Historically (in the 1960’s) econometric models with many variables and equations were developed.

- These models fit the data beautifully.
- They are not good at forecasting.

It turns out that low-dimensional ARMA(p,q) processes typically do a better job with forecasting.
The Box-Jenkins Principle for forecasting:

1. Transform the data, if necessary, so covariance stationarity is a reasonable assumption.
2. Make an initial guess at small values of $p$ and $q$ for an $ARMA(p, q)$ model that might describe the (transformed) series.
3. Estimate the parameters.
4. Perform diagnostic analysis to confirm that the model is consistent with the observed features of the data.

For Step 2 we need information on the correlation structure.
The Autocorrelation Function

- Recall that the autocorrelation function is defined as:

\[ \rho_s = \frac{\gamma_s}{\sqrt{V(y_{t+s})V(y_t)}} \]

- This is what these look like:
The Autocorrelation Function

- More examples
Clearly it is the case that it is hard to tell the processes apart from the autocorrelation function alone.

Another important tool is the partial autocorrelation function.

Definition:

- First demean $y$, such that $y_t^* = y_t - \mu$
- Then the first partial autocorrelation is $\phi_{11}$ such that

$$y_t^* = \phi_{11} y_{t-1}^* + u_t$$

- The second partial autocorrelation is $\phi_{22}$ where

$$y_t^* = \phi_{21} y_{t-1}^* + \phi_{22} y_{t-2}^* + u_t$$

Clearly for an $AR(1)$, $\phi_{22} = 0$.

These can be found as functions of the autocorrelations:

Note that the $m^{th}$ autocorrelation simply is the last coefficient in the linear regression of $y_t - \mu$ on its own $m$ most recent values.
Recall that we can write the coefficients of the projection of $y_{t+1}$ on $X_t$ as

$$\alpha' = E \left( y_{t+1} X'_t \right) \left[ E \left( X'_t X_t \right) \right]^{-1}$$

Here

$$X'_t = [y_{t-1} - \mu, y_{t-2} - \mu, \ldots, y_{t-m} - \mu] \Leftrightarrow X'_t = [y^*_t, y^*_t, \ldots, y^*_t]$$

And we’re projecting $y^*_t$, not $y_{t+1}$. 
Now note that

\[ E(X'_t X_t) = \begin{bmatrix}
E(y^*_{t-1}y^*_{t-1}) & E(y^*_{t-1}y^*_{t-2}) & \cdots & E(y^*_{t-1}y^*_{t-m}) \\
E(y^*_{t-2}y^*_{t-1}) & E(y^*_{t-2}y^*_{t-2}) & \cdots & E(y^*_{t-2}y^*_{t-m}) \\
\vdots & \vdots & \ddots & \vdots \\
E(y^*_{t-m}y^*_{t-1}) & E(y^*_{t-m}y^*_{t-2}) & \cdots & E(y^*_{t-m}y^*_{t-m})
\end{bmatrix} = 
\begin{bmatrix}
\gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\
\gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0
\end{bmatrix}
\]

and

\[ E(y^*_t X'_t) = \begin{bmatrix}
E(y^*_t y^*_{t-1}) & E(y^*_t y^*_{t-2}) & \cdots & E(y^*_t y^*_{t-m})
\end{bmatrix} \]
This means that we could find the partial autocorrelations by taking the \( m' \)th entry of \( \alpha \) where

\[
\alpha' = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_m \end{bmatrix} \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\
\gamma_1 & \gamma_0 & \cdots & \gamma_{m-2} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{bmatrix}^{-1}
\]

Note that we need to invert an \( m \times m \) matrix to get this one.....
The partial autocorrelation of different models look like:
The Autocorrelation Functions

- An overview of the theoretical properties of the correlation functions:

<table>
<thead>
<tr>
<th>Process</th>
<th>ACF</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>White noise</td>
<td>All $\rho_s = 0$ ($s = 0$)</td>
<td>All $\phi_{ss} = 0$</td>
</tr>
<tr>
<td>AR(1): $a_1 &gt; 0$</td>
<td>Direct exponential decay: $\rho_s = a_1^s$</td>
<td>$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$</td>
</tr>
<tr>
<td>AR(1): $a_1 &lt; 0$</td>
<td>Oscillating decay: $\rho_s = a_1^s$</td>
<td>$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$</td>
</tr>
<tr>
<td>AR(p)</td>
<td>Decays toward zero. Coefficients may oscillate.</td>
<td>Spikes through lag $p$. All $\phi_{ss} = 0$ for $s &gt; p$.</td>
</tr>
<tr>
<td>MA(1): $\beta &gt; 0$</td>
<td>Positive spike at lag 1. $\rho_s = 0$ for $s \geq 2$</td>
<td>Oscillating decay: $\phi_{11} &gt; 0$.</td>
</tr>
<tr>
<td>MA(1): $\beta &lt; 0$</td>
<td>Negative spike at lag 1. $\rho_s = 0$ for $s \geq 2$</td>
<td>Geometric decay: $\phi_{11} &lt; 0$.</td>
</tr>
<tr>
<td>ARMA(1, 1): $a_1 &gt; 0$</td>
<td>Exponential decay beginning at lag 1.</td>
<td>Oscillating decay beginning at lag 1. $\phi_{11} = \rho_1$.</td>
</tr>
<tr>
<td>ARMA(1, 1): $a_1 &lt; 0$</td>
<td>Oscillating decay beginning at lag 1.</td>
<td>Exponential decay beginning at lag 1. $\phi_{11} = \rho_1$ and $\text{sign} (\phi_{ss}) = \text{sign}(\phi_{11})$.</td>
</tr>
<tr>
<td>ARMA $(p, q)$</td>
<td>Decay (either direct or oscillatory) beginning at lag $q$.</td>
<td>Decay (either direct or oscillatory) after lag $p$.</td>
</tr>
</tbody>
</table>
We want to calculate the autocorrelations from the data and see how they match up to the theoretical autocorrelations of various processes. This will help us find the correct model.

