Assignment Project Exam Help

# Lecture 3 : Instrumental Variables

https://powcoder.com

Rigissa Megalokonomou
University of Queensland

Add WeChat powcoder

Assignment Project Exam Help

## Reading for lecture 3

- In the textbook (Wooldridge 2013): Chapter 15
- In Mostly Harmless Econometrics: Chapter 4

https://powcoder.com

Add WeChat powcoder

## Revision of Last' Lectures main points
## Our objective: Obtain causal effect of policy change

- In many economic models the **exogeneity assumption is violated**: that is at least one 'explanatory' variable is endogenous (explanatory variables are correlated with the error), or jointly determined with the 'dependent' variable. $y = \mathbf{x}\beta + u$, $E(u|\mathbf{x}) \neq 0$.
- There are four potential **sources of endogeneity** that we talked about last week:
  (i) omitted variables
  (ii) selection bias
  (iii) measurement error
  (iv) reverse causality or simultaneity
- In each of these four cases, the OLS is not capable of delivering unbiased/consistent parameter estimates.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Assignment Project Exam Help

## Review of IV

- RCT is one way to deal with endogeneity.
- Another way to deal with endogeneity is using **instrumental variables**.
- Typically, the point of IV is to allow causal inference in a non-experimental setting.

https://powcoder.com

Add WeChat powcoder

## This example AGAIN: Ability Bias in the Returns to Education

- Labour economists have been studying returns to education for a very long time.
- The correctly specified model that is analyzed is:

Assignment Project Exam Help

$$Y_i = a + \rho S_i + \gamma A_i + n_i (*)$$

- $Y_i$ = log of earnings.

https://powcoder.com

- $S_i$= schooling measured in years.
- $A_i$ = individual ability.
- Typically the econometrician cannot observe Ai .

Add WeChat powcoder

- Suppose you therefore estimate the short regression:

$$Y_i = a + \rho S_i + h_i$$

- where

$$h_i = \gamma A_i + n_i$$

## This example AGAIN: Ability Bias in the Returns to Education

- The OLS estimator for $\rho$ in this simple case is:

Assignment Project Exam Help

$$\widehat{\rho} = \frac{Cov(Y, S)}{Var(S)} \tag{1}$$

https://powcoder.com

- Plugging in the true model for Y

$$\widehat{\rho} = \rho + \gamma \frac{Cov(A, S)}{Var(S)} \tag{2}$$

Add WeChat powcoder

- This is the classic ability bias if $\gamma > 0$ and $Cov(A, S) > 0$, thus the coefficient on schooling in the short regression would be **upward biased**.

## How IV Can be Used to Obtain Unbiased and Consistent Estimates?

Assignment Project Exam Help

- How can we estimate the true $\rho$ if ability is unobserved?
- Use an instrument Z.
- 2 important conditions for a valid IV:

https://powcoder.com

- 1) $Cov(S_i, Z_i) \neq 0$ (1st stage exists). (hint: $S_i$ is the endogenous regressor)
- 2) $Cov(Z_i, h_i) = 0$ (exclusion restriction: Z is uncorrelated with any other determinants of the dependent variable).

Add WeChat powcoder

- While we can test whether the 1st condition is satisfied, the second condition cannot be formally tested. As a researcher you have to try to convince your audience that it is satisfied.

## First and Second Stages

- First Stage regression:

$$S_i = \alpha_1 + \rho_1 Z_i + \kappa_i$$

Assignment Project Exam Help

where $Z_i$ is the IV and $S_i$ is the endogenous variable

- Second Stage regression:

https://powcoder.com

$$Y_i = \alpha_2 + \rho S_i + h_i$$

Add WeChat powcoder

- Reduced Form equation:

$$Y_i = \alpha_3 + \rho_3 Z_i + \lambda_i$$

## How IV Can be Used to Obtain Consistent Estimates?

- With one endogenous variable and one instrument the IV estimator is:

$$\widehat{\rho}_{iv} = \frac{Cov(Y_i, Z_i)}{Cov(S_i, Z_i)} \tag{3}$$

Assignment Project Exam Help

- Hint: substitute the formula for Y from the correctly specified model (*)
- Substitute true model for Y:

https://powcoder.com

$$\widehat{\rho}_{iv} = \frac{Cov(a + \rho S_i + \gamma A_i + n_i], Z_i)}{Cov(S_i, Z_i)}$$

$$= \rho \frac{Cov(S_i, Z_i)}{Cov(S_i, Z_i)} + \gamma \frac{Cov(A_i, Z_i)}{Cov(S_i, Z_i)} + \frac{Cov(n_i, Z_i)}{Cov(S_i, Z_i)}$$

