

Problem Set 5

Empirical Problem - Duranton and Turner

Urban Economics, Spring 2022

Setup

In this problem set, we will explore the Duranton and Turner paper. Some of the questions require R, and all require a written response. For the empirical questions, you should use the file `dtData.RData`, which is available on Canvas.

Note: To compute “robust” standard errors for question 2 you can use the command `“coefTest(regRes, vcov = vcovHC(regRes, “HC1”))”` (where `regRes` is the object containing regression results) after installing the `lmtest` and `sandwich` libraries and calling them using the `library(lmtest)` and `library(sandwich)` commands.

Question 1: Understanding the Paper's Model

1. Given the theory illustrated in Figure 1 of the paper,
 - (a) What is the predicted effect of increasing the number of roads in an area (hint: there are 2 effects)?

As noted, there are two predicted effects. The first is that increasing the number of roads leads to a lower average cost of driving per VKT. The second is that increasing the number of roads increases the equilibrium number of cars.
 - (b) Given that theory, what is the predicted sign of ρ_R^Q in equation (2)?

According to the theory, the sign should be positive. An increase in lanes should lead to an increase in VKT.
2. The following questions refer to Equation 2:
 - (a) What are the authors' concerns about estimating this equation by OLS?

The authors state: "In practice, we hope that roads will be assigned to growing cities and fear that they are assigned to prop-up declining cities. In either case, the required orthogonality condition fails." (The orthogonality condition is that unexplained fluctuations in log VKT are not correlated with lane kilometers of road.) While population can be measured, there are other characteristics (political attempts to provide roads to cities that don't need it--"declining cities"--and expected future growth--"growing cities") that would not be observable to the econometrician.
 - (b) What are the three strategies they use to overcome this (these) problem(s) and how does each of these strategies deal with the authors' concern(s)? [Hint: Look at the tables in the paper and the accompanying discussion. Be concise in your response.]
 - i. The first strategy is to estimate the model via fixed effects. What this does is eliminate time-invariant city characteristics.

- ii. The second strategy is to estimate the model using first differences (of lane kilometers), including *initial* city characteristics as controls. Initial VKT is also included. It should be noted that this also only controls for time-invariant city characteristics.
- iii. Employ instrumental variables (IV) techniques to get quasi-experimental variation in right-hand-side variables thought to pose an endogeneity concern (lane kilometers built). These instruments must be uncorrelated with miles traveled except through their effect on the number of kilometers (conditioning on other variables on the right-hand side).

Question 2: Replicating and Interpreting Main Results

Preliminary notes on variable names (remember that all variables are at the MSA level):

Variable name	Variable description
vmt_IH_YR	Mean daily VKT for year YR (interstates)
ln_km_IH_YR	Mean lane km for year YR (interstates)
popYR	Population for year YR (in MSA)

In this question, you will need to *replicate* results. In this concept, replicate means to produce code that prints a results table on the screen when I run it. Note that you need to replicate *both the standard errors and the point estimates* (except as noted). Also, note that the authors exclude from these tables any MSA for which they do not have data on KM of highway in 1983, 1993, or 2003. (A missing data point for the mean lane km variable is one where the data table has a zero for that MSA for that year).

1. Table 1: Generate a data frame/matrix with the average across MSAs for the three variables listed above (for all years) [you do not need standard deviations]. [Hint: You can use either the lapply or summary functions]
2. Table 3: Replicate the results from columns (1), (4), (7). [Hint: Use the commands from the lmtest package mentioned above]
 - In your own words, what is the qualitative message from these regressions about the elasticity of VKT with respect to lane km?
For a 1% increase in roads, there is an increase in VKT of less than 1%.
 - What can we learn from here about the change in this relationship over time?
This elasticity has been decreasing over time.
3. Table 5: Replicate column 2 in panel A, except only use data from 1993-2003 in the regression.
 - What is your point estimate for the elasticity?
0.86
 - How does this differ from the one reported in the paper?
The estimate in the paper is higher at 1.05.
 - Can we learn anything from this difference?
We learn that the elasticity has been decreasing over time (also reflected in the paper).

Question 3: What is a Valid Instrument?

1. In Table 6, for which endogenous variables do the authors instrument?

They instrument for lane km of interstate highways.

2. What are the variables that they use as instruments?

They use planned highways in 1947, 1898 railroad route km, and the incidence of major explorations between 1835 and 1850.

3. Explain briefly why each of these is a valid instrument.

The authors discuss the two criteria for a valid instrument on page 2629. These two criteria are relevance and exclusion. The first is that instrument is correlated with the endogenous variable (this can be tested). The second is that the instrument does not affect the left-hand-side variable except through its effect on the endogenous right-hand-side variable. This cannot be tested, and must be established by the authors.

Because relevance can be tested, we will not dwell on it. The "first stage" regression can be used to establish relevance. The 1947 plan was used as the starting point for interstate highways, the railroad routes and explorations were often used planning and constructing highway routes.

In terms of exclusion, the authors make several arguments: The 1947 plan was largely based on military and not civilian needs, and is thus unlikely to be correlated with civilian demand for VKT. On the other hand, the rail network was built for an economy that was very different at the time, and thus uncorrelated with present-day VKT demand. There is a similar argument for explorer routes.

As noted above (and by the authors), instrumental variables only requires *conditional* exogeneity of the instruments. While the 1947 plan, the railroads in 1898, and the explorer routes may indeed be correlated with city characteristics, these will be effectively controlled for using other right-hand-side variables.