We use the following as sample autocorrelations. These are basically estimates of the parameters.

\[
\hat{\mu} = \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t
\]

\[
\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2
\]

\[
r_s = \frac{\sum_{t=s+1}^{T} (y_t - \bar{y}) (y_{t-s} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}
\]
Sample Autocorrelations II

- Just as with any other estimator we need to know the variance of these estimates.
- If \( \{y_t\} \) is stationary with normally distributed errors, Box and Jenkins found that

\[
V(r_s) = \begin{cases} 
\frac{1}{T} & \text{if } s = 1 \\
\frac{1}{T} \left( 1 + 2 \sum_{j=1}^{s-1} r_j^2 \right) & \text{if } s > 1
\end{cases}
\]

- How to use this?
  - For an MA\((s)\) process, \( \gamma_s \) is 0.
  - For large samples, \( r_s \), is normally distributed.
  - This is enough information to create a confidence interval.

- Use the confidence interval to test:
  - First look at \( r_1 \). If \( r_1 \) falls outside the confidence interval, we reject \( \gamma_1 = 0 \).
  - Then we can look at \( r_2 \) etc.
Sample Autocorrelations III

- Do enough of these and one **will** fall outside the confidence interval. (size!!)
- Ljung-Box Statistic:

  \[
  Q = T \left( T + 2 \right) \sum_{k=1}^{s} \frac{r_k^2}{T - k}
  \]

- If all the autocorrelations are zero this is \( \chi^2 (s) \).
- Can also be used to check that the residual are white noise. Then
  - under the null \( Q \sim \chi^2 (s - p - q) \), or, if a constant is included, \( Q \sim \chi^2 (s - p - q - 1) \)
Goodness of Fit

- Just as in standard models you can assess goodness of fit.
- Fit can always be improved by adding to \( p \) and \( q \).
- Criteria which trade off fit vs a more parsimonious model:
  - Akaike information criterion:
    \[
    AIC = T \ln(SSR) + 2n
    \]
  - Schwartz Bayesian criterion:
    \[
    SBC = T \ln(SSR) + n \ln T
    \]
- Where \( n \) : number of estimated parameters and \( T \) : number of usable observations.
- Keep \( T \) the same even if estimating first an \( AR(1) \) and then an \( AR(2) \).
- SBC punishes additional parameters harder.
- SBC is consistent, AIC is not.
In general, estimation of AR models can be done with simple regressions, but MA processes are more complicated. 

For now let the computer do the estimation. 

We will concern ourselves with selecting the right model. 

Suppose 100 data points are generated according to:

$$y_t = 0.7y_{t-1} + \varepsilon_t$$

First calculate the sample (partial) auto correlations. 

The first three sample autocorrelations are $r_1 = 0.74$, $r_2 = 0.58$ and $r_3 = 0.47$.

Note that $\gamma_1 = 0.7$, $\gamma_2 = 0.49$ and $\gamma_3 = 0.343$.

The first sample partial autocorrelation is 0.71.
Estimation of ARMA models. II
A data example

- Graph of sample correlations:

- Recall that

\[ V(r_s) = \begin{cases} \frac{1}{T} & \text{if } s = 1 \\ \frac{1}{T} \left(1 + 2 \sum_{j=1}^{s-1} r_j^2 \right) & \text{if } s > 1 \end{cases} \]

- And that if we have an MA(q) then \( \rho_{q+1} = 0 \).
Test $MA(0)$ vs $MA(1)$. If the process is an $MA(0)$, then the standard deviation of $r_1$ is 0.1.

We had $r_1 = 0.74$.

Clearly reject that it is 0.

$V(r_2) = 0.021$ and the standard deviation is 0.1449. Again we reject that $r_2$ is 0.

Seems like we might have some action.

For the partial autocorrelation only lags 1 and 12 seem to be significant.

Either $AR(1)$, $AR(12)$ or $ARMA(1,1)$.