Add WeChat powcoder

- Taking plims:

$$plim \widehat{\rho}_{iv} = \rho$$

- The exclusion restriction says that $Cov(Z_i, h_i) = 0$, but $h_i = \gamma A_i + n_i$ so $Cov(Z_i, A_i)$ should be 0 and the $Cov(Z_i, n_i)$ should be equal to 0. $Cov(S_i, Z_i) \neq 0$ (due to the 1rst stage).
- **The IV estimator is consistent if the IV assumptions are satisfied.**

## Wald Estimator

- With one endogenous variable and one instrument, the IV estimator is:

$$\rho^{iv} = \frac{cov(Y_i, Z_i)}{cov(S_i, Z_i)} =>$$

Wald Estimator:

$$\rho^{iv} = \frac{cov(Y_i, Z_i)/Var(Z_i)}{cov(S_i, Z_i)/Var(Z_i)} (= \frac{\rho 3}{\rho 1})$$

- The coefficient of interest is the ratio of the population regression of $Y_i$ on $Z_i$ (reduced form) to the population regression of $S_i$ on $Z_i$ (first stage).
- If there is a first stage, the denominator is different than zero.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Where can you find a good instrument?

Assignment Project Exam Help

- Good instruments come from a combination of institutional knowledge and ideas about the process that determines the variable of interest.
- One source of variation is institutional constraints.

https://powcoder.com

  - For schooling this can be compulsory schooling laws. Angrist and Krueger (1991) exploit the variation induced by compulsory schooling in a paper that uses an IV coming from a ”natural experiment” to eliminate OVB.

Add WeChat powcoder

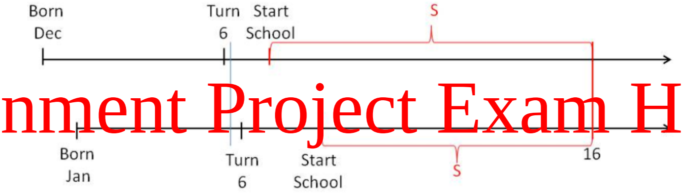
  - They would like to examine the effect of education on earnings. However, the naive regression has omitted variable bias.

## Instrument for Education using Compulsory Schooling Laws

- In practice it is often difficult to find convincing instruments
  - Many potential IVs do not satisfy the exclusion restriction.
- In the returns to education literature Angrist and Krueger (QJE, 1991) have a very influential study in which they looked at the effect of compulsory schooling requirement on schooling and earnings.
- This is done by exploring the fact that **school entry age** and **compulsory schooling laws** lead people born in different times of the year to have different average levels of education.
- In the US you could drop out of school once you turned 16.
- Typically schools require children to have turned 6 by January 1 of the year in which they enter school.
- Children have different ages when they start school and thus different lengths of schooling at the time they turn 16 when they can potentially drop out.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Instrument for Education using Compulsory Schooling Laws



- School start age is a function of date of birth. In particular, those born late in the year are young for their grades.
- In states with a December 31 birthday cutoff, children born in the 4th quarter enter school shortly before they turn 6, while those born in the first quarter enter school at around 6.5.
- So some stay more in schooling than others based on when they have to start school. Let's assume they drop out when they turn 16.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Is there a first stage?

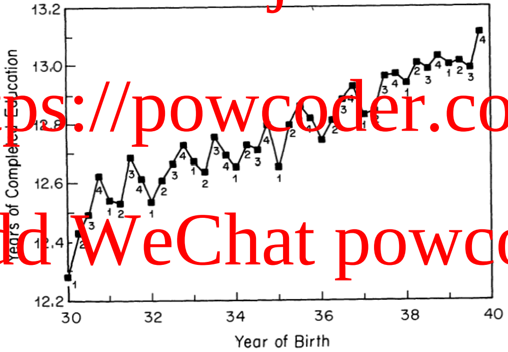
- First Stage Equation:

