LECTURE 13: TIME SERIES I

AUTOCORRELATION:

Consider $y = X\beta + u$ where $y$ is $T \times 1$, $X$ is $T \times K$, $\beta$ is $K \times 1$ and $u$ is $T \times 1$.

We are using $T$ and not $N$ for sample size to emphasize that this is a time series.

The natural order of observations in a time series suggest possible approaches to parametrizing the covariance matrix parsimoniously.

First order autoregression: $AR(1)$

This is the case where $u_t = \rho u_{t-1} + \varepsilon_t$ where $\varepsilon_t$ are independent and identically distributed with

$E\varepsilon_t = 0$ and $V(\varepsilon_t) = \sigma^2$. 
First order moving average: $MA(1)$

This is the case where $u_t = \varepsilon_t - \theta \varepsilon_{t-1}$.

Random walk: $(AR(1)$ with $p = 1)$

This is the case where $u_t - u_{t-1} = \varepsilon_t$.

Integrated moving average: $IMA(1)$

This is the case where $u_t - u_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1}$.

Autoregressive moving average $(1,1)$: $ARMA(1,1)$

$u_t - \rho u_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1}$
Autoregressive of order $p$: $AR(p)$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t.$$ 

Moving average of order $p$: $MA(p)$

$$u_t = \varepsilon_t - \sum_{i=1}^{p} \theta_i \varepsilon_{t-i}$$

**Proposition:** A first order autoregressive ($AR(1)$) process is an infinite order moving average ($MA(\infty)$) process.

**Proof:**

$$u_t = \rho (\rho u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = (\varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \ldots).$$

Thus

$$u_t = \sum_{r=0}^{\infty} \rho^r \varepsilon_{t-r}$$
$AR(1)$ arises frequently in economic time series.

Let $u_t = \rho u_{t-1} + \varepsilon_t$ which is an $AR(1)$ process.

Note that $E u_t = 0$ and $V(u_t) = \sigma^2(1 + \rho^2 + \rho^4 + \ldots) = \sigma^2/(1 - \rho^2)$.

Also note that

$$cov(u_t u_{t-1}) = \rho \sigma^2 + \rho^3 \sigma^2 + \rho^5 \sigma^2 + \ldots$$

$$= \rho \sigma^2/(1 - \rho^2) = \rho V(u_t),$$

and similarly

$$cov(u_t u_{t-s}) = \rho^s V(u_t) = \rho^s \sigma^2/(1 - \rho^2).$$

Thus

$$Euu' = \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ . & . & . & \ldots & . \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{bmatrix}$$
This is a symmetric matrix.

This is a variance-covariance matrix characterized by two parameters which fits into the GLS framework.

Consider the LS estimator $\hat{\beta}$ under the assumption of an $AR(1)$ process for the $u_t$'s:

1. What are the properties of $\hat{\beta}$?

2. What is the associated variance estimate?

In the LS method, $V(\hat{\beta})$ is estimated by $s^2(X'X)^{-1}$. Is this correct in the AR case?
Under the assumption of an $AR(1)$ error process, $V(\hat{\beta})$ should be $(\sigma^2/(1 - \rho^2))(X'X)^{-1}X'VX(X'X)^{-1}$.

with $V$ representing the variance-covariance matrix above.

If $X$ variables are trending up and $\rho > 0$ (usually $\approx 0.8$ or 0.9), the $s^2$ will probably underestimate $\sigma^2/(1 - \rho^2)$ and $(X'X)^{-1}X'VX(X'X)^{-1}$.

*Point:* We can seriously understate standard errors if we ignore autocorrelation.

Prof. N. M. Kiefer, Econ 620, Cornell University, Lecture 13. Copyright (c) N. M. Kiefer.
"SPURIOUS REGRESSIONS IN ECONOMETRICS":

(Granger-Newbold)

(Journal of Econometrics, 1974)

Consider a simple regression model.

Let \( y_t = \alpha + \beta x_t + \varepsilon_t \).

Suppose the true process with \( \varepsilon \) and \( \varepsilon^* \) independent are

\[
y_t = \rho y_{t-1} + \varepsilon_t \quad \text{and} \quad x_t = \rho^* x_{t-1} + \varepsilon^*_t
\]

The data are really independent \( AR(1) \) processes.

Prof. N. M. Kiefer, Econ 620, Cornell University, Lecture 13. Copyright (c) N. M. Kiefer.
Suppose we regress $y$ on $x$. Then if $T = 20$ and $\rho = \rho^* = 0.9$, then $ER^2 = 0.47$ and $F \approx 18$.

This falsely indicated a significant contribution of $x$.

Sampling experiments for $y_t = \alpha + \beta x_t + \varepsilon_t$ with $T = 50$ and $y, x$ independent random walks were carried out, and t-statistics on $\beta$ in 100 trials were calculated.

If these statistics were actually distributed as $t$, we would expect $t$ to be less than 2, 95 times. We actually observe $t$ to be less than 2, 23 times, and $t$ greater that 2, 77 times. There is spurious significance. The situation only becomes worse with more regressors.

**Point:** High $R^2$ does not "balance out" the effects of autocorrelation. Good time-series fits are not to be believed without diagnostic tests.
TESTING FOR AUTOCORRELATION:

The important thing is to look at the residuals.

