Reading Assignment

Serial Correlation and Heteroskedasticity

Chapters 12 and 11.

Kennedy: Chapter 8.

Serial Correlation or Autocorrelation

$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + e_t$

where $E(e_t\, e_s) \neq 0$ *for* $t \neq s$.

Common problem with samples over time.

Positive Serial Correlation

Negative Serial Correlation

Suppose

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + e_t$$

$$e_t = \rho\, e_{t-1} + u_t \qquad 0 \le |\rho| < 1$$

where $E(e_t) = 0$ $\qquad$ $E(u_t) = 0$

$E(e_t^2) = \sigma^2$ $\qquad$ $E(u_t^2) = \sigma_u^2$

$E(e_t\, e_{t-s}) \neq 0$ $\qquad$ $E(u_t\, u_{t-s}) = 0.$

This is a first-order autoregressive error.

We are assuming that we can define – and then estimate – a process that explains the serial correlation.

(Also need the errors to be stationary…)

$e_t = \rho\, e_{t-1} + u_t$ so

$$V(e_t) = \sigma^2 = V(\rho\, e_{t-1}) + V(u_t)$$

$$= \rho^2 V(e_{t-1}) + V(u_t)$$

$$= \rho^2 \sigma^2 + \sigma_u^2$$

$$\sigma^2 = \sigma_u^2 / (1 - \rho^2)$$

Constant and function of σ_u^2 and ρ.

Further, what does the magnitude of ρ imply about $V(e_t)$? (I.e., $\rho \geq 1$?)

$$Cov(e_t, e_{t-s}) = \rho^s \sigma^2 = \rho^s \sigma_u^2 / (1 - \rho^2)$$

Non-zero. This violates an OLS assumption.

Further, what does the magnitude of ρ imply about $Cov(e_t, e_{t-s})$?

Why does serial correlation occur?

Dependent variable is a dynamic process.

Inertia, lagged adjustments, lagged expectations.

Explicit dynamic models are improving...

Specification Bias: missing variable.

If missing variable follows dynamic process...

Specification Bias: Incorrect functional form.

Consequences of Serial Correlation

$\hat{\beta}$ ***are unbiased but not efficient. The variance-covariance matrix is incorrect.***

Variances of OLS estimators are biased and inconsistent.

Positive serial correlation: $\hat{\sigma}$ ***too small, t and F-tests too large.***

Negative serial correlation: $\hat{\sigma}$ ***too large, t and F-tests too small.***

<u>Detection of Serial Correlation</u>

1) ***Graph the residuals against time.***

2) ***Durbin-Watson d test***

H_0: $\rho = 0$

DW-d = $\Sigma (e_t - e_{t-1})^2 / \Sigma e_t^2$

DW-d = $2 (1 - \rho)$ or $\rho = 1 - (d/2)$

If	ρ	+1	0	-1
Then	DW-d	0	2	4

The exact distribution of DW-d under H_0 is unknown because DW-d depends on e_t which depends on the X's which are different for every model.

Upper and lower critical values.

