilon_t \)
1. MODEL IDENTIFICATION

MA1: \( Y_t = 0.4 \epsilon_{t-1} + \epsilon_t \)
MA2: $Y_t = 0.4\epsilon_{t-1} + 0.4\epsilon_{t-2} + \epsilon_t$
1. MODEL IDENTIFICATION

1.1.1 Using the ACF to determine a model

We can now begin to see how the ACF can be used to determine which model best fits a set of data.

Firstly we can see from the theoretical ACF and from the previous graphs that the ACF is very useful for identifying MA processes. This is because the first $q$ terms in the ACF of an MA($q$) process are non-zero and the remaining terms are all zero.

Secondly we have seen that the ACF of a non-stationary series displays a behaviour quite different from that of stationary series. In non-stationary series the terms of the ACF do not decay to zero as they do in a stationary series (c.f. the plots above).

We have also seen that the ACF of a stationary AR process decays exponentially to zero.

1.2 The PACF

In the previous section we have seen how the ACF can be used to identify MA processes, clearly indicating the order $q$ of the process by the number of non-zero terms in the ACF. It would be great if we could similarly identify the order of an AR process. There is another correlation function which allows us to do precisely that: The Partial Autocorrelation Function or PACF.

The PACF computes the correlation between two variables $y_t$ and $y_{t-k}$ after removing the effect of all intervening variables $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$. We can think of the PACF as a conditional correlation:
1. MODEL IDENTIFICATION

\[ \phi_{kk} = \text{Corr}(y_t, y_{t-k}|y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}). \] (1.4)

Another way to think of the PACF is to consider regressing \( y_t \) on \( y_{t-1}, y_{t-2}, \ldots, y_{t-k+1} \) to give the following least squares estimate:

\[ \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_{k-1} y_{t-k+1}. \] (1.5)

We define \( E_{\text{forward}} \) to be the error of prediction

\[ E_{\text{forward}} = y_t - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \ldots - \beta_{k-1} y_{t-k+1}. \] (1.6)

Consider also regressing \( y_{t-k} \) on \( y_{t-1}, y_{t-2}, \ldots, y_{t-k+1} \). It can be shown (however we will not prove this result) that the least squares estimate of \( y_{t-k} \) is

\[ \beta_1 y_{t-k+1} + \beta_2 y_{t-k+2} + \ldots + \beta_{k-1} y_{t-1}. \] (1.7)

Here the prediction error is

\[ E_{\text{backward}} = y_{t-k} - \beta_1 y_{t-k+1} - \beta_2 y_{t-k+2} - \ldots - \beta_{k-1} y_{t-1}. \] (1.8)

The Partial Autocorrelation Function is now

\[ \phi_{kk} = \text{Corr}(E_{\text{forward}}, E_{\text{backward}}). \] (1.9)
1. MODEL IDENTIFICATION

How does this PACF behave? Well, by convention we set

$$\phi_{11} = \rho_1.$$  \hspace{1cm} (1.10)

It is possible to show that

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}. $$  \hspace{1cm} (1.11)

Now let us consider some ARIMA models.

1.2.1 AR(1) Process

We have seen that for an AR(1) model

$$\rho_k = \phi^k$$  \hspace{1cm} (1.12)

hence

$$\phi_{22} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0 $$  \hspace{1cm} (1.13)

In fact $\phi_{kk} = 0$ for all $k > 1$ in the AR(1) model.

1.2.2 AR(p) Process

For the AR(p) model we can show that

$$\phi_{kk} = 0 \quad \forall \ k > p $$  \hspace{1cm} (1.14)
1. MODEL IDENTIFICATION

1.2.3 MA(1) Process

For the MA(1) process we find the following results

\[
\phi_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4} \tag{1.15}
\]

\[
\phi_{kk} = \frac{-\theta^k(1 - \theta^2)}{1 - \theta^{2(k+1)}} \text{ for } k \geq 1 \tag{1.16}
\]

It is clear from these equations that the PACF for an MA(1) process decays exponentially to zero.

1.2.4 MA(q) process

For a general MA(q) process we can use the following Yule-Walker equations to solve for the PACF \( \phi_{kk} \):

\[
\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \ldots + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \ldots, k \tag{1.17}
\]

Note that here the \( \rho \) terms are known and we are solving for \( \phi_{kk} \). Of all the terms \( \phi_{k1}, \phi_{k2}, \ldots, \phi_{kk} \) we are only interested in \( \phi_{kk} \).

