Time Series Analysis – Lecture 6

Model Identification

Dr. Kevin Burke

1 Model Identification

There are three basic steps in the so-called "Box-Jenkins" approach to modelling time series data (discussed in Lecture 1):

1. Identification: decide on initial models.
2. Fitting: estimate model parameters.
3. Checking: assess model adequacy using residuals.

If we have not found an appropriate model, we return to step 1 and consider other models.

Over the next three lectures we will consider each of these steps.

The focus of this lecture is **model identification** which is concerned with deciding on some preliminary values of p, d and q before we attempt to fit an ARIMA(p, d, q) model to the data.

- **First we deal with non-stationarity** by deciding on an appropriate value for d, the order of differencing. The series may also have non-constant variance which can be eliminated using a Box-Cox transformation.
- Once the series looks approximately stationary, we can **then investigate the correlation structure** in this transformed series to determine values for p and q.

2 Non-Stationarity

Evidence of non-stationarity is often relatively easy to determine from the time series plot, i.e., non-constant variance and trend in the series.

Non-constant variance can generally be handled by applying a log or Box-Cox transformation whereas trend can be dealt with through differencing (or by modelling the trend as described in Lecture 3).

Recall from Lecture 2 that the sample ACF is an estimator for the true ACF for a *stationary* series - thus, there is no basis for applying it to a non-stationary series. Nonetheless, a common symptom of non-stationarity is that **the sample ACF fails to decay quickly with increasing lag** (i.e., decay is linear rather than exponential) leading to significant sample autocorrelations at large lags.

2.1 Differencing

Our strategy for choosing the order of differencing, d, is to **find the lowest value of d which leads to a reasonably stationary series** as follows:

0. Start with $d = 0$.
1. Plot the series $\nabla^d Y_t$ and its sample ACF.
2. If this looks relatively stationary, then choose this value of d. Otherwise increase d by one unit and return to 1.

Note: typically $d \in \{0, 1, 2\}$ is enough but sometimes we may need more.

2.2 Dickey-Fuller Test

Consider the process

$$Y_t = \alpha Y_{t-1} + e_t$$

which is a random walk for $\alpha = 1$ and an AR(1) for $|\alpha| < 1$. Thus, assuming we can estimate α, we can test

$$H_0 : \alpha = 1 \Rightarrow \text{ random walk}$$
$$\textit{against}$$
$$H_1 : -1 < \alpha < 1 \Rightarrow \text{stationary AR(1)}.$$

this is called the **Dicky-Fuller (DF) test**. However, it is desirable to have something more general than this such as

H_0 : The series is non-stationary and differencing is required.

H_1 : The series is stationary.

This test is called the **augmented Dickey-Fuller (ADF) test**.

We can develop this test by first using the fact that any ARMA(p,q) process can be approximated by an AR(p^*) process for sufficiently large p^*. This follows from invertibility, i.e., the ARMA(p,q) process can be written as $\pi(B)\, Y_t = e_t$.

With the above in mind, we can develop the more general test by considering the process

$$Y_t = \alpha Y_{t-1} + X_t$$

where X_t is an AR(p^*) process, i.e., $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_{p^*} X_{t-p^*} + e_t$.

Note that we can write

$$\begin{aligned} \phi(B)X_t &= e_t \\ X_t + \phi(B)X_t &= X_t + e_t \\ X_t &= X_t - \phi(B)X_t + e_t \\ &= [1 - \phi(B)]X_t + e_t. \end{aligned}$$

Also note that

$$X_t = Y_t - \alpha Y_{t-1} = (1 - \alpha B)\, Y_t.$$

Putting this together we have

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + [1 - \phi(B)]\, X_t + e_t \\ &= \alpha B Y_t + [1 - \phi(B)]\, (1 - \alpha B)\, Y_t + e_t \end{aligned}$$

$$\begin{aligned} \Rightarrow Y_t - \alpha B Y_t - [1 - \phi(B)]\, (1 - \alpha B)\, Y_t &= e_t \\ \{1 - \alpha B - [1 - \phi(B)]\, (1 - \alpha B)\}\, Y_t &= e_t \\ \{(1 - \alpha B)(1 - [1 - \phi(B)])\}\, Y_t &= e_t \\ \{(1 - \alpha B)\phi(B)\}\, Y_t &= e_t \\ \phi(B)(1 - \alpha B)\, Y_t &= e_t. \end{aligned}$$

Clearly if $\alpha = 1$ this is an ARI($p^*, 1$), i.e., Y_t non-stationary but differencing once produces a stationary AR(p^*) process. If $|\alpha| < 1$ then the process is a stationary AR($p^* + 1$).

