Contents

References

- Wooldridge (2013) chapter 15
- Angrist and Pischke (2009) chapter 4
- Angrist and Pischke (2014) chapters 3, 6
- Angrist and Krueger (2001)
- Murray (2006)

1 Introduction

Introduction

- In the linear regression model,
$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$
the most important assumption for $\hat{\beta}_1^{\text{OLS}}$ to be consistent is exogeneity,
$$\mathrm{E}[x_i \epsilon_i] = 0$$
- Exogeneity is often an implausible assumption
- If we have an additional variable z_i with certain properties than we can still consistently estimate β_1 even when $\mathrm{E}[x_i \epsilon_i] \neq 0$
- New notation: $\hat{\beta}^{\text{OLS}}$ instead of just $\hat{\beta}$ for OLS estimates

---

To decide whether or not exogeneity is plausible, we must first be clear about what model we are trying to estimate. If we are simply interested in the population regression of y on x, then exogeneity automatically holds and OLS is consistent. However, much of the time we not interested in a population regression. Instead, we want to get at the causal relationship between or y and x or we want the linear model to represent some economic model (like a demand function of production function). In those cases, exogeneity is a strong and often implausible assumption.

As an example, we will investigate the causal effect of education on wages.

2 Example: return to education

Example: return to education

- Education (s_i) and log wages ($\log w_i$)
$$\log w_i = \beta_0 + \beta_1 s_i + \epsilon_i$$
- Suppose we want the causal effect of education on wages – then we want to hold constant everything else that affects wages
- We can never hold everything else constant, but we know that
$$\operatorname{plim} \hat{\beta}_1^{\text{OLS}} = \beta_1 + \frac{\mathrm{Cov}(s_i, \epsilon_i)}{\mathrm{Var}(s_i)}$$
so as long as whatever we are not holding constant (i.e. ϵ_i) is uncorrelated with s_i we are okay
- But it is very likely that there is unobserved ability, a_i, (IQ, work ethic, etc) that affects both education and wages
$$\log w_i = \beta_0 + \beta_1 s_i + \underbrace{\beta_2 a_i + u_i}_{=\epsilon_i}$$

	AK	Card
(Intercept)	4.6344***	5.5709***
	(0.0030)	(0.0391)
educ	0.0814***	0.0521***
	(0.0002)	(0.0029)
R^2	0.1371	0.0987
Adj. R^2	0.1371	0.0984
Num. obs.	1063634	3010
RMSE	0.6681	0.4214

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: OLS estimates

- Then,

$$\text{plim}\,\hat{\beta}_1^{\text{OLS}} = \beta_1 + \beta_2 \frac{\text{Cov}(s_i, a_i)}{\text{Var}(s_i)} + \frac{\text{Cov}(s_i, u_i)}{\text{Var}(s_i)}$$

OLS estimates of return to education

- Data from two papers:
  - Angrist and Krueger (1991): 1970 & 1980 U.S. census data 5% public use sample, men age 30-50
  - Card (1993): NLS young men 1966 cohort (wages measured in 1976 when age 24-34)
- We will start by looking at the OLS estimates even though we know that they are not consistent estimates of the causal effect of education on wages

---

OLS estimates of return to education

IV estimates of return to education

$$\log w_i = \beta_0 + \beta_1 s_i + \epsilon_i$$

[0] Code

- Suppose we observe a variable z_i such that $\mathrm{E}[z_i \epsilon_i] = 0$, then

$$
\begin{aligned}
0 =& \mathrm{E}[z_i \epsilon] \\
=& \mathrm{E}\left[z_i \left(\log w_i - \beta_0 - \beta_1 s_i\right)\right]
\end{aligned} \tag{1}
$$

  we also know $\mathrm{E}[\epsilon_i] = 0$, so

$$
0 = \mathrm{E}\left[(\log w_i - \beta_0 - \beta_1 s_i)\right] \tag{2}
$$

  replace the $\mathrm{E}[]$ with $\frac{1}{n}\sum_{i=1}^n$ and we have two equations to estimate two parameters β_0, β_1

$$
0 = \frac{1}{n}\sum_{i=1}^n z_i \left(\log w_i - \hat{\beta}_0^{\mathrm{IV}} - \hat{\beta}_1^{\mathrm{IV}} s_i\right) \tag{3}
$$

$$
0 = \frac{1}{n}\sum_{i=1}^n \left(\log w_i - \hat{\beta}_0^{\mathrm{IV}} - \hat{\beta}_1^{\mathrm{IV}} s_i\right) \tag{4}
$$

  - Note similarity to OLS first order conditions
  - This approach to estimation — start with an assumption about some expectations (moments) being zero and use them to derive an equation to use for estimation — is called the (generalized) method of moments
  - (1) and (2) are called the (population) moment conditions
  - (3) and (4) are called the sample (or empirical) moment conditions

- z_i is called an *instrumental variable*

- The solution to (3) and (4) is

$$
\hat{\beta}_1^{\mathrm{IV}} = \frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z}) \log w_i}{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z}) s_i} \text{ and } \hat{\beta}_0^{\mathrm{IV}} = \overline{\log w} - \hat{\beta}_1^{\mathrm{IV}} \bar{s}
$$

  they are called *instrumental variables (IV) estimators*

- Is $\hat{\beta}_1^{\mathrm{IV}}$ consistent?

$$
\begin{aligned}
\operatorname{plim} \hat{\beta}_1^{\mathrm{IV}} =& \operatorname{plim} \frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z}) \log w_i}{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z}) s_i} \\
=& \frac{\operatorname{plim} \frac{1}{n}\sum_{i=1}^n (z_i - \bar{z}) \log w_i}{\operatorname{plim} \frac{1}{n}\sum_{i=1}^n (z_i - \bar{z}) s_i} \\
=& \frac{\mathrm{Cov}(z, \log w)}{\mathrm{Cov}(z, s)} \text{ (assuming } \mathrm{Cov}(z, s) \neq 0) \\
=& \frac{\mathrm{Cov}(z, \beta_0 + \beta_1 s + \epsilon)}{\mathrm{Cov}(z, s)} \\
=& \frac{\mathrm{Cov}(z, \beta_0) + \mathrm{Cov}(z, \beta_1 s) + \mathrm{Cov}(z, \epsilon)}{\mathrm{Cov}(z, s)} \\
=& \beta_1
\end{aligned}
$$

  yes, as long as $\mathrm{Cov}(z, s) \neq 0$ (and $\mathrm{Cov}(z, \epsilon) = 0$, which we already assumed)

- How can we find such a z?

	Model 1
(Intercept)	3.7675 (0.3466)***
educ	0.1881 (0.0261)***
Num. obs.	3010

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Table 2: Card IV estimates

Card (1993) instrument: nearby college

- $nearc4_i = 1$ if i grew up in a county with a four-year college, else 0
- Two requirements to be a valid instrument:
  1. (exogenous) $\mathrm{E}[nearc4_i \epsilon_i] = 0$
  2. (relevant) $\mathrm{Cov}(nearc4_i, s_i) \neq 0$
- Relevance can be checked empirically
  - $\widehat{\mathrm{Cov}}(nearc4_i, s_i) = 0.18$
  - Regress s_i on $nearc4_i$

	Model 1
(Intercept)	12.70 (0.09)***
nearc4	0.83 (0.11)***
R^2	0.02
Adj. R^2	0.02
Num. obs.	3010

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

- Exogeneity cannot be tested empirically
  - Card (1993) discusses why maybe not $\mathrm{E}[nearc4_i \epsilon_i] = 0$
    * Families that value education might live near colleges
    * High schools and elementary schools might be higher quality near colleges
    * It's a challenge to show these concerns are not a problem (we will discuss it more later)

Card (1993) IV estimate

---

[0] Code

Angrist and Krueger (1991) instrument: quarter of birth

- In most of the U.S. must attend school until age 16 (at least during 1938-1967)
- Age when starting school depends on birthday, so grade when can legally drop out depends on birthday
- Plausible that quarter of birth uncorrelated with other factors affecting wages (there is some disagreement about this though)
- Is quarter of birth correlated with education?

