Time Series

Analysis

Jeffrey R. Russell
Winter 2017

• TA: An Qi: aqi@chicagobooth.edu
• Weekly homework due at the start of class.
• Midterm in week 6, final in finals week.
• Course web page:
  http://faculty.chicagobooth.edu/jeffrey.russell/teaching/timeseries/
• Software – Eviews or your choice
• Introduction to ARMA models
• Forecasting ARMA models.
• Maximum Likelihood Estimation
• GARCH models for time varying volatility
• GARCH models and forecasting applied to financial data.
• Models for multidimensional time series: VAR’s
• Unit Root inference
• Cointegration
• Introduction to Factor and DCC models.

A note on the target of the class

• Balance theory and application.
1. Introduction to Time Series and Autoregressive models.

1.1 Time-Series Data and Dependence
1.2 Checking for Dependence
1.3 The Autocorrelation Function
1.4 The AR(1) Model
1.5 More on the AR(1) Model
1.6 The AR(p) Model
1.7 The Partial Autocorrelation Function.

1.1 Time-Series Data and Dependence

Time-series data are simply a collection of observations gathered over time.

For now, consider only a single variable Y (X’s will come later). To emphasize that the data are time-series, we index observations by \( t \) rather than \( i \).
The data (number of observations = T):

\[ Y_1, Y_2, \ldots, Y_t, \ldots, Y_T \]

\( t \) indexes the order of the data, but the data may be Uniformly spaced (like weekly) or simply ordered (like arrival order).

**Unconditional moments**

- Consider a process \( y_t \) where \( y_t = \varepsilon_t \)

\[ \varepsilon_t \sim iid \ N(0, \sigma^2) \]

- We can have one gigantic realization of the series

\[ \ldots y_{t-1}, y_0, y_1, \ldots \]

- Clearly, the unconditional density for \( y_t \) is

\[ N(0, \sigma^2) \]
• The mean of $y_t$ is $E(y_t)$ if it exists.
• Alternatively, we could imagine taking a second independent realization of the time series generated by an independent set of Gaussian draws $...y^2_{-1}, y^2_0, y^2_1, ...$
• We could generate a lot of independent series and ask the question what is the expectation the $t^{th}$ observation, $y_t$?
• The unconditional mean of \( y_t \) (provided it exists) and can be thought of as the limit of an ensemble average:

\[
E(y_t) = \lim_{l \to \infty} \frac{1}{l} \sum_{i=1}^{l} y_{t}^{(i)}
\]

• Consider the trending series \( y_t = \beta t + \epsilon_t \)

• Here, \( E(y_t) = \beta t \) and the mean is different for each \( t \).

**Autocovariances**

• Similarly, given the joint distribution of \( y_{t-j}, y_{t-j+1}, \ldots, y_t \), we can compute

\[
\gamma_j = \text{Cov}(y_t, y_{t-j}) = E(y_t - \mu_t)(y_{t-j} - \mu_{t-j})
\]

• Again, we can also think of this as an ensemble average

\[
\gamma_j = \lim_{l \to \infty} \frac{1}{l} \sum_{i=1}^{l} (y_{t}^{(i)} - \mu_t)(y_{t-j}^{(i)} - \mu_{t-j})
\]
Stationary Time Series

• A time series is called (Strong) Stationary if
  \[ f(y_t, y_{t-1}, \ldots, y_{t-j}) = f(y_{s+t}, y_{s+t-1}, \ldots, y_{s+t-j}) \quad \text{for all } t \text{ and } s. \]

• A time series is called (Weakly) Stationary or Covariance Stationary if
  \[ E(y_t) = \mu \]
  and
  \[ E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j \quad \text{(only depends on } j, \text{ not } t) \]

Ergodic series

• Of course we (usually) only get to observe one realization of the time series, not an ensemble.

• If we can consistently estimate the mean of \( y_t \) using a single realization of the time series, that is, if \( \bar{y} = \frac{1}{T} \sum y \) converges to the ensemble average, then the series ergodic for the mean.