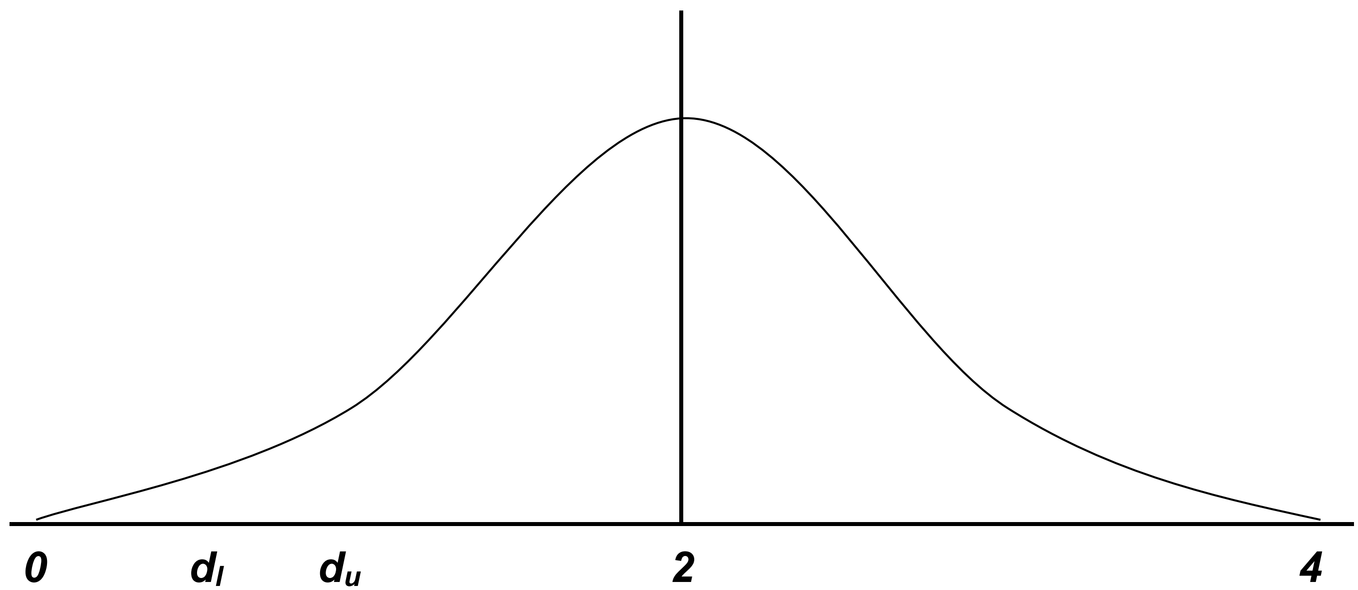

Table values of d_l and d_u depend on α, k, and T.

This is mainly important from the standpoint of econometric history…

Econometric research says use d_u – you have serial correlation unless you fail to reject the H_0 – you have it unless you clearly don't as it's too easy to correct for.

Further, some computer programs will calculate exact critical values.

Comments on DW-d statistic:

> ***Assumes first-order autoregressive error.***
> ***Invalid without intercept and with lagged dependent variable.***
> ***If DW-d < R^2 then regression results are spurious.***

Spurious?:

If you have $y_t = y_{t-1} + u_{1t}$ and $x_t = x_{t-1} + u_{2t}$ then run $y_t = \beta_0 + \beta_1 x_t + e_t$ you would think the slope would be close to zero. But that's not the case. The t-statistics have a 60-80% chance of being significant and the R^2 approaches 1 as the sample size increases. (Think consumption function and models from PS1.)

This is a digression, but this is major trouble. Spurious is bad.

However, if you run $\Delta y_t = \beta_0 + \beta_1 \Delta x_t + u_t$ then you get the anticipated result. The slope is close to zero, the t-statistic is small, and the R^2 is low. (Also think PS1 and remember that you have fundamentally changed the data and model.)

3) Breusch-Godfrey (LM) test

(addresses all DW-d limitations)

$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \ldots + \rho_p e_{t-p} + u_t$

$H_0: \rho_1 = \rho_2 = \ldots = \rho_p = 0$

Steps of test:

a) Run OLS regression and obtain $\hat{e}_t$

b) Regress $\hat{e}_t$ ***on X's in model and*** $\hat{e}_{t-1}, \ldots, \hat{e}_{t-p}$

c) Obtain R^2 ***of this "auxiliary" or "artificial" regression.***

If sample size is large

$(T - p)R^2 \sim \chi^2_p$

ex) $(32 - 5)\ 0.866 = 23.382 > 11.0705 = \chi^2_5\ (5\%)$

d) Econometric research shows an F-test on the auxiliary regression has better small sample properties.

$H_0: \rho_1 = \rho_2 = ... = \rho_p = 0$

Comments:

Examine t-statistics on auxiliary regression.

Use several lags lengths of autoregressive error.

Correcting for Serial Correlation

Procedure: transform the error term into a random variable that meets OLS assumption.

i.e., $E(e_t^* e_s^*) = 0$ for $t \neq s$.

Generalized Differencing: when ρ is known.

