Topic 8: Autocorrelated Errors

Consider the standard linear regression model

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad ; \qquad \boldsymbol{\varepsilon} \sim N[\mathbf{0}, \sigma^2 I_n]$$

- Among other things, because the off-diagonal elements of $V(\boldsymbol{\varepsilon})$ are all zero in value, we are assuming that the elements of the error vector are pair-wise *uncorrelated*.
- That is, they do not exhibit any *Autocorrelation*.
- Often, this assumption is unreasonable – especially with *time-series data*.
- Often, current values of the error term are correlated with past values.
- We often say they are "*Serially Correlated* ".
- In this case, the off-diagonal elements of $V(\boldsymbol{\varepsilon})$ will be non-zero.
- The particular values they take will depend on the *form of autocorrelation*.
- That is, they will depend on the *pattern of the correlations* between the elements of the error vector.
- $$V(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma^2 \end{bmatrix}$$
- If the *errors* themselves are autocorrelated, often this will be reflected in the regression *residuals* also being autocorrelated.
- That is, the residuals will follow some sort of *pattern*, rather than just being random.
- Typically, this reflects a mis-specification of the model *structure* itself.
- If the errors of our model are autocorrelated, then the OLS estimator of $\boldsymbol{\beta}$ usually will be unbiased and consistent, but it will be inefficient.
- In addition $V(\boldsymbol{b})$ will be computed incorrectly, and the standard errors, *etc*., will be ***inconsistent***.
- So, we need to consider formal methods for
    1. Testing for the presence/absence of autocorrelation.
    2. Estimating models when the errors are autocorrelated.
- It will be helpful to consider various specific forms of autocorrelation that may arise in practice.

- As we'll see, typically we can represent the important forms of autocorrelation with the addition of just a small number of parameters.
- That is, $V(\boldsymbol{\varepsilon})$ will be a function of σ^2, and just a small number of additional (unknown) parameters.

Autoregressive Process

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad ; \quad u_t \sim i.i.d.N[0\,,\sigma_u^2] \quad ; \quad |\rho| < 1$$

This is an AR(1) model for the error process.

More generally:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \cdots + \rho_p\varepsilon_{t-p} + u_t \quad ; \quad u_t \sim i.i.d.N[0\,,\sigma_u^2]$$

This is an AR(*p*) model for the error process. [*e.g.*, *p* = 4 with quarterly data.]

Moving Average Process

$$\varepsilon_t = u_t + \phi u_{t-1} \quad ; \quad u_t \sim i.i.d.N[0\,,\sigma_u^2]$$

This is an MA(1) model for the error process.

More generally:

$$\varepsilon_t = u_t + \phi_1\varepsilon_{t-1} + \cdots + \phi_q u_{t-q} \quad ; \quad u_t \sim i.i.d.N[0\,,\sigma_u^2]$$

This is an MA(*q*) model for the error process.

We can combine both types of process into an ARMA(*p* , *q*) model:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \cdots \rho_p\varepsilon_{t-p} + u_t + \phi_1 u_{t-1} + \cdots + \phi_q u_{t-q}$$

where: $u_t \sim i.i.d.N[0\,,\sigma_u^2]$.

- Note that in the AR(1) process, we said that $|\rho| < 1$.
- This condition is needed to ensure that the process is "stationary".
- Let's see what this actually means, more generally.
- Note – *all MA processes are stationary*.

Stationarity

Suppose that the following 3 conditions are satisfied:

1. $E[\varepsilon_t] = 0$; for all *t*
2. $var.[\varepsilon_t] = \sigma^2$; for all *t*
3. $cov.[\varepsilon_t , \varepsilon_s] = \gamma_{|t-s|}$; for all *t*, *s*; $t \neq s$

Then we say that the time-series sequence, $\{\varepsilon_t\}$ is "Covariance Stationary"; or "Weakly Stationary".

