# Problem Set #9: Instrumental Variables (revised 10/17/2011)

Economics 435: Quantitative Methods

Fall 2011

## 1 Public sector unionization and the size of government: Part VI

This time we will use an instrumental variables approach to accounting for the possible endogeneity of unionization rates. Caroline Hoxby has a well-known 1996 article in the *Quarterly Journal of Economics* on the effect of teacher unionization on student outcomes (she finds a negative effect). In it, she uses state-by-state data on laws relevant to teacher unionization to construct instruments for the state's unionization rate.

We will adapt her approach to our public sector unionization hypothesis. The structural model we have in mind is:

$$per.capita.payroll_i = \beta_0 + \beta_1(\%membership_i) + u_i$$

We will not assume $E(u_i|\%membership_i) = 0$, which is necessary for OLS to yield consistent estimates of $\beta_1$.

We will follow Hoxby and use her data on state-by-state labour laws as instruments for public sector[1] unionization rates.

The data on state-by-state laws are from her paper, and I have put them on the course website in CSV format at:
`http://www.sfu.ca/~bkrauth/econ435/restricted/hoxby.csv`. You can merge this data set in with your 2003 unionization data to do the regressions.

The three variables in the data set are:

- $duty_i$: Equals one if teachers' unions or associations have the explicit right to meet with school administration, or the explicit right to engage in collective bargaining. Equals zero otherwise.

- $agency_i$: Equals one if teachers' unions can have "agency shops." An agency shop is one where the union can collect dues from nonmembers as well as members.

- $closed_i$: Equals one if teachers' unions can have "closed shops." A closed shop is one in which union membership is mandatory for employment.

We will use the vector $z_i \equiv (duty_i, agency_i, closed_i)$ as an instrument for $\%membership_i$ in an IV/2SLS regression. By doing so, we are assuming that $E(u_i|z_i) = 0$ (exogeneity) and that $E(\%membership_i|z_i)$ is not a constant (relevance).

**a)** Fill in Table 5 in the paper. Leave the row for the "first-stage F-statistic" blank; you don't need to do it. You will find the `ivreg` function in the R package `AER` useful in estimating 2SLS regression. Please be

---

[1]Note that Hoxby's data are on labour laws for public school teachers, and we are using them as an instrument for unionization rates for the state's entire public sector. In other words, these aren't really good instruments for our purpose; their primary advantage is that the data set is easy to get.

aware that `ivreg` uses the term "instruments" to refer to all exogenous variables (whether they are in the model or not). So your eplanatory variables are $(\%membership, logpopulation)$ and your instruments are $(logpopulation, duty, agency, closed)$.

**b**) Fill in the discussion of these results in Section 4.4.

# 2  An alternative to the IV estimator

Imagine that you are helping a friend - we'll call him Clarence - with his 435 paper. Clarence has a data set $\{(y_i, x_i, z_i)\}_{i=1}^{n}$ generated by a random sample from the joint distribution of $(y, x, z, u)$. The relationship among the variables is given by the structural model:

$$y = \beta_0 + \beta_1 x + u$$

where $\beta_1$ is the effect of $x$ on $y$, and $u$ is the effect of all other variables on $y$. We are not able to assume that $E(u|x) = 0$, but we are able to assume that $z$ is a valid instrument:

$$cov(u, z) = 0$$
$$cov(x, z) \neq 0$$

Finally, we assume that there is variation in both $x$ and $z$ in Clarence's sample.

Clarence didn't pay attention in class and has never heard of the IV or 2SLS estimator. Instead of using an IV approach, he settles on a "run every possible OLS regression" approach to his paper. He estimates three regressions by OLS: one regression of $y$ on $x$, one of $y$ on $z$, and one of $x$ on $z$. Let the coefficients from those regressions be defined as:

$$\hat{\gamma}_{yx} = \frac{c\hat{o}v(y, x)}{v\hat{a}r(x)}$$
$$\hat{\gamma}_{yz} = \frac{c\hat{o}v(y, z)}{v\hat{a}r(z)}$$
$$\hat{\gamma}_{xz} = \frac{c\hat{o}v(x, z)}{v\hat{a}r(z)}$$

$$(1)$$

He has brought you the results from these regressions. Unfortunately, he has accidentally deleted the source data and cannot estimate any new regressions.

**a**) Is there a consistent estimator of $\beta_1$ that can be constructed using $(\hat{\gamma}_{yx}, \hat{\gamma}_{yz}, \hat{\gamma}_{xz})$? If so, find it and prove that it is consistent.

**b**) How does your candidate estimator relate to the standard IV estimator?