	Model 1
(Intercept)	12.69 (0.01)***
QOB	0.06 (0.00)***
R^2	0.00
Adj. R^2	0.00
Num. obs.	1063634

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Angrist and Krueger (1991) instrument: quarter of birth - relevance

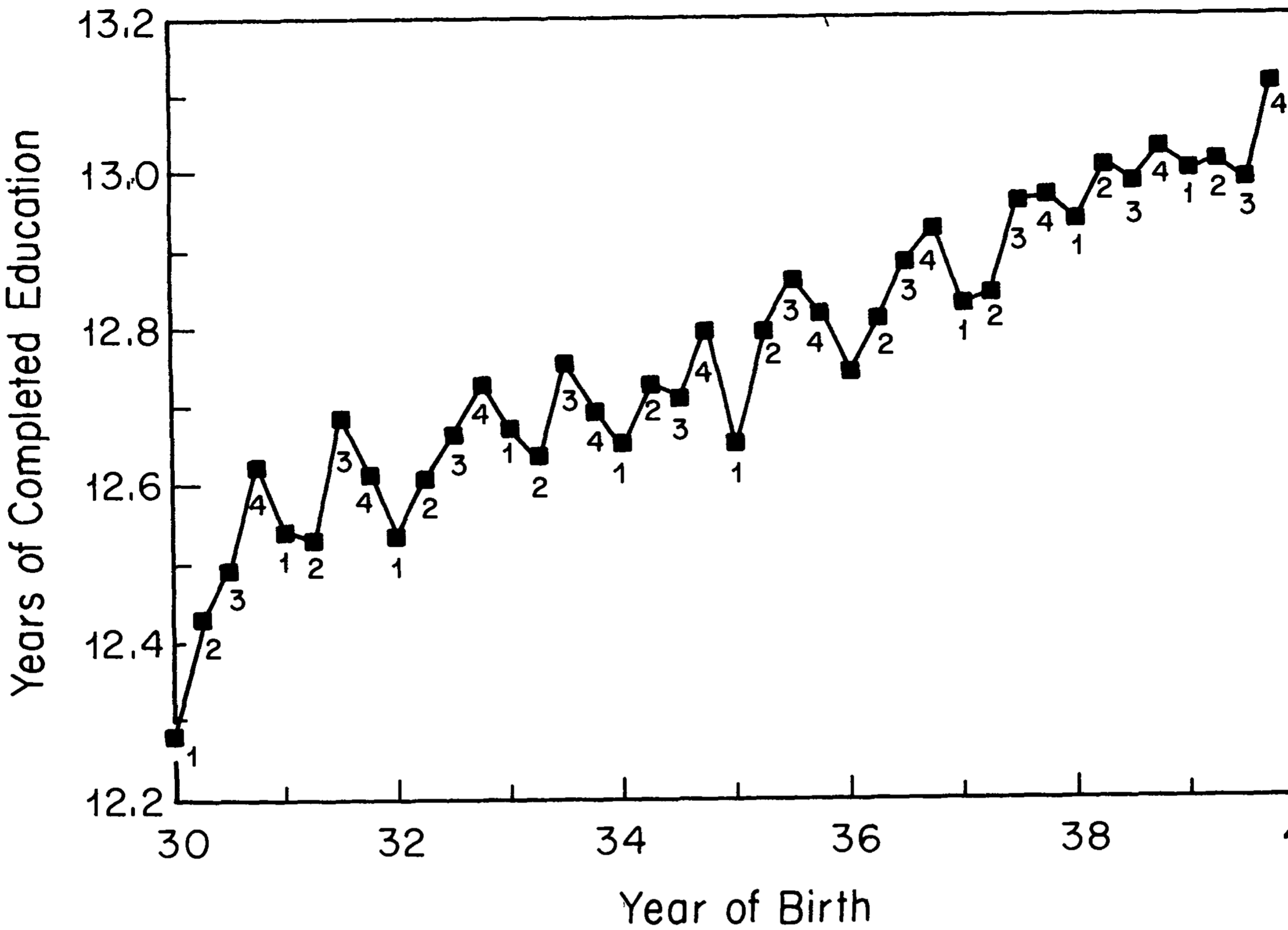


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Angrist and Krueger (1991) instrument: quarter of birth - relevance

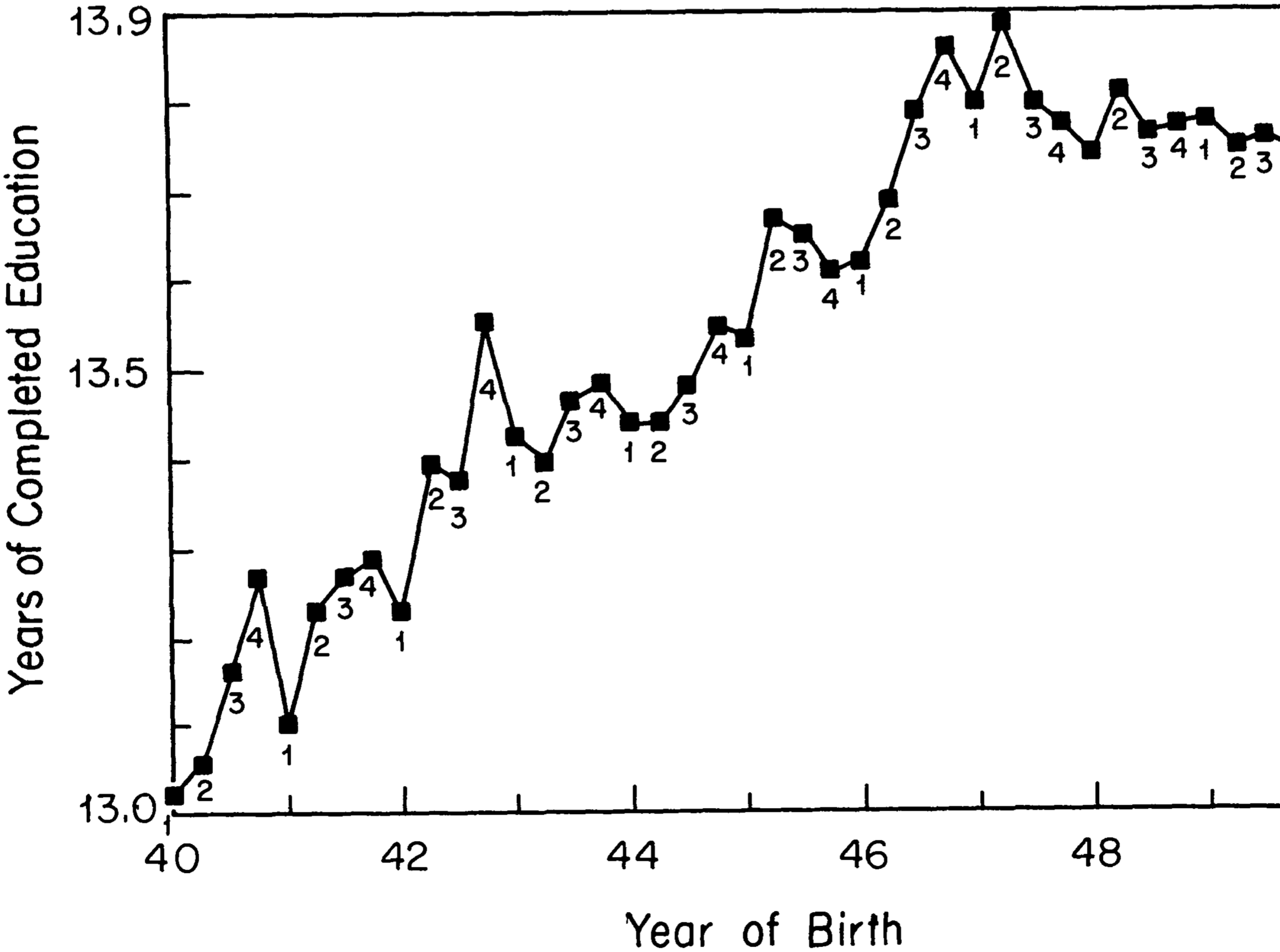


FIGURE II
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Angrist and Krueger (1991) instrument: quarter of birth - relevance