Box Jenkins would indicate looking only at $AR(1)$ and $ARMA(1,1)$ unless we have monthly data:
### Table 2.2 Estimates of an AR(1) Model

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$y_t = a_1 y_{t-1} + \varepsilon_t$</td>
<td>$y_t = a_1 y_{t-1} + \varepsilon_t + \beta_{12} \varepsilon_{t-12}$</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>Sum of squared residuals</td>
<td>85.21</td>
<td>85.17</td>
</tr>
<tr>
<td>Estimated $a_1$</td>
<td>0.7910</td>
<td>0.7953</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(0.0622)</td>
<td>(0.0638)</td>
</tr>
<tr>
<td>Estimated $\beta$</td>
<td></td>
<td>-0.033</td>
</tr>
<tr>
<td>(standard error)</td>
<td></td>
<td>(0.1134)</td>
</tr>
<tr>
<td>AIC; SBC</td>
<td>AIC = 441.9; SBC = 444.5</td>
<td>AIC = 443.9; SBC = 449.1</td>
</tr>
<tr>
<td>Ljung–Box $Q$-statistics for the residuals (significance level in parentheses)</td>
<td>$Q(8) = 6.43$ (0.490)</td>
<td>$Q(8) = 6.48$ (0.485)</td>
</tr>
<tr>
<td></td>
<td>$Q(16) = 15.86$ (0.391)</td>
<td>$Q(16) = 15.75$ (0.400)</td>
</tr>
<tr>
<td></td>
<td>$Q(24) = 21.74$ (0.536)</td>
<td>$Q(24) = 21.56$ (0.547)</td>
</tr>
</tbody>
</table>
For further diagnostics, we can plot the sample autocorrelations of the residuals:
More on the example:

Does there seem to be a model which allows most of PACF to be 0, but not at lag 12?

First consider the ACF and the PACF for the ARMA(1,1) model:

\[ y_t = a_1 y_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1} \]

First ACF. From Yule-Walker you get:

\[
\gamma_0 = a_1 \gamma_1 + \sigma^2 + b_1 E(y_t \varepsilon_{t-1})
\]

\[
E(y_t y_t) = a_1 E(y_t y_{t-1}) + E(y_t \varepsilon_t) + b_1 E(y_t \varepsilon_{t-1}) \iff
\]

\[
E(y_t \varepsilon_{t-1}) = E((a_1 y_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1}) \varepsilon_{t-1})
= a_1 \sigma^2 + b_1 \sigma^2 = (a_1 + b_1) \sigma^2
\]
\[ \gamma_0 = a_1 \gamma_1 + \sigma^2 + b_1 (a_1 + b_1) \sigma^2 = a_1 \gamma_1 + [1 + b_1 (a_1 + b_1)] \sigma^2 \]

\[ E(y_{t-1} y_t) = a_1 E(y_{t-1} y_{t-1}) + E(y_{t-1} \epsilon_t) + b_1 E(y_{t-1} \epsilon_{t-1}) \]

\[ \gamma_1 = a_1 \gamma_0 + b_1 \sigma^2 \]

Solve:

\[ \gamma_0 = a_1 \gamma_1 + [1 + b_1 (a_1 + b_1)] \sigma^2 = a_1^2 \gamma_0 + a_1 b_1 \sigma^2 + [1 + b_1 (a_1 + b_1)] \sigma^2 \]

\[ \gamma_0 = \frac{1 + 2a_1 b_1 + b_1^2}{1 - a_1^2} \sigma^2 \]

\[ \gamma_1 = a_1 \gamma_0 + b_1 \sigma^2 = \frac{a_1 + a_1^2 b_1 + a_1 b_1^2 + b_1}{1 - a_1^2} \sigma^2 \]

\[ \gamma_1 = \frac{(a_1 + b_1) (1 + a_1 b_1)}{1 - a_1^2} \sigma^2 \]
Estimation of ARMA models. III

\[ E(y_{t-2}y_t) = a_1 E(y_{t-2}y_{t-1}) + E(y_{t-2}\varepsilon_t) + b_1 E(y_{t-2}\varepsilon_{t-1}) \]

\[ \gamma_2 = a_1 \gamma_1 \]

- So, declining ACF.....
- Now PACF, Recall we need the \( m' \)th entry of \( \alpha \):

\[
\alpha' = \begin{bmatrix} \gamma_1 & \gamma_2 & \ldots & \gamma_m \end{bmatrix} \begin{bmatrix} \gamma_0 & \gamma_1 & \ldots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \ldots & \gamma_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \ldots & \gamma_0 \end{bmatrix}^{-1}
\]

\[
\phi_{11} = \frac{\gamma_1}{\gamma_0} = \frac{(a_1 + b_1)(1 + a_1 b_1)}{1 + 2a_1 b_1 + b_1^2}
\]
Estimation of ARMA models. IV

$$\alpha' = \begin{bmatrix} \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \gamma_0 \frac{\gamma_1}{\gamma_0^2 - \gamma_1^2} - \gamma_1 \frac{\gamma_2}{\gamma_0^2 - \gamma_1^2} & \gamma_0 \frac{\gamma_2}{\gamma_0^2 - \gamma_1^2} - \frac{\gamma_1^2}{\gamma_0^2 - \gamma_1^2} \end{bmatrix}$$

$$\phi_{22} = \frac{\gamma_0 \gamma_2 - \gamma_1^2}{\gamma_0^2 - \gamma_1^2} = \frac{\frac{\gamma_2}{\gamma_0} - \left(\frac{\gamma_1}{\gamma_0}\right)^2}{1 - \left(\frac{\gamma_1}{\gamma_0}\right)^2} = \frac{a_1 \frac{\gamma_1}{\gamma_0} - \left(\frac{\gamma_1}{\gamma_0}\right)^2}{1 - \left(\frac{\gamma_1}{\gamma_0}\right)^2}$$