*Definition:* The Durbin-Watson statistic ("d" or DW") is

\[
d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T}e_t^2} = \frac{e'Ae}{e'e}
\]

where

\[
A = \begin{pmatrix}
1 & -1 & 0 & \cdots \\
-1 & 2 & -1 & \cdots \\
0 & -1 & 2 & \cdots \\
& & & \ddots
\end{pmatrix}
\]

Which is a \( T \times T \) symmetric matrix
In other words, $d$ is the sum of squared successive differences divided by sum of squares.

The Durbin-Watson statistic is probably the most commonly used test for autocorrelation, although the Durbin h-statistic is appropriate in wider circumstances and should usually be calculated as well.

*Distribution of $d$:

Note: We want to calculate the distribution under the hypothesis that $\rho = 0$, i.e. no autocorrelation. Then a surprisingly large value indicated autocorrelation.
Intuition:

\[ E(\epsilon_t - \epsilon_{t-1})^2 = \sigma^2 + \sigma^2 - 2\text{cov}(\epsilon_t, \epsilon_{t-1}) = 2\sigma^2 \]

Then, why is \( Ed \neq 2 \)?

1. There is one less term in the numerator

2. The use if \( e \) rather that \( \epsilon \) makes the distribution depends on \( x \).

Note: \( d \) is a ratio of quadric forms in normals.

Why isn't it distributed a \( F \)?
Durbin-Watson test:

Durbin and Watson give bounds $d_L$ and $d_U$ which are both less than 2.

If $d > d_L$, then reject the null hypothesis of no autocorrelation. This indicated positive autocorrelation.

If $d_L < d < d_U$, then the result is ambiguous.

If the statistic $d$ calculated from the sample is greater than 2, the indication is negative autocorrelation. Then use the bounds of $d_L$ and $d_U$, and check against $4 - d$.

If $4 - d < d_L$, then reject the null.

If $4 - d > d_U$, then do not reject.
Interpretation of the Durbin-Watson test:

1. This is a test for general autocorrelation, not just for $AR(1)$ processes.

2. This test cannot be used when regressors include lagged values of $y$, for example,

\[ y_t = \alpha + \beta_0 y_{t-1} + \beta_1 x_t + \varepsilon_t. \]

Other tests are available in this case.
Other tests:

1. Wallis test: This is used for quarterly data. The test statistic is

\[ d_4 = \frac{\sum_{t=5}^{T} (e_t - e_{t-4})^2}{\sum_{t=1}^{T} e_t^2}. \]

2. Durbin's h test: This is used when there are lagged \( y \)'s. We regress \( e_t \) on \( e_{t-1}, x_t \) and as many lagged \( y \)'s as are included in the regression. Then test (with "t") the coefficient of \( e_{t-1} \). A significant coefficient on \( e_{t-1} \) indicates presence of autocorrelation. Note that this test is quite easy to do and it "works" when the Durbin-Watson test doesn't. This is a good test to use.
ESTIMATION WITH AN AR(1) ERROR PROCESS:

Consider \( y = X\beta + u \) where \( u_t = \rho u_{t-1} + \varepsilon_t \) with \( E(u) = 0 \) and

\[
Euu' = \frac{\sigma^2}{1-\rho^2} \begin{bmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\
\rho & 1 & \rho & \ldots & \rho^{T-2} \\
\rho & \rho^2 & \rho^{T-1} & \ldots & \rho^{T-2} \\
\rho & \rho^2 & \rho^{T-1} & \ldots & \rho^{T-3} \\
\rho & \rho^2 & \rho^{T-1} & \ldots & 1
\end{bmatrix} = \frac{\sigma^2}{1-\rho} \Omega.
\]
Thus

\[ \Omega^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & \ldots & 0 \\ -\rho & 1 + \rho^2 & \ldots & -\rho \\ \vdots & \vdots & \ddots & \vdots \\ -\rho & \ldots & 1 + \rho^2 & -\rho \\ 0 & \ldots & -\rho & 1 \end{bmatrix} = P'P \]

which is a "band" matrix. So,

\[ P = \frac{1}{\sqrt{1-\rho^2}} \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \ldots & 0 \\ -\rho & 1 & \ldots & 0 \\ 0 & -\rho & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ldots & -\rho & 1 \end{bmatrix}. \]
Matrix $P$ will be used to transform the model.

The first transformed observation is

$$\sqrt{1 - \rho^2} y_1 = \sum_{h=1}^{K} \beta_h x_{h,1} \sqrt{1 - \rho^2} + u_1 \sqrt{1 - \rho^2},$$

and all others are

$$y_t - \rho y_{t-1} = \sum_{h=1}^{K} \beta_h (x_{h,t} - \rho x_{h,t-1}) + u_t - \rho u_{t-1}.$$

Note that $x_{h,t}$ denotes the $t^{th}$ observation on the $h^{th}$ explanatory variable.

The GLS transformation puts the model back in standard form as expected.
Notes:

1. Given $\rho$, the estimation is by the LS method. We write the sum of squares as $S(\rho)$. Then minimization with respect to $\rho$ is a simple numerical problem.

2. ML can also be reduced to a one-dimensional maximization problem which is straightforward.

3. Early two-step methods which often dropped the first observation are less satisfactory. Never use the Cochrane-Orcutt (CORC) procedure.

4. The extension to higher-order AR or MA processes is straightforward.