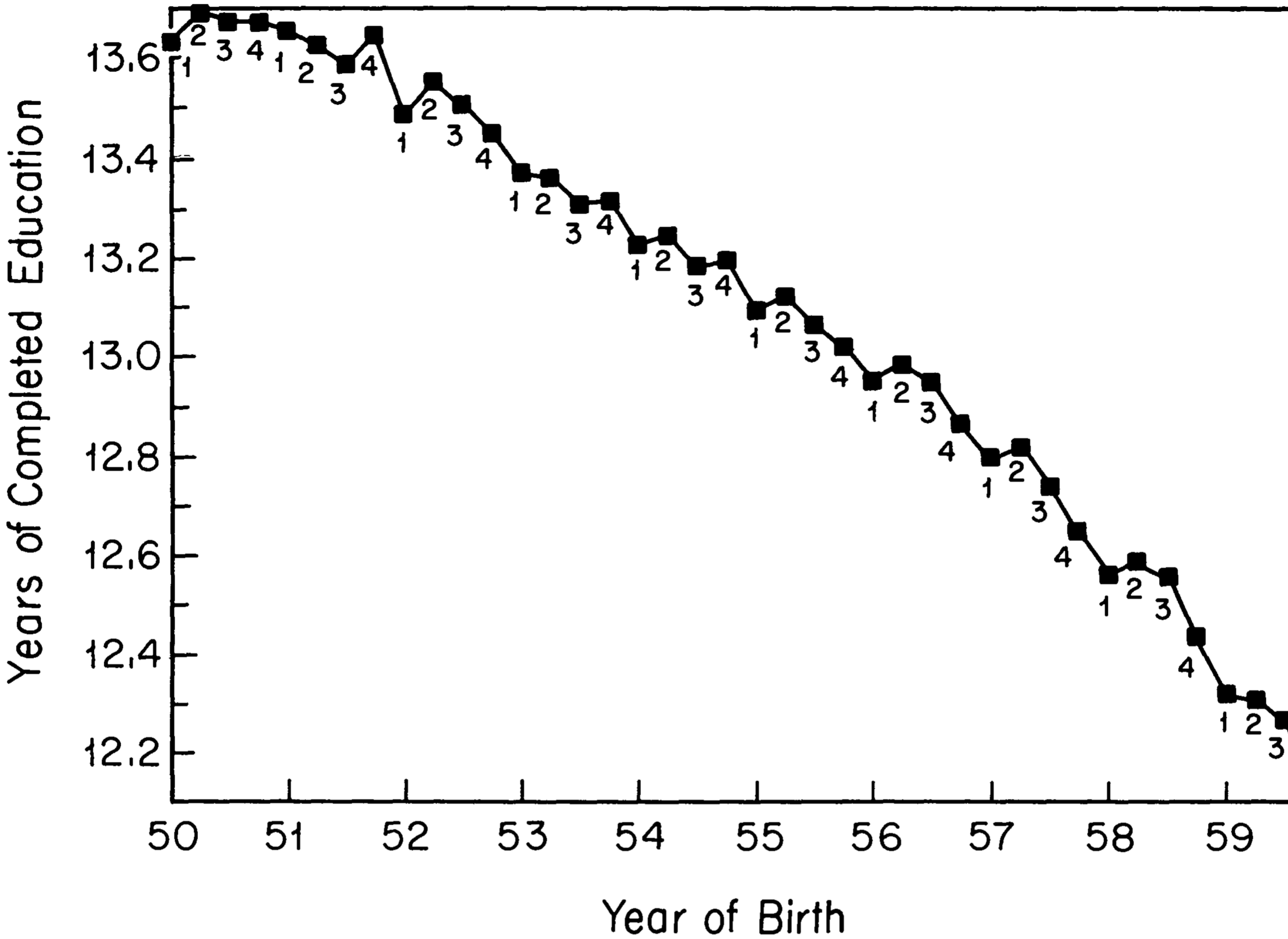
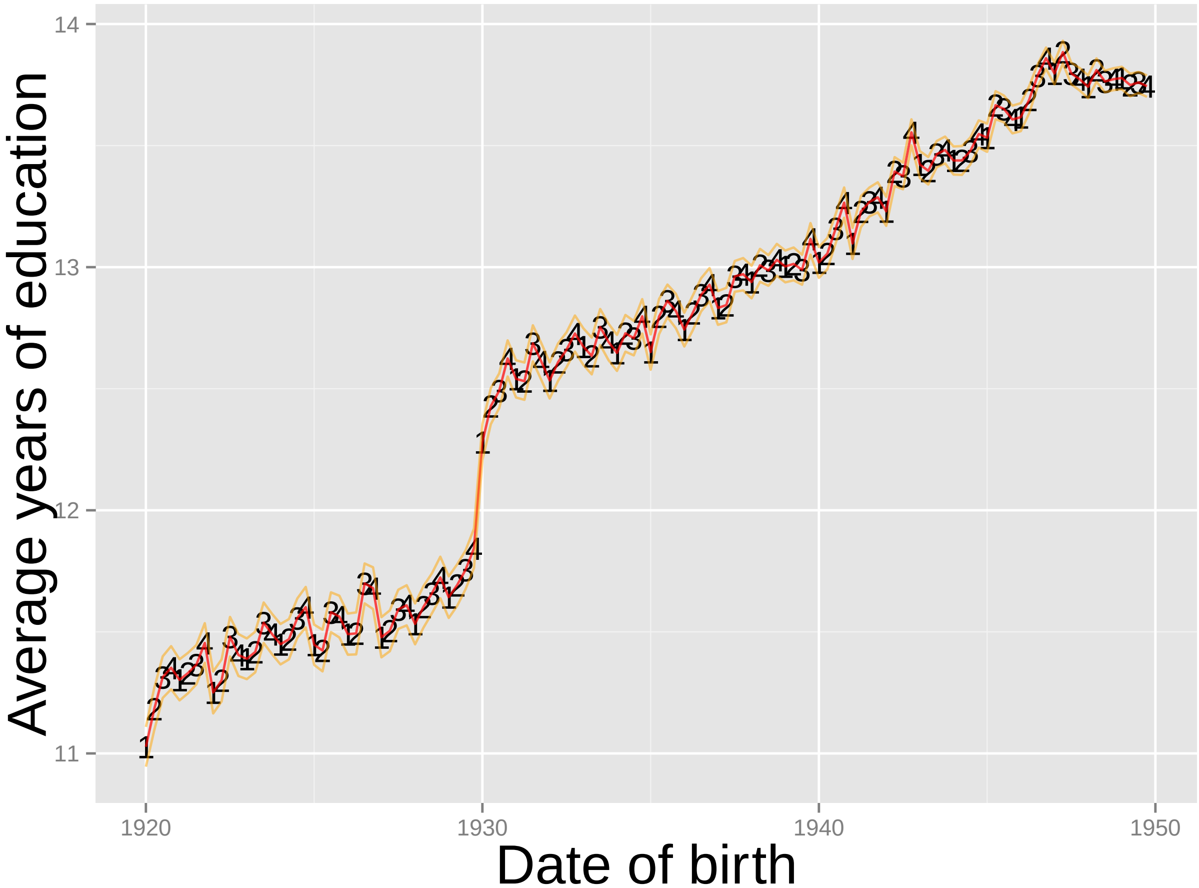


FIGURE III
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

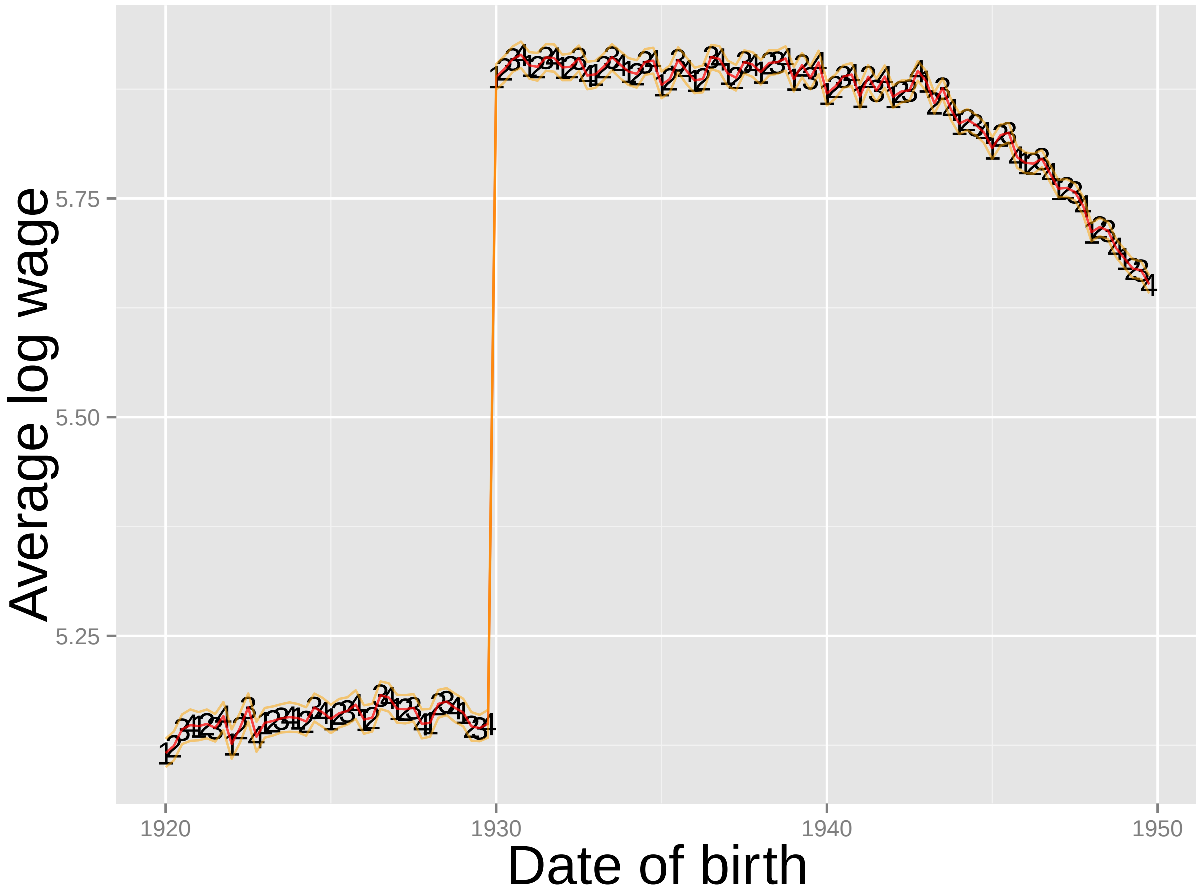


Angrist and Krueger (1991) reduced form

	All	1920-29	1930-39	1940-49
(Intercept)	4.6344***	4.2344***	4.9952***	5.0452***
	(0.0030)	(0.0048)	(0.0051)	(0.0049)
educ	0.0814***	0.0801***	0.0709***	0.0554***
	(0.0002)	(0.0004)	(0.0004)	(0.0004)
R^2	0.1371	0.1709	0.1173	0.0655
Adj. R^2	0.1371	0.1709	0.1173	0.0655
Num. obs.	1063634	247199	329509	486926

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Angrist & Krueger OLS estimates



Angrist and Krueger (1991) OLS estimate

---

[0] Code

	All	1920-29	1930-39	1940-49
(Intercept)	4.7056^{***}	4.4869^{***}	4.6329^{***}	6.6340^{***}
	(0.1247)	(0.1941)	(0.2505)	(0.3502)
educ	0.0759^{***}	0.0581^{***}	0.0992^{***}	-0.0616^{*}
	(0.0097)	(0.0169)	(0.0196)	(0.0258)
Num. obs.	1063634	247199	329509	486926