• If we can consistently estimate the covariance \( \gamma_j \) (for all \( j \)) then the series is ergodic for the second moments.
• We will see that stationary Gaussian models are ergotic for first and second moments if
  \[ \sum_{j=1}^{\infty} |y_j| < \infty \]

• Other conditions for ARMA models will be developed later.

• Let’s take a look at some data and sample statistics.
• The autocorrelation structure is a fundamental way to look at time series data.
• If the series is ergotic for second moments we can construct estimates of the autocovariances from our observed time series.
Example: Tbill rates

\( Y_t = \) annualized return on tbill

A time-series plot (\( Y_t \) versus \( t \)):

These data are not i.i.d.

The value of \( y_t \) is generally not far from the value of \( y_{t-1} \).

1.2 Checking for Dependence

It’s not always easy to just look at a time-series plot and say whether or not the series is independent.

Saying that the \( Y_t \) series is independent means that knowing previous values doesn’t help you to predict the next value.

So, knowing \( Y_{t-1} \) doesn’t help to predict \( Y_t \).

More generally, knowing \( Y_{t-1}, Y_{t-2}, \ldots, Y_{t-L} \) doesn’t help to predict \( Y_t \).
Corr(tbill, tbill(t-1))=0.913

Corr(tbill, tbill(t-2))=0.796
1.3 Sample Autocorrelations and the Autocorrelation Function (‘‘acf’’)

To summarize all of the plots of Y versus lagged Y’s, we compute the correlations between $Y_t$ and $Y_{t-L}$ for $L = 1, 2, 3, \ldots$

The sample autocovariances are given by:

$$\gamma_j = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})(y_{t-j} - \bar{y})$$

The sample autocorrelations are given by:

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}$$

These correlations between Y and lagged values of Y are called **autocorrelations**.

The **autocorrelation function** (or “acf”) is simply all of the autocorrelation values (for all possible $L$).
There seems to be dependence of \( y \) on past values. Also, the acf gets smaller as the lag gets larger --- \( y_t \) is more strongly related to \( y_{t-1} \) than \( y_{t-2} \), more strongly related to \( y_{t-2} \) than \( y_{t-3} \), and so on.

An autocorrelation is significantly different from zero at 5% level if*:
\[
|\text{autocorrelation}| > \frac{2}{\sqrt{T}}
\]

\( T=81 \) in tbill data--> cutoff around 0.22. Since many of the autocorrelations are above 0.22, there is strong evidence of dependence.
Testing that all of the first k autocorrelations are all zero

- We can test individual autocorrelations for significance, but which one should we look at? The first, the second, the tenth? We don’t know…
- Why not jointly test that the first k autocorrelations are all zero.

- We would like to test the null that
  \[ \rho_1 = \rho_2 = \ldots = \rho_k = 0 \]
  where \( \rho_j \) is the \( j \)th autocorrelation.
- The test statistic for this null is called the Q-stat. In the output from Eviews, each row has a Q-stat that corresponds to the Q-stat for each possible lag choice.
- The corresponding p-value is to the right of the test statistic.
• A simple form originally proposed by Box and Pierce is 
  \[ Q = T \sum_{j=1}^{k} \hat{\rho}_j^2 \]
  where \( \hat{\rho}_j \) is the \( j \)th sample autocorrelation.

• Under the null, if the data are iid, the test Q-statistic asymptotically has a Chi-Squared distribution with \( k \) degrees of freedom.

• The stat used in Eviews is an adjusted version proposed by Ljung and Box (1978) that has better finite sample properties:
  \[ Q = T(T+2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{T-j} \]
1.4 The AR(1) Model

If there is dependence in returns, we would like a model that allows us to predict future outcomes from the past outcomes.

Since there is information about $Y_t$ contained in the lagged values ($Y_{t-1}, Y_{t-2}, \ldots$), the obvious thing to try is a regression of $Y_t$ on its lags.

First, we need a model to describe the pattern in the data. A simple model, called AR(1) or the “autoregressive model of order 1,” is considered in this section.