$$= \frac{a_1 \phi_{11} - \phi_{11}^2}{1 - \phi_{11}^2}$$

$$\phi_{22} = 0 \iff \phi_{11} = a_1$$
Estimation of ARMA models. V

\[ \alpha' = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix} \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}^{-1} \]

\[ \phi_{33} = \gamma_1 \left( \frac{\gamma_1^2 - \gamma_0 \gamma_2}{(\gamma_0 - \gamma_2)(\gamma_0^2 - 2\gamma_1^2 + \gamma_0 \gamma_2)} \right) - \gamma_1 \frac{\gamma_2}{\gamma_0 + \gamma_2 \gamma_0 - 2\gamma_1^2} \\
+ \gamma_1 \frac{\gamma_3}{\gamma_0 \gamma_1 - \gamma_1 \gamma_2} \left( \frac{\gamma_0^2 - \gamma_1^2}{\gamma_0^2 + \gamma_2 \gamma_0 - 2\gamma_1^2} \right) \\
= \gamma_1 \left( \frac{\gamma_1^2 - \gamma_0 \gamma_2}{(\gamma_0 - \gamma_2)(\gamma_0^2 - 2\gamma_1^2 + \gamma_0 \gamma_2)} \right) - \gamma_2 + \frac{\gamma_3 (\gamma_0^2 - \gamma_1^2)}{\gamma_0 \gamma_1 - \gamma_1 \gamma_2} \right] \]
\[
\frac{\gamma_1^2 - \gamma_0 \gamma_2}{(\gamma_0 - \gamma_2)} - \gamma_2 + \frac{\gamma_3 (\gamma_0^2 - \gamma_1^2)}{\gamma_0 \gamma_1 - \gamma_1 \gamma_2}
\]

\[
= \frac{\gamma_3^2 + \gamma_1 \gamma_2^2 + \gamma_0 \gamma_3 - \gamma_1^2 \gamma_3 - 2 \gamma_0 \gamma_1 \gamma_2}{\gamma_1 (\gamma_0 - \gamma_2)}
\]

\[
= \frac{\phi_{11}^3 + a_1^2 \phi_{11}^2 + a_1^2 \phi_{11} - a_1^2 \phi_{11}^3 - 2 a_1 \phi_{11}^2}{\phi_{11} (1 - a_1 \phi_{11})}
\]

\[
= \gamma_0 \frac{\phi_{11}^3 + a_1^2 - 2 a_1 \phi_{11}}{\phi_{11} (1 - a_1 \phi_{11})} = \gamma_0 \frac{\phi_{11}^2 + a_1^2 - 2 a_1 \phi_{11}}{(1 - a_1 \phi_{11})}
\]

\[
= \gamma_0 \frac{(\phi_{11} - a_1)^2}{(1 - a_1 \phi_{11})}
\]
\[
\phi_{33} = \frac{\gamma_1}{(\gamma_0^2 - 2\gamma_1^2 + \gamma_0 \gamma_2)} \left[ \frac{\gamma_1^2 - \gamma_0 \gamma_2}{(\gamma_0 - \gamma_2)} - \gamma_2 + \frac{\gamma_3 (\gamma_0^2 - \gamma_1^2)}{\gamma_0 \gamma_1 - \gamma_1 \gamma_2} \right]
\]

\[
= \frac{\gamma_1 \gamma_0}{(\gamma_0^2 - 2\gamma_1^2 + \gamma_0 \gamma_2)} \frac{(\phi_{11} - a_1)^2}{(1 - a_1 \phi_{11})}
\]

So, for \( \phi_{33} = 0 \) and \( \phi_{11} > 0 \), we'd need

\[
\phi_{11} = a_1
\]

Why is this not compatible with our model?
You can also see this by solving:

\[ \phi_{11} = \frac{(a_1 + b_1)(1 + a_1 b_1)}{1 + 2a_1 b_1 + b_1^2} = a_1 \]

\[ \frac{(a_1 + b_1)(1 + a_1 b_1)}{1 + 2a_1 b_1 + b_1^2} = a_1 \Leftrightarrow \]

\[ (a_1 + b_1)(1 + a_1 b_1) - a_1 (1 + 2a_1 b_1 + b_1^2) = 0 \Leftrightarrow \]

\[ -b_1 (a_1 - 1)(a_1 + 1) = 0 \Leftrightarrow \]

\[ b_1 = 0 \]
Now calculate the ACF and the PACF for the model

\[ y_t = a_1 y_{t-1} + \varepsilon_t + b_{12} \varepsilon_{t-12} \]

