Problem Set #3: Private vs. Public Infrastructure

Note: This assignment is longer than the previous ones; please budget your time accordingly.

In this assignment you will examine the proposition that public capital has “spillover” to private output. The assumption that such spillovers exist underlies arguments in favor of programs proposed to increase public capital (bridges, highways, railroads, and the “information superhighway”). The idea is that public capital should increase public output directly (more government office buildings mean more working space for public employees who can then produce more) but also increase private output indirectly. A larger (or more modern) public infrastructure might increase the productivity of labor and private capital, thus increasing private output. In other words, if spillovers exist, we should find a positive, statistically significant relationship between private output (the dependent variable) and public capital stock (the independent variable). In this assignment you will investigate the extent to which such spillovers exist.

You will be using a data set that has both cross-sectional (across states) and time-series (within states, over time) aspects, and you will take advantage of both of these aspects. In particular, by using both the cross-sectional and time-series data you will be able to draw some conclusions about the extent to which one or the other approach alone might be misleading. In addition, past studies using only time-series data for the U.S. have been criticized for concluding that spillovers exist when in fact the correlation between public capital and total output may be induced by an omitted additional factor which varies over time.

The data consists of private output, private labor input, the stock of private capital, and the stock of public capital for each of the 48 contiguous states in the U.S. for each year from 1969-1986. Using data on private output rather than aggregate U.S. output allows us to isolate the spillover effect of public capital in increasing private output (clearly, public capital will increase public output, but the impetus for an infrastructure plan is to take advantage of the spillovers to private output). The data are contained in a Stata data set called infrast.dta and are described at the end of this assignment.

Economic theory suggests using a production function for this analysis. In particular, with private labor, the stock of private capital, and the stock of public capital as inputs to production, we can suppose that each state has a production function (aggregate production for the state, or gross state product (GSP)) for each time period in the following form:

\[
GSP_{it} = \beta_0(L_{it})^{\beta_1}(K_{pub,it})^{\beta_2}(K_{pri,it})^{\beta_3} \epsilon_{it}
\]

where \(GSP_{it}\) is private output in state \(i\) at time \(t\), \(L_{it}\) is private labor input, \(K_{pri,it}\) is the stock of private capital, and \(K_{pub,it}\) is the stock of public capital.
**Regression & Analysis**

1. Think about the production function:

   a) Rewrite the production function for state i at time t in a form in which linear regression analysis can be used (i.e., the answer is an equation). What are you assuming about the error term εi? (Hint: Just saying it is normal is not the right answer.) In your new regression equation, how should you interpret the β3 coefficient?

   b) Generate natural logged versions of the dependent and independent variables. Name these variables lngsp, lnemp, lnkpub, and lnkpri. Create three scatterplot graphs comparing lngsp to lnemp, lnkpub, and lnkpri. What do you expect the regression to find, given these graphs?

   c) Using the full sample of all states in all periods, test the null hypothesis of constant returns to scale from private labor and private capital. Constant returns to scale is when a 1% change in inputs results in a 1% change in outputs. In this instance, constant returns to scale from private labor and private capital is when β1 + β3 = 1. Test whether this is true without using the built-in Stata command by running two separate regressions, calculating the F statistic, and using the table in Pindyck and Rubinfeld. (You may, of course, check your answer using the Stata command.)

   d) Using the full sample of all states in all periods, test the null hypothesis that public and private capital have the same marginal contribution to output. (For an additional half point, provide the answer without using the built-in Stata command by running two separate regressions.)

2. Using data from 1985 only (i.e., a cross-section of the states in a single year), run a regression of ln(gsp) on a constant, ln(emp), and ln(kpub). Then run the same regression, but also include ln(kpri) as an explanatory variable. After you run this second regression, run the `hettest`.

   a) Interpret the different results on the coefficient of ln(kpub) in the two regressions. How can you provide evidence for your interpretation? What evidence do you find?

   b) Test the null hypothesis of no spillovers of public capital to private output; show how you would calculate the p value if Stata did not report it (using a t test). Which equation is more appropriate and why?

   c) Does `hettest` find evidence of heteroskedasticity in the second regression? What other method could you use to find heteroskedasticity (just describe what you would do; you don’t need to actually do it).

   d) Run the regression of ln(gsp) on a constant, ln(emp), ln(kpri), and ln(kpub) using the robust standard error computation (i.e., by adding “, robust” at the end of the regular “reg” statement). Compare the “robust” standard errors with the regression run without the robust option. What information does this give you about the validity of the homoskedasticity assumption for this subsample?

