

Keys to Lab 5

1 (a) The OLS estimation yields

$$\text{Log(wage)} = 5.97 + 0.061 \text{ edu} \\ (0.084) \quad (0.0062)$$

N=852, SSR=131.3, R2=0.11, F=100.6

(b) Using *sibs* as instrument for *edu*, IV estimation yields

$$\text{Log(wage)} = 5.07 + 0.127 \text{ edu} \\ (0.37) \quad (0.027)$$

N=852, SSR=149.1, Wald=21.5

The R2 is -0.0157. (Recall that in OLS, $R^2=1-SSR/SST$. In IV estimation, however, the equation $SST=SSR+SSE$ (or $\widehat{\text{var}}(\hat{y}) = \widehat{\text{var}}(\hat{x}'\hat{\beta}) + \widehat{\text{var}}(\hat{u})$) no longer holds, due to the covariance between x and u . In fact $\widehat{\text{var}}(\hat{u})$ may well exceed $\widehat{\text{var}}(\hat{y})$, which results in negative R2.)

(c) Birth order may be negatively correlated with *edu* since the younger ones may receive less education due to the limited budget of families, which may be allocated on “first come first served” basis. Indeed, if we regress *edu* on *brthord*, we obtain

$$\text{edu} = 14.1 - 0.28 \text{ brthord} \\ (0.13) \quad (0.042)$$

The coefficient on *brthord* is significantly negative.

(d) Using both *sibs* and *brthord* as instruments, TSLS yields,

$$\text{Log(wage)} = 5.06 + 0.129 \text{ edu} \\ (0.006) \quad (0.005)$$

N=852, SSR=149.8, Wald=74299.7

We can see that the IV and TSLS estimates are close and they differ markedly from the OLS estimates. The standard error of the TSLS estimator is much smaller than that of the IV estimator. This is reasonable since TSLS uses more information.

2 (a) Pooled OLS estimation yields

$$\text{mrdrte} = 0.348 + 0.165 \text{ exec} + 1.26 \text{ unem} \\ (2.32) \quad (0.11) \quad (0.49)$$

N=51, T=3, SSR = 12046, R2=0.06, F = 4.98

The sign on *unem* is reasonable, since higher unemployment rate may lead to higher murder rate. The positive sign on *exec* is counter-intuitive, since we may expect executions may deter murder. To understand the “wrong” sign, note that *exec* may be endogenous. States that have high murder rates (due to demographics, culture, history, etc.) tend to have more executions.

(b) To deal with the possible endogeneity problem, we estimate the fixed-effect panel data model and obtain

$$\text{mrdrte} = -0.11 \text{ exec} + 0.096 \text{ unem} \\ (0.062) \quad (0.22)$$

N=51, T=3, SSR = 1305.3, R²=0.90, F = 17.0

Now the coefficient on exec is negative and is statistically significant at 10% level. The fixed-effect panel data model allows the individual effects to be correlated with regressors (in this case, exec). Taking account of the possible endogeneity, we achieve a more reasonable estimate.