EC402: Regression with Autocorrelated Disturbances

Christian Julliard

Department of Economics and FMG
London School of Economics

What happens if errors are serially correlated?

Example 1:

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

where $t = 1, ..., T$, $|\phi| < 1$ (for stationarity) and ε_t is *iid* $\left(0, \sigma^2\right)$.
Since

$$u_t = \phi^s u_{t-s} + \sum_{j=0}^{s-1} \phi^j \varepsilon_{t-j}$$

we have that

$$
\begin{aligned}
\sigma_u^2 &= \frac{1}{1-\phi^2}\sigma^2 \\
E\left[u_t, u_{t-s}\right] &= \phi^s \sigma_u^2
\end{aligned}
$$

In then follows that

$$E\left[uu'\right] = \frac{\sigma^2}{1-\phi^2}\begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & 1 & \phi & \dots & \phi^{T-2} \\ \phi^2 & \phi & 1 & \dots & \phi^{T-3} \\ \dots & \dots & \dots & \dots & \dots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \dots & 1 \end{bmatrix} = \sigma^2\Omega$$

That is the covariance matrix of the residual is not diagonal due to the serial correlation.

Outline

Outline

Least Squares with Autocorrelated Disturbances

If we estimate model (1) by least square two cases can arise.

CASE 1: x is *process independent* i.e. x_t is independent of ε_s for all s and for all t ($E\left[x_t \varepsilon_s\right] = 0 \ \forall s, t$).
In this case:

- OLS estimates of β remain consistent (and unbiased) but are inefficient.
- The usual estimate of the variance covariance matrix of $\hat{\beta}$ (namely $\sigma^2 \left(X'X\right)^{-1}$) is wrong.
  The correct formula is

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

Note: for $\Omega = I$ this reduces to the standard formula

CASE 2: If the regressors are only contemporaneously independent of ε_t, things are much more serious.
$\Rightarrow$ x_t 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.

Example:

$$\begin{aligned} y_t &= \gamma y_{t-1} + u_t \quad |\gamma| < 1 \\ u_t &= \phi u_{t-1} + \varepsilon_t \quad |\phi| < 1,\ \varepsilon_t \sim iid\left(0, \sigma^2\right),\ E\left[y_{t-1}\varepsilon_t\right] = 0 \end{aligned}$$

In this case

$$\begin{aligned} E\left[y_{t-1}u_t\right] &= E\left[\left(\gamma y_{t-2} + u_{t-1}\right)\left(\phi u_{t-1} + \varepsilon_t\right)\right] \\ &= \gamma\phi E\left[y_{t-2}u_{t-1}\right] + \phi E\left[u_{t-1}^2\right] \end{aligned}$$

$$\rightarrow E\left[y_{t-1}u_t\right] = \frac{1}{1-\gamma\phi}\phi\sigma_u^2 = \frac{\phi\sigma^2}{\left(1-\gamma\phi\right)\left(1-\phi^2\right)} \neq 0$$

since ε_t is iid (therefore independent of past information), and by stationarity, $E\left[y_{t-1}u_t\right] = E\left[y_{t-2}u_{t-1}\right]$.

So OLS is not consistent!

Outline

Generalized Least Squares (when Ω is known)

If we know Ω, the following estimator has desirable properties:

GLS

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

whit

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

- If x and ε are process independent, $\hat{\beta}_{GLS}$ is unbiased and consistent.
- If x and ε are only contemporaneously independent, $\hat{\beta}_{GLS}$ is biased but consistent.

- Note that Ω^{-1} can be factorized as $\Omega^{-1} = L'L$ where L is $T \times T$ and nonsingular.
- Model (1) can be rewritten in matrix form as

$$y = X\beta + u, \quad E[u] = 0$$

- If we premultiply by L we get

$$Ly = LX\beta + Lu \tag{2}$$

- So, if we do OLS on this last expression we get

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

- Moreover, the variance of the residuals will be

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

$\Rightarrow$ This insures that the errors in the transformed model (2) are homoskedastic and serially uncorrelated.

$\Rightarrow$ If x and ε are process independent, GLS is BLUE.

Outline

Feasible GLS (when Ω is unknown)

In general, Ω is unknown. We can nevertheless proceed in two steps as follows:

1. obtain a consistent estimator of Ω
2. use $\hat{\Omega}$ in the GLS formulae

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 common methods of estimating Ω

Note: from now one we'll use Example 1 as working examples.

Method 1

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

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

- This technique combined with GLS omitting the first observation is called the **Cochrane Orcutt** two-step technique.

Warning: if the regressors are only contemporaneously independent of ε (the usual case in time series), first stage OLS is inconsistent and this causes both $\hat{\phi}$ and second stage estimates to be inconsistent.

Example: if lagged depended variables are among the regressors the method delivers inconsistent estimates.

Method 2

- Subtract ϕ times the lagged model from equation (1)

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

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

- So we can proceed as follows (Durbin):
  1. Estimate the second equation above by OLS ignoring the constraint on the third coefficient.
  2. Use $\hat{\phi}$ for the second stage GLS.

Note: the first stage estimates are consistent regardless of whether x is process or only contemporaneously independent of ε so second stage estimates are also consistent in both cases.

Problems:

1. the standard errors generated are typically incorrect, unless x is process independent.
2. suppose we have the standard case of contemporaneous independence, say

$$\begin{aligned} y_t &= \gamma y_{t-1} + w_t'\beta + u_t, \qquad |\gamma| < 1 \qquad (3) \\ u_t &= \phi u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{iid}N(0, \sigma^2),\ 0 < |\phi| < 1 \\ & w_t \qquad \text{stationary and process independent of } \varepsilon_t. \end{aligned}$$

- if we subtract ϕ times the lagged model we have

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

- 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 $\hat{\phi}\hat{\gamma}$, but since there is perfect symmetry we don't know what is $\hat{\gamma}$ and what is $\hat{\phi}$.
- We could go on to investigate the coefficients on w_{t-1} and w_t, but why not simply estimate ϕ, γ, β in (3) by MLE? (this would directly take into account the non-linear restrictions among coefficients)

Method 3: MLE

Let's use model (3) as our basic example. This implies as before

$$y_t = \phi y_{t-1} + \gamma(y_{t-1} - \phi y_{t-2}) + (w_t - \phi w_{t-1})'\beta + \varepsilon_t \sim \text{iid}N(0, \sigma^2)$$

- 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 conditional densities, and conditioning on y_1, y_2 fixed, the log likelihood is

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

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

- Then the FOC for a maximum are given by

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

where

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

- 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 for the variance of the MLE

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

where $\hat{z}_t$ is z_t evaluated at $\hat{\beta}, \hat{\gamma}, \hat{\phi}$.

Comparison of the 3 methods

- If all the regressors are process independent, then all three estimates are consistent and asymptotically efficient and provide consistent estimates of the variance covariance matrix of the estimates.
- If the x_t are only contemporaneously independent, Cochrane-Orcutt is 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.