From Yule-Walker you get:

\[
E(y_t y_t) = a_1 E(y_{t} y_{t-1}) + E(y_t \varepsilon_t) + b_{12} E(y_t \varepsilon_{t-12}) \]

\[ \gamma_0 = a_1 \gamma_1 + \sigma^2 + b_{12} E(y_t \varepsilon_{t-12}) \]

\[
E(y_t \varepsilon_{t-12}) = \mathbb{E} \left( \left[ \frac{1}{1 - a_1 L} (\varepsilon_t + b_{12} \varepsilon_{t-12}) \right] \varepsilon_{t-12} \right)
\]

\[
= \mathbb{E} \left( \left[ \left( \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} + b_{12} \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-12-i} \right) \right] \varepsilon_{t-12} \right)
\]

\[
= \mathbb{E} \left( a_{12}^2 \varepsilon_{t-12}^2 + b_{12} \varepsilon_{t-12}^2 \right)
\]

\[
= (a_{12}^2 + b_{12}) \sigma^2
\]
Estimation of ARMA models. X

\[ \gamma_0 = a_1 \gamma_1 + \gamma_0 + b_{12} E( y_t \varepsilon_{t-12}) \Leftrightarrow \]
\[ \gamma_0 = a_1 \gamma_1 + \sigma^2 + b_{12} (a_{12}^1 + b_{12}) \sigma^2 \]

\[ E( y_{t-1} y_t ) = a_1 E( y_{t-1} y_{t-1} ) + E( y_{t-1} \varepsilon_t ) + b_{12} E( y_{t-1} \varepsilon_{t-12}) \Leftrightarrow \]
\[ \gamma_1 = a_1 \gamma_0 + b_{12} E( y_{t-1} \varepsilon_{t-12}) \]

\[ E( y_{t-1} \varepsilon_{t-12}) = E \left( \left[ \frac{1}{1 - a_1 L} (\varepsilon_{t-1} + b_{12} \varepsilon_{t-13}) \right] \varepsilon_{t-12} \right) \]
\[ = E \left( \left[ \left( \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-1-i} + b_{12} \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-13-i} \right) \right] \varepsilon_{t-12} \right) \]
\[ = a_{11}^1 \sigma^2 \]

\[ \gamma_1 = a_1 \gamma_0 + b_{12} a_{11}^1 \sigma^2 \]
Plug into first equation

\[ \gamma_0 = a_1 \gamma_1 + \sigma^2 + b_{12} \left( a_{11}^1 + b_{12} \right) \sigma^2 \]

\[ \gamma_0 = a_1 \left( a_1 \gamma_0 + b_{12} a_{11}^1 \sigma^2 \right) + \sigma^2 + b_{12} \left( a_{11}^1 + b_{12} \right) \sigma^2 \]

\[ (1 - a_1^2) \gamma_0 = b_{12} a_{11}^1 \sigma^2 + \sigma^2 + b_{12} \left( a_{11}^1 + b_{12} \right) \sigma^2 \]

\[ \gamma_0 = \frac{1 + 2 b_{12} a_{11}^1 + b_{12}^2}{1 - a_1^2} \sigma^2 \]

\[ \gamma_1 = a_1 \gamma_0 + b_{12} a_{11}^1 \sigma^2 \]

\[ = a_1 \frac{1 + 2 b_{12} a_{11}^1 + b_{12}^2}{1 - a_1^2} \sigma^2 + b_{12} a_{11}^1 \sigma^2 \]

\[ = a_1 + 2 b_{12} a_{11}^1 + b_{12} a_{11}^1 + b_{12} a_{11}^1 - b_{12} a_{11}^1 \sigma^2 \]

\[ = \frac{a_1 + 2 b_{12} a_{11}^1 + b_{12} a_{11}^1 - b_{12} a_{11}^1}{1 - a_1^2} \sigma^2 \]
Estimation of ARMA models. XII

\[
E(y_{t-2}y_t) = a_1 E(y_{t-2}y_{t-1}) + E(y_{t-2}\epsilon_t) + b_{12} E(y_{t-2}\epsilon_{t-12}) \iff \\
\gamma_2 = a_1 \gamma_1 + b_{12} E(y_{t-2}\epsilon_{t-12})
\]

\[
E(y_{t-2}\epsilon_{t-12}) = E \left( \left[ \frac{1}{1 - a_1 L} (\epsilon_{t-2} + b_{12} \epsilon_{t-14}) \right] \epsilon_{t-12} \right) \\
= E \left( \left[ \sum_{i=0}^{\infty} a_1^i \epsilon_{t-2-i} + b_{12} \sum_{i=0}^{\infty} a_1^i \epsilon_{t-14-i} \right] \epsilon_{t-12} \right) \\
= a_1^{10} \sigma^2
\]
\[ \gamma_2 = a_1 \gamma_1 + b_{12} a_1^{10} \sigma^2 \]
\[ = a_1 \frac{a_1 + b_{12} a_1^{13} + b_{12}^2 a_1 + b_{12} a_1^{11}}{1 - a_1^2} \sigma^2 + b_{12} a_1^{10} \sigma^2 \]
\[ = \frac{a_1^2 + b_{12} a_1^{14} + b_{12}^2 a_1^2 + b_{12} a_1^{12} + b_{12} a_1^{10} - b_{12} a_1^{12}}{1 - a_1^2} \sigma^2 \]
\[ = \frac{a_1^2 + b_{12} a_1^{14} + b_{12}^2 a_1^2 + b_{12} a_1^{10}}{1 - a_1^2} \sigma^2 \]

\[ \gamma_j = 1 + \beta_{12} a_1^{12} + b_{12}^2 + \beta_{12} a_1^{12-2j} \frac{a_1^j \sigma^2}{1 - a_1^2}, \ 1 \leq j \leq 12 \]

\[ \gamma_j = a_1 \gamma_{j-1}, \ j \geq 13 \]
This provides autocorrelation coefficients:

\[ \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{a_1 + b_{12}a_1^{13} + b_{12}^2a_1 + b_{12}a_1^{11}}{1-a_1} \sigma^2 = a_1 \frac{1 + b_{12}a_1^{12} + b_{12}a_1^{10} + b_{12}^2}{1 + 2b_{12}a_1^{12} + b_{12}^2} \sigma^2 \]