We should note here that these Yule-Walker equations are slightly different from those we have seen before. These equations allow us to solve for \( \phi_{kk} \) for any stationary process not just an AR\( (p) \) process. If however we are dealing with an AR\( (p) \) process then these equations match exactly those that we ave already seen with \( \phi_{pp} = \phi_p \) and \( \phi_{kk} = 0 \) for \( k > p \).
1.3 The Sample PACF

When presented with actual data we can use equation (1.17) to compute the Sample PACF if we replace the true ACFs ($\rho$) by the sample ACFs ($r$). We will not go into any more details here however as in the case of the sample ACF we note that the behaviour of the sample PACF will be similar to the true PACF. We should not however expect complete agreement between the two and we will therefore not be surprised to see some ripples or trends in the Sample PACF which would not be present in the true ACF.
1. MODEL IDENTIFICATION

AR1: $Y_t = \phi Y_{t-1} + \epsilon_t$

PACF Simulated AR1 $\phi_1=0.9$ $T=1000$

PACF Simulated AR1 $\phi_1=0.7$ $T=1000$
1. MODEL IDENTIFICATION

MA1: $Y_t = (-\theta)\epsilon_{t-1} + \epsilon_t$
In the last chapter we examined how to determine values for $p, d, q$ and hence establish the appropriate ARIMA($p, d, q$) model to fit to a set of data. In this chapter we consider how to estimate the parameters of the ARIMA model. We will consider several different estimation techniques for estimating the various parameters in the model.

2.1 Estimating Drift and Trend terms

When we first encountered the Box-Jenkins ARIMA model the model contained a mean term $\mu$. Throughout this course so far we have set $\mu = 0$. Let us now allow the possibility that $\mu \neq 0$ and consider how we might estimate this non-zero mean.

\footnote{1Courtesy of Dr. P. Murphy, UCD}
In essence we are dealing with a time series

\[ x_t = \mu + y_t \]  

(2.1)

where \( E(y_t) = 0 \) and we wish to estimate \( \mu \) using the observed sample time series \( x_1, x_2, \ldots, x_n \).

### 2.1.1 Estimating Drift

Suppose that \( \mu \) is a constant then the obvious way to estimate \( \mu \) is use the sample mean:

\[ \bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t. \]  

(2.2)

This estimate is obviously unbiased. Let’s consider how precise an estimate it is by examining the variance of \( \bar{x} \).

If \( y_t \) is stationary then we may show that

\[ \text{Var}(\bar{x}) = \frac{\gamma_0}{n} \sum_{k=-n+1}^{n-1} \left( 1 - \frac{|k|}{n} \right) \rho_k \]  

(2.3)

\[ = \frac{\gamma_0}{n} \left[ 1 + 2 \sum_{k=1}^{n-1} \left( 1 - \frac{k}{n} \right) \rho_k \right] \]  

(2.4)

where \( \rho_k \) is the autocorrelation of \( y_t \) and in fact also of \( x_t \).

We note here that the variance is approximately inversely proportional to the sample size.

If \( y_t \) is non-stationary with zero mean then the precision of the estimate \( \bar{x} \) is very much worse.

Suppose \( y_t = y_{t-1} + \epsilon_t \) then it can be shown that
2. ESTIMATING THE MODEL PARAMETERS

\[
\text{Var}(\bar{x}) = \frac{1}{n^2} \left[ \sum_{t=1}^{n} t \sigma^2 + 2 \sum_{s=2}^{n} \sum_{t=1}^{s-1} t \sigma^2 \right] = \sigma^2 (2n + 1) \frac{n + 1}{6n} \tag{2.5}
\]

\[
\tag{2.6}
\]

2.1.2 Estimating a Linear Trend

Suppose now that the mean \( \mu \) is no longer a constant but instead takes the form:

\[
\mu = \beta_0 + \beta_1 t. \tag{2.7}
\]

We want to estimate the two parameters \( \beta_0 \) and \( \beta_1 \). One method is to use the ordinary least squares technique used in regression. Here we minimise

\[
Q(\beta_0, \beta_1) = \sum_{t=1}^{n} (x_t - \beta_0 - \beta_1 t)^2. \tag{2.8}
\]

