LECTURE 11

Introduction to Econometrics

Autocorrelation

November 29, 2016
ON PREVIOUS LECTURES

- We discussed the specification of a regression equation

- **Specification** consists of choosing:
  1. correct independent variables
  2. correct functional form
  3. correct form of the stochastic error term

- We talked about the choice of independent variables and their functional form

- We started to talk about the form of the error term - we discussed heteroskedasticity
ON TODAY’S LECTURE

▶ We will finish the discussion of the form of the error term by talking about **autocorrelation** (or **serial correlation**)

▶ We will learn

▶ what is the nature of the problem

▶ what are its consequences

▶ how it is diagnosed

▶ what are the remedies available
Nature of Autocorrelation

- Observations of the error term are correlated with each other
  \[ \text{Cov}(\varepsilon_i, \varepsilon_j) \neq 0 \quad , \quad i \neq j \]

- Violation of one of the classical assumptions

- Can exist in any data in which the order of the observations has some meaning - most frequently in time-series data

- Particular form of autocorrelation - AR(p) process:
  \[ \varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \ldots + \rho_p \varepsilon_{t-p} + u_t \]

  - \( u_t \) is a classical (not autocorrelated) error term
  - \( \rho_k \) are autocorrelation coefficients (between -1 and 1)
EXAMPLES OF PURE AUTOCORRELATION

- Distribution of the error term has autocorrelation nature
- First order autocorrelation

\[ \varepsilon_t = \rho_1 \varepsilon_{t-1} + u_t \]

- positive serial correlation: \( \rho_1 \) is positive
- negative serial correlation: \( \rho_1 \) is negative
- no serial correlation: \( \rho_1 \) is zero
- positive autocorrelation very common in time series data
- e.g.: a shock to GDP persists for more than one period

- Seasonal autocorrelation (in quarterly data)

\[ \varepsilon_t = \rho_4 \varepsilon_{t-4} + u_t \]
EXAMPLES OF IMPURE AUTOCORRELATION

▶ Autocorrelation caused by specification error in the equation:
  ▶ omitted variable
  ▶ incorrect functional form

▶ How can misspecification cause autocorrelation in the error term?
  ▶ Recall that the error term includes the omitted variables, nonlinearities, measurement error, and the classical error term.
  ▶ If we omit a serially correlated variable, it is included in the error term, causing the autocorrelation problem.

▶ Impure autocorrelation can be corrected by better choice of specification (as opposed to pure autocorrelation).
AUTOCORRELATION
CONSEQUENCES OF AUTOCORRELATION

1. Estimated coefficients ($\hat{\beta}$) remain unbiased and consistent

2. Standard errors of coefficients ($s.e.(\hat{\beta})$) are biased (inference is incorrect)
   - serially correlated error term causes the dependent variable to fluctuate in a way that the OLS estimation procedure attributes to the independent variable
   - Serial correlation typically makes OLS underestimate the standard errors of coefficients
   - therefore we find $t$ scores that are incorrectly too high

⇒ The same consequences as for the heteroskedasticity
**Durbin-Watson Test for Autocorrelation**

- Used to determine if there is a first-order serial correlation by examining the residuals of the equation

- Assumptions (criteria for using this test):
  - The regression includes the intercept
  - If autocorrelation is present, it is of $AR(1)$ type:
    \[ \varepsilon_t = \rho \varepsilon_{t-1} + u_t \]
  - The regression does not include a lagged dependent variable
Durbin-Watson Test for Autocorrelation

- Durbin-Watson $d$ statistic (for $T$ observations):

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

where $\hat{\rho}$ is the autocorrelation coefficient

- Values:
  1. Extreme positive serial correlation: $d \approx 0$
  2. Extreme negative serial correlation: $d \approx 4$
  3. No serial correlation: $d \approx 2$
 USING THE DURBIN-WATSON TEST

1. Estimate the equation by OLS, save the residuals

2. Calculate the $d$ statistic

3. Determine the sample size $T$ and the number of explanatory variables (excluding the intercept!) $k'$

4. Find the upper critical value $d_{U}$ and the lower critical value $d_{L}$ for $T$ and $k'$ in statistical tables

5. Evaluate the test as one-sided or two-sided (see next slides)
One-sided Durbin-Watson test

- For cases when we consider only positive serial correlation as an option

- Hypothesis:
  
  $H_0 : \rho \leq 0$ (no positive serial correlation)
  
  $H_A : \rho > 0$ (positive serial correlation)

- Decision rule:
  
  - if $d < d_L$ reject $H_0$
  
  - if $d > d_U$ do not reject $H_0$
  
  - if $d_L \leq d \leq d_U$ inconclusive
Durbin-Watson critical values for one-sided test

Panel A
One Tail

Reject

\[ \rho = 0 \]

Uncertain

Fail to Reject

\[ \rho = 0 \]

\[ d_L \]

2

\[ d_U \]

4
Two-sided Durbin-Watson Test

- For cases when we consider both signs of serial correlation

- Hypothesis:

  \[ H_0 : \rho = 0 \quad \text{(no serial correlation)} \]
  \[ H_A : \rho \neq 0 \quad \text{(serial correlation)} \]

- Decision rule:

  - if \( d < d_L \) reject \( H_0 \)
  - if \( d > 4 - d_L \) reject \( H_0 \)
  - if \( d > d_U \) do not reject \( H_0 \)
  - if \( d < 4 - d_U \) do not reject \( H_0 \)
  - otherwise inconclusive
Durbin-Watson critical values for two-sided test
EXAMPLE

- Estimating housing prices in the UK
- Quarterly time series data on prices of a representative house in UK (in £)
- Explanatory variable: GDP (in billions of £)
- Time span: 1975 Q1 - 2011 Q2
- All series are seasonally adjusted and in real prices (i.e. adjusted for inflation)
### Example

