POLS571 - Longitudinal Data Analysis
September 4, 2001

1 Outline of the course...

- First half on "time series" methods.
- Second half on methods for "panel" and "TSCS" data.

2 A quick review of autocorrelation in OLS regression...

One assumption of OLS is that of no autocorrelation in the errors:

$$E(u_i u_j) = 0 \text{ (for } i \neq j)$$

or...

$$E(\mathbf{uu}') = \sigma^2 \mathbf{I} = 0 \text{ for the off-diagonal elements}$$

Now if

$$E(u_i u_j) \neq 0 \text{ for } i \neq j$$

then we have autocorrelated errors.

For example, if we have data over time and

$$E(u_t u_{t-1}) \neq 0$$

then we have first-order autocorrelation (or **AR(1) errors**).

E.g.:

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 X_t + u_t \quad (1) \\
u_t &= u_{t-1} + e_t \quad (2)
\end{aligned}
$$

with e_t a mean-zero, unit variance white noise process.

Where does autocorrelation come from?

- Trends or inertia in the data
- MISSPECIFICATION
- OMITTED VARIABLES
- Aggregated or "smoothed" data

What does autocorrelation do?

- Estimator is still unbiased
- Still consistent
- NOT most efficient (why?)
  - If autocorrelation is positive, and the Xs are also positively autocorrelated, then overall variance estimates will be larger than w/o autocorrelation
- Gives biased standard errors
  - If $E(u_t u_{t-1}) > 0$ (say) then the estimated s.e.s will generally be underestimated...

How do we detect it?

- Graph of residuals on lagged residuals
- Runs test (Geary test)
- Chi-square test of positive vs. negative contemp. and lagged residuals

- Durbin-Watson d
  - Calculated as:

$$d = \frac{\sum_{t=2}^{N} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{N} \hat{u}_t^2} \tag{3}$$

  - Assumes that:
    * Regression Model contains an intercept
    * Fixed X variables (i.e. worthless if RHS variable is endogenous, e.g. lagged Y's)
    * Assumes first-order autocorrelation in the errors (says nothing about higher-order autocorrelation)

- Durbin's H
  - More general than d
  - Higher-order autocorrelation

What to do about it?
GLS, incorporating ρ into the equation

- is BLUE
- We don't normally know ρ...

A few different approaches...

- Difference equations
  - Assume that $\rho = 1$ (or ρ = -1)
  - Take first difference and then estimate
  - E.g.,

$$Y_t - Y_{t-1} = \beta_1 (X_t - X_{t-1}) + (u_t - u_{t-1}) \tag{4}$$

- Alternatively, use d to estimate ρ, then transform:

- Recall that $\rho = 1 - d/2$
- This means that we can use d to empirically transform the equation:

$$(Y_t - \rho Y_{t-1}) = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1}) \quad (5)$$

- Cochrane-Orcutt - an iterative procedure
  - Estimate the basic equation via OLS, and obtain residuals
  - Use the residuals to estimate $\hat{\rho}$ (i.e. the empirical correlation between u_t and u_{t-1}) – this is a biased but consistent estimate of ρ
  - Use this estimate of $\hat{\rho}$ to create the difference equation (5) and estimate
  - Save the residuals, and use them to estimate $\hat{\rho}$ again
  - Repeat this process until successive estimates of $\hat{\rho}$ differ by a very small amount
- Prais-Winsten - modification to Cochrane-Orcutt...
  - Cochrane-Orcutt "loses" the first observation...
  - transformation on first observation $\hat{Y_0} = Y_0(\sqrt{1-\hat{\rho}^2}), \hat{X_0} = X_0(\sqrt{1-\hat{\rho}^2})$
  - Will generally converge to global maximum (vs. local alternatives) since function is simple quadratic

3 Time Series Analysis: An Introduction...

We're typically talking about a series of observations on some variable Y over T time points:

$$\mathbf{Y} = \{Y_0, Y_1, ...Y_T\}$$

This series of observations can be thought of as a series of random variables (i.e., a stochastic process). We can consider the moments of this process in the usual way, e.g.:

$$
\begin{aligned}
\mu_t &= E(Y_t) &(7)\\
\sigma_t^2 \equiv Var(Y_t) &= E[(Y_t - \mu_t)^2] &(8)\\
\gamma_{t,t-s} \equiv Cov(Y_t, Y_{t-s}) &= E[(Y_t - \mu_t)(Y_{t-s} - \mu_{t-s})] &(9)
\end{aligned}
$$

Note, however, that we only have one realization of each data point Y_t; this will obviously make inferences about quantities like (7), um, difficult. So, we need to place some restrictions on the process generating the data. There are two such restrictions that are commonly used.

3.1 Stationarity

Stationarity is essentially a restriction on the data generating process over time. In particular, stationarity means that the fundamental form of the data generating process remains the same over time. This can be manifested in the moments of the process. For example, mean stationarity means that the expected value of the process is constant over time:

$$E(Y_t) = \mu \; \forall \; t \tag{10}$$

Similarly, variance stationarity means that the variance is temporally stable:

$$Var(Y_t) = E[(Y_t - \mu)^2] \equiv \sigma_Y^2 \; \forall \; t \tag{11}$$

and covariance stationarity is similar:

$$Cov(Y_t, Y_{t-s}) = E[(Y_t - \mu)(Y_{t-s} - \mu)] = \gamma_s \; \forall \; s \tag{12}$$

In the last case, this means that the autocorrelation of two observations $\{Y_t, Y_{t-s}\}$ depends only on the lag s, not on "where" in the series they fall.

