TSP-Problem Set 5: Mulitcollinearity and Panel Data

Preparations

Please create a new folder for this exercise session with your name in directory T:. Then go to L:\Applied Econometrics\ex05 and copy the files into your folder.

1) Multicollinearity

Consider the following regression model:

\[ y_i = 1 + 0.15 \cdot x_i + 0.15 \cdot z_i + u_i \]

with:

- \( u_i \) and \( x_i \) \( \sim \) \( N(0, 1) \)
- \( z_i = \lambda \cdot x_i + v_i, \lambda = const., v_i \sim N(0, 1) \)

(a) Which kind of distribution does \( z_i \) follow? How does the correlation between \( x_i \) and \( z_i \) depend on \( \lambda \)?

(b) Generate a dataset (N=100) with \( \lambda = 3.042 \). How large is the expected correlation between \( x \) and \( z \)?

(c) Estimate the model using OLS. Which problems might occur during the estimation? Which problems actually do occur in the case of your estimation?

(d) Repeat the data generating process and the OLS estimation 1000 times and save the estimated coefficients of \( x \) and \( z \). Display the estimated coefficients in a scatter diagram.
(e) Based on the scatter diagram, explain why, in the case of multicollinearity, the F-test often produces statistically significant results even though the t-tests for the coefficients of x and z do not suggest any statistically significant effects.

2) Effect of a Garbage Incinerator’s Location on Housing Prices

(Based on Wooldridge, Example 13.3, pp. 450-453)

Use the data in “ex05_1.xls” to study the effect that a new garbage incinerator had on housing values in North Andover, Massachusetts in the early 1980s. Two years of data are available for analysis: 1978 and 1981. In 1978 it was not yet publicly known, that a garbage incinerator was going to be constructed in the area. The rumor that a new incinerator would be built in North Andover began after 1978. Construction began in 1981 and the incinerator was expected to be in operation soon after the start of construction. For the years 1978 and 1981, the data set contains the following variables:

- \( rprice \)  house price in real terms (in 1978 dollars)
- \( nearinc \)  = 1 if house is near the incinerator (within three miles)
- \( age \)  age of the house
- \( intst \)  distance to the interstate in feet
- \( land \)  land area in feet
- \( area \)  house area in feet
- \( rooms \)  number of rooms
- \( baths \)  number of baths
- \( y81 \)  = 1 if year of data is 1981

(a) Estimate a simple regression model of \( rprice \) on \( nearinc \) using only the data for 1981. Does the result imply, that the siting of the incinerator is causing lower housing prices?

(b) Estimate the same regression using only the data for 1978. What do the results imply with regard to your answer to question (a)?
(c) Calculate and interpret the difference of the two coefficients from (a) and (b).

(d) Estimate the difference-in-difference estimation directly using the data pooled over both years to get the standard error. Under which assumption does this estimation give the causal effect?

(e) Why is it important to add covariates? Add the variables age, agesq, intst, land, area, rooms and baths to your regression.

3) Appendix: TSP-commands

<table>
<thead>
<tr>
<th>do</th>
<th>\textbf{do} \textit{indexname} = \textit{startvalue} \textbf{to} \textit{endvalue} \textbf{by} \textit{increment}; program code \textbf{endo}; → specifies a conventional \textbf{loop}. The statements between the \textit{do} and \textbf{endo} statement are executed repetitively as many times as specified by the information given on the \textit{do} statement.</th>
</tr>
</thead>
<tbody>
<tr>
<td>dot</td>
<td>→ creates a “dot” loop, which is like a regular “do” loop, except that the values of the index are a series of character strings (names), e.g.:</td>
</tr>
<tr>
<td></td>
<td>\textit{dot} \textit{variable1} \textit{variable2} \textit{variable3}; \textbf{genr} \textit{new.} = .*100; \textbf{enddot};</td>
</tr>
<tr>
<td></td>
<td>→ generates three new variables called newvariable1, newvariable2, etc.</td>
</tr>
<tr>
<td>random</td>
<td>\textbf{random} (mean=0, stdev=1) \textit{variable1}; \textbf{random} (mean=10, poisson) \textit{variable2}; etc. → draws a \textit{random variable} which follows a certain distribution; if no options are specified the default setting is normal distribution</td>
</tr>
</tbody>
</table>