Model 1: OLS, using observations 1975:1-2011:2 (T = 146)
Dependent variable: house_price

<table>
<thead>
<tr>
<th></th>
<th>coefficient</th>
<th>std. error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>-38409.8</td>
<td>6675.01</td>
<td>-5.754</td>
<td>5.04e-08 ***</td>
</tr>
<tr>
<td>gdp</td>
<td>737.065</td>
<td>31.4846</td>
<td>23.41</td>
<td>6.09e-51 ***</td>
</tr>
</tbody>
</table>

Mean dependent var 113072.8
S.D. dependent var 43254.80

Sum squared resid 5.65e+10
S.E. of regression 19799.38

R-squared 0.791921
Adjusted R-squared 0.790476

F(1, 144) 548.0434
P-value(F) 6.09e-51

Log-likelihood -1650.595
Akaike criterion 3305.191

Schwarz criterion 3311.158
Hannan-Quinn 3307.615

rho 0.984890
Durbin-Watson 0.023930
EXAMPLE

- We test for positive serial correlation:

  \[ H_0 : \rho \leq 0 \text{ (no positive serial correlation)} \]
  \[ H_A : \rho > 0 \text{ (positive serial correlation)} \]

- One-sided DW critical values at 95% confidence for \( T = 146 \) and \( k' = 1 \) are:

  \[ d_L = 1.72 \quad \text{and} \quad d_U = 1.74 \]

- Decision rule:
  - if \( d < 1.72 \) reject \( H_0 \)
  - if \( d > 1.74 \) do not reject \( H_0 \)
  - if \( 1.72 \leq d \leq 1.74 \) inconclusive

- Since \( d = 0.02 < 1.72 \), we reject the null hypothesis of no positive serial correlation
ALTERNATIVE APPROACH TO AUTOCORRELATION TESTING

▶ Suppose we suspect the stochastic error term to be $AR(p)$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \ldots + \rho_p \varepsilon_{t-p} + u_t$$

▶ Since OLS is consistent even under autocorrelation, the residuals are consistent estimates of the stochastic error term

▶ Hence, it is sufficient to:

1. Estimate the original model by OLS, save the residuals $e_t$
2. Regress $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \ldots + \rho_p e_{t-p} + u_t$
3. Test if $\rho_1 = \rho_2 = \ldots = \rho_p = 0$ using the standard $F$-test
Model 1: OLS, using observations 1975:1-2011:2 (T = 146)
Dependent variable: house_price

<table>
<thead>
<tr>
<th></th>
<th>coefficient</th>
<th>std. error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>-38409.8</td>
<td>6675.01</td>
<td>-5.754</td>
<td>5.04e-08 ***</td>
</tr>
<tr>
<td>gdp</td>
<td>737.065</td>
<td>31.4846</td>
<td>23.41</td>
<td>6.09e-51 ***</td>
</tr>
</tbody>
</table>

Mean dependent var 113072.8 S.D. dependent var 43254.80
Sum squared resid 5.65e+10 S.E. of regression 19799.38
R-squared 0.791921 Adjusted R-squared 0.790476
F(1, 144) 548.0434 P-value(F) 6.09e-51
Log-likelihood -1650.595 Akaike criterion 3305.191
Schwarz criterion 3311.158 Hannan-Quinn 3307.615
rho 0.984890 Durbin-Watson 0.023930
Dependent variable: e

<table>
<thead>
<tr>
<th></th>
<th>coefficient</th>
<th>std. error</th>
<th>t-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_1</td>
<td>1.75237</td>
<td>0.0843401</td>
<td>20.78</td>
<td>2.53e-44  ***</td>
</tr>
<tr>
<td>e_2</td>
<td>-1.05900</td>
<td>0.168179</td>
<td>-6.297</td>
<td>3.79e-09  ***</td>
</tr>
<tr>
<td>e_3</td>
<td>0.477195</td>
<td>0.168362</td>
<td>2.834</td>
<td>0.0053    ***</td>
</tr>
<tr>
<td>e_4</td>
<td>-0.190822</td>
<td>0.0848111</td>
<td>-2.250</td>
<td>0.0260    **</td>
</tr>
</tbody>
</table>

Mean dependent var  -443.8153  S.D. dependent var  19823.71
Sum squared resid   7.22e+08  S.E. of regression  2287.633
R-squared           0.986973  Adjusted R-squared  0.986690
F(4, 138)           2613.852  P-value(F)            5.8e-129
Log-likelihood      -1297.869  Akaike criterion     2603.739
Schwarz criterion   2615.562  Hannan-Quinn         2608.543
rho                 0.006283  Durbin-Watson        1.967108
**Remedy: White robust standard errors**

- Note that autocorrelation does not lead to inconsistent estimates, only to incorrect inference - similar to heteroskedasticity problem
- We can keep the estimated coefficients, and only adjust the standard errors
- The White robust standard errors solve not only heteroskedasticity, but also serial correlation
- Note also that all derived results hold if the assumption $\text{Cov}(x, \varepsilon) = 0$ is not violated
  - First make sure the specification of the model is correct, only then try to correct for the form of an error term!
SUMMARY

▶ Autocorrelation does not lead to inconsistent estimates, but it makes the inference wrong (estimated coefficients are correct, but their standard errors are not)

▶ It can be diagnosed using
  ▶ Durbin-Watson test
  ▶ Analysis of residuals

▶ It can be remedied by
  ▶ White robust standard errors

▶ Readings:
  ▶ Studenmund, Chapter 9
  ▶ Wooldridge, Chapter 12