Thus we have the following:

$$H_0 : \alpha = 1 \Rightarrow \text{ differencing leads to AR}(p^*)$$

$$H_1 : -1 < \alpha < 1 \Rightarrow \text{series is AR}(p^* + 1).$$

By noting that *any* stationary process can be approximated by an AR process of sufficiently high order, the above null and alternative hypotheses become

H_0 : Differencing needed to produce stationary series

H_1 : Series is stationary.

This test is implemented in `R` via the `adf.test` function.

Note: The reason for framing things in terms of a high-order AR process is that it allows α to be estimated using a linear regression procedure:

$$\nabla Y_t = \beta Y_{t-1} + \delta_1 \nabla Y_{t-1} + \cdots + \delta_{p^*} \nabla Y_{t-p^*}.$$

Here $\beta = \alpha - 1$ so that $\beta = 0$ is equivalent to $\alpha = 1$.

2.3 Overdifferencing

We must be careful not to **over-difference**, i.e., difference more than is required as this can introduce spurious autocorrelations into the series and increase the variance.

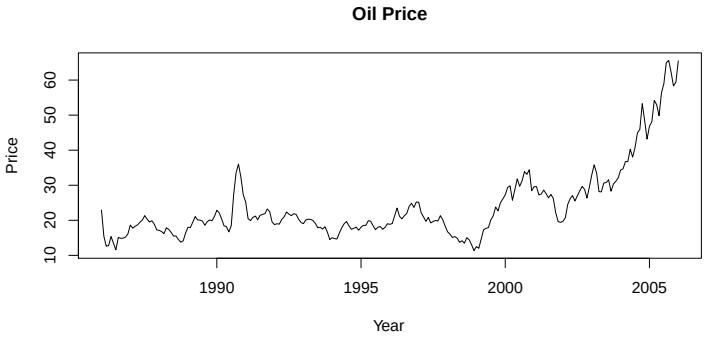
Consider the case where $Y_t = Y_{t-1} + e_t$, i.e., a random walk so that the differenced series is white noise: $\nabla Y_t = e_t$. So the appropriate model is an ARIMA(0,1,0). However, differencing again produces

$$\nabla^2 Y_t = \nabla Y_t - \nabla Y_{t-1} = e_t - e_{t-1}$$

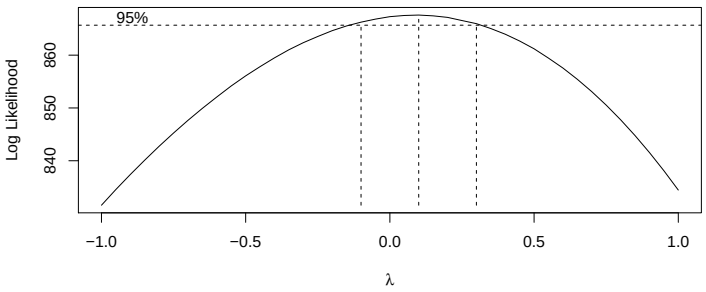
which is a (non-invertible) MA(1) and, therefore, we may then incorrectly consider an ARIMA(0,2,1).

Example 2.1. *Oil Price: Stationarity*

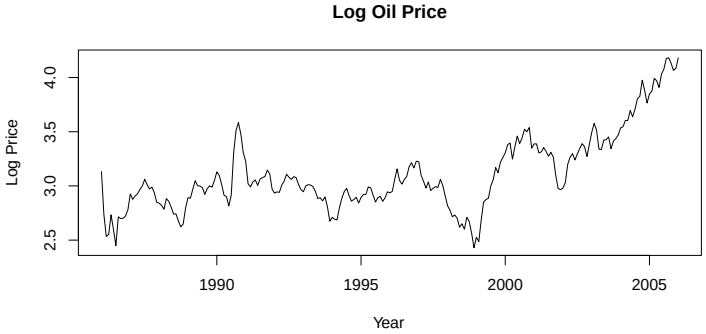
Consider the `oil.price` series which has clear trend and also some evidence of increasing variance.

