Serial Correlation

14.32 Spring 2007

Consider the generic model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

where all of the normal assumptions hold EXCEPT for random sampling, i.e., $E(\varepsilon_i \varepsilon_j) \neq 0$. This violates the Gauss-Markov assumptions, so OLS estimators will not be BLUE. BUT, they are still unbiased (or consistent in large samples), a feature we depend on for creating "good-guy" standard errors.

How to deal with standard errors under serial correlation?

1. Punt: Use Newey-West standard errors, which are robust to both heteroskedasticity and autocorrelation. You can't do this in SAS, but the Stata command newey will do it automatically. You know you've made it when you've got your own Stata command.

2. Model the serial correlation (GLS or FGLS). Just as with heteroskedasticity, the value of this procedure for correction will only be as good as the error model you use. Unlike heteroskedasticity, however, serial correlation usually involves some sort of autoregressive process, i.e.,

$$\varepsilon_t = \sum_{j=1}^{t} p_j \varepsilon_{t-j} + \xi_t$$

A specific (and intuitive) case that we often deal with is first-order autoregression, AR(1)

$$\varepsilon_t = \rho \varepsilon_{t-1} + \xi_t$$

where we (conveniently) assume that ξ_t has the properties we like, i.e., homoskedasticity and no serial correlation.

Aside: Macro-economists often think of the ε_t as "shocks." What sort of processes are implied when $\rho < 1, \rho = 1, \rho > 1$ in an AR(1) model?

GLS (Quasi-Differencing)

GLS with serial correlation involves a very similar process to the one we used with heteroskedasticity. We use the form of the standard errors to convert the model to one with a well-behaved error term. In the case of AR(1), we can subtract a multiple of observation $t-1$ from observation t:

$$\begin{aligned} y_t &= \alpha + \beta x_t + \rho\varepsilon_{t-1} + \xi_t \\ -(\rho y_{t-1} &= \rho\alpha + \rho\beta x_{t-1} + \rho\varepsilon_{t-1}) \\ y_t - \rho y_{t-1} &= (1-\rho)\alpha + \beta(x_t - \rho x_{t-1}) + \xi_t \\ \tilde{y}_t &= (1-\rho)\alpha + \beta\tilde{x}_t + \xi_t \end{aligned}$$

Aside: Prais-Winston and "getting back" the first observation.

If we know ρ, we can convert the variables using the process above, and estimate using OLS. The transformed equation now satisfies all of the G-M assumptions and will be BLUE (technically, you have to use the first observation for it to be BLUE, but it doesn't matter much). It will also be asymptotically efficient in large samples.

Without ρ (as is normally the case), we have to rely on FGLS.

FGLS

Just like with heteroskedasticity, FGLS relies on the fact that OLS estimators are consistent under serial correlation. We can therefore consistently estimate ρ using a the estimated OLS residuals. The full process is as follows:

1. Run the normal OLS regression and collect the residuals $\hat{\varepsilon}_t$.
2. Estimate the regression

   $$\hat{\varepsilon}_t = \rho\hat{\varepsilon}_{t-1} + \xi_t$$

   using OLS and get the estimate $\hat{\rho}$.
3. Use the estimated $\hat{p}$ to construct $\tilde{y}_t = y_t - \hat{\rho}y_{t-1}$ and $\tilde{x}_t = x_t - \hat{\rho}x_{t-1}$
4. Run the OLS regression

$$\tilde{y}_t = (1 - \hat{\rho})\alpha + \beta\tilde{x}_t + \xi_t$$

...and you're done! FGLS will be asymptotically efficient, and the standard errors you get will be valid.

Iterating

If you are just thinking about asymptotics, the process above will be just fine. In small samples, however, you could end up with very inefficient estimates of β. Iterating helps to produce more efficient estimates.

A. Run steps 1-4 above and collect the estimated $\hat{\beta}$.

B. Use that $\hat{\beta}$ to get the estimated residuals $\hat{\varepsilon}_t$ from the original estimating equation. Note: do not simply collect residuals from the transformed equation, as those are estimates of ξ_t and not ε_t!

C. Use the new $\hat{\varepsilon}_t$ and repeat steps 2-4 above.

D. Repeat steps A and B as many times as you like.

Eventually, your estimates of $\hat{\rho}$ and $\hat{\beta}$ should converge, i.e., after some point they won't change much if you iterate again. That's often a good time to stop.

We have seen that in SAS, the command for iterative quasi-differenced strategy (keeping the first observation) is

```
proc autoreg data=one;
model y=x / iter nlag=1;
```

Note that changing the *nlag=* option allows you to model a higher-order processes.

Testing for Serial Correlation

The above discussion suggests a very simple strategy for testing for serial correlation: check the magnitude and significance level of your estimated ρ. Economists that deal with time-series data often prefer the sophisticated-yet-unintuitive Durbin-Watson Statistic. It turns out that the D-W statistic is more or less a transform of $\hat{\rho}$:

$$DW \approx 2(1 - \hat{\rho})$$

We can reject serially uncorrelated errors if the DW statistic is close enough to 2. The annoying thing about DW is that there is a range of values for the DW statistic where we can neither reject or fail to reject. Part of the reason for this is that the underlying distribution changes based on the values of the of regressors in the original equation.