$$YearsofEducation = \alpha_1 + \rho_1 QuarterofBirth + \kappa_i$$

Assignment Project Exam Help

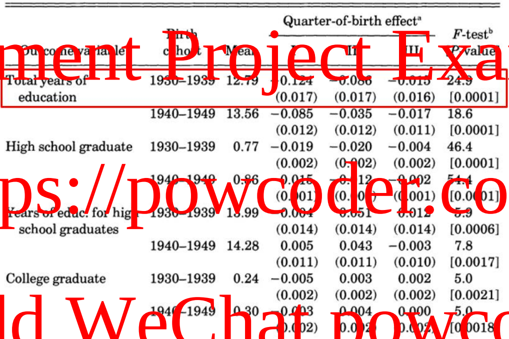
https://powcoder.com

Add WeChat powcoder



Men born earlier in the year have lower schooling. This indicates that there is a first stage.

## Is there a first stage?



| Outcome variable | Birth cohort | Mean | Quarter-of-birth effect[a] I | II | III | F-test[b] [P-value] |
|---|---|---|---|---|---|---|
| Total years of education | 1930–1939 | 12.79 | −0.124<br>(0.017) | −0.086<br>(0.017) | −0.015<br>(0.016) | 24.9<br>[0.0001] |
| | 1940–1949 | 13.56 | −0.085<br>(0.012) | −0.035<br>(0.012) | −0.017<br>(0.011) | 18.6<br>[0.0001] |
| High school graduate | 1930–1939 | 0.77 | −0.019<br>(0.002) | −0.020<br>(0.002) | −0.004<br>(0.002) | 46.4<br>[0.0001] |
| | 1940–1949 | 0.86 | −0.015<br>(0.001) | −0.012<br>(0.001) | −0.002<br>(0.001) | 54.4<br>[0.0001] |
| Years of educ. for high school graduates | 1930–1939 | 13.99 | −0.004<br>(0.014) | 0.051<br>(0.014) | 0.012<br>(0.014) | 5.9<br>[0.0006] |
| | 1940–1949 | 14.28 | 0.005<br>(0.011) | 0.043<br>(0.011) | −0.003<br>(0.010) | 7.8<br>[0.0017] |
| College graduate | 1930–1939 | 0.24 | −0.005<br>(0.002) | 0.003<br>(0.002) | 0.002<br>(0.002) | 5.0<br>[0.0021] |
| | 1940–1949 | 0.30 | −0.003<br>(0.002) | 0.004<br>(0.002) | 0.000<br>(0.002) | 5.0<br>[0.0018] |

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Completed years of schooling is lower for men born in the first quarter of the year than the forth quarter of the year. This indicates that there is a first stage.

**Do differences in schooling due to different quarter of birth translate into different earnings?**

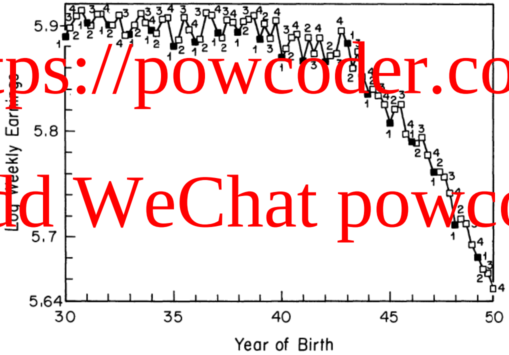
- Reduced Form Equation:

$$Earnings = \alpha_1 + \rho_3 QuarterofBirth + \kappa_i$$

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder



Younger cohorts tend to have lower earnings.

## Two Stage Least Square (2SLS)

- In practice one estimates IV as a Two-Stage-Least Squares estimator (2SLS).

Assignment Project Exam Help

- It is called 2SLS because you could estimate it as follows:
  1) Obtain the 1st stage fitted values:
- First Stage regression:

https://powcoder.com

$$\hat{S}_i = \hat{\alpha}_1 + \hat{\lambda} X_i + \hat{\rho}_1 Z_i + \kappa_i$$

where $\hat{\rho}_1$ and $\hat{\lambda}$ are OLS estimates of the 1st stage regression.
2) Plug the 1st stage fitted values into the ”second-stage equation”.