The Gaussian AR(1) model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$$

where

$$\varepsilon_t \sim N(0, \sigma^2) \text{ i.i.d.}$$

$\varepsilon_t$ independent of $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$
If the AR model is covariance stationary, then the unconditional mean can be inferred.

\[ y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t \]
\[ E(y_t) = E(\beta_0 + \beta_1 y_{t-1} + \epsilon_t) = \beta_0 + \beta_1 E(y_{t-1}) + E(\epsilon_t) \]
\[ \mu = \beta_0 + \beta_1 \mu \]

\[ \mu = \beta_0 + \beta_1 \mu \Rightarrow \mu = \frac{\beta_0}{1 - \beta_1} \]

The variance and autocorrelations can be computed as well.

\[ \gamma_0 = Var(y_t) = \beta_1 Var(y_{t-1}) + \sigma^2 \]
\[ Var(y_t) = \frac{\sigma^2}{1 - \beta^2_1} \]
\[ \gamma_1 = E[(y_t - \mu)(y_{t-1} - \mu)] = E[(\beta_1(y_{t-1} - \mu) + \epsilon_t)((y_{t-1} - \mu))] = \beta_1 E(y_{t-1} - \mu)^2 = \beta_1 \gamma_0 \]
\[ \rho_1 = \beta_1 \]
\[
\gamma_2 = E[(y_i - \mu)(y_{i-2} - \mu)] = E[(\beta_i (y_i - \mu) + \varepsilon_i)((y_{i-2} - \mu))] = \beta_i \gamma_1 = \beta_i^2 \gamma_0
\]
\[
\rho_2 = \beta_i^2
\]

- You can see that this recursive pattern continues as we get
\[
\gamma_j = \beta_i \gamma_{j-1} = \beta_i^j \gamma_0
\]
and \( \rho_j = \beta_i^j \)

We will go over maximum likelihood estimation later. For now, it is easy to see that we can estimate the model by least squares:

\[
\arg\min_{b_0, b_1} \sum_{t=2}^{T} (Y_t - b_0 - b_1 Y_{t-1})^2
\]

start at \( t=2 \) since \( Y_1 \) doesn't have a lagged value

Given \( b_0 \) and \( b_1 \), we have
\[
Y_t = b_0 + b_1 Y_{t-1} + e_t = \hat{Y}_t + e_t
\]

fitted value
residual
Let’s take a look at estimates for the tbills data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.043163</td>
<td>0.018371</td>
<td>2.349552</td>
<td>0.0213</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.913360</td>
<td>0.046127</td>
<td>19.80093</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.834070  Mean dependent var 0.041540
Adjusted R-squared 0.831942  S.D. dependent var 0.034881
S.E. of regression 0.014217  Akaike info criterion -5.644041
Sum squared resid 0.015766  Schwarz criterion -5.584491
Log likelihood 227.7616  F-statistic 392.0770
Durbin-Watson stat 1.559497  Prob(F-statistic) 0.000000

Inverted AR Roots .91

Example SP500 daily returns

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-3.38E-05</td>
<td>0.000230</td>
<td>-0.146087</td>
<td>0.8839</td>
</tr>
<tr>
<td>AR(1)</td>
<td>-0.086993</td>
<td>0.016094</td>
<td>-4.602213</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.007551  Mean dependent var -3.35E-05
Adjusted R-squared 0.007224  S.D. dependent var 0.013837
S.E. of regression 0.013767  Akaike info criterion -5.729558
Sum squared resid 0.570116  Schwarz criterion -5.725590
Log likelihood 8690.874  Hannan-Quinn criter. -5.728131
F-statistic 23.90125  Durbin-Watson stat 2.006714
Prob(F-statistic) 0.000002

Inverted AR Roots -0.99
If the model correctly captures the dependence in the data, then the residuals should look i.i.d. (There should be no dependence in the residuals. If there is dependence in the residuals, it means that we’ve failed to capture information from the past.)

As in regression, use the estimated residuals ($e_t$) as proxies for the true residuals ($\epsilon_t$).

