EC402: MEI. Spring Term: The Analysis of Time Series*

Lecture notes #1: Regression with Autocorrelated Disturbances

Up to this point of the course you have normally assumed that the regression errors are iid $(0, \sigma^2)$. Now let us consider what happens if they are serially correlated.

Consider the following simple model with $AR(1)$ errors

$$\begin{aligned} y_t &= x_t'\beta + u_t \qquad (t = 1, ..., T) \\ u_t &= \phi u_{t-1} + \varepsilon_t, \end{aligned} \tag{0.1}$$

$|\phi| < 1$, ε_t $iid(0, \sigma^2)$. Extension to more complex $ARMA(p, q)$ is feasible, but we will focus on the $AR(1)$ case as illustrative setting.

Note that

$$\begin{aligned} Var\left(u_t\right) &= \phi^2 Var\left(u_{t-1}\right) + Var\left(\varepsilon_t\right) + \phi 2 Cov\left(u_{t-1}, \varepsilon_t\right) \\ &\rightarrow \ \sigma_u^2 = \phi^2 \sigma_u^2 + \sigma^2 \text{ since } Cov\left(u_{t-1}, \varepsilon_t\right) = 0 \end{aligned}$$

impliying that $\sigma_u^2 = \frac{\sigma^2}{(1-\phi^2)}$. Also note that

$$u_t = \phi^s u_{t-s} + \sum_{j=0}^{s-1} \phi^j \varepsilon_{t-j} \rightarrow E(u_t, u_{t-s}) = \phi^s \sigma_u^2$$

Then we have

$$
\begin{aligned}
Var(u) &= E(u\, u') \\
&= \frac{\sigma^2}{(1-\phi^2)} f \begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{T-1} \\ \phi & 1 & \phi & \cdots & \phi^{T-2} \\ \phi^2 & \phi & 1 & \cdots & \phi^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \cdots & 1 \end{bmatrix} \\
&= \sigma^2 \Omega.
\end{aligned}
$$

Obviously this is no longer diagonal as in the case of *iid* errors.

1. Least Squares with Autocorrelated Disturbances

What happens if we estimate the model (0.1) with ordinary least squares?

Unsurprisingly this depends on whether the regressors are *process independent* of ε_t or if there is lagged feedback and they are only *contemporaneously independent*.[1]

If the regressors are *process independent* of ε_t, then this is basically the GLS problem discussed last term in the context of heteroskedasticity and the results are the same.

1. OLS estimates of β remain consistent (and unbiased) but are inefficient.

2. The usual estimate of the variance covariance matrix of $\hat{\beta}$ (namely $\sigma^2(X'X)^{-1}$) is wrong. The correct formula is (check that it is so)

$$Var(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}.$$

If the regressors are only *contemporaneously independent* of ε_t, things are much more serious. If x_t is only contemporaneously independent of ε_t, it may be correlated with lags of ε_t, and hence with u_t. So x_t is in effect endogenous and the OLS estimates are inconsistent.[2]

[1] If you don't remember these definitions from the first part of this course, check the appendix at the end of this handout.

[2] Since

$$p\lim \hat{\beta}_{OLS} = p\lim \left[(X'X)^{-1} X'y \right] = \beta + p\lim \left[\left(\frac{1}{N} X'X \right)^{-1} \right] p\lim \left[\frac{1}{N} X'u \right]$$

and the last term does not go to zero if x and u are correlated.

For example, consider the simple model with a lagged dependent variable

$$\begin{aligned} y_t &= \gamma y_{t-1} + w_t'\beta + u_t, \qquad |\gamma| < 1 \\ u_t &= \phi u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \text{ iid } (0, \sigma^2),\ 0 < |\phi| < 1 \\ & w_t \qquad \text{stationary and process independent of } \varepsilon_t. \end{aligned} \tag{1.1}$$

As a consequence of these assumptions, y_t and u_t are stationary and $E(w_t u_s) = 0$ all t, s.