Add WeChat powcoder

$$Y_i = \alpha_2 + \beta X_i + \gamma \hat{S}_i + error$$

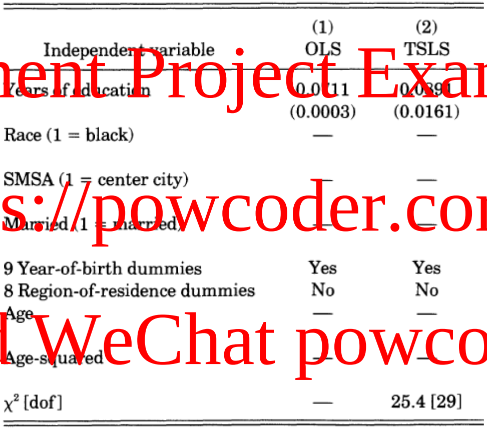
## Two Stage Least Square (2SLS)

- Despite its name the estimation is usually not done in two steps (because then the standard errors would be wrong).
- STATA or other regression softwares are usually doing the job for you (and get the standard errors right).
- The intuition of 2SLS, however, is very useful: 2SLS only keeps the variation in S that is generated by quasi-experimental variation (and thus hopefully exogenous).
- Angrist and Krueger use more than one instrument to instrument for schooling: they include a dummy for each quarter of birth.

$$Si = X_0\pi_{10} + \pi_{11}Z_{1i} + \pi_{12}Z_{2i} + \pi_{13}Z_{3i} + \psi_{1i}$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

# IV Estimates Birth Cohorts 20-29, 1980 Census

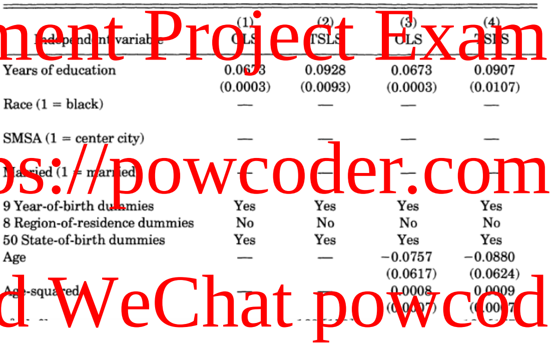


| Independent variable | (1)<br>OLS | (2)<br>TSLS |
|---|---|---|
| Years of education | 0.0711<br>(0.0003) | 0.0891<br>(0.0161) |
| Race (1 = black) | — | — |
| SMSA (1 = center city) | — | — |
| Married (1 = married) | — | — |
| 9 Year-of-birth dummies | Yes | Yes |
| 8 Region-of-residence dummies | No | No |
| Age | — | — |
| Age-squared | — | — |
| $\chi^2$ [dof] | — | 25.4 [29] |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Additional years of education increase earnings.

# Including More Covariates



| Independent variable | (1)<br>OLS | (2)<br>TSLS | (3)<br>OLS | (4)<br>TSLS |
|---|---|---|---|---|
| Years of education | 0.0673<br>(0.0003) | 0.0928<br>(0.0093) | 0.0673<br>(0.0003) | 0.0907<br>(0.0107) |
| Race (1 = black) | — | — | — | — |
| SMSA (1 = center city) | — | — | — | — |
| Married (1 = married) | — | — | — | — |
| 9 Year-of-birth dummies | Yes | Yes | Yes | Yes |
| 8 Region-of-residence dummies | No | No | No | No |
| 50 State-of-birth dummies | Yes | Yes | Yes | Yes |
| Age | — | — | −0.0757<br>(0.0617) | −0.0880<br>(0.0624) |
| Age-squared | — | — | 0.0008<br>(0.0007) | 0.0009<br>(0.0007) |

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Including more covariates should not affect much the coefficient of interest.

## Including more Instruments

Assignment Project Exam Help

- They also included specifications where they use 30 (quarter of birth x year) dummies and 150 (quarter of birth x state) dummies as IVs (intuition: the effect of quarter of birth may vary by birth year or state).

https://powcoder.com

- This reduces standard errors.
- But also comes at the cost of potentially having a weak instrument problem.