When \( 2 \leq j \leq 12 \)

\[ \rho_j = \frac{\gamma_j}{\gamma_0} = \frac{1 + b_{12}a_1^{12} + b_{12}^2 + b_{12}a_1^{12-2j}}{1-a_1} \sigma^2 a_1^j \]

\[ = \frac{1 + b_{12}a_1^{12} + b_{12}^2 + b_{12}a_1^{12-2j}}{1 + 2b_{12}a_1^{12} + b_{12}^2} \sigma^2 a_1^j \]

and

\[ \rho_j = \frac{\gamma_j}{\gamma_0} = \frac{a_1 \gamma_{j-1}}{\gamma_0} = a_1 \rho_{j-1}, \ j \geq 13 \]
Estimation of ARMA models. XV

- So, decaying autocorrelations after lag 12.
- Now, the PACF. Starting with the second:

\[
\alpha' = \begin{bmatrix} \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}^{-1}
\]

\[
= \begin{bmatrix} \frac{\gamma_0 \gamma_1 - \gamma_1 \gamma_2}{\gamma_0^2 - \gamma_1^2} & \frac{\gamma_0 \gamma_2 - \gamma_1^2}{\gamma_0^2 - \gamma_1^2} \end{bmatrix}
\]

- Thus,

\[
\phi_{22} = \frac{\gamma_0 \gamma_2 - \gamma_1^2}{\gamma_0^2 - \gamma_1^2} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}
\]

\[
\phi_{33} = \frac{\gamma_1}{(\gamma_0^2 - 2\gamma_1^2 + \gamma_0 \gamma_2)} \left[ \frac{(\phi_{11} - a_1)^2}{(1 - a_1 \phi_{11})} \right]
\]

For this to be 0 we need

\[
\rho_2 = \rho_1^2
\]
but that implies

\[
\frac{1 + b_{12} a_1^{12} + b_{12}^2 + b_{12} a_1^8}{1 + 2 b_{12} a_1^{12} + b_{12}^2} a_1 = a_1^2 \left( \frac{1 + b_{12} a_1^{12} + b_{12} a_1^{10} + b_{12}^2}{1 + 2 b_{12} a_1^{12} + b_{12}^2} \right)^2 \Leftrightarrow
\]

\[
(1 + b_{12} a_1^{12} + b_{12}^2 + b_{12} a_1^8) (1 + 2 b_{12} a_1^{12} + b_{12}^2) = (1 + b_{12} a_1^{12} + b_{12} a_1^{10} + b_{12}^2)^2
\]

\[
\beta_{12} a_1^8 (a_1 - 1)^2 (a_1 + 1)^2 (b_{12}^2 + b_{12} a_1^{12} + 1) = 0 \Leftrightarrow b_{12} = 0 \lor a_1 = 0
\]

- Not 0 for this model.
- Clearly PACF 2 – 11 are NOT 0 for this model.
Another model candidate would be

\[ y_t = a_1 y_{t-1} + a_2 y_{t-12} + \varepsilon_t \]

Calculate ACF:

\[
E(y_t y_t) = a_1 E(y_t y_{t-1}) + a_2 E(y_t y_{t-12}) + E(y_t \varepsilon_t) \\
\gamma_0 = a_1 \gamma_1 + a_2 \gamma_{12} + \sigma^2
\]

\[
E(y_{t-1} y_t) = a_1 E(y_{t-1} y_{t-1}) + a_2 E(y_{t-1} y_{t-12}) + E(y_{t-1} \varepsilon_t) \\
\gamma_1 = a_1 \gamma_0 + a_2 \gamma_{11}
\]

\[
E(y_{t-2} y_t) = a_1 E(y_{t-2} y_{t-1}) + a_2 E(y_{t-2} y_{t-12}) + E(y_{t-2} \varepsilon_t) \\
\gamma_2 = a_1 \gamma_1 + a_2 \gamma_{10}
\]

\[
E(y_{t-3} y_t) = a_1 E(y_{t-3} y_{t-1}) + a_2 E(y_{t-3} y_{t-12}) + E(y_{t-3} \varepsilon_t) \\
\gamma_3 = a_1 \gamma_2 + a_2 \gamma_9
\]
Skip to the 12th:

\[ E(y_{t-12}y_t) = a_1 E(y_{t-12}y_{t-1}) + a_2 E(y_{t-12}y_{t-12}) + E(y_{t-12} \varepsilon_t) \]
\[ \gamma_{12} = a_1 \gamma_{11} + a_2 \gamma_0 \]

\[ E(y_{t-13}y_t) = a_1 E(y_{t-13}y_{t-1}) + a_2 E(y_{t-13}y_{t-12}) + E(y_{t-13} \varepsilon_t) \]
\[ \gamma_{13} = a_1 \gamma_{12} + a_2 \gamma_1 \]