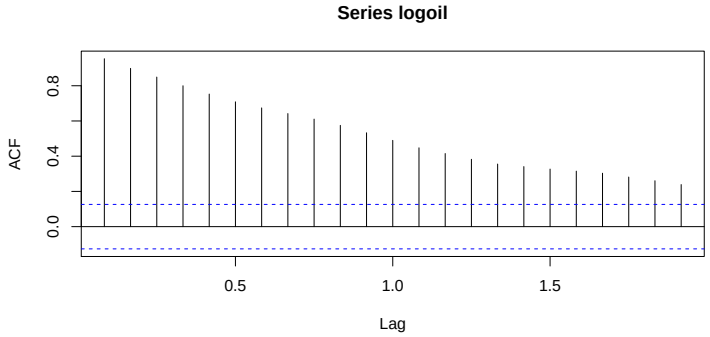


Applying `BoxCox.ar` shows that $\lambda = 0$ is well within the confidence interval which supports a log-transformation.



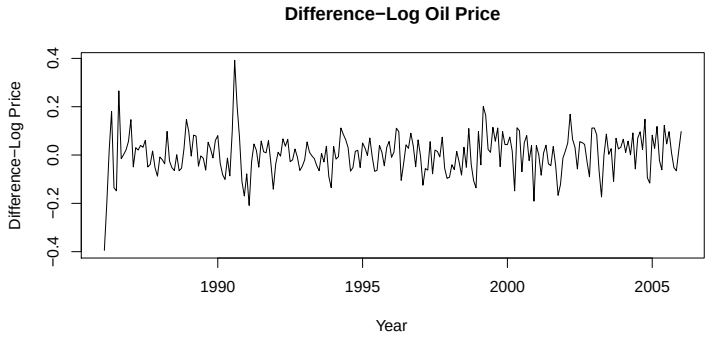
The log-series clearly has trend which suggests that differencing is also required.

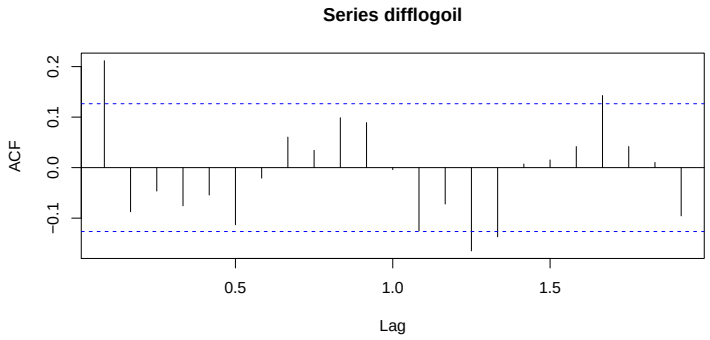




The persisting autocorrelations (failing to decay quickly) also suggest trend and carrying out `adf.test` leads to a p-value of 0.9189 supporting the null hypothesis that differencing is needed.

A plot of the difference-log-series below looks reasonably stationary (apart from a few outliers).





Furthermore, the ACF suggests stationarity and this is confirmed by the Dicky-Fuller test (p-value = 0.01).

---

3 ARMA Order

Once we have reduced the series to a stationary series through differencing (and Box-Cox transformation), i.e., we have decided on d in ARIMA(p, d, q), we can then decide on reasonable values of p and q.

Note that at this stage we do not need to be so precise about p and q. We only require a few preliminary values so that we can start estimating the parameters in the ARIMA model - and later zone in on a final model.

3.1 Autocorrelation Function (ACF)

For a pure MA(q) process, we know that the theoretical autocorrelation function is non-zero for the first q lags and zero thereafter. Thus, on average the sample **ACF cuts off after lag q for a *pure* MA(q) process**.

Note: Due to sampling error, and the fact that sample-autocorrelations can be highly correlated with each other, the we will never expect to see this behaviour exactly, i.e., some autocorrelations will be statistically significant just by chance.

Recall from Lecture 2 that

$$r_k \to N(0, \sigma^2 = \tfrac{1}{n})$$

when n is large under the assumption that the process is white noise. Therefore, any sample autocorrelations bigger than $\frac{1.96}{\sqrt{n}}$ in magnitude may be deemed *statistically significant*. We can use this as a **rough guide** for selecting the order of an MA process.

As an example, looking at the sample ACF for the difference-log oil series in Example 2.1 (the last plot), an MA(1) model looks plausible, i.e., an ARIMA(0,1,1) model for $\log Y_t$ where Y_t is the original oil price series.