Now consider the correlation between y_{t-1} and u_t. Intuitively this is unlikely to be zero because y_{t-1} is correlated with u_{t-1} and u_{t-1} is correlated with u_t. More formally

$$\begin{aligned} E(y_{t-1}u_t) &= E\left[\left(\gamma y_{t-2} + w_{t-1}'\beta + u_{t-1}\right)\left(\phi u_{t-1} + \varepsilon_t\right)\right] \\ &= \gamma\phi E(y_{t-2}u_{t-1}) + E\left(u_{t-1}w_{t-1}'\beta\phi\right) + \phi E(u_{t-1}^2) \\ &\quad + \gamma E(y_{t-2}\varepsilon_t) + E(\varepsilon_t w_{t-1}'\beta) + E(\varepsilon_t u_{t-1}). \end{aligned}$$

Consider each of these terms in turn:

$$\begin{array}{ll} E(y_{t-2}u_{t-1}) = E(y_{t-1}u_t) & \text{(by stationarity)} \\ E(u_{t-1}w_{t-1}'\beta\phi) = 0 & (w_t \text{ process independent}) \\ \phi E(u_{t-1}^2) = \frac{\phi\sigma^2}{(1-\phi^2)} & \text{(as shown in the previous section)} \\ \gamma E(y_{t-1}\varepsilon_t) = 0 & (\varepsilon_t \text{ iid, so independent of past information}) \\ E(\varepsilon_t w_{t-1}'\beta) = 0 & \text{ditto} \\ E(\varepsilon_t u_{t-1}) = 0 & \text{ditto} \end{array}.$$

From this we have

$$E(y_{t-1}u_t) = \gamma\phi E(y_{t-1}u_t) + \frac{\phi\sigma^2}{(1-\phi^2)}$$

or equivalently

$$E(y_{t-1}u_t) = \frac{\phi\sigma^2}{(1-\phi^2)(1-\gamma\phi)} \neq 0.$$

So y_{t-1} is correlated with the error in (1.1), therefore OLS applied to (1.1) will generate inconsistent parameter estimates.

2. Ω Known: Generalized Least Squares

If we know Ω (that is, we know ϕ), then using results proved last term, the GLS estimator is

$$\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y \tag{2.1}$$

with

$$Var(\hat{\beta}_{GLS}) = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}.$$

If the x variables are process independent of ε, $\hat{\beta}_{GLS}$ is unbiased and consistent.[3] If the x variables are only contemporaneously independent, $\hat{\beta}_{GLS}$ is biased, but consistent.[4]

In our example with $AR(1)$ errors

$$\Omega^{-1} = \begin{bmatrix} 1 & -\phi & 0 & \cdots & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \ddots & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{bmatrix}.$$

Check that multiplying this by Ω gives the unit matrix. Next you can verify that Ω^{-1} can be factored into $L'L$ where L is the $T \times T$ nonsingular matrix,

$$L = \begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi & 1 & & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{bmatrix}.$$

An easy way to carry out GLS is to premultiply the basic model by L and perform OLS. Why does this work? Writing (0.1) in matrix form gives

$$y = X\beta + u \qquad E(u) = 0.$$

Premultiply by L,

$$Ly = LX\beta + Lu. \tag{2.2}$$

Two points are worth noting. First, OLS on (2.2) gives

$$\begin{aligned} \hat{\beta} &= \left[(LX)'(LX)\right]^{-1}(LX)'Ly \\ &= \left(X'L'LX\right)^{-1}X'L'Ly \\ &= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y \\ &= \hat{\beta}_{GLS}. \end{aligned}$$

[3] $E\hat{\beta}_{GLS} = \beta$, and $p\lim \hat{\beta}_{GLS} = \beta$

[4] $E\hat{\beta}_{gls} \neq \beta$ but $p\lim \hat{\beta}_{gls} = \beta$

Second, note that the variance matrix of Lu is

$$\begin{aligned} E\left((Lu)(Lu)'\right) &= E\left(Lu\ u'L'\right) = LE(u\ u')L' \\ &= \sigma^2 L\Omega L' = \sigma^2 L(L'L)^{-1}L' \\ &= \sigma^2 LL^{-1}L'^{-1}L' \\ &= \sigma^2 I. \end{aligned}$$

This ensures that the errors in (2.2) are homoskedastic and serially uncorrelated.

In order to use OLS on (2.2), we need to compute Ly and LX. Let $Ly = y^*$, $LX = X^*$. Given the definition of L, we have

$$\begin{aligned} y_1^* &= \left(\sqrt{1-\phi^2}\right) y_1 \\ y_t^* &= y_t - \phi y_{t-1}, \qquad t = 2, ..., T \\ x_1^* &= \left(\sqrt{1-\phi^2}\right) x_1 \\ x_t^* &= x_t - \phi x_{t-1}, \qquad t = 2, ..., T. \end{aligned}$$

The transformed model satisfies the usual assumptions of the linear regression model and hence as for least squares in the independent homoskedastic case. If the regressors are process independent of ε, the GLS estimator is Best Linear Unbiased. If there is lagged feedback and the regressors are only contemporaneously independent of ε, the GLS estimator is biased but consistent.