Add WeChat powcoder

## Weak Instruments

Assignment Project Exam Help

- As pointed out by Bound, Jaeger, and Baker (1993, 1995), the ”Cure can be worse than the disease” when the instruments are only weakly correlated with the endogenous variables.

https://powcoder.com

- Staiger and Stock (1997) formalized the definition of ”weak instruments”, and for the case of one IV and one endogenous variable if the **F-statistic** on the instruments in the **first stage** is **greater than 10**, one need worry no further about weak instruments. This is a rule-of-thumb. (Stock, Wright, and Yogo (2002))

Add WeChat powcoder

## Weak Instruments

- In the early 1960s a number of papers highlighted that IV estimates can be severely biased. In particular, if the IV is weak (first stage relationship is weak) and if you use many instruments to instrument for one endogenous variable.
- Adding more weak instruments will increase the bias of 2SLS. **By adding further instruments without predictive power the first stage F-statistic goes towards 0 and the bias will increase.**

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Weak Instruments in Angrist and Krueger

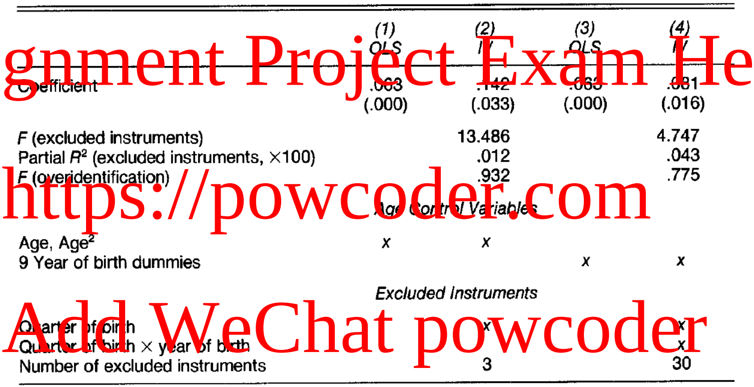
Assignment Project Exam Help

- Angrist and Krueger present findings while using different sets of instruments:
- 1) quarter of birth dummies, 3 instruments.
- 2) quarter of birth + (quarter of birth) x (year of birth) dummies, 30 instruments.
- 3) quarter of birth + (quarter of birth) x (year of birth) + (quarter of birth) x (state of birth), 180 instruments.

https://powcoder.com

Add WeChat powcoder

# Adding Instruments in Angrist and Krueger



|  | *(1)*<br>*OLS* | *(2)*<br>*IV* | *(3)*<br>*OLS* | *(4)*<br>*IV* |
|---|---|---|---|---|
| Coefficient | .063<br>(.000) | .142<br>(.033) | .063<br>(.000) | .081<br>(.016) |
| *F* (excluded instruments) |  | 13.486 |  | 4.747 |
| Partial $R^2$ (excluded instruments, ×100) |  | .012 |  | .043 |
| *F* (overidentification) |  | .932 |  | .775 |
|  | *Age Control Variables* |  |  |  |
| Age, Age$^2$ | x | x |  |  |
| 9 Year of birth dummies |  |  | x | x |
|  | *Excluded Instruments* |  |  |  |
| Quarter of birth |  | x |  | x |
| Quarter of birth × year of birth |  |  |  | x |
| Number of excluded instruments |  | 3 |  | 30 |

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

**Adding more weak instruments reduced the 1rst stage F-statistic and moves the coefficient towards the OLS coefficient.**

## Another Example: Angrist (1990) Veteran Draft Lottery

- In the following, we will often refer to an example from Angrist's paper on the effects of military service on earnings.
- Consider the following equation:

$$Y_i = \beta + \gamma S_i + h$$

  where $Y_i$ are **earnings** and $S_i$ is a **military/veteran dummy**.
- If one wants to interpret $\gamma$ as being the direct effect of having served the armed forces, OLS will be biased.
- Why? Many omitted factors might be correlated with one's willingness to serve the military, and these factors are likely to be correlated with earnings.
- If we want to estimate the causal effect of Si on Yi, **we need to find an instrument**.
- This instrument should be correlated with military/veteran status, but it should not affect earning in any other way (only through the effect on veteran status).
- Why is this important? Debate about whether veterans are adequately compensated for their service.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Angrist (1990) Veteran Draft Lottery

