

ECON 8686 Problem Set #2- Instrumental Variables

This problem set is based on:

Black, Dan, Terra McKinnish and Seth Sanders. 2003. "Does the availability of high-wage jobs for low-skilled men affect welfare expenditures? Evidence from shocks to the steel and coal industries." *Journal of Public Economics* 87:1921-1942.

At least since the publication of Charles Murray's *Losing Ground* (1984), many commentators and policy makers have argued that the United States' welfare system has created a "culture of poverty," encouraging labor force detachment and marital instability. Women, it is argued, become dependent on transfer payments and remain on welfare even given opportunities to work or marry. Many commentators point to rising levels of welfare receipt, joblessness, and out-of-wedlock childbearing as evidence of the harm done by welfare.

In contrast, William Julius Wilson (1980, 1987, and 1996) offers a different explanation for the rise of such "underclass" behaviors. He writes in *When Work Disappears*, "Many of today's problems in the inner-city ghetto neighborhoods—crime, family dissolution, welfare, low levels of social organization, and so on—are a consequence of the disappearance of work" (p.xiii). Wilson argues that pervasive welfare receipt occurs in the absence of manufacturing jobs, which have traditionally provided relatively high-wage jobs to low-skilled male workers. Wilson and Neckerman (1986) argue the decline of manufacturing made it difficult for low-skilled men to support a family. As men became less attractive marriage partners, marriage rates decreased, and out-of-wedlock child bearing and welfare receipt increased. Implicit in this argument is the belief that if better economic opportunities were available for low-skilled men, welfare participation rates would decline.

In this paper we study the effect of a boom and bust in the coal industry during the 1970s and 1980s on welfare receipt and family structure in the coal-producing regions of rural Appalachia. During the 1970s, regulatory changes and the Organization of Petroleum Exporting Countries (OPEC) oil embargo drove up the price of coal and generated an enormous boom in localities with coal reserves. The coal boom reversed into a bust in the 1980's. These exogenous shocks to coal industry caused the creation and destruction of high-paying jobs for low-skilled male workers, the exact sort of economic events central to Wilson's hypothesis.

Our analysis of the coal economy focuses on the states of Kentucky, Ohio, Pennsylvania, and West Virginia. The Appalachian regions of these states have traditionally been poor and dependent on the coal economy. Some counties have substantial coal reserves and benefited from the boom in the coal industry and suffered from the bust. More than half of the counties in this region, however, have virtually no coal and were not directly affected by the coal boom and bust.

The data set coalIV.dta contains data on all counties in the 4 states of Kentucky, Ohio, Pennsylvania, and West Virginia for the years 1969-93. The variables are:

year

state: (1=Kentucky, 2=Ohio, 3=Pennsylvania, 4=West Virginia)

county: (first digit is state code, remaining 3 digits are a county code that is unique within state)

earn: total real earnings in the county

afdc: total real AFDC expenditures in the county

pop: county population

coalprice: real price of coal (varies by year)

coalres: total coal reserves in county (in millions of tons)

cat: (0=coalres<100 million, 1= 100<coalres<1000, 2=coalres>1000)

t: (1=1969-77, 2=1978-83, 3=1984-93)

1) Plot price of coal over time to confirm a big spike in coal prices in the 1970's.

scatter coalprice year

or

line coalprice year if county==1001

2) Do some descriptive analysis using only the counties in Kentucky.

Keep if state==1

Calculate mean earnings by coal reserve category [light (cat=0), medium (cat=1), or heavy (cat=2)] for 1970, 1977 and 1993.

tabstat earn if year==1970, by(cat)

etc

Use these results to create a small table (can do this by hand) of the % change in earnings from 1970-77 and 1977-93 for light, medium and heavy coal counties in Kentucky.

3) Repeat the analysis in problem #2 for AFDC expenditures and add to your table.

Discuss the significance of the findings.

4) Return to the full data set. Restrict the sample to:

keep if cat<=1&t==1

and create the following variables:

gen lafdc=ln(afdc)

gen learn=ln(earn)

sort county

quietly by county: gen cafdc=lafdc-lafdc[_n-1]

quietly by county: gen cearn=learn-learn[_n-1]

and then calculate the following statistics:

tabstat cearn, by(cat)

tabstat cafdc, by(cat)

5)Estimate:

reg cearn cat

interpret the intercept and the coefficient, relating them back to the statistics in #4 above.

Then:

```
predict cearnhat
```

```
tabstat cearnhat, by(cat) stat(mean sd)
```

explain the values of *cearnhat*

6) Estimate:

```
reg cafdc cearnhat
```

Interpret the coefficient. Explain where the coefficient estimate comes from, relating as appropriate back to #4 and #5.

7) Return to the full sample, recreate *cafdc* and *cearn* as above. Estimate a basic first-differences model with state-year effects and controls for population size, clustering the standard errors by county.

```
gen lpop=ln(pop)
```

```
sort county
```

```
quietly by county: gen cpop=lpop-lpop[_n-1]
```

```
gen id=state*100+(year-1900)
```

```
tab id, gen(sty)
```

```
reg cafdc cearn cpop sty*, cluster(county)
```

a) Interpret the coefficient on *cearn*.

b) Do we need the population control? Under what conditions is it unnecessary and when would it be important?

8) Estimate first-stage regressions appropriate for the main regression model described in problem #4. For the instrument, interact logged coal reserves with the change in logged coal price: $cvalue = \log(\text{coalres} + 1) * \Delta \log(\text{coalprice})$, to measure the change in the value of coal in the county. Create and include a lag of this variable (as it can take time for a mine to open).

Estimate the appropriate first-stage regressions and calculate the partial F-stat on the instrument using:

```
testparm cvalue cvalueL1 or testparm cvalue*
```

after estimating the 1st stage regression.

Discuss the results.

9) Then:

```
predict cearnhat
```

```
summ cearn cearnhat if cearnhat!=.
```

a) Comment on the means and standard deviations.

b) Estimate the 2nd stage regression (don't worry about adjusting the standard errors). Interpret and discuss the results.

10) Now estimate the IV model using the same instruments:

```
ivreg cafdc cpop sty* (cearn=cvalue cvalueL1), cluster(county)
```

Compare the coefficient estimates to the results in problem #6. Compare the standard errors.

11) Re-estimate the IV model using the following subsets:

`ivreg cafdc cpop sty* (cearn=cvalue cvalueL1) if cat>0, cluster(county)`

`ivreg cafdc cpop sty* (cearn=cvalue cvalueL1) if cat<2, cluster(county)`

`ivreg cafdc cpop sty* (cearn=cvalue cvalueL1) if cat==2, cluster(county)`

Explain the difference in identification between these models and the one estimated in problem #7.