- More generally, this can apply to *any* time-series – not just the error process.
- Unless a time-series is stationary, we can't identify & estimate the parameters of the process that is generating its values.
- Let's see how this notion relates to the AR(1) model, introduced above.
- We have: $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

$$E[u_t] = 0$$

$$var.[u_t] = E[u_t^2] = \sigma_u^2$$

$$cov.[u_t, u_s] = 0 \quad ; \quad t \neq s$$

- So,

$$\begin{aligned} \varepsilon_t &= \rho[\rho\varepsilon_{t-2} + u_{t-1}] + u_t \\ &= \rho^2\varepsilon_{t-2} + \rho u_{t-1} + u_t \\ &= \rho^2[\rho\varepsilon_{t-3} + u_{t-2}] + \rho u_{t-1} + u_t \\ &= \rho^3\varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t \\ & etc. \end{aligned}$$

- Continuing in this way, eventually, we get:

$$\varepsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots \qquad \textbf{(1)}$$

  [This is an infinite-order MA process.]

  The value of ε_t embodies the entire past history of the u_t values.

- From **(1)**, $E(\varepsilon_t) = 0$, and

  $$var.(\varepsilon_t) = var.(u_t) + var.(\rho u_{t-1}) + var.(\rho^2\varepsilon_{t-2}) + \cdots$$

  $$= \sigma_u^2 + \rho^2\sigma_u^2 + \rho^4\sigma_u^2 + \cdots$$

$$= \sigma_u^2 \sum_{s=0}^{\infty} \rho^{2s} = \sigma_u^2 \sum_{s=0}^{\infty} (\rho^2)^s$$

- Now, under what conditions will this series converge?
  The series will converge to $\sigma_u^2(1-\rho^2)^{-1}$, as long as $|\rho^2| < 1$, and this in turn requires that $|\rho| < 1$.
- This is a necessary condition needed to ensure that the process, $\{\varepsilon_t\}$ is stationary, because if this condition isn't satisfied, then $var.[\varepsilon_t]$ is ***infinite***.
- So, for the AR(1) process, as long as $|\rho| < 1$, then $var.[\varepsilon_t] = \sigma_u^2(1-\rho^2)^{-1}$.
- In addition, stationarity implies that $var.[\varepsilon_t] = var.[\varepsilon_{t-s}]$, for all '*s*'.
- So, now consider the covariances of terms in the process:

  $cov.[\varepsilon_t, \varepsilon_{t-1}] = E[(\varepsilon_t - E(\varepsilon_t))(\varepsilon_{t-1} - E(\varepsilon_{t-1}))]$
  $= E[\varepsilon_t \varepsilon_{t-1}]$
  $= E[\varepsilon_{t-1}(\rho\varepsilon_{t-1} + u_t)]$
  $= \rho E[\varepsilon_{t-1}^2] + 0$
  $= \rho var.[\varepsilon_{t-1}] = \rho\sigma_u^2/(1-\rho^2)$
- Similarly,
  $cov.[\varepsilon_t, \varepsilon_{t-2}] = E[(\varepsilon_t - E(\varepsilon_t))(\varepsilon_{t-2} - E(\varepsilon_{t-2}))]$
  $= E[\varepsilon_{t-2}(\rho\varepsilon_{t-1} + u_t)]$
  $= E[\varepsilon_{t-2}(\rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t)]$
  $= \rho^2 E[\varepsilon_{t-2}^2] + 0$
  $= \rho^2 var.[\varepsilon_{t-2}] = \rho^2\sigma_u^2/(1-\rho^2)$
- In general, then, for the AR(1) process:
  $cov.[\varepsilon_t, \varepsilon_s] = \rho^{(t-s)}\sigma_u^2/(1-\rho^2)$; depends on $(t-s)$, not values of t, s ; and we can reverse t and s, so it actually depends on $|t-s|$.
- Also, recall that
  $var.[\varepsilon_t] = \sigma_u^2/(1-\rho^2)$
- So, the full covariance matrix for $\boldsymbol{\varepsilon}$ is:

$$V(\boldsymbol{\varepsilon}) = \sigma_u^2 \Omega = \frac{\sigma_u^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \ddots & \rho^{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}$$

If we can find a matrix, P, such that $\Omega^{-1} = P'P$, and if the value of ρ were *known*, then we could apply GLS estimation.

- More likely, in practice, find P, which will depend on ρ, and then estimate ρ consistently, and we can implement *feasible* GLS estimation.
- Before we consider GLS estimation any further, let's first see what implications autocorrelation of the errors has for the OLS estimator of $\boldsymbol{\beta}$.