We have 13 equations with 13 unknowns.
Computer or much patience can solve these.

\[
\begin{align*}
\gamma_0 & = a_1 \gamma_1 + a_2 \gamma_{12} + \sigma^2 \\
\gamma_1 & = a_1 \gamma_0 + a_2 \gamma_{11} \\
\gamma_2 & = a_1 \gamma_1 + a_2 \gamma_{10} \\
\gamma_3 & = a_1 \gamma_2 + a_2 \gamma_9 \\
\gamma_4 & = a_1 \gamma_3 + a_2 \gamma_8 \\
\gamma_5 & = a_1 \gamma_4 + a_2 \gamma_7 \\
\gamma_6 & = a_1 \gamma_5 + a_2 \gamma_6 \\
\gamma_7 & = a_1 \gamma_6 + a_2 \gamma_5 \\
\gamma_8 & = a_1 \gamma_7 + a_2 \gamma_4 \\
\gamma_9 & = a_1 \gamma_8 + a_2 \gamma_3 \\
\gamma_{10} & = a_1 \gamma_9 + a_2 \gamma_2 \\
\gamma_{11} & = a_1 \gamma_{10} + a_2 \gamma_1 \\
\gamma_{12} & = a_1 \gamma_{11} + a_2 \gamma_0
\end{align*}
\]
Estimation of ARMA models. XX

\[ \phi_{11} = \frac{\gamma_1}{\gamma_0} \]

\[ \phi_{22} = \frac{\gamma_0 \gamma_2 - \gamma_1^2}{\gamma_0^2 - \gamma_1^2} = 0 \iff \gamma_0 \gamma_2 = \gamma_1^2 \iff \frac{\sigma^4 a_{10} a_2}{K} = 0 \]

- Again this one cannot be 0 and be compatible with the model in question.....
Now consider data generated by

\[ y_t = 0.7y_{t-1} - 0.49y_{t-2} + \varepsilon_t \]

The estimation results are:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-Statistic</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1)</td>
<td>0.692389807</td>
<td>0.089515769</td>
<td>7.73484</td>
<td>0.00000000</td>
</tr>
<tr>
<td>(a_2)</td>
<td>-0.480874620</td>
<td>0.089576524</td>
<td>-5.36831</td>
<td>0.000000055</td>
</tr>
</tbody>
</table>

AIC = 219.87333  SBC = 225.04327

Model looks good.
AR(2) example

Here there are a few that are too large. ACF 16 and PACF 17.
This might have tempted us to estimate a different model.
ACF for residuals is large at 14 and 17
Also Ljung-Box(16) is significant.

<table>
<thead>
<tr>
<th>Lags:</th>
<th>1–12</th>
<th>13–24</th>
</tr>
</thead>
</table>
| ACF   | 0.466–0.161–0.322–0.108–0.052–0.165
|       | –0.0100.1280.1800.034–0.087–0.113
|       | –0.164–0.0580.1150.2540.046–0.175
|       | 0.1500.0100.032–0.089–0.0460.052
| PACF  | 0.466–0.4820.0230.045–0.253–0.121
|       | 0.1010.037–0.0760.023–0.020–0.139
|       | –0.1670.2070.0070.085–0.2160.013
|       | –0.022–0.0320.015–0.0610.038–0.184
AR(2) example

- Try the following model:

\[ y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \beta_{16} \varepsilon_{t-16} \]

- Estimation results are:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-Statistic</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1)</td>
<td>0.716681247</td>
<td>0.091069451</td>
<td>7.86961</td>
<td>0.00000000</td>
</tr>
<tr>
<td>(a_2)</td>
<td>-0.464999924</td>
<td>0.090958095</td>
<td>-5.11224</td>
<td>0.00000165</td>
</tr>
<tr>
<td>(\beta_{16})</td>
<td>0.305813568</td>
<td>0.109936945</td>
<td>2.78172</td>
<td>0.00652182</td>
</tr>
</tbody>
</table>

- This looks like a better model!
The simple-minded approach would be to run regression of $y$ on $p$ lags of $y$ and use OLS estimates.

Let's consider the properties of this estimator.

- Use $AR(1)$ as example
- Assume $y_0$ is available

Then:

$$ \hat{\rho} = \frac{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1}) (y_t - \bar{y}_t)}{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1}) (y_{t-1} - \bar{y}_{t-1})} $$

$$ = \rho + \frac{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1}) (\varepsilon_t - \bar{\varepsilon}_t)}{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1}) (y_{t-1} - \bar{y}_{t-1})} $$

We would like to answer questions about bias, consistency, variance etc.
We have to regard ‘regressor’ as stochastic.

It is not clear that

\[ E \left\{ \frac{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1}) (\varepsilon_t - \bar{\varepsilon}_t)}{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1}) (y_{t-1} - \bar{y}_{t-1})} \right\} = 0 \]

We can’t derive explicit expression for bias, but it can be shown that:

- The OLS estimate is biased and bias is negative. This bias often called Hurwicz bias – it can be sizeable in small samples
- Hurwicz bias goes to zero as \( T \to \infty \)
- The OLS estimate is consistent
- The OLS estimator is asymptotically normal with usual formulae for asymptotic variance.
Note that to estimate an $AR(p)$ model by OLS does not use information contained in first $p$ observations.