As the first observation is handled differently from the rest, there is a temptation to drop it. If this is done the resulting transform is known as the Cochrane Orcutt transform. In large samples dropping the first observation has no effect. In small samples it can result in a substantial loss of efficiency.

3. Ω Unknown: Feasible GLS, Two Step Procedures

In general we do not know Ω. This leads to the obvious feasible GLS two step estimator: (i) obtain a consistent estimator of ϕ, (ii) use it in the GLS formula. Under standard conditions we would expect that the resulting estimator would have similar properties to

GLS, consistent and efficient in large samples. There are three main methods of estimating ϕ.

3.1. Method 1

The first method uses the residuals from the OLS regression of y on X, $\hat{u}$. Provided that the x are process independent of ε, then the least squares estimates of β are consistent and the residuals, $\hat{u}$ are consistent estimates of the errors, u. A consistent estimate of ϕ is obtained by regressing $\hat{u}$ on $\hat{u}_{-1}$,

$$\hat{\phi} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}.$$

This technique combined with GLS omitting the first observation is called the Cochrane Orcutt two-step technique. Unfortunately if the regressors are only contemporaneously independent of ε (the usual case in time series),[5] the first stage OLS estimates are inconsistent and this in turn causes both $\hat{\phi}$ and the second stage estimates to be inconsistent as well.

3.2. Method 2

If we take the original model (0.1) and subtract off ϕ times the lagged model, we obtain

$$\begin{aligned} y_t - \phi y_{t-1} &= (x_t - \phi x_{t-1})'\beta + u_t - \phi u_{t-1} \\ &= (x_t - \phi x_{t-1})'\beta + \varepsilon_t \end{aligned}$$

or

$$y_t = \phi y_{t-1} + x_t'\beta - x_{t-1}'(\beta\phi) + \varepsilon_t.$$

Durbin suggested estimating the second equation by OLS ignoring the constraint on the third coefficient. Then use the estimate of the coefficient on lagged y, $\hat{\phi}$ as the estimate of ϕ for the second stage GLS. Unlike the first procedure, the first stage estimates are consistent

[5] For example when lagged depended variables are among the regressors.

regardless of whether x is process or only contemporaneously independent of ε which means that the second stage estimates are also consistent in both cases. However, there are two problems here. First, the standard errors generated by following this procedure are typically incorrect, unless the x variables are process independent. Second, suppose we have the standard case of contemporaneous independence, that is where there is a lagged dependent variable as in (1.1). So if we take (1.1) and subtract off (1.1) lagged times ϕ, we have

$$y_t - \phi y_{t-1} = \gamma(y_{t-1} - \phi y_{t-2}) + (w_t' - \phi w_{t-1}')\beta + \varepsilon_t$$

or

$$y_t = (\phi + \gamma)y_{t-1} + \phi\gamma y_{t-2} + w_t'\beta - w_{t-1}'(\phi\beta) + \varepsilon_t. \tag{3.1}$$

In this case, following Durbin's procedure, we estimate by OLS and find that the coefficient on y_{t-1} is $(\hat{\phi} + \hat{\gamma})$ and that on y_{t-2} is $\hat{\phi}\hat{\gamma}$. We can solve these for $\hat{\phi}$ and $\hat{\gamma}$, but since there is perfect symmetry between $\hat{\phi}$ and $\hat{\gamma}$ here, we find two values but we don't know which is $\hat{\phi}$ and which is $\hat{\gamma}$. Of course, we could go on to investigate the ratios of the coefficients on w_{t-1} (a vector) to those on w_t, all of which generate estimates of ϕ. But this would be half baked. Why not simply estimate ϕ, γ, β in (3.1) using maximum likelihood, which will take account of the nonlinear restrictions on the coefficients? So we have the next method.