3. Create a new data set that contains only the observations for Pennsylvania. Next, tell Stata which variable (in this case, “year”) contains the time index using the “`tsset`” command (i.e., type in the command “`tsset year`”). Then run a regression of ln(gsp) on a constant, ln(emp), ln(kpri), and ln(kpub). Then type the command “`dwsat`” to get the Durbin-Watson statistic. (For more
information on this command, see the *Stata Reference Manual* pages for “regress” and “prais,” which are available through ERes.)

a) Is there evidence of public capital spillovers to private output?

b) Test the null hypothesis of no serial correlation of the error against the alternative of positive first-order serial correlation.

c) Correct for first-order serial correlation using the Cochrane-Orcutt method. This method is used within another regression command: “prais.” This command has the form: `prais <dependent> <independents>, corc`. (For more information on this command, see the *Stata Reference Manual* pages for “prais” that are available through ERes.) Explain what effect this procedure should have on the estimated coefficients and standard errors of these coefficients. What actually does happen to the coefficients and their standard errors when you make this adjustment?

4. Explain the difference between the cross-sectional results on spillovers and the time-series results (you may assume for this purpose that all states behaved as Pennsylvania did, rather than estimating the spillover for each state). *Hint:* Plot ln(gsp) against ln(kpub) for California, Pennsylvania, New Jersey, South Carolina, and Vermont together (these states show a disparity in level of public capital). Based on this plot, what would a cross-section regression tell you about a spillover? What would a time-series regression for any individual state tell you?

5. Now try estimating the results using a fixed effects model and a random effects model.

   a) First, estimate the fixed effects model using the following regression command:

   ```
xtre glngsp lnemp lnkpri lnkpub, fe i(state)
   ```

   This regression controls for fixed state effects. Does it appear that there are constant returns to scale from private capital and labor in this model? Does public capital have any effect on state output?

   b) Now generate a dummy variable for each year in the model; name each one d1969, d1970, etc. Then run the following fixed effects regression:

   ```
   ```

   This regression controls for fixed state and fixed year effects. How do the coefficients from a) compare to the findings here? Does it appear that there are constant returns to scale from private capital and labor in this model? Does public capital have any effect on state output?

   c) Finally, run a random effects model using the following regression command:

   ```
xtre glngsp lnemp lnkpri lnkpub, re i(state)
   ```

   Random effects control for a mixture of fixed year and state effects. How do the findings from this command compare to the two previous regressions?
d) Of the models run in a), b), and c), which do you find most persuasive? Give a brief explanation. (You might wish to review pages 650-652 in Gujarati to answer this question. This section is available from ERes.)

Extra Credit: Analytic Problem
(Worth 1 additional point.)

Constant returns to scale of labor and private capital in production suggests that doubling labor and doubling private capital (holding all else constant) will double output. In terms of the regression coefficients, constant returns to scale would imply that $\beta_2 + \beta_4 = 1$. Show that $\beta_2 + \beta_4 = 1$ yields constant returns to scale in private labor and private capital.

Data Description

INFRAST.dta

The data set consists of annual output, labor, private capital, and state and local government capital for the 48 contiguous states in the United States over the years 1969-1986. Output is measured by the U.S. Department of Commerce, Bureau of Economic Analysis (BEA) state-by-state series on Gross State Product of private industries. Estimates of private sector capital are from A. Munnell (ed.), Is There a Shortfall in Public Capital Investment?, published by the Federal Reserve Bank of Boston in 1990. Labor input is the BEA measure of full-time and part-time wage and salary employees in private industries. Public sector capital data are the sum of state government capital and local government capital, by state, for all government functions. The public sector capital stocks are estimated using the perpetual inventory method. Benchmark capital stocks were imputed to each state, and constrained to sum to equal the BEA estimate of aggregate capital in 1960.

The data are organized by year and then by state. Thus observation 1 is Alabama in 1969, observation 2 is Arizona in 1969, observation 49 is Alabama in 1970, etc.

Variable Definitions

- year: Year
- state: Numerical value denoting state (1-48, in alphabetical order. See correspondence below)
- gsp: Gross state product (current dollars, private sector)
- emp: Employment
- kpri: Private capital stock (current dollars)
- kpub: Public (state and local government) capital stock (current dollars)
Correspondence between “state” variable and state name

1: Alabama 25: Nebraska
2: Arizona 26: Nevada
3: Arkansas 27: New Hampshire
4: California 28: New Jersey
5: Colorado 28: New Mexico
7: Delaware 31: North Carolina
8: Florida 32: North Dakota
9: Georgia 33: Ohio
10: Idaho 34: Oklahoma
11: Illinois 35: Oregon
12: Indiana 36: Pennsylvania
13: Iowa 37: Rhode Island
14: Kansas 38: South Carolina
15: Kentucky 39: South Dakota
16: Louisiana 40: Tennessee
17: Maine 41: Texas
18: Maryland 42: Utah
19: Massachusetts 43: Vermont
20: Michigan 44: Virginia
21: Minnesota 45: Washington
22: Mississippi 46: West Virginia
23: Missouri 47: Wisconsin
24: Montana 48: Wyoming