This causes a loss of efficiency from this.

There are a number of methods which use this information.
Now consider estimation of an $AR(1)$ process using maximum likelihood.

Assume the errors are normal white noise.

The model is:

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

We wish to estimate $c$, $\phi$ and $\sigma^2$.

The first observation $y_1$ is normally distributed with

$$E(y_1) = \mu = \frac{c}{1 - \phi}$$

$$V(y_1) = \frac{\sigma^2}{1 - \phi^2}$$
The density of this observation is

\[
 f_{Y_1}(y_1; c, \phi, \sigma^2) = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{1-\phi^2}}} \exp\left( - \frac{(y_1 - \frac{c}{1-\phi})^2}{2 \frac{\sigma^2}{1-\phi^2}} \right)
\]

Now consider \( y_2 \). We know that

\[
 y_2 = c + \phi y_1 + \varepsilon_2
\]

We can easily see that

\[
 f_{Y_2|Y_1}(y_2|y_1; c, \phi, \sigma^2) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left( - \frac{(y_2 - (c + \phi y_1))^2}{2 \sigma^2} \right)
\]
Recall that
\[ f_{Y_1, Y_2} = f_{Y_2|Y_1} \cdot f_{Y_1} \]

Similarly
\[ y_3 = c + \phi y_2 + \varepsilon_3 \]

and
\[
f_{Y_3|Y_1, Y_2} (y_3|y_1, y_2; c, \phi, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y_3 - (c + \phi y_2))^2}{2\sigma^2} \right)
\]

Continuing, we get
\[
f_{Y_1, Y_2, \ldots, Y_T} (y_1, y_2, \ldots, y_T; c, \phi, \sigma^2) = f_{Y_1} (y_1; c, \phi, \sigma^2) \prod_{t=2}^{T} f_{Y_t|Y_{t-1}} (y_t|y_{t-1}; c, \phi, \sigma^2)
\]
Then the log-likelihood is

\[
L(c, \phi, \sigma^2) = -\frac{1}{2} \log (2\pi) - \frac{1}{2} \log \left( \frac{\sigma^2}{1 - \phi^2} \right) - \frac{(y_1 - \frac{c}{1-\phi})^2}{2\frac{\sigma^2}{1-\phi^2}}
\]

\[
- \frac{T - 1}{2} \log (2\pi) - \frac{T - 1}{2} \log (\sigma^2)
\]

\[
- \sum_{t=2}^{T} \frac{(y_t - (c + \phi y_{t-1}))^2}{2\sigma^2}
\]

This one must be maximized numerically.

Another option is to do maximum likelihood conditional on the first observation. This turns out to give the same result as the OLS estimation with the same drawbacks.
We will consider a MA(1) process:

\[ y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1} \]

Assume the errors are normal white noise.

We wish to estimate \( \mu, \theta \) and \( \sigma^2 \).

If we knew the value of \( \varepsilon_{t-1} \), we could write

\[
f_{Y_t|\varepsilon_{t-1}} (y_t|\varepsilon_{t-1}; \mu, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y_t - \mu - \theta \varepsilon_{t-1})^2}{2\sigma^2} \right)\]

Now, if we knew with certainty that \( \varepsilon_0 = 0 \),

\[ y_1 = \mu + \varepsilon_1 \sim N(\mu, \sigma^2) \]
Note that this trick is similar to assuming the first observation was known for the AR(1) process.

Also, when $y_1$ is known, so is $\epsilon_1 = y_1 - \mu$. Therefore

$$f_{y_2 | y_1, \epsilon_0 = 0} (y_2 | y_1, \epsilon_0 = 0; \mu, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y_2 - \mu - \theta \epsilon_1)^2}{2\sigma^2} \right)$$

The, since

$$y_2 = \mu + \epsilon_2 + \theta \epsilon_1$$

$\epsilon_2$ is also known when $\epsilon_1$ is.
Continue like this to find the density functions.

\[ f_{Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_1, \varepsilon_0=0} \left( y_t | y_{t-1}, y_{t-2}, \ldots, y_1, \varepsilon_0 = 0; \mu, \theta, \sigma^2 \right) \]

\[ = f_{Y_t|\varepsilon_{t-1}} \left( y_t | \varepsilon_{t-1}; \mu, \theta, \sigma^2 \right) \]

\[ = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{\varepsilon_t^2}{2\sigma^2} \right) \]

From this we get the log-likelihood function

\[ \mathcal{L} (\mu, \theta, \sigma^2) = -\frac{T}{2} \log (2\pi) - \frac{T}{2} \log (\sigma^2) - \sum_{t=2}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \]

where \( \varepsilon_t \) is calculated recursively from the data.

It is not simple to get the MLE from this expression, so numerical methods must be used to maximize the likelihood function.
It turns out that if $|\theta| < 1$, the effect of assuming $\varepsilon_0 = 0$ dies out as $T$ becomes big.

There are also many methods for calculating the MLE for MA processes without this assumption, but these too are non-trivial numerical methods.