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Angrist & Krueger IV estimates

Angrist and Krueger (1991) IV estimate

Issues raised

- Statistical properties of $\hat{\beta}^{\mathrm{IV}}$
  - Unbiased? Consistent? Asymptotic distribution? Standard error?
- How to use instrumental variables in multiple regression
- Why are Angrist and Krueger (1991) and Card (1993) results so different?
- What happens if IV assumptions not true? Assumptions that might be wrong:
  - $\mathrm{E}[z_i \epsilon_i] = 0$
  - $\mathrm{Cov}(z, s) \neq 0$
  - Linear model
- See Card (2003) for a review of many papers about the returns to education

---

3 Statistical properties

Statistical properties

- Model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{5}$$

- Assumptions:

  IV.1 Linearity: (5) holds

[0] Code

IV.2 Independent observations

IV.3 Relevance (rank condition): $\mathrm{Cov}(z, x) \neq 0$

IV.4 Exogeneity: $\mathrm{E}[z_i \epsilon_i] = 0$

Note: these are the same as for OLS except the rank condition and exogeneity assumptions are now about the instrument, z, instead of the regressor, x

- Relevance + exogeneity = z affects y only through x
- Terminology:
  - z_i is an *instrument* or *instrumental variable*
  - (5) is the *structural equation*
  - x_i is an *endogenous regressor*
  - The regression of x on z is the *first stage*
  - The regression of y on z is the *reduced form*
- Properties to look at:
  - Bias
  - Consistency
  - Asymptotic distribution

3.1 Bias

IV is biased

- Consider $\mathrm{E}[\hat{\beta}_1^{\mathrm{IV}}]$

$$\begin{aligned}
\mathrm{E}[\hat{\beta}_1^{\mathrm{IV}}] =& \mathrm{E}\left[\frac{\sum_{i=1}^n (z_i - \bar{z}) y_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i}\right] \\
=& \mathrm{E}\left[\frac{\sum_{i=1}^n (z_i - \bar{z})(\beta_0 + \beta_1 x_i + \epsilon_i)}{\sum_{i=1}^n (z_i - \bar{z}) x_i}\right] \\
=& \beta_1 + \mathrm{E}\left[\frac{\sum_{i=1}^n (z_i - \bar{z})\epsilon_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i}\right] \\
\neq& \beta_1
\end{aligned}$$

- Cannot show $\mathrm{E}\left[\frac{\sum_{i=1}^n (z_i - \bar{z})\epsilon_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i}\right] = 0$ because of x_i in denominator and

$$\mathrm{E}\left[\frac{\sum_{i=1}^n (z_i - \bar{z})\epsilon_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i}\right] \neq \frac{\mathrm{E}\left[\sum_{i=1}^n (z_i - \bar{z})\epsilon_i\right]}{\mathrm{E}\left[\sum_{i=1}^n (z_i - \bar{z}) x_i\right]}$$

3.2 Consistency

Showing consistency and asymptotic normality of IV uses almost the exact same steps as we used for OLS. For OLS, we had

$$\hat{\beta}_1^{OLS} = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})\epsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and for IV we have a similar expression,

$$\hat{\beta}_1^{IV} = \beta_1 + \frac{\sum_{i=1}^n (z_i - \bar{z})\epsilon_i}{\sum_{i=1}^n (z_i - \bar{z})x_i}.$$

In fact IV with $z = x$ is OLS. In either case consistency and asymptotic normality will involve working with the sum of stuff times ϵ_i.

IV is consistent

- As in the education example,

$$\begin{aligned}
\operatorname{plim} \hat{\beta}_1^{\mathrm{IV}} = & \operatorname{plim} \frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})y_i}{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i} \\
= & \beta_1 + \operatorname{plim} \frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})\epsilon_i}{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i} \\
= & \beta_1 + \frac{\operatorname{plim} \frac{1}{n}\sum_{i=1}^n z_i\epsilon_i - \operatorname{plim} \bar{z} \operatorname{plim} \bar{\epsilon}}{\operatorname{plim} \frac{1}{n}\sum_{i=1}^n z_i x_i - \operatorname{plim} \bar{z} \operatorname{plim} \bar{x}} \\
= & \beta_1 + \frac{\mathrm{E}[z\epsilon]}{\mathrm{E}[zx] - \mathrm{E}[z]\mathrm{E}[x]} = \beta_1 + \frac{\mathrm{Cov}(z,\epsilon)}{\mathrm{Cov}(z,x)} \\
= & \beta_1
\end{aligned}$$

- So IV is biased but consistent
- Another useful way of expressing $\hat{\beta}_1^{\mathrm{IV}}$ is as the reduced form divided by the first stage:
  - Reduced form:
  $$y_i = \pi_{y,0} + \pi_{y,1}z_i + u_i$$
  OLS estimate $= \hat{\pi}_{y,1} = \frac{\widehat{\mathrm{Cov}}(z,y)}{\widehat{\mathrm{Var}}(z)}$
  - First stage:
  $$x_i = \pi_{x,0} + \pi_{x,1}z_i + v_i$$
  OLS estimate $= \hat{\pi}_{x,1} = \frac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)}$
  - Then,
  $$\hat{\beta}^{\mathrm{IV}} = \frac{\widehat{\mathrm{Cov}}(z,y)}{\widehat{\mathrm{Cov}}(z,x)} = \frac{\widehat{\mathrm{Cov}}(z,y)/\widehat{\mathrm{Var}}(z)}{\widehat{\mathrm{Cov}}(z,x)/\widehat{\mathrm{Var}}(z)} = \frac{\hat{\pi}_{y,1}}{\hat{\pi}_{x,1}}$$

3.3 Asymptotic distribution

Asymptotic distribution

- We will allow for heteroskedasticity
- As when looking at bias and consistency of $\hat{\beta}^{\text{IV}}$,

$$\hat{\beta}_1^{\text{IV}} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})\epsilon_i}{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i}$$

- To apply CLT we look at $\sqrt{n}(\hat{\beta}_1^{\text{IV}} - \beta_1)$,

$$\sqrt{n}(\hat{\beta}_1^{\text{IV}} - \beta_1) = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})\epsilon_i}{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i}$$

- As for OLS with heteroskedasticity,

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})\epsilon_i \xrightarrow{d} N\left(0, \mathrm{E}\left[(z - \mathrm{E}[z])^2\epsilon^2\right]\right)$$

- Previous slide showed

$$\operatorname{plim}\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i = \mathrm{Cov}(x, z)$$

- So using Slutsky's theorem, we can conclude

$$\sqrt{n}(\hat{\beta}_1^{\text{IV}} - \beta_1) \xrightarrow{d} N\left(0, \frac{\mathrm{E}\left[(z - \mathrm{E}[z])^2\epsilon^2\right]}{\mathrm{Cov}(x, z)^2}\right)$$

- We can estimate the asymptotic variance by $\frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})^2\epsilon_i^2}{\left(\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i\right)^2}$
- t-statistic

$$t = \frac{\hat{\beta}_1^{\text{IV}} - \beta_1}{\sqrt{\frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})^2\hat{\epsilon}_i^2}{n\left(\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i\right)^2}}} \xrightarrow{d} N(0, 1)$$

- Standard error:

$$s.e.(\hat{\beta}_1^{\text{IV}}) = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})^2\hat{\epsilon}_i^2}{n\left(\frac{1}{n}\sum_{i=1}^n (z_i - \bar{z})x_i\right)^2}}$$

3.4 IV when exogeneity fails

IV without exogeneity

- People sometimes defend an instrument by saying: "even though it might not be true that $\mathrm{E}[z\epsilon] = 0$, it is likely that the correlation between z and ϵ is smaller than the correlation between x and ϵ. Therefore we prefer the IV estimate to the OLS estimate." Is this argument correct?

- We showed earlier that

$$\operatorname{plim} \hat{\beta}_1^{\mathrm{IV}} = \beta_1 + \frac{\mathrm{E}[z\epsilon]}{\mathrm{Cov}(z, x)}$$

  IV is consistent only when $\mathrm{E}[z\epsilon] = 0$

- We also showed that

$$\operatorname{plim} \hat{\beta}_1^{\mathrm{OLS}} = \beta_1 + \frac{\mathrm{E}[x\epsilon]}{\mathrm{Var}(x)}$$

- Express in terms of correlations:[1]

$$\operatorname{plim} \hat{\beta}_1^{\mathrm{IV}} - \beta_1 = \frac{\rho_{z,\epsilon}\sqrt{\mathrm{Var}(z)\mathrm{Var}(\epsilon)}}{\rho_{z,x}\sqrt{\mathrm{Var}(z)\mathrm{Var}(x)}} = \frac{\rho_{z,\epsilon}}{\rho_{z,x}}\sqrt{\frac{\mathrm{Var}(\epsilon)}{\mathrm{Var}(x)}}$$

  and

$$\operatorname{plim} \hat{\beta}_1^{\mathrm{OLS}} - \beta_1 = \frac{\rho_{x,\epsilon}\sqrt{\mathrm{Var}(x)\mathrm{Var}(\epsilon)}}{\mathrm{Var}(x)} = \rho_{x,\epsilon}\sqrt{\frac{\mathrm{Var}(\epsilon)}{\mathrm{Var}(x)}}$$

- So IV is "less inconsistent" than OLS only if

$$\left|\frac{\rho_{z,\epsilon}}{\rho_{z,x}}\right| < |\rho_{x,\epsilon}|$$

  - Just z being "less endogenous" or less correlated with ϵ is not enough

- No, the proposed argument is not correct

---

4 IV for multiple regression

IV for multiple regression

- Model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \beta_{k+1} w_{1,i} + \cdots + \beta_{k+r} w_{r,i} + \epsilon_i \tag{6}$$

  with instruments $z_{1,i}, \ldots, z_{m,i}$

- Assumptions:

  IV.1 Linearity: (6) holds

[1] Let $\rho_{x,y}$ denote the correlation of x and y, and note that $\rho_{x,y} = \mathrm{Cov}(x, y)/\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}$.