3.3. Method 3: MLE

As our basic example, we use (1.1), that is

$$\begin{aligned} y_t &= \gamma y_{t-1} + w_t'\beta + u_t, && |\gamma| < 1 \\ u_t &= \phi u_{t-1} + \varepsilon_t, && |\phi| < 1,\ t = 1, ..., T \\ & \quad \varepsilon_t, \text{ iid } N(0, \sigma^2), && w_t \text{ process independent of } \varepsilon_t. \end{aligned}$$

Lagging the equation, multiplying by ϕ and subtracting gives

$$y_t = \phi y_{t-1} + \gamma(y_{t-1} - \phi y_{t-2}) + (w_t - \phi w_{t-1})'\beta + \varepsilon_t.$$

So the conditional density of y_t given information $(y_{t-1}, y_{t-2}, w_t, w_{t-1})$ is

$$N\left(\phi y_{t-1} + \gamma(y_{t-1} - \phi y_{t-2}) + (w_t - \phi w_{t-1})'\beta, \sigma^2\right)$$

Taking the product of the conditionals in the usual way, and conditioning on y_1, y_2 fixed, the log likelihood is

$$\begin{aligned} \log L\left(\beta,\gamma,\phi,\sigma^2\right) &= -\frac{T-2}{2}\log 2\pi - \frac{(T-2)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=3}^{T}\varepsilon_t^2 \\ \varepsilon_t &= y_t - \phi y_{t-1} - \gamma(y_{t-1} - \phi y_{t-2}) - (w_t - \phi w_{t-1})'\beta. \end{aligned}$$

This is nonlinear least squares since maximizing $\log L$ wrt β,γ,ϕ is equivalent to minimizing $\sum_{t=3}^{T}\varepsilon_t^2$.

If we define z_t by

$$z_t = \begin{bmatrix} -\frac{\partial\varepsilon_t}{\partial\beta} \\ -\frac{\partial\varepsilon_t}{\partial\gamma} \\ -\frac{\partial\varepsilon_t}{\partial\phi} \end{bmatrix} = \begin{bmatrix} w_t - \phi w_{t-1} \\ y_{t-1} - \phi y_{t-2} \\ y_{t-1} - \gamma y_{t-2} - w_{t-1}'\beta \end{bmatrix}$$

then the FOC for a maximum are given by

$$\sum_{t=3}^{T} z_t\varepsilon_t = 0.$$

The solutions are the ML estimates $\hat\beta,\hat\gamma,\hat\phi$ which can be used to compute

$$\hat\sigma^2 = \frac{1}{T-2}\sum_{t=3}^{T}\left[y_t - \hat\phi y_{t-1} - \hat\gamma\left(y_{t-1} - \hat\phi y_{t-2}\right) - \left(w_t - \hat\phi w_{t-1}\right)'\beta\right]^2.$$

Then, by the usual formula MLE formula

$$Var\begin{bmatrix}\hat\beta \\ \hat\gamma \\ \hat\phi\end{bmatrix} = \hat\sigma^2\left(\sum_{t=3}^{T}\hat z_t\hat z_t'\right)^{-1}$$

where $\hat z_t$ is z_t evaluated at $\hat\beta,\hat\gamma,\hat\phi$.

3.4. Comparison of the Three Methods

If all the right-hand-side variables are process independent, then all three estimates are consistent and asymptotically efficient. They also provide consistent estimates of the variance covariance matrix of the estimates.

If the x_t are only contemporaneously independent, then the first stage of Cochrane-Orcutt is inconsistent, which makes the whole procedure inconsistent. If the x_t are only contemporaneously independent, the Durbin procedure remains consistent, but the simple estimates of the variance covariance matrix are wrong. Only ML maintains its properties, consistency and asymptotic efficiency and provides a consistent estimate of the variance covariance matrix. In practice it is probably best to use the ML method.

4. Appendix

Recall the following definitions:

1. *Process independence*: a variable x is process independent of the errors if each element x_t is independent of the errors, ε_s for all s and for all t. Process independence is closely related to the idea of *strong exogeneity* and is often referred to as such.

2. *Contemporaneous independence*: a variable x is contemporaneously independent if x_t is independent of ε_t for all t, but it may be correlated with ε_s, where $s < t$. The standard example of a right hand side variable that is contemporaneously independent but not process independent is a lagged endogenous variable (i.e. y_{t-1}). Contemporaneous independence is closely related to the idea of *weak exogeneity* and is often referred to as such. An alternative term that is frequently used to describe contemporaneously independent variables is *predetermined*. The simplest example of a linear model with a contemporaneously independent regressor is the $AR(1)$,

$$y_t = \phi y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \text{ iid}, \ |\phi| < 1.$$

   It is directly implied by the model that y_{t-1} is a function of ε_{t-1}, indeed, since $y_{t-1} = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j-1} = \varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + ...$, y_{t-1} depends on $\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, ...$

3. *Endogenous*: a variable x is endogenous if x_t is correlated with ε_t, i.e. contemporaneously correlated.