- Angrist (1990) uses the **Vietnam draft lottery** as in **IV for military service**.
- In the 1960s and early 1970s, young American men were drafted for military service to serve in Vietnam. However, there were **concerns about the fairness of the conscription policy** (certain types of men used to serve the army). This leads to the introduction of a **draft lottery** in 1970.
- From 1970 to 1972 **random** sequence numbers were assigned to each birth date in cohorts of 19-year-olds.
- Men with lottery numbers below a cutoff were drafted while men with numbers above the cutoff could not be drafted.
- *Nevertheless, the draft did not perfectly determinate military service:*
  - Many draft-eligible men were exempted for health and other reasons.
  - Exempted men volunteered for service.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- First stage results:
- Having a low lottery number (being eligible for the draft) increases veteran status by about 16 percentage points (the mean of veteran status is about 27 percent).
- Second stage results: Serving in the army lowers earnings by between 2,050 and 2,741 dollars per year, even long after the service in Vietnam was ended.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- Up to this point we only considered models where the causal effect was the same for all individuals (**homogenous treatment effects**):
- Be aware that the treatment effects might be heterogeneous for each population group.
- This will inform us about two types of validity characterizing research designs:
  1 Internal validity: Does the design successfully uncover causal effects for the population studied?
  2 External validity: Do the study's results inform us about different populations?

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- Different subpopulations:
- 1) **Compliers**: treated because of being selected. All the rest are **non-compliers**.
- 2) **Always-takers**: They always take the treatment independently of the draft.
- 3) **Never-takers**: They never take the treatment independently of the draft.
- 4) **Defiers**: Non-treated when selected and treated when non-selected.
- Under some assumptions, IV estimates focus on the average effect of military service on earnings for the sub-population who enrolled in military service, because of the draft but would not have served otherwise.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Card (1995) Use geographical variation in college proximity to estimate the returns to schooling

Assignment Project Exam Help

- Estimating the returns to schooling using the following regression:

$$lwage = \beta_0 + \beta_1 educ + u$$

https://powcoder.com

- educ is endogenous and E(educ,u) is not 0.
- We need to find an Instrument Zi, that is highly correlated with educ
- Thus, the naive OLS estimate is biased.

Add WeChat powcoder

## Card (1995) paper

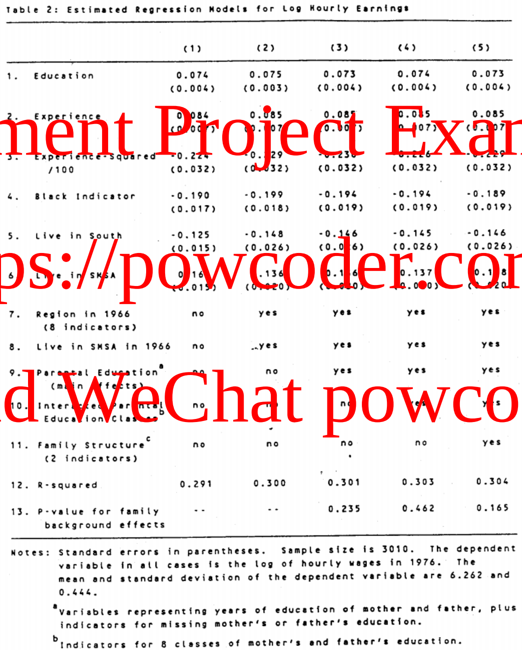


Table 2: Estimated Regression Models for Log Hourly Earnings

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| 1. Education | 0.074<br>(0.004) | 0.075<br>(0.003) | 0.073<br>(0.004) | 0.074<br>(0.004) | 0.073<br>(0.004) |
| 2. Experience | 0.084<br>(0.007) | 0.085<br>(0.007) | 0.085<br>(0.007) | 0.085<br>(0.007) | 0.085<br>(0.007) |
| 3. Experience-Squared /100 | -0.224<br>(0.032) | -0.229<br>(0.032) | -0.230<br>(0.032) | -0.226<br>(0.032) | -0.229<br>(0.032) |
| 4. Black Indicator | -0.190<br>(0.017) | -0.199<br>(0.018) | -0.194<br>(0.019) | -0.194<br>(0.019) | -0.189<br>(0.019) |
| 5. Live in South | -0.125<br>(0.015) | -0.148<br>(0.026) | -0.146<br>(0.026) | -0.145<br>(0.026) | -0.146<br>(0.026) |
| 6. Live in SMSA | 0.16[illegible]<br>(0.015) | 0.136<br>(0.020) | 0.1[illegible]6<br>(0.020) | 0.137<br>(0.020) | 0.1[illegible]8<br>(0.020) |
| 7. Region in 1966 (8 indicators) | no | yes | yes | yes | yes |
| 8. Live in SMSA in 1966 | no | yes | yes | yes | yes |
| 9. Parental Education[a] (main effects) | no | no | yes | yes | yes |
| 10. Interacted Parental Education Classes[b] | no | no | no | yes | yes |
| 11. Family Structure[c] (2 indicators) | no | no | no | no | yes |
| 12. R-squared | 0.291 | 0.300 | 0.301 | 0.303 | 0.304 |
| 13. P-value for family background effects | -- | -- | 0.235 | 0.462 | 0.165 |

Notes: Standard errors in parentheses. Sample size is 3010. The dependent variable in all cases is the log of hourly wages in 1976. The mean and standard deviation of the dependent variable are 6.262 and 0.444.

