Assignment 2

Basic Econometrics Fall 2022

Andreas Dzemski

Due **Sep 15, 18:00h**.
Only hand-ins following the submission guidelines will be accepted.
Starred (*) problems will not be graded.

Problem 6

For this problem we will be using the data set "airtravel00.dta" which contains data on airline routes. Each observation corresponds to a domestic airline route (identified by origin and destination) operated by at least one US airline in the year 2000. Note that — unlike in other examples — the units in the population are not people. We have data on the following variables

`passen` average number of passengers,

`fare` average fare,

`dist` distance between origin and destination (in miles),

`route` unique route identifier.

Compute robust standard errors in all regressions.

1. Consider the following model of demand on a given airline route

$$\log(\texttt{passen}) = \beta_0 + \beta_1 \log(\texttt{fare}) + U. \tag{1}$$

   a) Interpret the coefficient β_1.
   b) Estimate the model.

2. Our application here is an example of demand estimation in a *many markets* framework. What is a market in our application?

3. In the model above, we do not control explicitly for the distance between the airport of origin and the destination. Suppose that model (1) can be rewritten as

$$\log(\texttt{passen}) = \beta_0 + \beta_1 \log(\texttt{fare}) + \beta_2 \log(\texttt{dist}) + V, \tag{2}$$

where the demand shock V satisfies the exogeneity assumption

$$E[V \mid \log(\texttt{fare}), \log(\texttt{dist})] = E[V \mid \texttt{fare}, \texttt{dist}] = 0. \tag{3}$$

a) Rewrite U from equation (1) as a function of `dist` and V.

b) Use this representation of U to state the exogeneity condition for OLS estimation of model (1).

c) Is it plausible that this exogeneity assumption is satisfied? Explain.

4. We will now try to sign the bias that we face if the true model is (2) and we estimate β_1 by fitting (1) by OLS. You may assume that we have a "large sample".

   a) Adapt the formula for omitted variable bias to this scenario.

   b) What sign do you expect for β_2?

   c) What sign do you expect for the covariance between $\log(\texttt{fare})$ and $\log(\texttt{dist})$?

   d) What sign do you expect for the omitted variable bias?

5. We have data on `dist` and can estimate (2). Compare the the OLS estimates for the effect of $\log(\texttt{fare})$ from the two models. Is your conjecture about the sign of the omitted variable bias confirmed?

6. Verify that the slope coefficient from fitting model (1) by OLS is equal to

$$\hat{\beta}_1 + \hat{\delta}_{\log(\texttt{dist})|\log(\texttt{fare})} \hat{\beta}_2,$$

   where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimates obtained by fitting model (2) and

$$\hat{\delta}_{\log(\texttt{dist})|\log(\texttt{fare})} = \frac{\widehat{\text{cov}}(\log(\texttt{dist}), \log(\texttt{fare}))}{\widehat{\text{var}}(\log(\texttt{fare}))}$$

   is the slope coefficient from a regression of $\log(\texttt{dist})$ on $\log(\texttt{fare})$.

7. (*) The standard error for the effect of $\log(\texttt{fare})$ is smaller in regression (1) then in regression (2). Assume homoscedastic errors and explain what part of the formula for the standard error of an estimated slope coefficient is responsible for this.

Problem 7

For this problem we continue to use the data from Problem 6.

1. Generate new variables `lpassen`, `lfare` and `ldist` for logged `passen`, `fare` and `dist`, respectively (if you haven't done so already).

2. Fares for air travel are determined in a competitive market by the laws of supply and demand. Explain intuitively why this renders the exogeneity assumption (3) implausible.

3. The data set “airconc00.dta” contains for each route the variable `bmktshr` which gives the market share of the airline with the largest market share on this route. This is a measure of the concentration of market power on a given route.

   a) Explain in intuitive terms (drawing from your knowledge of economic theory) how `bmktshr` influences the supply side of the market.

   b) Suppose that one carrier stops providing service on a route so that `bmktshr` for this route increases. What happens to the supply curve for this route? What happens to the ticket price? Is the price change informative about the demand shock?

4. We want to import the variable `bmktshr` into our dataset.

   a) Type

```
help merge
```

   to learn about the syntax of the stata command `merge`.

   b) Run

```
merge 1:1 route using airconc00.dta
```

   and explain the meaning of this command.

   c) How does your dataset change?

   d) Why do you have to specify “route”?

5. We want to estimate the structural equation (2) by instrumental variable regression.

   a) State the instrument relevance condition for using `bmtshr` as an instrument for `lfare`. *Hint: The structural equation contains a control variable!*

   b) Do you expect it to be satisfied? Explain.

   c) Run the first stage and use the regression output to conduct a statistical test of the instrument relevance condition ($\alpha = 0.05$).

   d) Now we want to develop a hypothetical setting in which the instrument relevance assumption is *not* satisfied and in which we still have

$$\mathrm{cov}(\texttt{bmtshr}, \log(\texttt{fare})) \neq 0.$$

   Assume that passengers are willing to spend more money the longer the distance travelled. What would the joint distribution of $\log(\texttt{dist})$ and `bmtshr` have to look like to ensure that that

$$\mathrm{cov}(\texttt{bmtshr}, \log(\texttt{fare})) < 0.$$

6. a) State the appropriate instrument exogeneity assumption.

   b) Do you expect it to be satisfied? Explain.

7. Instruments like market share that affect the supply side of the market are called "supply shifters". Supply shifters are often used to estimate demand parameters. Explain briefly in intuitive terms
   a) why a supply shifter is a relevant instrument,
   b) under what conditions a supply shifter satisfies the instrument exogeneity assumption.
8. The Stata command `ivregress` computes the IV estimates $\hat{\beta}_1^{\text{iv}}(\omega_0)$ and $\hat{\beta}_2^{\text{iv}}(\omega_0)$. The output of `ivregress` also reports *correct* standard errors for the estimated coefficients. The Stata command to fit the IV regression is

```
ivregress 2sls lpassen ldist (lfare = bmktshr), robust
```

   a) Run the IV regression and compare the estimated coefficient $\hat{\beta}_1^{\text{iv}}(\omega_0)$ to the OLS estimate $\hat{\beta}_1^{\text{ols}}(\omega_0)$ that we obtained by fitting (2) by OLS.
   b) Give an econometric explanation for the different estimates.
9. Now we want to compute the IV estimate $\hat{\beta}_1^{\text{iv}}(\omega_0)$ by implementing the two-stage least squares (2SLS) protocol.
   a) Run the first-stage regression and store the fitted values from this regression in a new variable `lfare_hat`. (*Hint*: use the command `predict` with the option `xb`).
   b) Run the second-stage regression and verify that you get the same estimates as you get by using the `ivregress` command.
   c) Compare the standard errors reported at the second stage to the standard errors computed by `ivregress`. Explain.

What have you learned?

Econometric skills: After working through this problem set, you can ✎ ☒

- ☐ discuss possible endogeneity in a demand model
- ☐ make an educated guess about the bias from omitting a relevant variable
- ☐ assess the validty of an instrumental variable
- ☐ interpret an IV regression.

Programming skills: After working through this problem set, you can ✎ ☒

- ☐ merge data from different files into one dataset
- ☐ use the command `ivregress` to run an IV regression
- ☐ use the command `predict` to store predicted values from an OLS regression.