3.2 Partial ACF

The sample ACF is useful for determining the order of an MA process. However, for an AR process, recall that the ACF decays rather than cutting off. In practice it is very difficult to determine the order of an AR process based on the ACF.

Thankfully, the so-called **partial ACF (PACF) cuts off after lag p for a *pure* AR(p) process**.

We will not study the properties of partial autocorrelation in detail apart from showing the above result. First we need to formally define partial correlation.

Partial correlation is the **residual (or adjusted) correlation between two random variables after removing the effects of other variables**. Consider three random variables X, Y and Z. The partial correlation is defined as

$$\text{Corr}[X - \hat{X}(Z),\ Y - \hat{Y}(Z)]$$

where $\hat{X}(Z) = E(X|Z)$ and $\hat{Y}(Z) = E(Y|Z)$ are the predicted values of X and Y given Z. Note that these predicted values are random variables as they depend on the random variable Z.

The partial correlation is the correlation between residuals $X - \hat{X}(Z)$ and $Y - \hat{Y}(Z)$.

In the context of time series analysis, **partial *auto*correlation is the residual correlation between Y_t and Y_{t-k} after removing the effects of all variables in between $Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}$**, i.e.,

$$\begin{aligned}\tau_k &= \text{Corr}(Y_t - \hat{Y}_t,\ Y_{t-k} - \hat{Y}_{t-k}) \\ &\quad \text{where} \\ \hat{Y}_t &= E(Y_t \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}) \\ \hat{Y}_{t-k} &= E(Y_{t-k} \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}).\end{aligned}$$

Note that, by the definition of partial autocorrelation,

$$\begin{aligned}\tau_1 &= \text{Corr}(Y_t - E(Y_t),\ Y_{t-1} - E(Y_{t-1})) \\ &= \text{Corr}(Y_t,\ Y_{t-1}) \\ &\quad \text{(since } E(Y_t) \text{ and } E(Y_{t-1}) \text{ are constants)} \\ &= \rho_1\end{aligned}$$

i.e., the lag-1 partial autocorrelation is equal to the lag-1 autocorrelation.

AR(1)

The AR(1) process is defined by

$$Y_t = \phi Y_{t-1} + e_t$$

and therefore, Y_t is correlated with Y_{t-1} as this is a component of Y_t. Furthermore, Y_{t-2} is a component of Y_{t-1} and, hence, Y_t is correlated with Y_{t-2} through Y_{t-1}. Continuing this argument, Y_t is correlated with all Y_{t-k} (although the correlations decay).

Clearly the link between Y_t and previous series values occurs through Y_{t-1}. This suggests that we can cut off this train of dependence by eliminating the Y_{t-1} dependence, i.e., remove the effect of Y_{t-1}.

For $k > 1$ we have

$$\begin{aligned}\hat{Y}_t &= E(Y_t \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}) \\ &= E(\phi Y_{t-1} + e_t \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}) \\ &= \phi E(Y_{t-1} \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}) \\ &\quad + E(e_t \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}) \\ &= \phi Y_{t-1} + E(e_t) \\ &= \phi Y_{t-1}.\end{aligned}$$

Note that the second last line follows since the expected value of Y_{t-1} *given* that we know Y_{t-1} is just Y_{t-1} and e_t is independent of the past.

Thus, we have the residual

$$\begin{aligned}Y_t - \hat{Y}_t &= \phi Y_{t-1} + e_t - \phi Y_{t-1} \\ &= e_t\end{aligned}$$

and, hence, for $k > 1$, the partial autocorrelation function is

$$\begin{aligned}\tau_k &= \text{Corr}(Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k}) \\ &= \text{Corr}(e_t, Y_{t-k} - \hat{Y}_{t-k}) \\ &= 0\end{aligned}$$

since e_t is independent of the past and

$$\begin{aligned}&Y_{t-k} - \hat{Y}_{t-k} \\ &= Y_{t-k} - E(Y_{t-k} \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1})\end{aligned}$$

is a function of the past values in the series $Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}, Y_{t-k}$.

Thus, $\tau_1 = \rho_1$ and $\tau_k = 0$ for $k > 1$.