[a] Variables representing years of education of mother and father, plus indicators for missing mother's or father's education.

[b] Indicators for 8 classes of mother's and father's education.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Card (1995) paper

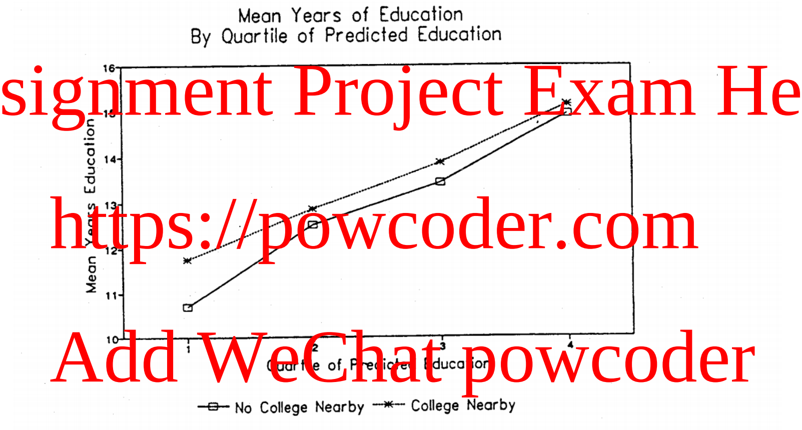
- Card (1990) used the **proximity to a four-year college** ($nearc4_i$) in high school as an IV.
  - One who lives in an area without a college faces a higher cost of college education, since the option of living at home is precluded (i.e., they will have to move). One would expect this to reduce investment in higher education, at least for low income families.
- So $nearc4_i$ is a binary indicator, equal to 1 if a man is near a four-year college in high school.
- Would expect $x_i = educ_i$ and $z_i = nearc4_i$ to be positively related. They have to be for the first stage to exist.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

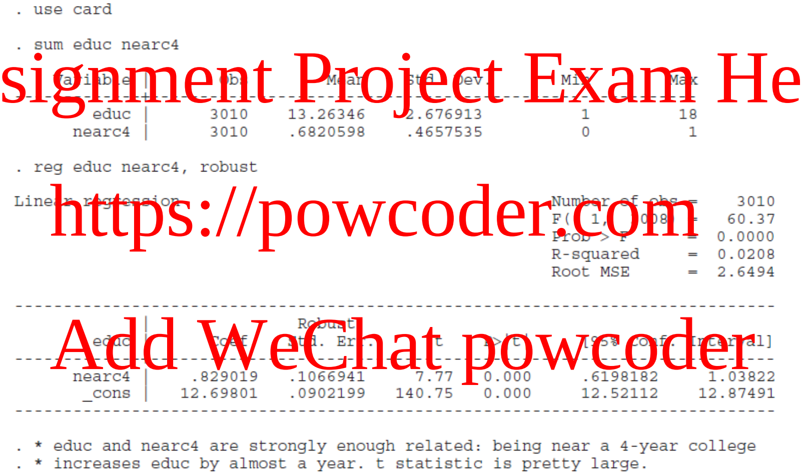
## Always plot the First Stage!



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

For every quartile the mean level of education is higher for those who grew up near a college.

