Causal Inference: Problem Set 10

Bezirgen Veliyev (Aarhus University)

November 12, 2020

Based on Chapter 10 problems of Wooldridge.

1. Decide if you agree or disagree with each of the following statements and give a brief explanation of your decision:

   (i) Like cross-sectional observations, we can assume that most time series observations are independently distributed.

   (ii) The OLS estimator in a time series regression is unbiased under the first three Gauss-Markov assumptions.

   (iii) A trending variable cannot be used as the dependent variable in multiple regression analysis.

   (iv) Seasonality is not an issue when using annual time series observations.

2. Suppose you have quarterly data on new housing starts, interest rates, and real per capita income. Specify a model for housing starts that accounts for possible trends and seasonality in the variables.

3. Use the data in FERTIL3 for this exercise:

   (i) Regress gfr on t and t^2 and save the residuals. This gives a detrended gfr_t, say $\ddot{g}_t$.

   (ii) Regress $\ddot{g}_t$ on all of the variables in equation (10.35) of Wooldridge, including t and t^2. Compare the R-squared with that from (10.35). What do you conclude?

   (iii) Reestimate equation (10.35) but add t^3 to the equation. Is this additional term statistically significant?

4. Use the data in VOLAT for this exercise. The variable rsp500 is the monthly return on the Standard & Poor's 500 stock market index, at an annual rate (this includes price changes as well as dividends). The variable $i3$ is the return on three-month US Treasury bills, and $pcip$ is the percentage change in US industrial production (these are also at an annual rate).

   (i) Consider the equation

   $$rsp500_t = \beta_0 + \beta_1 pcip_t + \beta_2 i3_t + u_t.$$

   What signs do you think β_1 and β_2 should have?

(ii) Estimate the previous equation by OLS, reporting the results in standard form. Interpret the signs and magnitudes of the coefficients.

(iii) Which of the variables is statistically significant?

(iv) Does your finding from part (iii) imply that the return on the S & P 500 is predictable? Explain.

5. Use the data in APPROVAL to answer the following questions. The data set consists of 78 months of data during the presidency of George W. Bush. (The data end in July 2007, before Bush left office.) In addition to economic variables and binary indicators of various events, it includes an approval rate, *approve*, collected by Gallup.

(i) What is the range of the variable *approve*? What is its average value?

(ii) Estimate the model

$$approve_t = \beta_0 + \beta_1 lcpifood + \beta_2 lrgasprice_t + \beta_3 unemploy_t + u_t,$$

where the regressors are the log of CPI for food, the log of the real gasoline price, and the unemployment rate.

(iii) Interpret the coefficients in the estimates from part (ii). Comment on the signs and sizes of the effects, as well as statistical significance.

(iv) Add the binary variables X11.Sep (a dummy to indicate the three months during and after the September 11 terrorist attacks) and iraqinvade (a dummy to indicate the three months after the announcement to invade Iraq) to the equation from part (ii). Interpret the coefficients on the dummy variables. Are they statistically significant?

(v) Does adding the dummy variables in part (iv) change the other estimates much? Are any of the coefficients in part (iv) hard to rationalize?

(vi) Add lsp500 to the regression in part (iv). Controlling for other factors, does the stock market have an important effect on the presidential approval rating?