IV.2 Independent observations

IV.3 Relevance (rank condition): $m \geq k$ and (loosely speaking) each $x_{j,i}$ is correlated with some $z_{l,i}$

IV.4 Exogeneity: $\mathrm{E}[w_{s,i}\epsilon_i] = 0$ for $s = 1, .., r$ and $\mathrm{E}[z_{l,i}\epsilon_i] = 0$ for $l = 1, ..., m$

- Terminology:
  - $w_{s,i}$ are exogenous controls
  - $x_{j,i}$ are endogenous regressors
  - $z_{l,i}$ are instruments

IV for multiple regression

- Example: returns to education:

$$\log w_i = \beta_0 + \beta_1 s_i + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \beta_4 \text{region}_i + \epsilon_i$$

  - s_i is endogenous
  - age_i, age_i^2, and region_i are exogenous
- How to estimate β?
  - As before, could find a consistent estimator by using the assumed moment conditions

$$\begin{aligned}
0 =& \mathrm{E}[\epsilon] && = \mathrm{E}\left[(y - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k - \beta_{k-1} w_1 - \cdots - \beta_{k+r} w_r)\right] \\
0 =& \mathrm{E}[w\epsilon] && = \mathrm{E}\left[w(y - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k - \beta_{k-1} w_1 - \cdots - \beta_{k+r} w_r)\right] \\
0 =& \mathrm{E}[z\epsilon] && = \mathrm{E}\left[z(y - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k - \beta_{k-1} w_1 - \cdots - \beta_{k+r} w_r)\right]
\end{aligned}$$

    but easier to describe in a different way - two stage least squares

---

In multiple regression, the idea behind IV is the same — take the assumed exogeneity consumptions and use them to get some equations to solve for the unknown β_j. With a single x and z, writing the explicit solution to this system of equations was easy. With multiple x's and z's (and the added w's), the solution is complicated to write down (but easy for a computer to calculate). One useful way of expressing the solution to this system is two stage least squares.

Two stage least squares

- *Two stage least squares*: $\hat{\beta}^{2SLS}$
  1. Estimate (by OLS) the first stage

$$\hat{x}_{j,i} = \hat{\pi}_{x_j,0} + \hat{\pi}_{x_j,z_1} z_{1,i} + \cdots + \hat{\pi}_{x_j,z_m} z_{m,i} + \hat{\pi}_{x_j,w_1} w_{1,i} + \cdots + \hat{\pi}_{x_j,w_r} w_{r,i}$$

     to get predicted values $\hat{x}_j$
  2. Regress (using OLS) y on $\hat{x}_1, ..., \hat{x}_k, w_1, ..., w_r$ the coefficients are $\hat{\beta}_j^{2SLS}$

- Exercise: show that for bivariate regression $\hat{\beta}_1^{\text{IV}} = \hat{\beta}_1^{\text{2SLS}}$

To understand why two stage least squares works, it's useful to think about how the second stage, regressing y on $\hat{x}$ (and w) compares to just the OLS regression of y on x. The OLS regression of y on x is not consistent because we think x might be correlated with ϵ. $\hat{x}$ is the part of x that can be explained by z. We assume that z is uncorrelated with ϵ, so $\hat{x}$ is part of x that is uncorrelated with ϵ.

Here we will show that $\hat{\beta}_1^{\text{IV}} = \hat{\beta}_1^{\text{2SLS}}$, where $\hat{\beta}_1^{\text{IV}} = \frac{\sum_{i=1}^n (z_i - \bar{z}) y_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i}$ and $\hat{\beta}_1^{\text{2SLS}}$ is defined as on the previous slide. The second stage of two-stage least squares is the OLS regression of y on $\hat{x}$, so

$$\hat{\beta}_1^{\text{2SLS}} = \frac{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}}) y_i}{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}})^2}.$$

Recall that one of properties of OLS fitted values is that the covariance of the fitted value and actual value is equal to the variance of the fitted value. We showed this earlier in the course. Using that on the denominator here, we have

$$\hat{\beta}_1^{\text{2SLS}} = \frac{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}}) y_i}{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}}) x_i}$$

Using the fact that $\hat{x}_i = \hat{\pi}_0 + \hat{\pi}_1 z_i$ where $\hat{\pi}_0$ and $\hat{\pi}_1$ are OLS estimates, we have

$$\begin{aligned}
\hat{\beta}_1^{\text{2SLS}} = & \frac{\sum_{i=1}^n (\hat{\pi}_0 + \hat{\pi}_1 z_i - \overline{\hat{\pi}_0 + \hat{\pi}_1 z}) y_i}{\sum_{i=1}^n (\hat{\pi}_0 + \hat{\pi}_1 z_i - \overline{\hat{\pi}_0 + \hat{\pi}_1 z}) x_i} \\
= & \frac{\hat{\pi}_1 \sum_{i=1}^n (z_i - \bar{z}) y_i}{\hat{\pi}_1 \sum_{i=1}^n (z_i - \bar{z}) x_i} \\
= & \frac{\sum_{i=1}^n (z_i - \bar{z}) y_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i} = \hat{\beta}_1^{\text{IV}}.
\end{aligned}$$

Two stage least squares

- $\hat{\beta}^{\text{2SLS}}$ is consistent and asymptotically normal
- Essential that x be regressed on both z and w in the first stage
- When calculating $\hat{\beta}^{\text{2SLS}}$ best not to preform two regressions
  - OLS standard errors of second stage regression are not correct for $\hat{\beta}^{\text{2SLS}}$
  - In R use `ivreg` or `felm`
- Test relevance condition: look at the F-statistic in for $H_0 : \pi_{x_j,z_1} = \cdots = \pi_{x_j,z_m} = 0$ in the first stage
  - Rule of thumb: $F \geq 10$ is okay, $F < 10$ need to use another method (weak instruments)

The logic for this rule of thumb is a bit different than what we have seen. It comes from thinking about a different sort of asymptotic approximation, called weak instruments asymptotics, than what we have been using. See Stock, Wright, and Yogo (2002) for more information.

Understanding 2SLS

- In bivariate regression $\hat{\beta}_1^{\text{IV}} = \hat{\beta}_1^{\text{2SLS}}$
- With one endogenous variable and one instrument, ($k = m = 1$),

$$\hat{\beta}_1^{\text{2SLS}} = \frac{\hat{\pi}_{y,z_1}}{\hat{\pi}_{x_1,z_1}} = \frac{\text{reduced form coefficient on instrument}}{\text{first stage coefficient on instrument}}$$

  - First stage:

$$\begin{aligned} x_{j,i} =& \pi_{x_j,0} + \pi_{x_j,z_1} z_{1,i} + \cdots + \pi_{x_j,z_m} z_{m,i} + \\ &+ \pi_{x_j,w_1} w_{1,i} + \cdots + \pi_{x_j,w_r} w_{r,i} + v_{j,i} \end{aligned}$$

  - Reduced form:

$$y_i = \pi_{y,0} + \pi_{y,z_1} z_{1,i} + \cdots + \pi_{y,z_m} z_{m,i} + \pi_{y,w_1} w_{1,i} + \cdots + \pi_{y,w_r} w_{r,i} + u_i$$

Understanding 2SLS

- Control function interpretation:
  - 2SLS is equivalent to the following:
    1. Regress x_j on z and w, calculate the *residuals*, $\hat{v}_{j,i}$
    2. Regress y on x, w *and* $\hat{v}_{j,i}$ estimated coefficient on x_j is equal to $\hat{\beta}_j^{\text{2SLS}}$

Two stage least squares and this control function procedure give exactly the same estimate. If $\hat{x}$ is the part of x that is uncorrelated with ϵ, then the remaining $\hat{v}$ is part of x that is correlated with ϵ. If we can control (i.e. hold constant) this "bad" part of x, then we can consistently estimate the coefficient on x. Multiple regression does exactly what we want. The regression of y on x, w, and $\hat{v}$ estimates the relationship between y and x, holding w and $\hat{v}$ constant.