Plot of residuals

No obvious pattern
Here is the ACF of the residuals

<table>
<thead>
<tr>
<th>Date: 12/18/13  Time: 19:16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample: 1/01/2004 12/30/2005</td>
</tr>
<tr>
<td>Included observations: 564</td>
</tr>
<tr>
<td>Q-statistic probabilities adjusted for 1 ARMA term(s)</td>
</tr>
<tr>
<td>Autocorrelation</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
</tbody>
</table>

The AR(1) model seems to fit the S&P500 data pretty well. Once yesterday’s return is used to explain today’s return, the part left over (the residual) has no significant dependence structure.

If the model were wrong, there would be evidence of dependence in the residuals.

To see this phenomenon, consider what happens when we try to fit an AR(1) model to the tbills data.
Dependent Variable: TBILL
Method: Least Squares
Date: 12/14/07 Time: 13:35
Sample (adjusted): 1927 2006
Included observations: 80 after adjusting endpoints
Convergence achieved after 3 iterations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.043163</td>
<td>0.018371</td>
<td>2.349552</td>
<td>0.0213</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.913380</td>
<td>0.046127</td>
<td>19.80093</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.834070 Mean dependent var 0.041540
Adjusted R-squared 0.831942 S.D. dependent var 0.034681
S.E. of regression 0.014217 Akaike info criterion -5.644041
Sum squared resid 0.015766 Schwarz criterion -5.564491
Log likelihood 227.7616 F-statistic 392.0770
Durbin-Watson stat 1.559497 Prob(F-statistic) 0.000000

Inverted AR Roots .91

Correlogram of Residuals

<table>
<thead>
<tr>
<th>Date: 12/14/07 Time: 14:04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample: 1927 2006</td>
</tr>
<tr>
<td>Included observations: 80</td>
</tr>
<tr>
<td>Q-statistic probabilities adjusted for 1 ARMA term(s)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Autocorrelation</th>
<th>Partial Correlation</th>
<th>AC</th>
<th>PAC</th>
<th>Q-Stat</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.211</td>
<td>0.211</td>
<td>3.6826</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.213</td>
<td>-0.269</td>
<td>7.4899</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.086</td>
<td>0.031</td>
<td>8.1152</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.123</td>
<td>-0.186</td>
<td>9.4220</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.050</td>
<td>0.133</td>
<td>6.6426</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.048</td>
<td>-0.088</td>
<td>9.8447</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.002</td>
<td>0.059</td>
<td>9.8450</td>
<td>0.131</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.086</td>
<td>0.030</td>
<td>10.240</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.009</td>
<td>0.012</td>
<td>10.247</td>
<td>0.248</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.139</td>
<td>0.193</td>
<td>12.068</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.203</td>
<td>0.135</td>
<td>15.993</td>
<td>0.100</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-0.194</td>
<td>-0.222</td>
<td>19.619</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-0.162</td>
<td>0.047</td>
<td>22.185</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.030</td>
<td>-0.010</td>
<td>22.277</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.071</td>
<td>0.081</td>
<td>22.783</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.166</td>
<td>0.061</td>
<td>25.615</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.042</td>
<td>0.015</td>
<td>25.902</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-0.118</td>
<td>-0.102</td>
<td>27.277</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>-0.007</td>
<td>0.071</td>
<td>27.282</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-0.022</td>
<td>-0.071</td>
<td>27.336</td>
<td>0.097</td>
<td></td>
</tr>
</tbody>
</table>
• Since the errors are possibly autocorrelated the AR(1) model may not fit the tbill data well.

1.6 The general AR(p) model

If the AR(1) model doesn’t capture the dependence in the data, you can try a multiple regression with higher lags of Y thrown in:

\[ Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t \]

(This model is an “autoregressive model of order p.”)
Unconditional mean of AR(p)

- Again, if the process is covariance stationary we can take expectations of both sides of the equation and get:

\[ \mu = \frac{\beta_0}{1 - \beta_1 - \ldots - \beta_p} \]

- Notice also, we can write

\[ \beta_0 = \mu \left(1 - \beta_1 - \ldots - \beta_p \right) \]

What about the variance and autocovariances?