Partial differentiation gives us the least squares estimators:

\[
\hat{\beta}_1 = \frac{\sum_{t=1}^{n} (x_t - \bar{x})(t - \bar{t})}{\sum_{t=1}^{n} (t - \bar{t})} \tag{2.9}
\]

\[
\hat{\beta}_0 = \bar{x} - \hat{\beta}_1 \bar{t} \tag{2.10}
\]

where \( \bar{t} \) is the average of 1, 2, 3, \ldots, \( n \):

\[
\bar{t} = (n + 1)/2. \tag{2.11}
\]
2. Method of Moments Estimators for AR, MA and ARMA models

2.2.1 AR Models

In the AR(1) model $\rho_1 = \phi_1$ therefore we could estimate $\phi_1$ by

$$\hat{\phi}_1 = r_1.$$  \hfill (2.12)

In the AR(2) model we have the Yule-Walker equations:

$$\rho_1 = \phi_1 + \rho_1 \phi_2$$  \hfill (2.13)
$$\rho_2 = \rho_1 \phi_1 + \phi_2.$$  \hfill (2.14)

The method of moments then instructs us to replace each $\rho$ by the corresponding $r$ giving us

$$r_1 = \phi_1 + r_1 \phi_2$$  \hfill (2.15)
$$r_2 = r_1 \phi_1 + \phi_2.$$  \hfill (2.16)

When we solve these we get

$$\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2},$$  \hfill (2.17)
$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}.$$  \hfill (2.18)

The exact same technique gives us the method of moments estimators of the $p$ parameters in an AR($p$) model, namely we replace each of the $p$ ACFs “$\rho$” by the corresponding sample ACFs “$r$” in the Yule-Walker equations and then we solve
2. ESTIMATING THE MODEL PARAMETERS

for the method of moments estimators: \( \hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p \).

2.2.2 MA Models

In the MA(1) process we know that

\[
\rho_1 = -\frac{\theta_1}{1 + \theta_1^2}
\]  
(2.19)

Now we replace \( \rho_1 \) by \( r_1 \) and then we solve this equation to get two roots

\[
\hat{\theta}_1^+ = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}
\]  
(2.20)

\[
\hat{\theta}_1^- = \frac{-1 - \sqrt{1 - 4r_1^2}}{2r_1}
\]  
(2.21)

We now must consider three situations:

1. If \( |r_1| < 0.5 \) then we have real roots but only \( \hat{\theta}_1^+ \) is invertible.

2. If \( r_1 = \pm 0.5 \) then neither of the solutions are invertible.

3. If \( |r_1| > 0.5 \) then no real solutions exist and so we have no estimator for \( \theta_1 \).

The MA(q) model can be considered in the same way but the solution requires solving sets of non-linear equations. In fact it is only possible to solve these equations numerically, analytical solutions do not exist in most cases.

2.2.3 ARMA(1,1) Model

We have seen that for the ARMA(1,1) model

\[
\rho_k = \frac{(1 - \theta_1 \phi_1)(\phi_1 - \theta_1)}{1 - 2\phi_1 \theta_1 + \theta_1^2} \theta^{k-1}.
\]  
(2.22)
From this we see that
\[
\frac{\rho_2}{\rho_1} = \phi_1
\] (2.23)
and so we can estimate \( \phi_1 \) by
\[
\hat{\phi}_1 = \frac{r_2}{r_1}.
\] (2.24)
Substituting this estimate into (2.22) we obtain
\[
\hat{r}_1 = \frac{(1 - \theta_1 \hat{\phi}_1)(\hat{\phi}_1 - \theta_1)}{1 - 2\phi_1 \theta_1 + \theta_1^2}.
\] (2.25)
We can now solve this quadratic equation for \( \hat{\theta}_1 \), remembering that of the two solutions thus obtained we only keep the invertible solution.

### 2.3 Alternative Estimation Techniques

We have just seen that the method of moments cannot estimate the parameters in a Moving Average model. We now note that there are several alternative estimation procedures that can be used to estimate the parameters in an ARIMA model. The first is based on the principle of least squares, the second uses maximum likelihood methods and the technique of ”unconditional least squares” is a compromise between these two methods. We will not go into the detail of any of these estimation procedures here but in the following we will provide a brief introduction to the basic ideas.