This is a "weak" form of stationarity (that is, *stationarity in the moments*); a stricter form of stationarity requires that the joint probability distribution (in other words, *all* the moments) of series of observations $\{Y_1, Y_2, ...Y_t\}$ is the same as that for $\{Y_{1+s}, Y_{2+s}, ...Y_{t+s}\}$ for all t and s. In general, people deal with weak stationarity, and that's the sense in which we'll use it here.

(Note that, because it is fully characterized by its first two moments, if a normally-distributed series is weakly stationary, it is also strongly stationary).

3.2 Asymptotic Independence

We might generally expect that the covariance between two observations Y_t and Y_{t-s} would decrease as the distance between them (the lag) increases (that is, as s increases). In the limit, we could say that two observations in a series are asymptotically uncorrelated if:

$$Cov(Y_t, Y_{t-s}) = \gamma_s \rightarrow 0 \;\; as \;\; s \rightarrow \infty \tag{13}$$

This condition is often referred to as *ergodicity*.

4 Autocovariance and Autocorrelation

If a series is stationary and ergodic, then there are easy, consistent estimates for the mean, variance and autocovariance:

$$\hat{\mu} = \bar{Y} = T^{-1} \sum_{t=1}^{T} Y_t \tag{14}$$

$$\hat{\sigma^2} = T^{-1} \sum_{t=1}^{T} (Y_t - \bar{Y})^2 \tag{15}$$

$$\hat{\gamma_s} = T^{-1} \sum_{t=s+1}^{T} (Y_t - \bar{Y})(Y_{t-s} - \bar{Y}), \; s = 1, 2, 3, ... \tag{16}$$

One way to characterize a series as to the extent of dependence over time is by plotting its autocovariance against the number of lags s. In practice, however, we often want to "standardize" this measure, to get the *autocorrelation* function. We do so by dividing the autocovariance by the estimated variance:

$$\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\sigma}^2},\ s = 0, \pm 1, \pm 2, ... \tag{17}$$

Note that $\rho_0 = 1$, by definition. Plotting ρ_s against s gives the *autocorrelation function* (often abbreviated ACF). This plot is also occasionally referred to as a *correlogram*.

EXAMPLE: Consider the example of a simple *moving average process*:

$$Y_t = e_t + \theta e_{t-1},\ t = 1, 2, ...T \tag{18}$$

where e_t is i.i.d. with mean zero and constant variance σ_e^2. Its easy to calculate the properties of this series. The mean is:

$$\begin{aligned} \mu &= E(Y_t) \\ &= E(e_t) + \theta E(e_{t-1}) \\ &= 0 \end{aligned} \tag{19}$$

The variance is:

$$\begin{aligned} \sigma_Y^2 &= E[(e_t + \theta e_{t-1})(e_t + \theta e_{t-1})] \\ &= E(e_t^2) + \theta^2 E(e_{t-1}^2) + 2\theta E(e_t e_{t-1}) \\ &= (1 + \theta^2)\sigma_e^2 \end{aligned} \tag{20}$$

The one-lag covariance is:

$$\begin{aligned} \gamma_1 &= E[(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2})] \\ &= E(e_t e_{t-1}) + \theta E(e_{t-1}^2) + \theta E(e_t e_{t-2}) + \theta^2 E(e_{t-1} e_{t-2}) \\ &= \theta E(e_{t-1}^2) \\ &= \theta \sigma_e^2 \end{aligned} \tag{21}$$

and the two- and higher-lag covariances are all zero (why?).

For this series, then:

1. the means, variances and autocovariances are all independent of t,
2. the ACF is zero for all lags greater than one,
3. the ACF at one lag is dependent on the degree of dependence in the moving average process. In particular, since we defined $\hat{\rho}$ in (17) as $\frac{\hat{\gamma_s}}{\hat{\sigma^2}}$, this suggests that

$$\begin{aligned}\hat{\rho_1} &= \frac{\theta\sigma_e^2}{(1+\theta^2)\sigma_e^2} \\ &= \frac{\theta}{1+\theta^2} \qquad (22)\end{aligned}$$

(see Figure 1).

This in turn means that, for the MA(1) series given here,

- $\theta = 0$ corresponds to $\rho_1 = 0$, and is a "white noise" process (i.e., no temporal dependence in Y_t).
- When θ is positive,
    1. ρ_1 will be greater than zero,
    2. successive values of Y_t will be *positively* related, and
    3. the series will be "smoother" than a white noise series.
- In contrast, when θ is less than zero,
    1. ρ_1 will be less than zero as well,
    2. successive values of Y_t will be *negatively* correlated, and
    3. the series will be less "smooth" than a white noise sequence.

Next Time: More discussion of AR and MA processes, and ARIMA modeling in general...

Figure 1: Relationship between θ and ρ_1 for an MA(1) series