- First, notice that if we substitute * from the previous slide into the AR(p) representation we get:

\[ Y_t - \mu = \beta_1 (Y_{t-1} - \mu) + \ldots + \beta_p (Y_{t-p} - \mu) + \varepsilon_t \]

- So without loss of generality, we can think of a (covariance stationary) AR(p) as mean 0. We just subtract the mean off of every observation.
Consider the de-meaned AR(p) model

\[ Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t \]

If we multiply both sides by \( Y_t \) and take expectations of both sides we get:

\[ \gamma_0 = \beta_1 \gamma_1 + \beta_2 \gamma_2 + \cdots + \beta_p \gamma_p + \sigma^2 \]

If we multiply both sides by \( Y_{t-1} \) and take expectations of both sides we get:

\[ \gamma_1 = \beta_1 \gamma_0 + \beta_2 \gamma_1 + \cdots + \beta_p \gamma_{p-1} \]

Continuing this through \( Y_{t-p} \) we get \( p+1 \) equations with \( p+1 \) unknowns

\[ \gamma_0 = \beta_1 \gamma_1 + \beta_2 \gamma_2 + \cdots + \beta_p \gamma_p + \sigma^2 \]
\[ \gamma_1 = \beta_1 \gamma_0 + \beta_2 \gamma_1 + \cdots + \beta_p \gamma_{p-1} \]
\[ \vdots \]
\[ \gamma_p = \beta_1 \gamma_{p-1} + \beta_2 \gamma_{p-1} + \cdots + \beta_p \gamma_0 \]

We can solve these equations to find \( \gamma_0, \ldots, \gamma_p \)
• We can get \( \gamma_{p+1} \) solved for as a function of \( \gamma_1, \ldots, \gamma_p \)

\[
\gamma_{p+1} = \beta_1 \gamma_p + \beta_2 \gamma_{p-1} + \cdots + \beta_p \gamma_1
\]

and more generally, for \( j > p \)

\[
\gamma_j = \beta_1 \gamma_{j-1} + \beta_2 \gamma_{j-2} + \cdots + \beta_p \gamma_{j-p}
\]

1.7 The Partial Autocorrelation Function.

• How do we figure out how many lags to include in the model?
• The simple strategy outlined was to start with the AR(1) model and then continue to increase the order of the model until the added regression coefficient is not significant (and check the residual series).
• The partial autocorrelation function (PACF) is one more useful tool for model selection.
PACF

• The first partial autocorrelation is simply the estimated regression coefficient $\beta_1$ from the regression

$$Y_t = \beta_0 + \beta_1 Y_{t-1}$$

• The second partial autocorrelation is the coefficient $\beta_2$ from the regression

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$$

• In general, the $j$th partial autocorrelation is the $j$th regression coefficient in a regression of $Y_t$ on $Y_{t-1}$, $Y_{t-2}$, $\ldots$, $Y_{t-j}$

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_j Y_{t-j}$$

Run this regression

<table>
<thead>
<tr>
<th>PACF</th>
<th>1st PAC</th>
<th>2nd PAC</th>
<th>3rd PAC</th>
<th>$j$th PAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_t = \beta_0 + \beta_1 Y_{t-1}$</td>
<td>$b_1$</td>
<td>$b_2$</td>
<td>$b_3$</td>
<td>$b_j$</td>
</tr>
<tr>
<td>$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_t = \beta_0 + \sum_{i=1}^{j} \beta_i Y_{t-i}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$j$th partial autocorrelation (the last coefficient of regression of $Y_t$ on $Y_{t-1}$, $Y_{t-2}$, $\ldots$, $Y_{t-j}$)
How to use the PACF

• Suppose that the true time series process is an AR(1) model

\[ Y_t = \beta_0 + \beta Y_{t-1} + \epsilon_t \]

• What should the PACF look like?

• The first PACF, \( b_1 \), should be close to \( \beta_1 \), the true slope coefficient. The second PACF, \( b_2 \), should be close to zero.

• Why? Because in the AR(1) model, \( Y_t \) only depends on \( Y_{t-1} \), not \( Y_{t-1} \) and \( Y_{t-2} \).

• Of course, even if the true model is an AR(1) the second PAC will not be exactly zero, but it should be close.

• We would reject the null that the \( j \)th PAC is zero if

\[ |b_j| > \frac{2}{\sqrt{T}} \]

• It is therefore common to plot the first \( L \) PAC’s and compare their magnitude to \( \frac{2}{\sqrt{T}} \).