AR(2)

The AR(2) process is defined by

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

For $k > 2$ we have

$$\begin{aligned}\hat{Y}_t &= E(Y_t \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}) \\ &= E(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \,|\, Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}) \\ &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2}\end{aligned}$$

which follows from the same line of argument used in the AR(1) case.

Thus, we have that

$$Y_t - \hat{Y}_t = e_t$$

and, hence, for $k > 2$, the partial autocorrelation function is

$$\begin{aligned}\tau_k &= \text{Corr}(Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k}) \\ &= \text{Corr}(e_t, Y_{t-k} - \hat{Y}_{t-k}) \\ &= 0\end{aligned}$$

since e_t is independent of the past and $Y_{t-k} - \hat{Y}_{t-k}$ is is a function of the past values in the series $Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}, Y_{t-k}$.

We know that $\tau_1 = \rho_1$ and that $\tau_k = 0$ for $k > 2$. What about $k = 2$? It can be shown that

$$\begin{aligned}\hat{Y}_t &= E(Y_t \,|\, Y_{t-1}) = \rho_1 Y_{t-1} \\ \hat{Y}_{t-2} &= E(Y_{t-2} \,|\, Y_{t-1}) = \rho_1 Y_{t-1}\end{aligned}$$

and, furthermore, the partial autocorrelation is

$$\begin{aligned}\tau_2 &= \text{Corr}(Y_t - \hat{Y}_t,\ Y_{t-2} - \hat{Y}_{t-2}) \\ &= \text{Corr}(Y_t - \rho_1 Y_{t-1},\ Y_{t-2} - \rho_1 Y_{t-1}) \\ &= \phi_2.\end{aligned}$$

AR(p)

We have seen that the partial autocorrelation cuts off after lag-1 for an AR(1) process and after lag-2 for an AR(2) process. It is not difficult to show that the **partial autocorrelation function cuts off after lag-p for an AR(p) process**. This follows along the same lines as the AR(1) and AR(2) cases.

Thus, the PACF for an AR process behaves in the same way as the ACF does for an MA process. Hence, the sample PACF is useful for determining the order of a pure AR process. Furthermore, like the sample ACF, it can be shown that

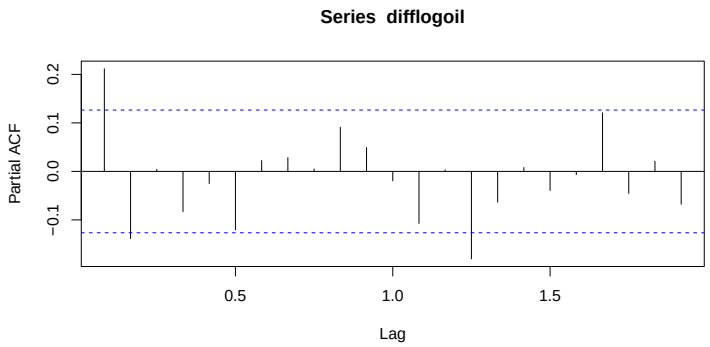
$$\hat{\tau}_k \to N(0, \sigma^2 = \tfrac{1}{n})$$

when n is large under the assumption that the process is white noise (i.e., an AR(0)).

Example 3.1. *Oil Price: PACF*
In Example 2.1, we saw that the sample ACF for the difference-log oil series suggested an MA(1) model may be appropriate.

We now consider the sample PACF.

```
pacf(difflogoil)
```



This appears to suggest that an AR(1) or AR(2) model may be appropriate.

3.3 Summary of ACF and PACF

The following table is a guide for recognising the model based on the ACF and PACF:

	AR(p)	MA(q)	ARMA(p,q)
ACF	Decays	Cuts off after lag q	Decays
PACF	Cuts off after lag p	Decays	Decays

3.4 Extended ACF

From the above summary of ACF and PACF, it should be clear that it will not be easy to determine the order of an ARMA process based on these functions. The **extended ACF (EACF) is used for determining the orders p and q of an ARMA process**.

The theory surrounding the EACF is more complicated than the ACF and PACF but we will gives some basic insight into this procedure which is based on *filtering out* the AR part of the ARMA process.

As an example consider the ARMA(1,1) process:

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}$$

Assuming we can estimate ϕ without estimating θ (Tsay and Tiao (1984) developed such a procedure), then we could define a type of residual given by

$$\begin{aligned}\varepsilon_t &= Y_t - \hat{\phi} Y_{t-1} \\ &\approx Y_t - \phi Y_{t-1} \\ &= e_t - \theta e_{t-1}\end{aligned}$$

Therefore, ε_t would approximately be an MA(1) process and, hence, its sample ACF cuts off after lag-1.

