Problem Set 1

Econometrics II

Due: Tuesday, February 4, 2025

Instructions

- PS1 will be submitted by mail to djsanchez@colmex.mx with CC to egameren@colmex.mx and fernando.garcia@colmex.mx.
- Your submission should include an R-script and a LaTeX compiled PDF document.
  - The R-script should be divided into subsections, each corresponding to a problem in the PS.
  - The PDF document should include answers to both theoretical and computer exercises. The document should be compiled in LaTeX, no other forms of submission will be allowed (I recommend using Overleaf).
- Computer excercises require the use of companion datasets to your textbook. You can access these datasets by installing the R-package **wooldridge** following the procedure seen in Lab Session 1.
- Any required visualization of data should be directly included in the PDF document. When the problem asks to include the output of a regression, you should include a table with those results (check R-package **stargazer**).
- Problems are based in *Introductory Econometrics: A Modern Approach, 4th Edition* by Jeffrey Wooldridge.

1 Theoretical Excercises

1.1 W10.1

Let $gGDP_t$ denote the annual percentage change in gross domestic product and let int_t denote a short-term interest rate. Suppose that $gGDP_t$ is related to interest rates by:

$$\alpha_0 + \delta_0 \cdot int_t + \delta_1 \cdot int_{t-1} + u_t$$

Where u_t is uncorrelated with int_t, int_{t-1}, and all other past values of interest rates. Suppose that the Federal Reserve follows the policy rule:

$$int_t = \gamma_0 + \gamma_1(gGDP_{t-1} - 3) + vt$$

where $\gamma_1 > 0$. (When last year's GDP growth is above 3%, the Fed increases interest rates to prevent an "overheated" economy.) If v_t is uncorrelated with all past values of int_t and u_{t-1}, argue that int_t must be correlated with u_{t-1}. Which Gauss-Markov assumption does this violate?

1.2 W10.8

Consider the linear model described in equation 10.8 of the textbook. The explanatory variables $\mathbf{x}_t = (x_{t,1}, ..., x_{t,k})$ are said to be sequentially exogenous (sometimes called weakly exogenous) if:

$$E[u_t|\mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_1] = 0, t = 1, 2, ...$$

So that the errors are unpredictable given current and all past values of the explanatory variables.

1. Proof that sequential exogeneity is implied by strict exogeneity.
2. Explain why contemporaneous exogeneity is implied by sequential exogeneity.
3. Are the OLS estimators generally unbiased under the sequential exogeneity assumption? Explain.
4. Consider a model to explain the annual rate of HIV infections as a distributed lag of per capita condom usage for a state, region, or province. Explain why this model satisfies the sequential exogeneity assumption. Does it seem likely that strict exogeneity holds too?

$E(HIVrate \mid pccon, pccon_{t-1}, \ldots) = \alpha_0 + \delta_0 pccon + \delta_1 pccon_{t-1} + \delta_2 pccon_{t-2} + \delta_3 pccon_{t-3}$

2 Computer Excercises

2.1 WC10.2

Consider Equation 10.22 of the textbook, based on *C.M. Krupp and P.S. Pollard (1999), Market Responses to Antidumpting Laws: Some Evidence from the U.S. Chemical Industry, Canadian Journal of Economics 29, 199-227.* Use the dataset **barium**.

1. Add a linear time trend to equation (10.22). Are any variables, other than the trend, statistically significant?

2. In the equation estimated in part (1), test for joint significance of all variables except the time trend. What do you conclude? (Hint: Estimate the F-statistic by computing the R squared of the restricted and unrestricted model).

3. Add monthly dummy variables to this equation and test for seasonality. Does including the monthly dummies change any other estimates or their standard errors in important ways?

2.2 WC10.5

Use the dataset **ezanders** for this exercise, based on *L.E. Papke (1994), "Tax Policy and Urban Development: Evidence from the Indiana Enterprise Zone Program," Journal of Public Economics 54, 37-49.* Observations are on monthly unemployment claims in Anderson Township in Indiana, from January 1980 through November 1988. In 1984, an enterprise zone (EZ) was located in Anderson (as well as other cities in Indiana).

1. Regress $\log(uclms)$ on a linear time trend and 11 monthly dummy variables.What was the overall trend in unemployment claims over this period? (Interpret the coefficient on the time trend.) Is there evidence of seasonality in unemployment claims?

2. Add ez, a dummy variable equal to 1 in the months Anderson had an EZ, to the regression in part (1). Does having the enterprise zone seem to decrease unemployment claims? By how much? [You should use formula (7.10) from Chapter 7.]

3. What assumptions do you need to make to attribute the effect in part (ii) to the creation of an EZ?

2.3 WC10.6

Consider Equation 10.35 of the textbook. Use the dataset **fertil3** for this excercise, based on *L.A. Whittington, J. Alm, and H.E. Peters (1990), "Fertility and the Personal Exemption: Implicit Pronatalist Policy in the United States," American Economic Review 80, 545- 556.*

1. Regress gfr_t on t and t^2 and save the residuals. This gives a de-trended gfr_t, say, $\widehat{gfr}_t$.

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

2.4 K1

The *Indice de Precios y Cotizaciones* is a weighted index of 35 stocks traded in the Mexican Stock Exchange[1], an equivalent of the American S&P500 but tropicalized. IPC.RData (attached in this mail, loaded in R with **load()**) contains monthly observations on the levels of the IPC.

1. Obtain the monthly rates of change of the IPC by estimating log-differences as seen on Lab 2.

2. Plot the levels and distribution of these rates of change. Comment on their distribution and run the Jarque-Bera Test.

3. Run the following models[2]:

   - $gIPC_t$ on t (a linear time trend).
   - $gIPC_t$ on IPC_{t-1} (the lagged value).
   - $gIPC_t$ on t and IPC_{t-1} (both).

4. Include 11 monthly dummy variables (look for command **case_when()** on R). Is there any evidence of seasonality on this time series? Which frequencies of data might require seasonal adjustments?

[1] As a fun fact, the three largest components are *América Movil*, *Walmex* (Walmart Mexico) and *FEMSA* (oxxo and so on).

[2] $gIPC_t$ reffers to the monthly rates of change in IPC.