OLS Estimation

- Given that the error term in our model now has a non-scalar covariance matrix, we know that the OLS estimator, ***b***, is still linear and unbiased, but it is *inefficient*.
- In general, ***b*** will still be a consistent estimator. However, there is one important situation where it will be *inconsistent*.
- This will be the case if the errors are autocorrelated, *and* one or more lagged values of the dependent variable enter the model as regressors.
  [The GLS estimator will also be inconsistent in this case.]
- A quick way to observe that inconsistent estimation will result in this case is as follows:

- Suppose that

$$y_t = \beta y_{t-1} + \varepsilon_t \quad ; \quad |\beta| < 1 \qquad \textbf{(2)}$$
$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad ; \quad u_t \sim i.i.d.[0\,, \sigma_u^2] \quad ; \ |\rho| < 1$$

Now subtract ρy_{t-1} from the expression for y_t in equation **(2)**:

$$(y_t - \rho y_{t-1}) = (\beta y_{t-1} + \varepsilon_t) - \rho(\beta y_{t-2} + \varepsilon_{t-1})$$

or,

$$y_t = (\beta + \rho)y_{t-1} - \beta\rho y_{t-2} + (\varepsilon_t - \rho\varepsilon_{t-1})$$
$$= (\beta + \rho)y_{t-1} - \beta\rho y_{t-2} + u_t$$

- So, if we estimate the model with just y_{t-1} as the only regressor, then we are effectively omitting a relevant regressor, y_{t-2}, form the model.
- This amounts to imposing a false (zero) restriction on the coefficient vector, and we know that this causes OLS to be not only biased, but also *inconsistent*.
- As was noted when we were discussing the general situation involving a regression model whose error vector has a non-scalar covariance matrix (in Topic 6), the estimated $V(\boldsymbol{b})$ will be *inconsistent*, regardless of the form of the regressors.
- So, to get consistent standard errors for the elements of $\boldsymbol{b}$, we can use the Newey-West correction when estimating $V(\boldsymbol{b})$.

Testing for Serial Independence

- Let's consider the problem of testing the hypothesis, H_0: "The errors in our regression model are serially independent".
- We'll need to formulate both the null, and an alternative hypothesis, expressing them in terms of the underlying parameters of the model.
- First, consider the possibility that the errors follow an AR(1) process, if they are not serially independent.
- That is:

$$y_t = \boldsymbol{x}'_t\boldsymbol{\beta} + \varepsilon_t \quad ; \quad t = 1, 2, \ldots, n \qquad \textbf{(3)}$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad ; \quad u_t \sim i.i.d.[0, \sigma_u^2] \quad ; \; |\rho| < 1$$

Then, we have $\quad H_0: \rho = 0 \quad vs. \quad H_A: \rho \neq 0 \quad (> 0 \quad ; \quad < 0\ )$

- Notice that, as usual, we can learn something about the behaviour of the *errors* in our regression model by looking at the *residuals* obtained when we estimate the model.
- So, estimate **(3)** by OLS (ignoring any possibility of serial correlation), and get the residuals, $\{e_t\}$.

- Then, fit the following "*auxiliary regression*":

$$e_t = re_{t-1} + v_t \qquad ; \qquad t = 2, 3, \ldots, n$$

- The OLS estimator of the coefficient, *"r"*, is:

$$\hat{r} = \left[\sum_{t=2}^{n} e_t e_{t-1}\right] / \left[\sum_{t=2}^{n} e_{t-1}^2\right]$$

- We could think of using a "z-test" to test if $r = 0$. This test will be valid, *asymptotically*:

$$z = \frac{(\hat{r} - 0)}{s.e.(\hat{r})} \xrightarrow{d} N[0\,,1]$$

- Now, testing for serial independence, against the alternative hypothesis that the process is AR(1) is very interesting.
- Anderson (1948) proved that **there does not exist any UMP test** for this problem!
- So, historically, there were lots of attempts to construct tests that were "approximately" most powerful.
- These days we generally use tests from the so-called "**Lagrange Multiplier Test**" family. Also called the family of "**Score Tests**".
- Tests of this type can be used for all sorts of testing problems – not just for testing for serial independence.
- They are especially useful when it is relatively easy to estimate the model under the assumption that the null hypothesis is true.
- Here, such estimation involves just OLS.
- LM tests have only *asymptotic validity*. Asymptotically, the distribution of the test statistic is Chi-Square, with d.o.f. equal to the number of restrictions being tested, if the null hypothesis is true.
- The pay-off is that the test can be applied under *very general conditions*.
- We don't need to have normally distributed errors in our regression model.
- The regressors can be random; *etc*.

