Problem set 6

Welcome

Our final problem set, covering chapters 10 and 12

See the exercises below

What do I submit?

- Your written up answers to exercise questions. If you work on a piece of paper, please scan using some sort of phone software (like Microsoft Lens or Adobe Scan) rather than just taking a picture.
- A do-file that runs your Stata analysis.
- A log file that includes the output from running your do-file.

Exercises

1. In 1985, neither Florida nor Georgia had laws banning open alcohol containers in vehicle passenger compartments. By 1990, Florida had passed such a law, but Georgia had not.

   a. Suppose you collect random samples of the driving-age population in both states, for 1985 and 1990. Let *arrest* be a binary variable equal to one if a person was arrested for drunk driving during the year. Without controlling for any other factors, write down a linear probability model that allows you to test whether the open container law reduced the probability of being arrested for drunk driving. Which coefficient measures the effect of the law?

   b. Why might you want to control for other factors in the model? What might some of these factors be?

   c. Now, suppose that you can only collect data for 1985 and for 1990 at the county level for the two states. The dependent variable would be the fraction of licensed drivers arrested for drunk driving during the

year. How does this data structure differ from the individual-level data described in part (a)? What econometric method would you use?

2. For this exercise, use `JTRAIN.dta` to determine the effect of a job training grant on hours of job training per employee. The basic model for the three years is the following:

$$hrsemp_{it} = \beta_0 + \delta_1 d88_t + \delta_2 d89_t +$$
$$\beta_1 grant_{it} + \beta_2 grant_{i,t-1} + \beta_3 log(employ_{it}) + a_i + u_{it}$$

a. Estimate the equation using first differencing. How many firms are used in the estimation? How many total observations would be used if each firm had data on all variables (in particular, $hrsemp$) for all three time periods?

b. Interpret the coefficient on $grant$, and comment on its significance.

c. Is it surprising that $grant_{-1}$ is insignificant? Explain.

d. Do larger firms train their employees more or less, on average? How big are the differences in training?

3. Use `CRIME4.dta` for this exercise, and see example 13.9 in this poor-quality scanned upload.

a. Replicate the results in Example 13.9.

b. Re-estimate the unobserved effects model for crime in Example 13.9, but use fixed effects rather than differencing. Are there any notable sign or magnitude changes in the coefficients? What about statistical significance?

c. Add the logs of each wage variable in the data set and estimate the model by fixed effects. How does including these variables affect the coefficient on the criminal justic variables in part (b)?

d. Do the wage variables in part (c) have the expected sign? Are they jointly significant?

4. **SW-12.6** In an instrumental variable regression model with one reressor, X_i, and one instrument, Z_i, the regression of X_i onto Z_i has $R^2 = 0.05$ and $n = 100$. Is Z_i a strong instrument?[1] Would your answer change if $R^2 = 0.05$ and $n = 500$?

5. **SW-12.9** A researcher is interested in the effect of military service on human capital. She collects data from a random sample of 4000 workers aged 40 and runs the OLS regression $Y_i = \beta_0 + \beta_1 X_i + u_i$, where Y_i is a worker's annual earnings and X_i is a binary variable equal to 1 if the person served in the military and is equal to 0 otherwise.

---

[1] *Hint:* See equation 7.14 in your textbook.

a. Explain why the OLS estimates are likely to be unreliable. (*Hint:* Which variables are omitted from the reression? Are they correlated with military service?)

b. During the Vietnam war there was a draft in which priority for the draft was determined by a national lottery. The days of the year were randomly re-ordered 1 through 365. (Those whose birthdays were ordered first were drafted before those with birthdates ordered second, and so forth.) Explain how the lottery might be used as an instrument to estimate the effect of military service on earnings. For more about this issue, see Joshua D Angrish's paper "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records," *American Economic Review*, June 1990: 313–336.

6. **SW-E12.2** Does viewing a violent movie lead to violent behavior? If so the incidence of violent crimes, such as assault, should rise following the release of a violent movie that attracts many viewers. Alternatively, movie viewing may substitute for other activities, such as alcohol consumption, that lead to violent behavior, so that assaults should fall more when more viewers are attracted to the cinema. Use the data file `Movies.dta`, which contains data on the number of assaults and movie attendance for 516 weekends from 1995 through 2004[2]. A detailed description is given here. The data set includes weekend US attendance for strongly violent movies (such as *Hannibal*), mildly violent movies (such as *Spiderman*), and non-violent movies (such as *Finding Nemo*). The data also includes the count of the number of assaults for the same weekend in a subset of counties in the United States. Finally, the data set includes indicators for year, month, whether the weekend is a holiday, and various measures of the weather.

a. Regress the logarithm of the number of assaults ($ln\_assaults = ln(assaults)$) on the year and month indicators. Is there evidence of seasonality in assaults? That is, do there tend to be more assaults in some months than others? Explain.

b. Now, regress total movie attendance ($attend = attend\_v + attend\_m + attend\_n$) on the year and month indicators. Is there evidence of seasonality in movie attendance? Explain.

c. Regress $ln\_assaults$ on $attend\_v$, $attend\_m$, $attend\_n$, the year and month indicators, and the weather and holiday control variables available in the data set.

   1. Based on the regression, does viewing a strongly violent movie increase or decrease assaults? By how much? Is the estimated affect statistically significant?

_____

[2]These are aggregated versions of data provided by Gordon Dahl and Stefano DellaVigna, used in their paper, "Does Movie Volence Increase Violent Crime?"

2. Does attendance at strongly violent movies affect us all differently than attendance at moderately violent movies? Differently than attendance at non-violent movies?
3. A strongly violent blockbuster movie is released and the weekends attendance is at strongly violent movies increases by 6 million; meanwhile, attendance falls by 2 million for moderately violent movies and by 1 million for non-violent movies. What is the predicted effect on assault? Construct a 95% confidence interval for the change in assault.[3]

d. It is difficult to control for all the variables that affect assaults and that might be correlated with movie attendance. For example, the effect of the weather on assaults and movie attendance is only crudely approximated by the weather variables in the data set. However, the data set does include a set of instruments *pr_attend_v*, *pr_attend_m*, and *pr_attend_n*, that are correlated with attendance but are (arguably) uncorrelated with weekend-specific factors such as the weather that affects both assaults add movie attendance. These instruments use historical attendance patterns, not information on a particular weekend, to predict a film's attendance in a given weekend. For example, if a film's attendance is high in the second week of its release, then this could be used to predict that attendance was also high in the first week of its release. The details of the construction of these instruments are available on in the Dahl and DellaVina paper. Run the regression from part c, including year, month, holiday, and weather controls, but now using the instruments for attendance. Use this regression to now re-answer the questions from part c: c(1)- c(3).

e. The intuition underlined the instruments in part 4 is that attendance in a given week is correlated with attendance and surrounding weeks. For each movie category, the data set includes attendance in surrounding weeks. Run the regression using the instruments *attend_v_f*, *attend_m_f*, *attend_n_f*, *attend_v_b*, *attend_m_b*, and *attend_n_b* instead of the instruments used in part d, then use this regression to answer part c: c(1)- c(3).

f. Based on your analysis, what do you conclude about the effects of violent movies on short-run violent behavior?

---

[3] *Hint:* Review section 7.3 and material surrounding equations 8.7 and 8.8