## Run regressions to see if there is a First Stage!

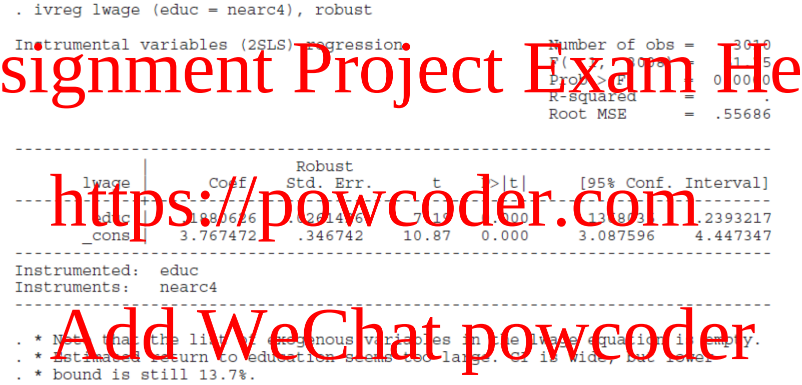


```
. use card

. sum educ nearc4

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        educ |      3010    13.26346    2.676913          1         18
      nearc4 |      3010    .6820598    .4657535          0          1

. reg educ nearc4, robust

Linear regression                                      Number of obs =    3010
                                                       F(  1,  3008) =   60.37
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0208
                                                       Root MSE      =  2.6494

------------------------------------------------------------------------------
             |               Robust
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      nearc4 |    .829019   .1066941     7.77   0.000     .6198182     1.03822
       _cons |   12.69801   .0902199   140.75   0.000     12.52112    12.87491
------------------------------------------------------------------------------

. * educ and nearc4 are strongly enough related: being near a 4-year college
. * increases educ by almost a year. t statistic is pretty large.
```

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder



```
. ivreg lwage (educ = nearc4), robust

Instrumental variables (2SLS) regression               Number of obs =    3010
                                                       F(  1,  3008) =   51.75
                                                       Prob > F      =  0.0000
                                                       R-squared     =       .
                                                       Root MSE      =  .55686

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1880626   .0261426     7.19   0.000     .1368035    .2393217
       _cons |   3.767472    .346742    10.87   0.000     3.087596    4.447347
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   nearc4
------------------------------------------------------------------------------

. * Note that the list of exogenous variables in the lwage equation is empty.
. * Estimated return to education seems too large. CI is wide, but lower
. * bound is still 13.7%.
```

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

```
. * For comparison, OLS:

. reg lwage educ, robust

Linear regression                                      Number of obs =    3010
                                                       F(  1,  3008) =  321.16
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0987
                                                       Root MSE      =  .42139

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0520942   .0029069    17.92   0.000     .0463946    .0577939
       _cons |   5.570882   .0390935   142.50   0.000      5.49423    5.647535
------------------------------------------------------------------------------

. * 18.8% versus 5.2%
```

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

- **Card(1995) paper**
  - Card in this paper reveals that men who grew up in local labor markets with a nearby college have significantly higher education and earnings than other men.
  - These gains are concentrated among men with poorly educated parents, who would otherwise stop schooling at low levels.
  - Effects may vary by family background.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- **Why are OLS and IV some cases so different in the Card case?**
    - A common explanation is that *educ* is **measured with error**, so there is attenuation bias (the estimate is biased towards 0) when using OLS.
    - Another explanation is that the **return to schooling is not constant** and IV is picking up the effect for a certain subgroup.
    - Another explanation is the possibility that **the instrument is somewhat endogenous**. (i.e. $Corr(distancetocollege, u) \neq 0$)
        - Families that place a strong emphasis on education might choose to live near a college.
        - The presence of a college might be associated with higher school quality at nearby schools (they could control for school quality.)
        - College proximity may be correlated with unobserved geographic wage premiums.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Two stage least squares in Stata

Assignment Project Exam Help

- (In Stata) ivregress 2sls depvar exogvars (endogvars = insts)

where exogvars is the list of exogenous regressors; endogvars is the list of endogenous variables (there can be more than one); and insts is the list of instruments.

https://powcoder.com

- ivregress 2sls depvar exogvars (endogvars = insts), **first**
- The ‘first’ option yields estimates of the 1st-stage equation.

Add WeChat powcoder

## Why not always use IV?

- It is hard to find variables that meet the definitions of valid instruments: conceptually, most variables that have an effect on endogenous variables $x$ may also have a direct effect on the dependent variable $y$.
- The standard errors on IV estimates are likely to be larger than OLS estimates, and much larger if the instrumental variables are only weakly correlated with the endogenous regressors.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

## Practical Tips for IV Papers

- Report the first stage
  - Does it make sense?
  - Do the coefficients have the right magnitude and size?
- Report the F-statistic.
  - Stock, Wright and Yogo (2002) suggest that F-statistics above 10 indicates that you do not have a weak instrument problem.
- Look at the Reduced Form:
  - The reduced form is estimated with OLS.
  - **If you can't see the causal relationship of interest in the reduced form it is probably not there.**

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder