Graded Assignment 2

FIN 403: Econometrics

November 15, 2023

1 General information

- This is the second and last graded practical assignment this semester.
- It is due at **noon (12:00) on Tuesday, November 28** — no late submissions will be accepted.
- Please work in the same teams of up to three students as for the first assignment. (If you need to change your team, e.g. because other team members dropped the course, please let us know.) Do not share code or solutions with other students outside your team.
- Please submit through Moodle (one submission per team). Your submission should consist of ONE zip file containing (i) a concise write-up of your solutions ($\leq$ 3 pages of text; this limit does not include tables/figures, meaning your total submission can be more than 3 pages), and (ii) the code(s) used to generate the solutions. As for the first assignment, we strongly recommend to use Stata or R.
- Grading: there will be a total of 100 points, distributed as follows:
  - 70 points: Content of the write-up — whether the presented solutions & explanations are correct. Each of the 14 questions below accounts for 5 points.
  - 30 points: Presentation and coding style — whether output from your analysis is presented clearly and in **professional-looking figures and tables** (e.g. as discussed for the regression tables in Chapter 3 of the lectures) and whether the submitted code is well commented and easy to follow.

2 Task

The file fish.dta (or .csv) contains 97 daily price and quantity observations on fish prices at the Fulton Fish Market in New York City.[1] It contains the following variables:

avgprc	average price
lavgprc	log(avgprc)
totqty	total quantity sold
ltotqty	log(totqty)
speed2	min past 2 days wind speeds at sea
speed3	3 day lagged max wind speed
wave2	avg max last 2 days wave height at sea
wave3	avg max wave heights of 3 & 4 day lagged
mon	=1 if Monday
tues	=1 if Tuesday
wed	=1 if Wednesday
thurs	=1 if Thursday
t	time trend

[1] The data originally come from Graddy (1995); `http://www.jstor.org/stable/2556036`.

Part 1. For the first few questions, use the variable `lavgprc` (= the log of the average price for fish) as the dependent variable.

(a) Regress `lavgprc` on four daily dummy variables, with Friday as the omitted base. Include a linear time trend. Is there evidence that price varies systematically within a week?

(b) Now, add the independent variables `wave2` and `wave3`, which are measures of wave heights out at sea over the past several days. Are these variables individually and jointly statistically significant? Describe a mechanism by which stormy seas would increase the price of fish.

(c) What happened to the estimated coefficient on the time trend when `wave2` and `wave3` were added to the regression? What do you think is the explanation for this?

(d) Plot a time series of the residuals from the regression in (b). Does it look like they are serially correlated? Test formally for AR(1) serial correlation. What do you conclude about whether OLS is unbiased/consistent? And what about the usual OLS standard errors in this setting?

(e) Obtain the Newey-West standard errors using four lags. What happens to the t-statistics on `wave2` and `wave3`? Do you expect them to be more reliable or less reliable than the usual OLS t-statistics in this application?

(f) Now, obtain the Prais-Winsten estimates for the same model. Are `wave2` and `wave3` jointly statistically significant? What is the estimated magnitude of the autocorrelation in the error term?

Part 2. Now, consider the following demand equation:

$$\log(quantity)_t = \beta_1 + \beta_2 \log(price)_t + \beta_3 mon_t + \beta_4 tues_t + \beta_5 wed_t + \beta_6 thurs_t + \beta_7 t + \varepsilon_t$$

(g) Run this regression using OLS and report the results. Interpret the coefficient on log(price) in terms of its economic meaning. Why would we be concerned that the estimated coefficient on log(price) is not a consistent estimate of the causal effect of price on the demand for fish?

(h) Consider using `wave2`, the average over the past two days of the observed maximum wave height for the day, as an instrument for the price of fish. What conditions does it need to fulfill to be a valid instrument? Do you think these conditions are plausibly fulfilled?

(i) Run the first-stage regression using OLS:

$$\log(price)_t = \gamma_1 + \gamma_2 wave2_t + \gamma_3 mon_t + \gamma_4 tues_t + \gamma_5 wed_t + \gamma_6 thurs_t + \gamma_7 t + \varepsilon_t$$

Is `wave2` a weak or a strong instrument?

(j) Report the IV estimate of the effect of price on quantity using `wave2` as an instrumental variable. How does this compare to the OLS estimate of the association between price and quantity (is it the same, larger, smaller)? Provide an intuitive explanation of why your OLS estimate is different from your IV estimate. Furthermore, compare the magnitude of the standard errors and explain why the IV standard errors may be larger/smaller than the OLS ones.

(k) Consider using `speed3`, the maximum wind speed from three days earlier, as the instrument for price instead of `wave2`. Report the IV results and discuss based on the first-stage regression whether `speed3` is a strong instrument.

(l) Now use both `wave2` and `speed3` as instruments; report the IV results and test based on the first-stage regression whether they jointly fulfill the criterion for being sufficiently strong instruments.

(m) Test the overidentifying restriction in the model from the previous item. What do you conclude?

(n) Test for the endogeneity of log(price) based on the same model. What do you conclude?

<u>Bonus question (5 points)</u>

Figure out how to adjust the IV estimates of the model in item (l) for potential serial correlation in the error terms (which we ignored so far in part 2). (We did not cover this in class.)