In practice we will not know the order of AR part so we try various orders. Let's say we try an AR(2) for example. Since the reality is an AR(1), we should expect $\hat{\phi}_1 \approx \phi$ and $\hat{\phi}_2 \approx 0$. In this case our residual is

$$\begin{aligned}\varepsilon_t &= Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} \\ &\approx Y_t - \phi Y_{t-1} - \hat{\phi}_2 Y_{t-2} \\ &= e_t - \theta e_{t-1} - \hat{\phi}_2 Y_{t-2}.\end{aligned}$$

Now, since Y_t is an ARMA(1,1), it can be written as a general linear process (see Lecture 4):

$$\begin{aligned}Y_t &= e_t + \sum_{j=1}^{\infty} \phi^{j-1}(\phi - \theta) e_{t-j} \\ &\approx e_t + \underbrace{(\phi - \theta) e_{t-1}}_{\text{1st term in sum}}\end{aligned}$$

Note: this is a very loose approximation, i.e., its use is not very formal but simply provides some intuition.

Thus, $Y_{t-2} \approx e_{t-2} + (\phi - \theta) e_{t-3}$ and, hence,

$$\begin{aligned}\varepsilon_t &\approx e_t - \theta e_{t-1} - \hat{\phi}_2 Y_{t-2} \\ &\approx e_t - \theta e_{t-1} - \hat{\phi}_2 e_{t-2} - \hat{\phi}_2(\phi - \theta) e_{t-3}.\end{aligned}$$

Therefore, the result of *incorrectly* filtering out an AR(2) component from an ARMA(1,1) process produces residuals which look like an MA(2) or MA(3) process.

Therefore, if we filter out:

- Nothing $\Rightarrow$ ARMA(1,1) $\Rightarrow$ ACF decays.

- AR(1) ⇒ MA(1) ⇒ ACF cuts off after lag-1.
- AR(2) ⇒ MA(2) ⇒ ACF cuts off after lag-2.
- AR(3) ⇒ MA(3) ⇒ ACF cuts off after lag-3.
- AR(4) ⇒ MA(4) ⇒ ACF cuts off after lag-4.
- ⋮

If we keep track of the significant sample autocorrelations we will expect to see something like:

True model ARMA(1,1)

Filter	ACF cut-off after lag										
AR(?)	0	**1**	2	3	4	5	6	7	8	9	10
0	x	x	x	x	x	x	x	x	x	x	x
1	x	**0**	0	0	0	0	0	0	0	0	0
2	x	x	0	0	0	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0	0	0
4	x	x	x	x	0	0	0	0	0	0	0
5	x	x	x	x	x	0	0	0	0	0	0
6	x	x	x	x	x	x	0	0	0	0	0
7	x	x	x	x	x	x	x	0	0	0	0

This leads to a triangle whose **upper-left vertex identifies the order of the ARMA model**.

For an ARMA(2,3) model we would expect:

True model ARMA(2,3)

Filter	ACF cut-off after lag										
AR(?)	0	1	2	**3**	4	5	6	7	8	9	10
0	x	x	x	x	x	x	x	x	x	x	x
1	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	**0**	0	0	0	0	0	0	0
3	x	x	x	x	0	0	0	0	0	0	0
4	x	x	x	x	x	0	0	0	0	0	0
5	x	x	x	x	x	x	0	0	0	0	0
6	x	x	x	x	x	x	x	0	0	0	0
7	x	x	x	x	x	x	x	x	0	0	0

For an ARMA(4,1) model we would expect:

True model ARMA(4,1)

Filter	ACF cut-off after lag										
AR(?)	0	**1**	2	3	4	5	6	7	8	9	10
0	x	x	x	x	x	x	x	x	x	x	x
1	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	x	x	x	x	x	x	x	x
3	x	x	x	x	x	x	x	x	x	x	x
4	x	**0**	0	0	0	0	0	0	0	0	0
5	x	x	0	0	0	0	0	0	0	0	0
6	x	x	x	0	0	0	0	0	0	0	0
7	x	x	x	x	0	0	0	0	0	0	0

In practice the sample EACF will not be this clear cut. Many estimated correlations will be statistically significant just by chance. However, we try to spot this basic triangular pattern.

