**Group 23**

Ishaan Dey

Jedidiah Park

Angela Li

Austin Rule

**Lab 11**

4/26/19

ECON 3720 INTRODUCTORY ECONOMETRICS

**Prof. Ron Michener**

**University of Virginia**

The problem set deals with instrumental variables and simultaneous equations. The data set was gathered by Kathryn Graddy, who studied the Fulton Fish Market in New York and published her findings in a number of papers. The history, importance, and functioning of the market are described in her paper “The Fulton Fish Market,” The Journal of Economic Perspectives vol. 20, no. 2 (Spring 2006), pp. 207-220. The data she gathered were used in a coauthored article estimating the demand for Whiting (a particular kind of comparatively inexpensive fish used in fish cakes, etc.) using simultaneous equations. Joshua D. Angrist, Kathryn Graddy and Guido W. Imbens, “The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to The Demand for Fish,” The Review of Economic Studies vol. 67, no. 3 (July 2000), pp. 499-527.

Step I

**(7 points) The first of these articles on the Fulton Fish Market provides a brief non-technical description of how the market operated. In the 6th edition, Studenmund advised (Rule 3, p. 385) that one should be “intimately familiar with the subject being investigated – its history, institutions, operating constraints, measurement peculiarities, cultural customs, and so on . . .” I would like you to begin by looking at this article. If you access Virgo you will see a menu along the top bar. Click on “Databases.” Under the heading “Popular Links” you will find JSTOR. Click on JSTOR. You can focus your search best by using the “Advanced Search” option. Using the bibliographic information given above you should quickly be able to access a pdf of the *Journal of Economic Perspectives* article. According to this article, what are the factors that chiefly determine the quantity of fish supplied? What daily purchasing pattern was commonly observed among the customers who came to the market? Sp9b**

“Quantity supplied was primarily determined by weather conditions. Wind and waves are the greatest determinant of the quantity of fish likely to be caught.” (pg 211)

“Most buyers followed distinct purchasing patterns. Monday, Thursday and Friday were big days, but Tuesday and Wednesday were relatively quiet.” (pg 211)

Step II

**(3 points) The data set consists of 111 daily observations. Code the string variables as integer variables. Then summarize your data and include a copy of the summary in your report. The quantity and price data are for Whiting. Fit a “demand curve” to this data using ordinary least squares. Let the dependent variable be Lnqty, which is the natural log of quantity, and the right-hand-side variables be Lnprice, which is the natural log of price, and a set of day-of-the-week dummies. Use Friday as the left-out category. Include a copy of the results in your report. Sp9b**



Step III

**(6 points) Ignoring for the moment any concerns about simultaneous equations bias, how would you interpret the estimated coefficient you obtained in Step II for Lnprice? The estimated coefficient you obtained in Step II for the dummy variable for Wednesday? Be precise.**

We expect to see a -.563% shift in quantity demanded following a 1% increase in price, holding all else constant. We expect a -.559% shift in quantity demanded on Wednesdays relative to Fridays, holding all else constant.  **Sp9b**

Step IV

**(6 points) Would we expect the simultaneous equations bias present in Step II to have biased the coefficient of Lnprice in a negative or positive direction? Work out the direction of the bias from first principles, taking the discussion in Studenmund that occurs in conjunction with equations (14.2) and (14.3) as a model. You may assume the supply curve has a positive slope. Sp9b**

We would expect a positive (upwards) bias on the coefficient of Lnprice, as the error terms are correlated with the endogenous variable.

Step V

**(4 points) We wish to use two dummy variables based on sea\_conditions as instruments for Lnprice in estimating the demand equation. Although we could do so by using the instrumental variable command, we will illustrate the basic idea behind 2SLS first by using the two stage procedure to get coefficient estimates, as outlined by Studenmund on p. 423. Begin by fitting the first stage regression explaining the endogenous variable (price) with the exogenous and predetermined variables (there are no predetermined variables in this case). Include a copy in your report. Sp9b**



Step VI

**(6 points) Are the sea\_conditions dummy variables strongly correlated with Lnprice, or is there a danger that they are weak instruments? Compute an appropriate F-statistic and reach a conclusion based on that F-statistic. Include the output in your report. Hint: It is not the F-statistic testing overall significance on the Step V output. Sp9b**

****

The sea condition variables appear to be sufficiently associated with Lnprice. This is because the overall F-statistic (15.83) is greater than 10, so we can thus reject the null hypothesis that the instruments are weakly associated with Lnprice.

Step VII

**(6 points) Use the regression fit in Step V to predict Lnprice. Then repeat Step II using the predicted value of Lnprice in place of Lnprice. (That is, perform Stage Two as described on p. 423.) Include a copy of the output in your report. Sp9b**



Step VIII

**(7 points) Carefully interpret the estimated coefficient you obtained on the price variable in Step VII. Did the coefficient change in the direction you’d have expected based on your analysis in Step IV? Can you use the t-statistics, p-values, and confidence intervals obtained in the Step VII regression to do statistical inference? Explain your answer. Sp9b**

We expect a -.946% change in quantity consumed following a 1% change in price, holding all else constant. This coefficient is smaller relative to the initial OLS estimate (-.563%), indicating a positive bias on the OLS estimate.