- The Breusch-Godfrey Test for serial independence of the errors can be implemented as follows:

  1. Estimate the model, $y_t = \boldsymbol{x}'_t\boldsymbol{\beta} + \varepsilon_t$; $t = 1, 2, \ldots, n$ by OLS, and get the residuals $\{e_t\}$.
  2. If the Alternative Hypothesis is that the errors follow *either* an AR(p) process, *or* an MA(p) process, then estimate the following auxiliary regression:

$$e_t = \boldsymbol{x}'_t\boldsymbol{\gamma} + \delta_1 e_{t-1} + \cdots + \delta_p e_{t-p} + v_t \qquad \textbf{(4)}$$

  3. The test statistic is $LM = nR^2$, where R^2 is the "uncentered" coefficient of determination from **(4)**.

  4. Reject $H_0 : \varepsilon_t\ serially\ independent$; if $LM > \chi^2_{(p)}$ critical value.

- If we reject H_0 , we're left with ***incomplete information*** about the particular form of the autocorrelation.

Estimation Allowing for Autocorrelation

- Suppose we have a regression model with AR(1) errors:

$$y_t = \boldsymbol{x}'_t\boldsymbol{\beta} + \varepsilon_t \quad ; \quad t = 1, 2, \ldots, n$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad ; \quad u_t \sim i.i.d.[0, \sigma_u^2] \quad ; \quad |\rho| < 1$$

- So, the full covariance matrix for $\boldsymbol{\varepsilon}$ is:

$$V(\boldsymbol{\varepsilon}) = \sigma_u^2\Omega = \frac{\sigma_u^2}{(1-\rho^2)}\begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \ddots & \rho^{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \ldots & 1 \end{bmatrix}$$

- We need to find a matrix, P, such that $\Omega^{-1} = P'P$, and then we can apply GLS estimation.
- In the AR(1) case, we can show that:

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & & & -\rho & 1 \end{bmatrix}$$

- GLS is simply OLS, using the data $\boldsymbol{y}^*$ and X^*, where:

$$y^* = \begin{bmatrix} y_1\sqrt{1-\rho^2} \\ y_2 - \rho y_1 \\ \vdots \\ \vdots \\ y_n - \rho y_{n-1} \end{bmatrix} \quad ; \quad x_j^* = \begin{bmatrix} x_{1j}\sqrt{1-\rho^2} \\ x_{2j} - \rho x_{1j} \\ \vdots \\ \vdots \\ x_{nj} - \rho x_{n-1,j} \end{bmatrix} \quad ; \quad j = 1, 2, \ldots, k$$

- What if ρ is unknown, as is likely to be the case?
- We can apply feasible GLS – this is essentially what Cochrane & Orcutt (1949) did, except that they "dropped" the first observation as they didn't know the leading (1 , 1) element of the P matrix.
- The steps are:
  1. Estimate the model, $y_t = \boldsymbol{x}'_t\boldsymbol{\beta} + \varepsilon_t$, by OLS and get the residuals, $\{e_t\}$.
  2. Estimate ρ, using

     $$\hat{\rho} = \left[\sum_{t=2}^{n} e_t e_{t-1}\right] / \left[\sum_{t=2}^{n} e_{t-1}^2\right]$$

  3. Construct $\boldsymbol{y}^*$ and X^*, using $\hat{\rho}$ in place of ρ .
  4. Apply OLS using the transformed data. This is **feasible GLS** estimation.
  5. Iterate Steps 1 through 4.
  6. Continue until convergence is achieved.

- Convergence is guaranteed in a *finite number of steps*, unless the model includes lagged values of the dependent variable.
- The same approach can be used if the errors follow a ("simple") AR(p) process: $\varepsilon_t = \rho\varepsilon_{t-p} + u_t \quad ; \quad u_t \sim i.i.d.[0\,, \sigma_u^2]$
- Things are more complicated if the errors follow an MA(q) or ARMA(p , q) process.