• So for the AR(1) model, the first PAC should be significant and the remaining PAC’s should all be insignificant.
ACF and PACF for AR models

- An AR model generally has a slowly decaying autocorrelation function.

- An AR(p) model generally the first p partial autocorrelations non-zero and nearly zero thereafter.

2 Some matrix algebra facts and tricks for analyzing higher order models and more analysis of dynamics
2.1 Some matrix algebra facts

• As we move back to time series analysis, there are some useful matrix algebra tricks of the trade.

• The (nxn) matrix A is non-singular if Ax=0 only when x equals the vector of all zeros.

Eigenvalues and eigenvectors

• Let A be a nxn matrix, x be a nx1 non-zero vector and \( \lambda \) a scalar. Then x is an eigenvector and \( \lambda \) an eigenvalue if

\[
Ax = \lambda x
\]

• The eigenvalues are obtained by solving for the values of \( \lambda \) such that \( |A - \lambda I_n| = 0 \) where \(|.|\) denotes the determinant and \( I_n \) is the nxn identity matrix.
• Fact: the (nxn) matrix A will have n unique eigenvalues if A is non-singular (i.e. Ax=0 is only satisfied for the vector x composed of all zeros).

Neat trick of the trade

• Let A be an nxn and nonsingular matrix. Then we can always decompose A into:

$$A = T\Lambda T^{-1}$$

where $\Lambda$ is the diagonal matrix

$$\begin{bmatrix}
\lambda_1 & 0 \\
& \ddots & \ddots \\
0 & \cdots & \lambda_n
\end{bmatrix}$$

and $T$ is the matrix of associated eigenvectors: $T=[X_1 \cdots X_n]$
2.2 Impulse Response Functions

- Consider again the AR(1) model.
  \[ y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \]

- We might want to know what we should expect the future value of \( y_{t+k} \) to look like if \( y_t \) were one unit larger (holding all \( y_{t-j}, j>0 \) fixed). This is equivalent to asking how we should expect \( y_{t+k} \) to change given a one unit change in \( \varepsilon_t \).

Consider the recursive substitution:

\[
Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \\
Y_{t-1} = \beta_0 + \beta_1 Y_{t-2} + \varepsilon_{t-1} \\
\text{so } y_t = \beta_0 + \beta_1 \left( \beta_0 + \beta_1 Y_{t-2} + \varepsilon_{t-1} \right) + \varepsilon_t \\
\quad = \beta_1^2 Y_{t-2} + \left(1+\beta_1\right) \beta_0 + \beta_1 \varepsilon_{t-1} + \varepsilon_t \\
Y_{t-2} = \beta_0 + \beta_1 Y_{t-3} + \varepsilon_{t-2} \\
\text{so } y_t = \beta_1^2 \left( \beta_0 + \beta_1 Y_{t-3} + \varepsilon_{t-2} \right) + \left(1+\beta_1\right) \beta_0 + \beta_1 \varepsilon_{t-1} + \varepsilon_t \\
\quad = \beta_1^3 Y_{t-3} + \left(1+\beta_1 + \beta_1^2 \right) \beta_0 + \beta_1^2 \varepsilon_{t-2} + \beta_1 \varepsilon_{t-1} + \varepsilon_3
\]
\begin{itemize}
  \item Hence we end up with:
  \[ Y_t = \beta_1' Y_0 + (1 + \beta_1 + \cdots + \beta_1^{t-1}) \beta_0 + \beta_1^{t-1} \varepsilon_t + \cdots + \beta_t \varepsilon_{t-1} + \varepsilon_t \]
  \[ \text{and if } |\beta_1| < 1 \text{ we get:} \]
  \[ Y_t = \beta_1' Y_0 + \left(1 - \beta_1^t\right) \beta_0 / (1 - \beta_1) + \sum_{j=0}^{t-1} \beta_j \varepsilon_{t-j} \]
  \item The impact of a one unit one unit increase in \( \varepsilon_{t-k} \) on \( y_t \) is given by:
  \[ \frac{dy_t}{d\varepsilon_{t-k} \mid y_{t-k-1}, y_{t-k-2}, \ldots} = \beta_1^k \]
  \item When we plot the change in \( y_t \) given a one unit shock to \( \varepsilon_{t-k} \) we call this an impulse response function (IRF).
\end{itemize}
Example: IRF for dollars traded

• Let $y_t$ be dollar volume traded on day $t$ for a given asset.