The t-statistics, p-values, or confidence intervals obtained in Step VII cannot be used for statistical inference because the second stage in 2SLS uses fitted values of the endogenous variables, so the standard errors are larger and cannot accurately reflect the variance in the true values of the endogenous variable. Because the t-statistic, and thus p-value and CI, use the SEs, we cannot conduct statistical inference.

Step IX

**(6 points) Now estimate the demand function from Step II with STATA’s instrumental variables estimation routine using the two sea\_conditions dummy variables as instruments for Lnprice. Include the output in your report. Use the degrees of freedom adjustment Studenmund uses. Sp9b**



Step X

**(6 points) Compare the** ** associated with Lnprice in Step IX with the comparable number estimated in Step II. Which one is larger? How do you reconcile this with the alleged superiority of 2SLS? Sp9b**

The SE(BetaLnPrice) of OLS(.168) is smaller than in the 2SLS estimate (.353). This is reconciled by the fact that the 2SLS estimates are primarily used to reduce the bias and produce consistent estimates of the slope coefficient. The larger SEs are due to the use of the predicted values of the endogenous variable, as opposed to the precise values of the variable itself.

Step XI

**(8 points) Is the equation estimated in Step IX exactly identified or over identified? If over identified, test the overidentifying restrictions, interpret the results of the test, and include the output in your report. If the test had gone decisively in the other direction, what would such a result suggest? Sp9b**



The equation is overidentified because the number of instruments are greater than the number of endogenous variables.

Both p-values (.3844 and .3982) are greater than alpha of 0.05, indicating that we must accept the null hypothesis that the instrumental variables provide comparable results to the endogenous variable, thus suggesting that all of the instrumental variables are valid. If the test had gone in the other direction, this would indicate that at least one of the instrumental variables are invalid.

Step XII

**(8 points) Economic theory strongly suggests that Lnprice is endogenous in the demand schedule, but is there clear statistical evidence in this data set to confirm this? Do a Wu-Hausman test to see if we can reject the null that Lnprice is not endogenous. Include the result in your report, and interpret it. Is there clear statistical evidence in this data set to confirm that Lnprice is endogenous? Sp9b**



The Wu-Hausman test indicates that there is not sufficient evidence to suggest that Lnprice is not exogenous, as p(.2263) > alpha(.05), which means that Lnprice is not endogenous.

Step XIII

**(3 points) In his textbook *Econometrics: A Modern Introduction*, author Michael P. Murray estimates the demand curve using the Fulton Fish Market data in much the way that we have.[[1]](#footnote-1) He then goes on to suggest that the day-of-the-week dummy variables appear in the demand equation, but not the supply equation. Whether this is true or not is debatable because fishing boat captains, knowing that certain days exhibit slack demand, may spend less time fishing on those days. However, let’s follow Murray’s lead and see where this takes us. He wants to estimate the following supply function: Sp9b**

.

**He does so using 2SLS, with day-of-the-week dummies serving as instruments for the endogenous variable. For purposes of comparison, begin by estimating Murray’s supply curve using OLS and report the result in your output. Sp9b**



Step XIV

**(8 points) Estimate the first stage regression expressing the endogenous variableas a function of all exogenous and instrumental variables and include the result in your output. Run an appropriate F that will help you to judge whether the correlation between the day-of-the-week instruments and the endogenous variable is reassuringly strong. Include the output in your project. What is your conclusion? What problems result if the relationship between the instruments and the endogenous variable is weak?**

We observe a F-statistic of 4.36 which is below 10, thus suggesting that the association between day\_of\_week and Lnqty is weak. This creates the problem of not reducing the bias in the 2SLS slope coefficient estimates, making them unreliable.

 **Sp9b**

****

Step XV

**(8 points) Use the instrumental variables estimation routine in STATA to estimate the supply curve using 2SLS in the fashion Murray suggests and report the result in your report. Use the degrees of freedom adjustment Studenmund uses. Then continue by generating test statistics to test whether there is clear statistical evidence is endogenous and to test the over identifying restrictions. Include the post estimation results in your report. Sp9b**







Step XVI

**(8 points) What do you make of the computer output created in Step XV? Does the coefficient of  seem “better” when estimated by 2SLS? Explain. Interpret the results of your post estimation tests. What do they suggest about the endogenity of and the validity of your instruments? What would seem to be the chief concern we have uncovered about the estimated supply function? Sp9b**

The coefficient of Lnqty does appear to be better, in that the estimate of the coefficient is positive. However, the large standard error makes the 95% CI contain negative values, thus indicating that the 2SLS was not sufficiently better after all.

The postestimation test for endogeneity suggests that we must fail to reject the null that Lnqty is exogenous, indicating that Lnqty is not sufficiently endogenous.

The postestimation test for overidentification suggests that we must fail to reject the null hypothesis that the instrument variables (being day of the week) are not valid. The chief concern here would be that the estimated supply function is that the slope coefficient estimates from the 2SLS regression would be biased and inconsistent.

1. He includes a handful of additional variables that prove to be statistically insignificant. Because their absence has only a slight effect on our results I have, for simplicity, omitted them. [↑](#footnote-ref-1)