

## Homework 22: ECO220Y

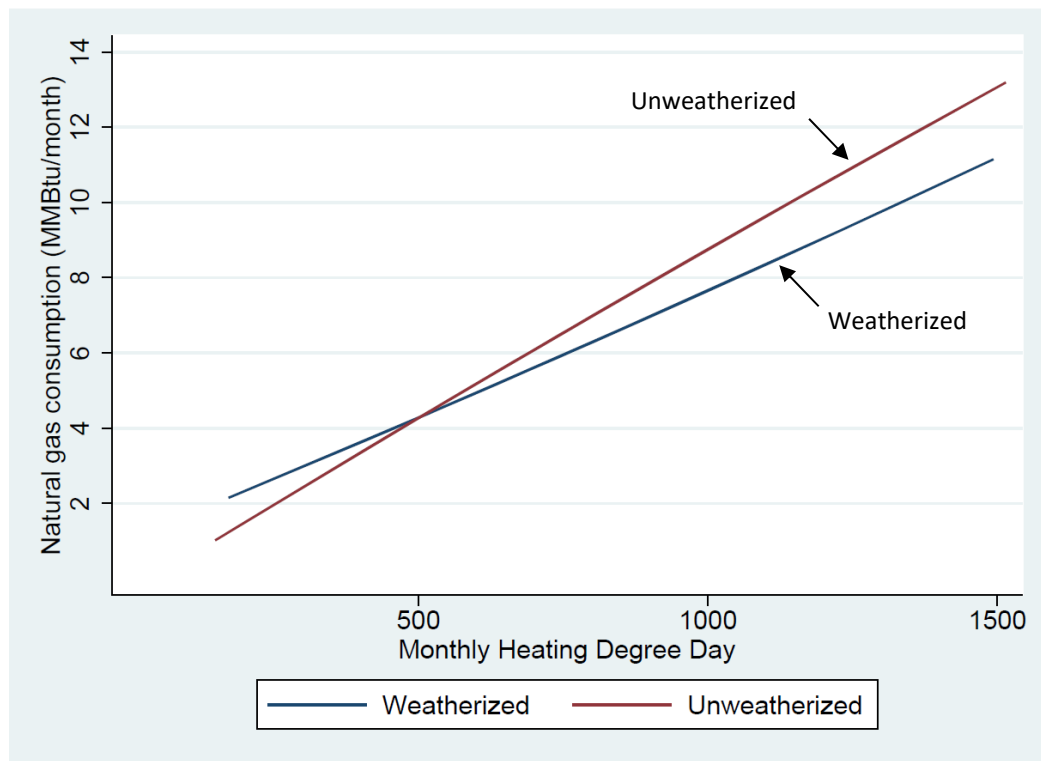
**Required Exercises:** Chapter 21: 1, 19

### Required Problems:

**(1)** Consider the NBER Working Paper “Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program” (<http://www.nber.org/papers/w21331.pdf>). An excerpt and Figure 4 are reproduced below. Note that “heating degree days (HDD)” is a way to measure how much energy is needed to heat a home to a comfortable temperature. For example, a home in Toronto would have a lower measure of HDD than a similar home in Québec City where the winter months are substantially colder. Also note that “weatherized” refers to taking steps to improve the energy efficiency of your home in winter months: for example, adding insulation, addressing drafty windows, etc.

**EXCERPT (p. 29):** Figure 4 summarizes the relationship between energy consumption (during winter months) and HDD separately for weatherized and unweatherized observations. The estimated slope of the relationship is less steep among weatherized homes. That is, weatherizations effectively reduce the marginal cost of indoor space heating during the winter.

Figure 4: Building energy performance at weatherized versus unweatherized homes



Notes: This figure plots the estimated relationship between monthly natural gas consumption and heating degree days during winter months at weatherized and unweatherized households, respectively. See Equation (3) in the text.

- (a)** Using reasonable approximation, what is the equation for the weatherized line? The unweatherized line?
- (b)** Making sure to define your variables, write down *one* model that captures the situation in Figure 4.
- (c)** Using Figure 4, which summarizes some estimation results, approximate the parameter *estimates* for the model you gave in Part (b). (Note: Figure 4 reports the *estimates* of the parameters, *not* the parameters themselves. While a model may be  $\hat{y} = \alpha + \beta x + \varepsilon$ , the estimates are  $a$  and  $b$ ;  $a$  and  $b$  are sample statistics.)
- (d)** Fully interpret the estimation results.

**(2)** A social scientist is interested in empirically testing the “Gateway Theory,” which posits that smoking marijuana as a teenager increases the chance of becoming a “hard drug” (cocaine, heroine etc.) user as an adult. Suppose you had a random sample of adults who report how often they currently use hard drugs (number of times per year) and whether or not they ever used marijuana as a teenager. What is the research question? Suggest a regression model approach to this problem. How would you test if you have a statistically significant result? What kind of results overall would you expect to obtain? Indicate any problems you would encounter in answering the research question.

**(3)** Suppose you wanted to model research output by faculty members at U of T. Consider measuring research output as: number of pages published in peer-reviewed academic journals. You suppose that research output may be effected by the amount of grant money a faculty member receives and the title of the faculty member. There are currently five different titles of regular full-time faculty members at U of T: Assistant Professor, Associate Professor, Full Professor, Assistant Professor, Teaching Stream, and Associate Professor, Teaching Stream.