• Suppose we estimate the following AR(1) model:
  \[ y_t = 10,200 + .8y_{t-1} \]

IRF in pictures

\[ \frac{dy_t}{d\epsilon_{t-k}} = .8^k \]

Interpretation? If there is an unexpected increase of one additional dollar traded, how much does dollar volume change $k$-periods ahead?
2.3 Lag operators and companion form

- Clearly the IRF’s are important in understanding the dynamics of a time series process.

- How do we find the IRF for an AR(p) model in general?

- There are two equivalent approaches.

- The first approach converts the AR(p) model back into an AR(1) model.
How do we convert an AR(p) back into an “AR(1)” type model?

• Consider the AR(p) model:

\[ y_t = \beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + \epsilon_t \]

Companion Form representation of an AR(p) model

• Next we define the new vectors and matrix:

\[ \xi_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad v_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad F = \begin{bmatrix} \beta_1 & \beta_2 & \beta_{p-1} & \beta_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix} \]

• Then we can write the AR(p) model as the following first order model:

\[ \xi_t = B + F \xi_{t-1} + v_t \]
\[ \xi_t = B + F \xi_{t-1} + v_t \]

- We can then recursively substitute in just as we did for the AR(1) model to get:

\[ \xi_t = F^t \xi_0 + \left( 1 + F + F^2 + \cdots + F^{t-1} \right) B + F^{t-1} v_1 + F^{t-2} v_2 + \cdots v_t \]

\[ = F^t \xi_0 + \left( 1 + F + F^2 + \cdots + F^{t-1} \right) B + \sum_{j=0}^{t-1} F^j v_{t-j} \]

- We can now see that

\[ \frac{dy_t}{dy_{t-k}} \mid y_{t-k-1}, y_{t-k-2}, \ldots = \frac{dy_t}{d\xi_{t-k}} \mid y_{t-k-1}, y_{t-k-2}, \ldots \]

given by the (1,1) element of \( F^k \)

- Recall our decomposition of the (nxn) matrix \( A \):

\[ A = T \Lambda T^{-1} \]

- Notice that \( A^2 = T \Lambda T^{-1} T \Lambda T^{-1} = T \Lambda^2 T^{-1} \)

- More generally: \( A^j = T \Lambda^j T^{-1} \)

- Since \( \Lambda \) is diagonal \( A^j = T \begin{bmatrix} \lambda_1^j & 0 \\ \vdots & \ddots \\ 0 & \lambda_n^j \end{bmatrix} T^{-1} \)
• Hence, $F^j$ can be written as:

$$F^j = T \begin{bmatrix} \lambda_1^j & 0 \\ \vdots & \ddots \\ 0 & \lambda_p^j \end{bmatrix} T^{-1}$$

• The impact of a past shock on future values of $y$ is determined by powers of the matrix $F$.
• The rate of decay of a shock is determined by the eigenvalues.
• If the eigenvalues are smaller than 1 then the impact of past shocks dies out. The process mean reverts and is stationary.

If all the eigenvalues lie inside the unit circle the process is covariance stationary.
• Fact (*), the eigenvalues $\lambda$ of the matrix $F$ satisfy:
\[ \lambda^p - \beta_1\lambda^{p-1} - \beta_2\lambda^{p-2} - \cdots - \beta_{p-1}\lambda - \beta_p = 0 \]

• Hence we can solve for the roots of this polynomial to obtain the eigenvalues. If all the roots of this polynomial lie inside the unit circle then the process is covariance stationary.

* See appendix of Chapter 1 of Hamilton for proof.