Example 3.2. *Oil Price: EACF*

In Examples 2.1 and 3.1, we looked at the ACF and PACF for the difference-log oil series. We now consider the sample EACF.

```
eacf(difflogoil)
```

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o o o o o o  o  o  o
1 x x o o o o o o o o x  o  o  o
2 o x o o o o o o o o o  o  o  o
3 o x o o o o o o o o o  o  o  o
4 o x x o o o o o o o o  o  o  o
5 o x o x o o o o o o o  o  o  o
6 o x o x o o o o o o o  o  o  o
7 x x o x o o o o o o o  o  o  o
```

This appears to suggest that an ARMA(0,1), i.e., an MA(1), model may be appropriate.

3.5 Information Criteria

Another way to determine the order of the ARMA(p,q) model is to fit an array of different ARMA models to the data and compare them using some measure of fit.

Let $\ell(\hat{\phi}, \hat{\theta})$ be the maximum log-likelihood value for a fitted ARMA. If we add more parameters, we can always achieve a higher likelihood value (i.e., better fit) but our aim is to find the simplest model which adequately fits the data. Thus, a penalty term is included which accounts for the number of parameters, m. So-called **information criteria (IC)** provide a measure of fit with a penalty for complexity.

The most popular information criteria are the **Akaike Information Criterion (AIC)**

$$\text{AIC} = -2\ell(\hat{\phi}, \hat{\theta}) + 2m$$

and the **Bayesian Information Criterion (BIC)**

$$\text{BIC} = -2\ell(\hat{\phi}, \hat{\theta}) + m\ln(n)$$

where n is the sample size. Both of these measures have the following form

$$-2\ell(\hat{\phi}, \hat{\theta}) + \delta\, m$$

where δ is the level of penalty: $\delta = 2$ for AIC and $\delta = \ln(n)$ for BIC. Note that $\ln 7 \approx 2$ and, therefore, the BIC has a larger penalty than the AIC for sample sizes

usually observed in practice, e.g., $\ln 100 = 4.6$. Note that the penalty grows with the sample size for BIC but is fixed for AIC.

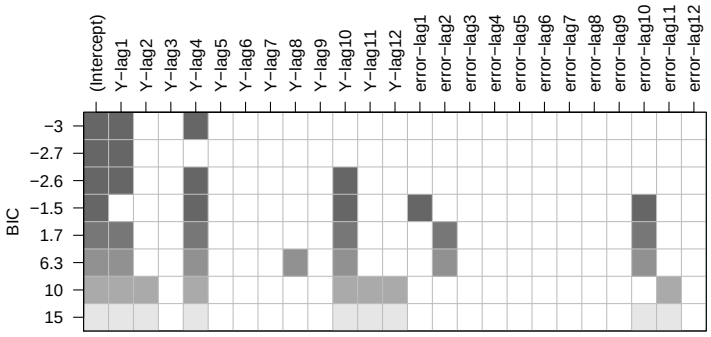
The aim is to **minimize the IC** (whichever one we decide to use). As a rule of thumb, **two models with IC values differing by less than 2 units are both plausible** based on the data.

Example 3.3. *Oil Price: Best Subsets*

The `armasubsets` function in `R` fits an array of different models to the data and displays the top 8 models (ranked based on BIC).

```
oilsubs <- armasubsets(difflogoil, nar=12, nma=12)
plot(oilsubs)
```

The above code produces the plot below which ranks the top 8 models from lowest BIC (best model) at the top.



Firstly note that in the code we have to specify values for nar and `nma` which are the maximum AR and MA orders to be explored - the choice is very arbitrary but we typically will not have extremely high order models. As this data is monthly, it seems reasonable to try up to order 12 in both components.

As for the plot produced, each row corresponds to a model and the shaded boxes indicate which terms were significant in that model. So, for example, the top three models in the above display are:

1. AR(4) with significant Y_{t-1} and Y_{t-4} terms
2. AR(1) with significant Y_{t-1} term
3. AR(10) with significant Y_{t-1}, Y_{t-4} and Y_{t-10} terms.

Based on this result and the previous examples we would certainly be looking at MA(1), AR(1), AR(2) and AR(4) models for this series. Remember that this is a difference-log series so these are in fact ARIMA(0,1,1), ARIMA(1,1,0), ARIMA(2,1,0) and ARIMA(4,1,0) models for the log oil series.