#### 2.3.1 Least Squares Estimation

The AR(1) model
\[
Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t
\] (2.26)
can be viewed as a regression with predictor variable \( Y_{t-1} \) and response variable \( Y_t \). The least squares estimator of the parameter \( \phi \) is clearly then obtained by
minimising the following sum of squares:

$$S_*(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2. \quad (2.27)$$

This function is called the “conditional sum of squares function”. Minimising \( S_* \) we get

$$\hat{\mu} = \frac{\sum_{t=2}^{n} Y_t - \phi \sum_{t=2}^{n} Y_{t-1}}{(n-1)(1 - \phi)} \quad (2.28)$$

For large values of \( n \) this equation reduces to

$$\hat{\mu} \approx \frac{\overline{Y} - \phi \overline{Y}}{1 - \phi} = \overline{Y}. \quad (2.29)$$

And substituting \( \overline{Y} \) for \( \mu \) we find that when dealing with large samples the estimator for \( \phi \) is given by

$$\hat{\phi} = \frac{\sum_{t=2}^{n} (Y_t - \overline{Y})(Y_{t-1} - \overline{Y})}{\sum_{t=2}^{n} (Y_{t-1} - \overline{Y})^2}. \quad (2.30)$$

In the AR\((p)\) model we again find that for large sample sizes

$$\hat{\mu} = \overline{Y}. \quad (2.31)$$

More interestingly we can show that the estimates of \( \phi_1, \phi_2, \ldots, \phi_p \) are given by the solutions of the sample Yule-Walker equations. For example in the AR\((2)\) model we solve

$$r_1 = \phi_1 + r_1 \phi_2 \quad (2.32)$$

$$r_2 = r_1 \phi_1 + \phi_2 \quad (2.33)$$

to obtain \( \hat{\phi}_1 \) and \( \hat{\phi}_2 \).
2. ESTIMATING THE MODEL PARAMETERS

Suppose now we are trying to estimate the parameter $\theta$ in the MA(1) model:

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1}. \quad (2.34)$$

We remember that if this model is invertible then it can be written as an infinite order AR model:

$$Y_t = -\theta_1 Y_{t-1} - \theta_1^2 Y_{t-2} - \ldots + \epsilon_t \quad (2.35)$$

And again we have a regression model that in principle allows us to estimate $\theta_1$ using least squares. We note however that this regression is not only of infinite order but is non-linear in $\theta_1$. In fact we cannot minimise the appropriate sum of squares

$$S_*(\theta_1) = \sum \epsilon_t \quad (2.36)$$

analytically but must use numerical methods to find a solution.

Actually for MA($q$) and ARMA($p,q$) models we will have to resort to numerical techniques to find solutions to the least squares equations.

2.3.2 Maximum Likelihood Estimation

For the AR(1) model

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + \epsilon_t \quad (2.37)$$

if we assume that the error terms $\epsilon_t$ are iid normal white noise then we can show that the likelihood function is given by

$$L(\phi_1, \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2}(1 - \phi_1^2)^{1/2} \exp \left[ -\frac{S(\phi_1, \mu)}{2\sigma^2} \right] \quad (2.38)$$
where

\[ S(\phi_1, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi_1(Y_{t-1} - \mu)]^2 + (1 - \phi_1^2)(Y_1 - \mu)^2. \] (2.39)

The term \( S(\phi_1, \mu) \) is called the “unconditional sum of squares function” and should be compared with the previous “conditional sum of squares”. Clearly they are related as follows:

\[ S(\phi_1, \mu) = S_*(\phi_1, \mu) + (1 - \phi_1^2)(Y_1 - \mu)^2. \] (2.40)

To obtain maximum likelihood estimators for \( \phi_1, \mu \) and \( \sigma^2 \) we need to minimise \( L(\phi_1, \mu, \sigma^2) \) with respect to these parameters.

We note that we can extend this procedure to other ARIMA models.

### 2.3.3 Unconditional Least Squares

In previous sections we have seen that when estimating parameters using Least Squares we must minimise the conditional sum of squares \( S_*(\phi_1, \mu) \). We have also just seen that the maximum likelihood estimators are obtained by minimising \( L(\phi_1, \mu, \sigma^2) \) and we have seen \( L(\phi_1, \mu, \sigma^2) \) is a function of the unconditional sum of squares \( S(\phi_1, \mu) \). An alternative estimation procedure tells us that instead of minimising the full likelihood function \( L(\phi_1, \mu, \sigma^2) \) we could just minimise the unconditional sum of squares \( S(\phi_1, \mu) \). This procedure is then a compromise between the conditional least squares estimates and the full maximum likelihood procedure.

We will finish our discussion here and will not investigate any of these estimation procedures further.