---

**Time series trick**

• The lag operator is denoted by $L$ and is extremely convenient for manipulating time series. We will use it a lot.
\[ L(y_t) = y_{t-1} \]

• The lag operator shares the same properties of multiplication so we write it like multiplication.
\[ L(y_t) = Ly_t = y_{t-1} \]
The lag operator shares the properties of multiplication

- It commutes: \( L(\beta x_t) = \beta L(x_t) \)

- It’s distributive: \( L(x_t + y_t) = L(x_t) + L(y_t) \)

- Applying it twice:
  \[
  L(L(x_t)) = L^2(x_t) = x_{t-2}
  \]

- Using this notation, an AR(1) model can be written as:
  \[
  (1 - \beta_1 L) y_t = \beta_0 + \epsilon_t
  \]

- For a given \( t \), pre-multiply both sides of the AR(1) model above by: \( (1 + \beta_1 L + \beta_1^2 L^2 \ldots \beta_1^{t-1} L^{t-1}) \)
  \[
  (1 + \beta_1 L + \beta_1^2 L^2 + \ldots + \beta_1^{t-1} L^{t-1})(1 - \beta_1 L)y_t
  \]
  \[
  = (1 + \beta_1 L + \beta_1^2 L^2 + \ldots + \beta_1^{t-1} L^{t-1})(\beta_0 + \epsilon_t)
  \]
\[
(1 + \beta_1 L + \beta_1^2 L^2 + \ldots + \beta_1^{t-1} L^{t-1})(1 - \beta_1 L) y_t \\
= (1 + \beta_1 L + \beta_1^2 L^2 + \ldots + \beta_1^{t-1} L^{t-1})(\beta_0 + \epsilon_t) \\
(1 - \beta_1^t L) y_t = (1 + \beta_1 L + \beta_1^2 L^2 + \ldots + \beta_1^{t-1} L^{t-1})(\beta_0 + \epsilon_t)
\]

or

\[
Y_t = \beta_1^t Y_0 + \frac{(1 - \beta_1^t)}{(1 - \beta_1)} \beta_0 + \sum_{j=0}^{t-1} \beta_1^j \epsilon_{t-j}
\]

Look familiar?

• We say that \(1 - \beta_1 L\) has an inverse if \(|\beta_1| < 1\).
• The inverse is \((1 - \beta_1 L)^{-1} = (1 + \beta_1 L + \beta_1^2 L^2 + \ldots)\)
  \((1 + \beta_1 L + \beta_1^2 L^2 + \ldots)(1 - \beta_1 L) \rightarrow 1\)
• Invertibility of the AR model is the same as stationarity.
  \[Y_t = (1 - \beta_1 L)^{-1} (\beta_0 + \epsilon_t) = \mu + \sum_{j=0}^{\infty} \beta_1^j \epsilon_{t-j}\]
• If the epsilons are iid normal then the y’s are normal and strong form stationary.
• Using this notation we can write the $p$th order AR model as:

$$y_t = \beta_0 + \beta_1 L y_t + \beta_2 L^2 y_t + \ldots + \beta_p L^p y_t + \epsilon_t$$

or

$$\left(1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_p L^p\right) y_t = \beta_0 + \epsilon_t$$

• The lag operator shares the rules of multiplication.
• Thinking of $L$ as a variable (but its not!) we can “factor” the polynomial as in:

$$\left(1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_p L^p\right) = (1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L)$$

• The process has an inverse (and is stationary) if all the $\lambda_i$ have modulus less than 1.

$$(1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1}\ldots(1 - \lambda_p L)^{-1}\left(1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_p L^p\right)y_t = y_t$$

$$= (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1}\ldots(1 - \lambda_p L)^{-1}(\beta_0 + \epsilon_t)$$
• The lamdas are the inverse of the roots of the lag polynomial \((1-\beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p)\)

• Alternatively, with the change of variable \(\lambda = L^{-1}\) and multiplying both sides by \(\lambda^p\)

\[
\lambda^p - \beta_1 \lambda^{p-1} - \beta_2 \lambda^{p-2} - \cdots - \beta_{p-1} \lambda + \beta_p = 0
\]

• The roots of this polynomial are the eigenvalues. This is the same polynomial we found for the eigenvalues of the companion form matrix \(F\).

Examples