5 Example: return to education

Example: return to education (continued)

$$\log w_i = \beta_0 + \beta_1 s_i + \epsilon_i$$

- Card (1993) and Angrist and Krueger (1991) estimates very different

	Card	**AK**
Sample	NLS66	Census 1970 & 1980
Instrument	$nearc4$	QOB
$\hat{\beta}_1^{\text{OLS}}$	0.052	0.071
$\hat{\beta}_1^{\text{IV}}$	0.188	0.099

	AK 20-29	AK 30-39	Card
(Intercept)	2.7055***	5.1251***	3.2677***
	(0.3086)	(0.2849)	(0.6940)
educ	0.0802***	0.0711***	0.0522***
	(0.0004)	(0.0004)	(0.0028)
age	0.0673***	−0.0107	0.1222*
	(0.0138)	(0.0128)	(0.0488)
I(age^2)	−0.0007***	0.0002	−0.0014
	(0.0002)	(0.0001)	(0.0008)
R^2	0.1710	0.1177	0.1821
Adj. R^2	0.1710	0.1177	0.1813
Num. obs.	247199	329509	3010

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: OLS estimates

- Why?

Adding controls

- Card's sample features younger men than Angrist and Krueger's
- Use multiple regression to control for age

OLS controlling for age

IV controlling for age

[1] Code

[1] Code

	AK 20-29	AK 30-39	Card
(Intercept)	2.9315***	3.8145***	3.4221***
	(0.3796)	(0.5794)	(0.8800)
educ	0.0567*	0.1660***	0.1736***
	(0.0226)	(0.0349)	(0.0240)
age	0.0704***	−0.0121	−0.0029
	(0.0143)	(0.0143)	(0.0662)
I(age^2)	−0.0008***	0.0003	0.0008
	(0.0002)	(0.0002)	(0.0011)
Num. obs.	247199	329509	3010

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: IV estimates

Controlling for urban

- Card instrument = being in same county as a college
- Colleges are more common in urban areas
- Wages are also higher in urban areas
- Should control for urban (and any other available geographic variables)

IV controlling for age and urban

Using multiple instruments

- Quarter of birth $= 1, 2, 3, 4$
- If assume $\mathrm{E}[\epsilon_i | QOB_i] = 0$, then can use quarter of birth dummies as instruments $z_i = (qob_i^1, qob_i^2, qob_i^3)$ where $qob_i^q = 1$ if $QOB_i = q$, else 0
- Since relationship between quarter of birth and education seems to change with year of birth, can use $QOB \times YOB$ dummies as instruments
  - $d_i^{q,y} = 1$ if $QOB_i = q$ and $YOB_i = y$
  - $3 \times 9 = 27$ dummies for 1930-1939 cohort
- In our linear model $\operatorname{plim} \hat{\beta}^{2SLS}$ is the same whether we use QOB or dummies as instrument; in a richer model it can matter

[1] Code

	AK 20-29	AK 30-39	Card	Card (geo 1966)
(Intercept)	3.0608***	3.7768***	3.2469***	3.0334***
	(0.3802)	(0.5678)	(0.7049)	(0.7134)
educ	0.0672**	0.1680***	0.0955*	0.0905
	(0.0207)	(0.0342)	(0.0481)	(0.0473)
age	0.0626***	−0.0104	0.0816	0.1028
	(0.0139)	(0.0144)	(0.0702)	(0.0739)
I(age^2)	−0.0007***	0.0002	−0.0007	−0.0011
	(0.0002)	(0.0002)	(0.0012)	(0.0013)
smsa	−0.1246***	−0.0589	0.1039*	
	(0.0108)	(0.0360)	(0.0472)	
south	−0.1418***	−0.0431*	−0.1278**	
	(0.0199)	(0.0214)	(0.0479)	
smsa66				0.0882**
				(0.0299)
south66				−0.1061
				(0.0543)
Num. obs.	247199	329509	3010	3010

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7: IV estimates

AK estimates with dummy instruments

5.1 Lemieux and Card (2001)

Lemieux and Card (2001) "Education, earnings, and the 'Canadian G.I. Bill' "

- Question: what is the causal effect of education on earnings?
- Strategy: use VRA as instrument for education
- Veteran Rehabilitation Act (1944)
  - Tuition + living expenses allowance of \$60 ($\approx$ \$500 today) per month for university or vocational training
  - Different impact in Ontario and Quebec
  - Ontario had compulsory schooling until age 16, more universities, higher average education at start of WWII

[1]Code

	QOB	QOB × YOB
(Intercept)	4.0885***	4.7347***
	(0.5072)	(0.3858)
educ	0.1451***	0.0977***
	(0.0296)	(0.0188)
age	−0.0094	−0.0074
	(0.0137)	(0.0129)
I(age^2)	0.0002	0.0001
	(0.0002)	(0.0001)
smsa	−0.0829**	−0.1325***
	(0.0311)	(0.0199)
south	−0.0573**	−0.0867***
	(0.0185)	(0.0119)
Num. obs.	329509	329509

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8: AK 1930-1939 IV estimates

  - Quebec had no compulsory schooling, few universities, lower average education at start of WWII; lower portion of veterans
  - VRA had smaller impact in Quebec than Ontario

- Instrument = Ontario × university age in 1945
- Data: 1971 Census
  - Observations: 11,163 Ontario + 10,078 Quebec

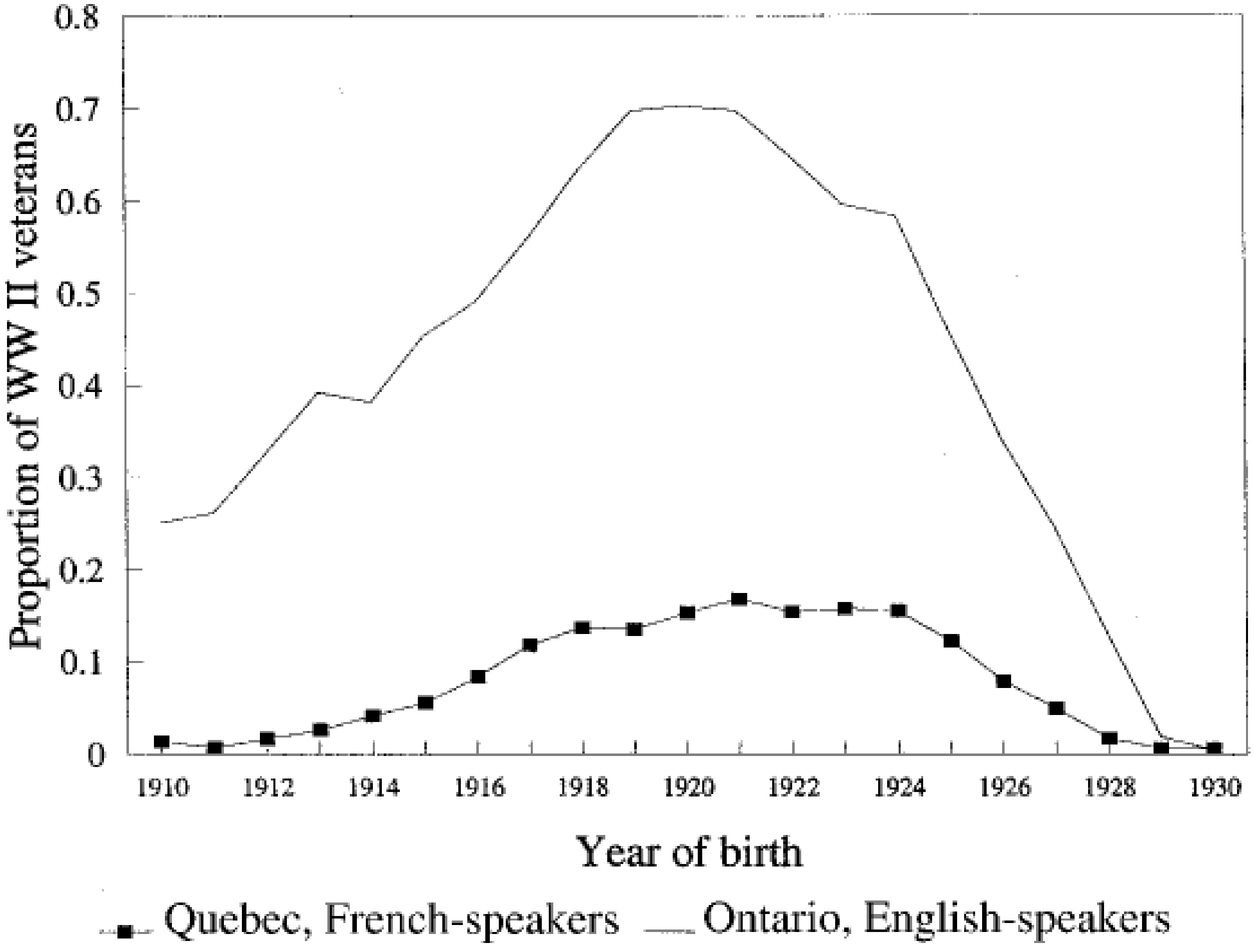
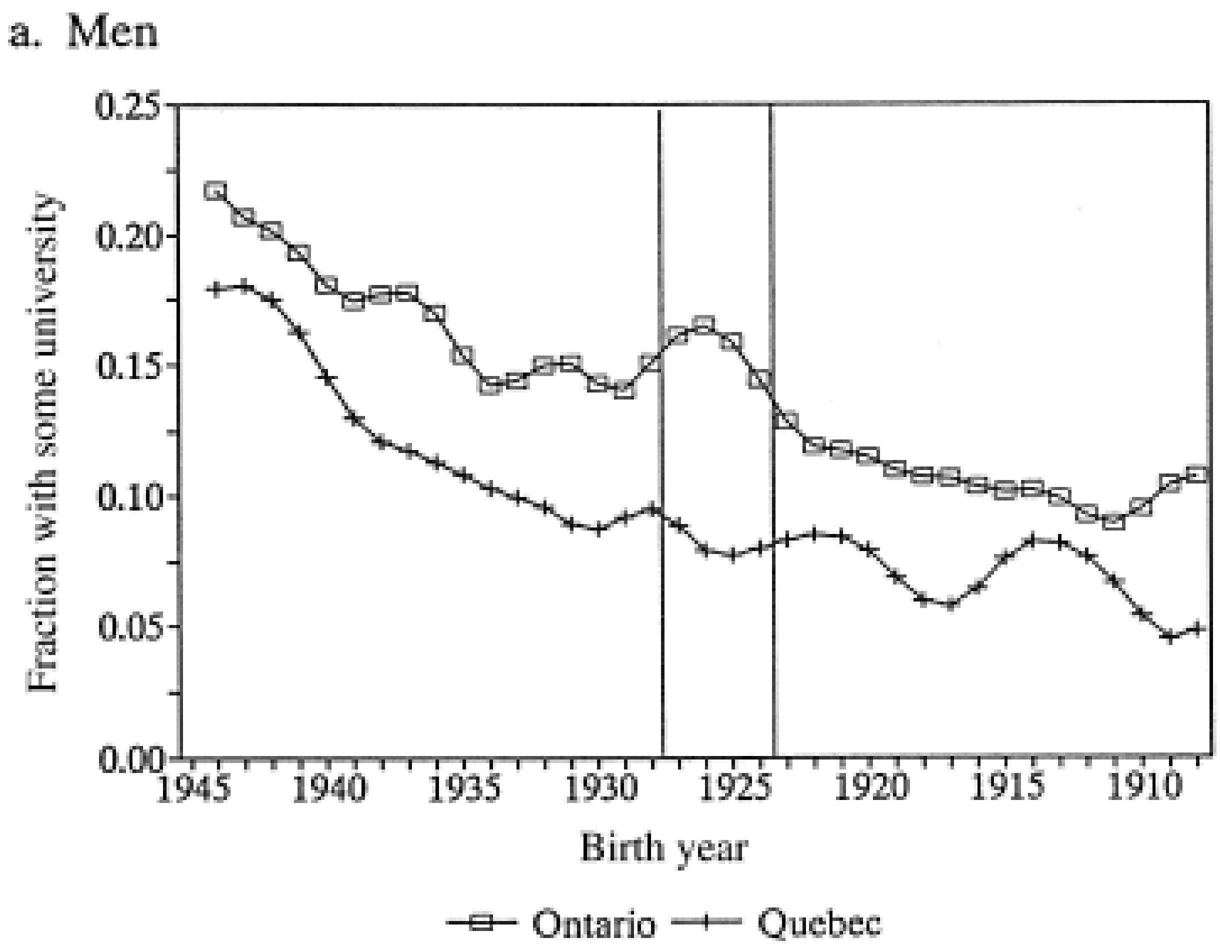