$y_t = \beta_0 + \beta_1 x_t + e_t$ ***and*** $e_t = \rho e_{t-1} + u_t$

$y_t = \beta_0 + \beta_1 x_t + \rho e_{t-1} + u_t$

$y_t = \beta_0 + \beta_1 x_t + \rho (y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + u_t$

$(y_t - \rho y_{t-1}) = (\beta_0 - \rho\beta_0) + \beta_1(x_t - \rho x_{t-1}) + u_t$

$y_t^* = \beta_0^* + \beta_1 x_t^* + e_t^*$

Run OLS on this model. $\hat{\beta}$ are BLUE.

So, if we knew ρ we could fix the problem.

Cochrane-Orcutt procedure

$y_t = \beta_0 + \beta_1 x_t + e_t$ and $e_t = \rho e_{t-1} + u_t$

a) Estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ via OLS using y_t and x_t.

b) Calculate $\hat{e}_t$ using $\hat{\beta}_0$ and $\hat{\beta}_1$.

c) Estimate $\hat{\rho}$ via OLS using $\hat{e}_t$ and $\hat{e}_{t-1}$.

d) Calculate y_t^* and x_t^* using $\hat{\rho}$

$$y_t^* = (y_t - \hat{\rho} y_{t-1}) \text{ and } x_t^* = (x_t - \hat{\rho} x_{t-1}).$$

e) Estimate $\hat{\hat{\beta}}_0$ and $\hat{\hat{\beta}}_1$ via OLS using y_t^* and x_t^*.

Repeat until estimate of ρ converges.

We are replacing an unknown parameter with a consistent estimate of that unknown parameter and then iterating – the estimates should improve.

EViews: $y_t = \beta_0 + \beta_1 x_t + \rho e_{t-1} + u_t$

Comment on generalized differencing. Lose one observation.
This is not efficient.

<u>Prais-Winsten transformation</u> for recovering first observation.

$$y_1^* = y_1 \sqrt{(1 - \rho^2)}$$

$$x_1^* = x_1 \sqrt{(1 - \rho^2)}$$

$$e_1^* = e_1 \sqrt{(1 - \rho^2)}$$

Most programs will do this. Use with caution –first observation could be influential – especially with a "small" sample. If P-W improves efficiency then how should the results – β's and V(β)'s – change when the transformation is used? What might they do if the observation is influential?

Maximum Likelihood Approach

The Cochrane-Orcutt procedure is an Estimated Generalized Least Squares (EGLS) method and relies on a consistency argument. We can consistently estimate β and then ρ and σ_u^2. Alternatively, we can use maximum likelihood where all parameters are estimated simultaneously.

The likelihood function looks like

$$\boldsymbol{lnL = \left[-\frac{1}{2}ln2\pi - \frac{1}{2}ln\sigma_u^2 - \frac{1}{2}ln(1 - \rho^2) - \frac{(1 - \rho^2)(y_1 - x_1'\beta)^2}{2\sigma_u^2}\right] + \sum_{t=2}^{T}\left[-\frac{1}{2}ln2\pi - \frac{1}{2}ln\sigma_u^2 - \frac{[(y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})'\beta]^2}{2\sigma_u^2}\right]}$$

and unlike the ML example earlier there are no closed-form expressions for the set of parameters – β, σ_u^2, and ρ. Changing one parameter will change the others. So, we must use nonlinear methods. (Choose starting value and iterate until convergence.)

How does your software do this? (Usually, start with OLS estimates…) There are also several ways to simplify the Hessian in the optimization problem. What does your software do, and can you choose options?

Example (dated version of problem set)

OLS: $\textit{NYSE Index}_t = -146.36 + 2.6193\ CPI_t + e_t$

(Std err)	*(22.39)*	*(0.1969)*
(p-value)	*(0.0001)*	*(0.0001)*

$R^2 = 0.9123$, F-test = 176.8868, P-Value = 0.0001, $\sigma = 23.354$

DW-d = 0.8492 where $d_l = 1.180$ and $d_u = 1.401$

EGLS: $\textit{NYSE Index}_t = -246.05 + 3.2319\ CPI_t + e_t$

(Std err)	*(52.06)*	*(0.4136)*
(P-value)	*(0.0001)*	*(0.0001)*

$e_t = 0.6721\ e_{t-1} + u_t$

(0.0305)

(0.0001)

$R^2 = 0.9713$, F = 67.7668, P-Value = 0.0001, $\sigma = 13.773$

Focus on the changes in significance.