**(a)** Write down a theoretical model where research output depends on grant money (in \$’s) and the title of the faculty member. Explain the interpretation of the parameters in the theoretical model you have written.

**(b)** Describe the endogeneity problem that will plague your attempts to estimate the parameter in the theoretical models you have written down. In other words, how the fact that the only available data will be observational will affect your regression results and their link to the theoretical model. Will you be able to answer the research question: How effective is grant money in promoting academic research as measured by pages published in academic peer-reviewed journals?

**(c)** Write down a theoretical model where research output depends on grant money (in \$’s) and title AND allow the impact of grant money to depend on title. For example, suppose you wanted to test the research hypothesis that effectiveness in translating grant money into research output depends on title. Explain the interpretation of the parameters in the theoretical model you have written. Explain how you would test the research hypothesis.

**(4)** Term Test #4 from March 2017 ([http://homes.chass.utoronto.ca/~murdockj/eco220/TT220\\_4\\_MAR17.pdf](http://homes.chass.utoronto.ca/~murdockj/eco220/TT220_4_MAR17.pdf)) has some excellent questions for you to work on that cover this week’s material.

**(a)** To practice the three methods of handling outliers discussed in lecture, answer Question (2), all parts.

**(b)** To make sure you can connect material from the last two weeks of classes, answer Question (3).

**(c)** To test your understanding of the “Waterloo 2016 Salary Disparities” case study required readings (available on portal), answer Question (4), all parts *except for* part (d), which is a topic (“Quadratic terms”) covered next week.

**(5)** Term Test #5 from April 2018 ([http://homes.chass.utoronto.ca/~murdockj/eco220/TT220\\_5\\_APR18.pdf](http://homes.chass.utoronto.ca/~murdockj/eco220/TT220_5_APR18.pdf)) has some excellent questions for you to work on that cover this week’s material.

**(a)** To practice working with dummy variables (and connecting it to previous week’s material), answer Question (1), all parts.

**(b)** To practice working with dummy variables and interaction terms, answer Question (2).

(6) Consider the 2018 *NBER Working Paper* “Platforms, Promotion, and Product Discovery: Evidence from Spotify Playlists” (<http://www.nber.org/papers/w24713>). The excerpt and the figure below describe one aspect of their data.

**Excerpt (p. 12):** Before turning to regression approaches, a simple look at some data is instructive. Figure 1 shows the evolution of playlist followers and U.S. daily streams for a song added to Today’s Top Hits during 2017. The song “What Ifs” by Kane Brown was added to the Today’s Top Hits playlist on October 5, 2017. On or about that date, the number of playlist followers for the song jumped from 11.6 to 29.2 million. The number of playlist followers then fluctuated about 30 million for about a month. On November 2, the song was removed from Today’s Top Hits, and its number of followers fell from 30.8 million to just 10.8 million. In subsequent months the number of followers continued to generally decline, sometimes rapidly as particular playlists removed the song.

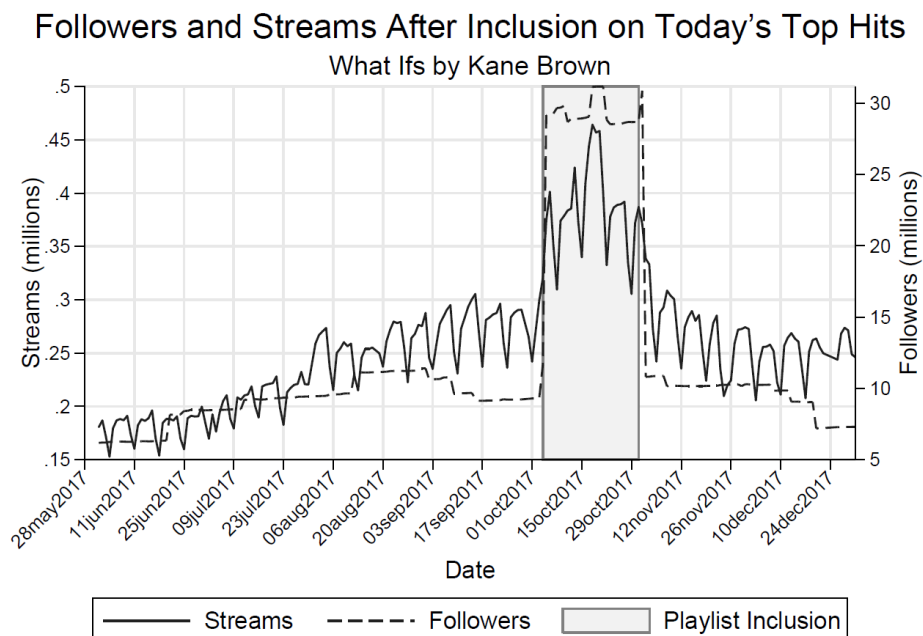


Figure 1: Daily Followers and US Streams for a Song added to Today’s Top Hits.

- (a) Focusing on streams in millions (not followers), describe how you would use *dummy variables* on the right hand side of the regression to help fit the pattern in the data shown in Figure 1 above. (Do not worry about non-dummy variables.) Write down your best idea for a regression model with dummy variables. (Hint #1: June 11, 2017 is a Sunday as is June 25, 2017 and July 9, 2017. Hint #2: Don’t forget about the area shaded grey.)
- (b) Given Figure 1, report the expected sign (positive or negative) for as many coefficients in your model in Part (a) as possible.