FIGURE 1 Proportion of men who served in WW II by year of birth ~f ve-year moving average!

First stage



Reduced form

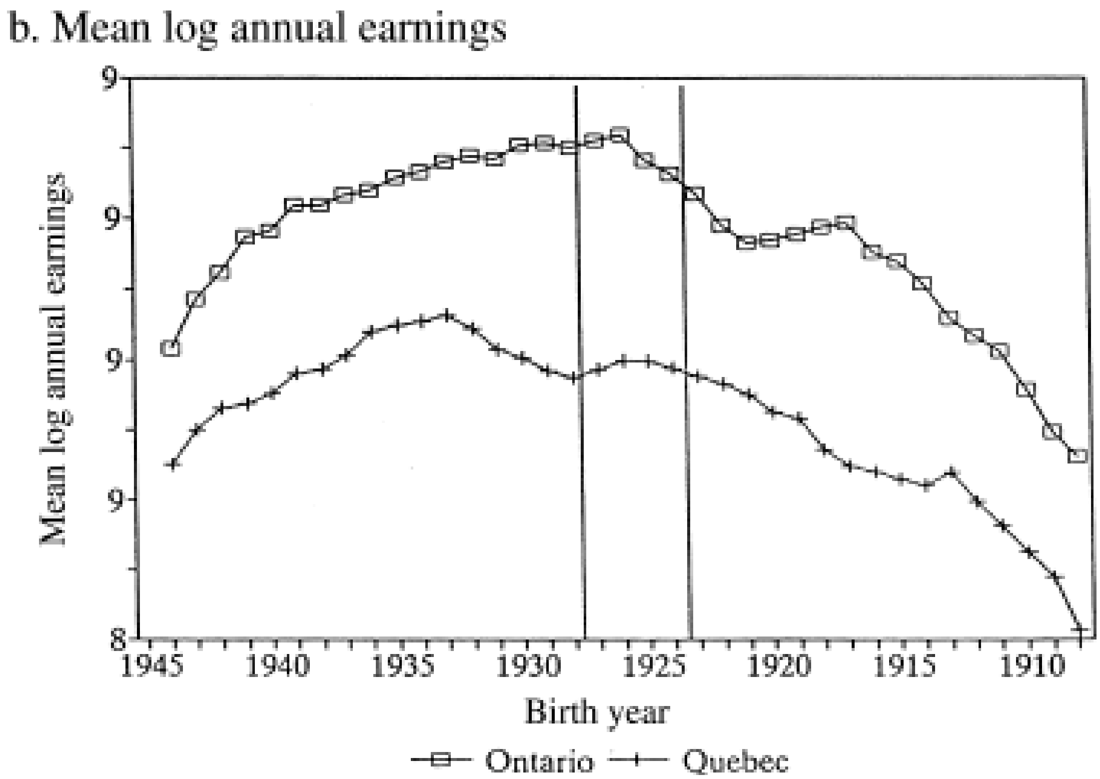


FIGURE 5 Labour market outcomes of men, 1971 Census (five-year moving average)

Model

$$
\begin{aligned}
y_i = & s_i\beta + \gamma_0 + \gamma_1 exper_i + \gamma_2 exper_i^2 + \gamma_3 exper_i^3 + \gamma_4 exper_i^4 + \\
& + \gamma_5 Quebec_i + \gamma_6 weeks_i + \gamma_7 fulltime_i + \epsilon_i
\end{aligned}
$$

- y_i = log annual earnings in 1970
- $weeks_i$ = weeks worked in 1970
- $fulltime_i = 1$ if full-time worker in 1970
- $exper_i$ = potential experience = $age - education - 6$
- Some specifications add interactions between $Quebec$ and $exper_i$
  - I.e. add $\gamma_8 exper_i \times Quebec_i + \gamma_9 exper_i^2 \times Quebec_i + \cdots$
  - Results on next slide include interactions

Results: education coefficient

Model	*Coefficient*
OLS	0.070
	(0.002)
Using $z =$ Ontario $\times$ age 18-21 in 1945	
First stage	0.465
	(0.101)
Reduced form	0.073
	(0.023)
IV	0.157
	(0.051)
IV using Ontario $\times$ age 18-24 in 1945	0.080
	(0.044)
IV for women using Ontario $\times$ age 18-24 in 1945	-0.111
	(0.524)

5.2 Fang et al. (2012)

Fang et al. (2012) "The Returns to Education in China: Evidence from the 1986 Compulsory Education Law"

- Question: what is the causal effect of education on earnings in China?
- Strategy: use China Compulsory Education Law of 1986 as instrument
- China Compulsory Education Law of 1986
  - 9 years of education compulsory
  - Education begins at age 6
  - National law, but variation across provinces in date of implementation and strength of enforcement
  - Ages 15+ at implementation date unaffected

Fang et al. (2012)

- Structural model:

$$\log(earnings)_i = \beta_0 + \beta_1 s_i + \text{ other controls } + \epsilon_i$$

- First stage:

$$S_i = \alpha_0 + \alpha_1 IV_i + + \text{ other controls } + u_i$$

- Instrument:

$$IV_i = \begin{cases} 1 & \text{if age}_i < 15 \text{ on law's effective date} \\ 0 & \text{otherwise} \end{cases}$$

Table 1: Descriptive statistics

Variable[a]	All	Control cohort[b]	Treatment cohort[b]	P value[c]
Sample size	N = 11271	N = 7380	N = 3891	
Treatment[b]	0.35	0.00	1.00	N/A
School years completed	8.88	8.66	9.28	<0.01
	(3.07)	(3.17)	(2.84)	
Yearly earnings in natural log	8.44	8.56	8.21	<0.01
	(1.22)	(1.09)	(1.39)	
Age	31.83	35.57	24.74	<0.01
	(7.12)	(4.93)	(4.88)	
Male	0.51	0.50	0.52	0.07
Race minority	0.13	0.12	0.15	<0.01
Married	0.75	0.90	0.47	<0.01
Urban	0.25	0.27	0.21	<0.01
Health status				<0.01
Excellent	0.19	0.17	0.24	
Good	0.58	0.58	0.58	
Fair	0.21	0.22	0.17	
Poor	0.02	0.02	0.01	
Province				<0.01
Heilongjiang	0.14	0.13	0.15	
Liaoning	0.07	0.08	0.07	
Jiangsu	0.12	0.12	0.13	
Shandong	0.09	0.09	0.11	
Henan	0.11	0.10	0.14	
Hubei	0.11	0.12	0.10	
Hunan	0.09	0.11	0.05	
Guangxi	0.13	0.16	0.08	
Guizhou	0.13	0.11	0.16	
CHNS wave				<0.01
1997	0.27	0.28	0.26	
2000	0.21	0.22	0.19	
2004	0.26	0.25	0.28	
2006	0.26	0.25	0.27	

Data source: China Health and Nutrition Survey (CHNS) 1997, 2000, 2004, and 2006.

