Lecture 15: Serial Correlation

1. Serial Correlation

The fourth “classical assumption” of the linear regression model is

The stochastic error terms, $\varepsilon_i$, are uncorrelated.

Stated differently, we require there to be no serial correlation or autocorrelation in $\varepsilon_i$. That is to say that $\varepsilon_i$ must not be correlated with $\varepsilon_{i-1}$ for any $i$.

The interpretation of this requirement is slightly different for cross-sectional data and for time series data. In a cross-section, the assumption is less strict because typically we think of the observational units as being independent of each other. If, for example, individuals who are geographically close have similar unobserved characteristics, then we might worry about correlation in $\varepsilon_i$ across $i$. However, in cross-sectional data we are typically content that the no serial correlation assumption is satisfied.

Time series data are typically more problematic in terms of this assumption. The unobservables which determine time series variables, say GDP, income, or consumption, are likely to be persistent over time.

Suppose that we have a time series model where $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$ with $|\rho| < 1$ and $u_t$ being a classical error term with mean zero and no serial correlation. If $\rho > 0$, then we say $\varepsilon_t$ exhibits positive autocorrelation. Similarly, if $\rho < 0$, then there is negative autocorrelation. We will only focus on the case where $|\rho| < 1$, and the process $\varepsilon_t$ is said to be stable. If $|\rho| > 1$ then $\varepsilon_t$ is said to be explosive.

Other forms of serial correlation include second-order serial correlation, $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + u_t$, and (if $t$ represents one quarter) quarterly autocorrelation: $\varepsilon_t = \rho \varepsilon_{t-4} + u_t$.

1.1. Consequences of Serial Correlation

In the case of pure serial correlation (because the error term is actually persistent, rather than through some sort of specification error), then:

1. The coefficient estimates will remain unbiased.

2. Standard errors and $t$-statistics will be incorrect, leading to invalid inference.

3. OLS is no longer the “best” estimator (in terms of minimizing the variance of the estimated coefficients).
1.2. Testing for Serial Correlation

We can use the Durbin-Watson test statistic, denoted $d$, to test for first-order serial correlation. This statistic is defined as

$$d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \approx 2(1 - \hat{\rho}).$$

This statistic always lies between 0 and 4. When $d \approx 2$, this is evidence of no autocorrelation, $2 < d < 4$ indicates negative autocorrelation, and $d < 2$ indicates positive autocorrelation.

The null hypothesis of the Durbin-Watson test is $\rho \leq 0$, or no (positive) autocorrelation. The rejection rule is not as simple as with $t$-tests of $F$-tests. Upper and lower critical values of this distribution are tabulated, and serve as boundaries between the rejection region, the not reject region, and the inconclusive region.

To perform a test of $H_0 : \rho \leq 0$ (no positive autocorrelation) against $H_A : \rho > 0$ (positive autocorrelation) at level $\alpha$, compare $d$ to the upper and lower critical values ($d_L$ and $d_U$):

1. If $d < d_L$, reject $H_0$.
2. If $d > d_U$, fail to reject $H_0$.
3. If $d_L < d < d_U$, then the test is inconclusive.

A quick rule of thumb for the 5% test is: if $d < 1.6$ and $n \geq 100$, then reject $H_0$, no (positive) autocorrelation, against $H_A$, positive autocorrelation.

1.3. Accounting for Serial Correlation

A practical way to deal with serial correlation is to include lags of $Y_t$ and $X_t$ as explanatory variables. For example, rather than estimating the model

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t,$$

estimate

$$Y_T = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 Y_{t-1} + \varepsilon_t$$

instead.

Another method, called the Cochrane-Orcutt technique, is to estimate $\hat{\rho}$ and adjust the dependent and independent variables accordingly. Consider the following two identities:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t,$$
$$\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{t-1} + \rho \varepsilon_{t-1}.$$

Subtracting these two equations gives

$$Y_t - \rho Y_{t-1} = \beta_0 (1 - \rho) + \beta_1 (X_t - \rho X_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1}).$$

So, we have

$$\tilde{Y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_t + u_t$$
where
\[ \tilde{Y}_t = Y_t - \rho Y_{t-1}, \]
\[ \tilde{X}_t = X_t - \rho X_{t-1}, \]
\[ \tilde{\beta}_0 = \beta_0 - \rho \hat{\beta}_0, \]
\[ u_t = \varepsilon_t - \rho \varepsilon_{t-1}. \]

Therefore, if we only knew \( \rho \), then we could use the transformed model above to estimate \( \beta_0 \) and \( \beta_1 \) almost as usual. However, we can use an estimate of \( \rho \), \( \hat{\rho} \), obtained using the following method:

1. Regress \( Y_t \) on \( X_t \) and obtain the residuals \( e_t \).

2. Regress \( e_t \) on \( e_{t-1} \) and obtain \( \hat{\rho} \).

Once we have \( \hat{\rho} \), we can use it to calculate \( \tilde{Y}_t \) and \( \tilde{X}_t \) and use OLS to obtain \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). This process can be iterated to obtain successively better estimates of \( \hat{\beta}_0 \), \( \hat{\beta}_1 \), and \( \hat{\rho} \).