[a]Standard deviations are reported in parentheses for continuous variables.

[b]The control cohort includes respondents that were not affected by the 1986 China Compulsory Education Law, and the treatment cohort includes respondents that were affected by the 1986 China Compulsory Education Law. The effective dates of the 1986 China Compulsory Education Law in the different provinces varied. We define the sample so that a treatment respondent was less than 15 years old on the law's effective date in the province where he or she lived, and a control respondent was 15 years or older on the effective date.

[c]Chi-square tests for categorical variables and students' t tests for continuous variables between the control cohort and treatment cohort.

N/A: not applicable.

Table 2: The impact of the compulsory schooling law on years of schooling: Selected results from the first stage of the 2-stage least squares estimation (2SLS)

First stage estimation in 2SLS	School years completed is the dependent variable (OLS coefficient)[a]			
	All	Two-year control and two-year treatment cohort	Two-year control cohort	Two-year treatment cohort
Age on the date the law was implemented	N/A	13 - 16 years old	15 - 16 years old	13 - 14 years old
Instrumental variable				
Less than 15 years old by the effective date	0.79***	0.66***		
(Treatment dummy of compulsory education law)	(0.11)	(0.14)		
Less than 16 years old by the effective date			0.12	
(Year dummy variable)			(0.21)	
Less than 14 years old by the effective date				0.38**
(Year dummy variable)				(0.18)
Test of excluded instruments				
F statistic	54.78***	21.85***	0.33	4.23**
Under-identification tests				
Kleibergen-Paap rk LM statistic	55.15***	21.79***	0.33	4.28**
Kleibergen-Paap rk Wald statistic	54.89***	22.06***	0.33	4.32**
Weak identification test				
Kleibegen-Paap Wald rk F statistic[b]	54.78*	21.85*	0.33	4.23
Weak-instrument-robust inference				
Anderson-Rubin Wald test: F statistic	10.69***	9.97***	0.41	0.03
Anderson-Rubin Wald test: Chi-square statistic	10.71***	10.07***	0.40	0.03
Stock-Wright LM S statistic	10.69***	9.98***	0.40	0.03

* significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

[a] All the estimations have controlled for other explanatory variables in Table 1.

[b] 10% maximal IV size as the Stock-Yogo weak ID test critical values is 16.38, and smaller maximal IV sizes are not available in Stock-Yogo (2005).

N/A: not applicable.

Table 3: The impact of the compulsory schooling law by gender and location: Selected results of the first stage estimation in 2SLS for various sub-populations

First stage estimation in 2SLS	School years completed is the dependent variable (OLS coefficient)[a]					
	Female	Male	Rural	Urban	Inland	Coastal
Instrumental variable						
Less than 15 years old by the effective date	1.17***	0.40***	0.82***	0.76***	0.72***	0.83***
(Treatment dummy of compulsory education law)	(0.15)	(0.15)	(0.16)	(0.21)	(0.12)	(0.22)
Test of excluded instruments						
F statistic	59.84***	7.36***	45.96***	21.68***	35.02***	14.43***
Under-identification tests						
Kleibergen-Paap rk LM statistic	60.00***	7.43***	46.26***	12.91***	35.34***	14.49***
Kleibergen-Paap rk Wald statistic	60.06***	7.39***	46.07***	12.78***	35.10***	14.51***
Weak identification test						
Kleibegen-Paap Wald rk F statistic[b]	59.84*	7.36	45.96*	12.68	35.02*	14.43
Weak-instrument-robust inference						
Anderson-Rubin Wald test: F statistic	3.33*	8.56***	6.47***	2.05	2.21	15.18***
Anderson-Rubin Wald test: Chi-square statistic	3.34*	8.59***	6.48***	2.06	2.22	15.26***
Stock-Wright LM S statistic	3.34*	8.54***	6.47***	2.06	2.22	15.01***

* significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

[a] All estimations have controlled for the other explanatory variables in Table 1. Descriptive statistics for these subpopulations are given in the appendix table.

[b] 10% maximal IV size as the Stock-Yogo weak ID test critical values is 16.38, and smaller maximal IV sizes are not available in Stock-Yogo (2005).

Table 4: Returns to schooling results by OLS and 2SLS for CHNS respondents born after 1961 (the "All" sample)

Variable	OLS			2SLS		
	Coeff.		S.E.	Coeff.		S.E.
School years completed	0.09	***	(0.004)	0.20	***	(0.06)
Age	0.25	***	(0.02)	0.22	***	(0.02)
Age squared	0.00	***	(0.0003)	0.00	***	(0.0003)
Male	0.21	***	(0.02)	0.15	***	(0.04)
Race minority	-0.12	***	(0.05)	-0.09	*	(0.05)
Married	-0.14	***	(0.04)	-0.06		(0.06)
Urban	0.19	***	(0.03)	-0.10		(0.16)
Health status						
Excellent (reference)						
Good	-0.04		(0.03)	-0.05		(0.03)
Fair	-0.15	***	(0.04)	-0.14	***	(0.04)
Poor	-0.34	***	(0.10)	-0.24	**	(0.12)
Province						
Heilongjiang (reference)						
Liaoning	0.11	**	(0.06)	0.03		(0.07)
Jiangsu	0.50	***	(0.04)	0.44	***	(0.05)
Shandong	0.14	***	(0.05)	0.11	**	(0.05)
Henan	-0.19	***	(0.05)	-0.20	***	(0.05)
Hubei	-0.17	***	(0.05)	-0.16	***	(0.05)
Hunan	0.02		(0.05)	-0.05		(0.07)
Guangxi	-0.04		(0.05)	-0.02		(0.05)
Guizhou	-0.21	***	(0.05)	-0.14	**	(0.07)
CHNS wave						
1997 (reference)						
2000	-0.01		(0.03)	-0.07		(0.05)
2004	0.15	***	(0.04)	0.04		(0.08)
2006	0.43	***	(0.04)	0.28	***	(0.10)
Constant	3.24	***	(0.28)	2.64	***	(0.48)

* significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 5: Robustness check on instrument

Yearly earning in natural log as the dependent variable[a]	Two-year control and Two-year treatment cohorts			Two-year control cohort			Two-year treatment cohort		
Age by the effective date	13 - 16 years old			15 - 16 years old			13 - 14 years old		
	Coeff.		S.E.	Coeff.		S.E.	Coeff.		S.E.
OLS									
School years completed	0.09	***	(0.01)	0.09	***	(0.01)	0.10	***	(0.02)
2SLS[b]									
School years completed	0.26	***	(0.09)	0.54		(1.11)	0.04		(0.22)

* significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

[a] All estimations have controlled for the other explanatory variables in Table 1.

[b] Using the instrumental variables as those in Table 3 for various study cohorts respectively.

Table 6: Selected results by gender, urbanization, and province location (the "All" sample)

The dependent variable is the annual income in natural log	Coefficients on "School years completed"					
	OLS			2SLS		
	Coeff.		S.E.	Coeff.		S.E.
By gender						
Female	0.09	***	(0.01)	0.10	*	(0.05)
Male	0.09	***	(0.01)	0.51	**	(0.23)
By urbanization						
Rural	0.09	***	(0.01)	0.18	***	(0.07)
Urban	0.08	***	(0.01)	0.14		(0.09)
By province location						
Inland provinces	0.09	***	(0.01)	0.12		(0.08)
Coastal provinces	0.09	***	(0.01)	0.37	***	(0.12)

* significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

References

References

Angrist, J.D. and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, Joshua D. and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4):pp. 979–1014. URL http://www.jstor.org/stable/2937954.

———. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15 (4):69–85. URL http://www.aeaweb.org/articles.php?doi=10.1257/jep.15.4.69.

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

Card, David. 1993. "Using geographic variation in college proximity to estimate the return to schooling." Tech. rep., National Bureau of Economic Research. URL http://www.nber.org/papers/w4483.

———. 2003. "Estimating the return to schooling: Progress on some persistent econometric problems." *Econometrica* 69 (5):1127–1160. URL http://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00237/full.

Fang, Hai, Karen N. Eggleston, John A. Rizzo, Scott Rozelle, and Richard J. Zeckhauser. 2012. "The Returns to Education in China: Evidence from the 1986 Compulsory Education Law." Working Paper 18189, National Bureau of Economic Research. URL http://www.nber.org/papers/w18189.

Lemieux, Thomas and David Card. 2001. "Education, earnings, and the 'Canadian G.I. Bill'." *Canadian Journal of Economics/Revue canadienne d'économique* 34 (2):313–344. URL http://dx.doi.org/10.1111/0008-4085.00077.

Murray, Michael P. 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspectives* 20 (4):111–132. URL http://www.aeaweb.org/articles.php?doi=10.1257/jep.20.4.111.

Stock, James H, Jonathan H Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics* 20 (4):518–529. URL http://dx.doi.org/10.1198/073500102288618658.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. South-Western.