Instrumental Variable

簡錦漢
中研院 經濟所

人文社會科學研究中心 制度與行爲研究專題中心
應用個體計量研習營

June 12, 2014

Why instrumental variables?

The Question of Causality

Why instrumental variables?

The Question of Causality

- Simply establishing a relationship between variables is rarely sufficient
- Want the effect to be considered causal
- If we have truly controlled for enough other variables, then the estimated *ceteris paribus* effect can often be considered to be causal
- Can be difficult to establish causality
- *ceteris paribus* —other things being equal—required for identification of causality.
- In practice *ceteris paribus* is difficult to attain. The IV methods are developed for that purpose.

Why instrumental variables?

Why "ceteris paribus" so difficult?

Why instrumental variables?

Why "ceteris paribus" so difficult?

- In social science, we rarely have experimental data.
- We use observational data.
- Unobservability is usual.

The Question of Causality

Some Examples

The Question of Causality

Some Examples

- **Returns to Education**
  - A model of human capital investment implies getting more education should lead to higher earnings
  - In the simplest case, this implies an equation like
    $$Earnings_i \quad = \quad \beta_0 + \beta_1 education_i \; (\; ability) + u_i \; (\; ability)$$
  - The estimate of β_1 is the return to education, but can it be considered causal?
  - While the error term u includes other factors affecting earnings, want to control for as much as possible
  - Some things are still unobserved, which can be problematic E.g., ability is omitted variables.

The Question of Causality

Some Examples

- **Effects of Fertilizer on Crop Yield**
  - Model:

$$yield_i = \alpha_0 + \alpha_1 fertilizer_i \; (\; land \;\; quality) + u_i \; (\; land \;\; quality) \quad (1)$$

  - In a nonexperimental setting the quantity of fertilizer applied may vary with the quality of land.

The Question of Causality

Some Examples

- **Law Enforcement and Crime Rate**
  - Model:

$$crime\ rate_{it} = \alpha_0 + \alpha_1 police_{it} + \alpha_2 \boldsymbol{x}_{it} + u_{it} \tag{2}$$

  - Will cities with higher crime rate spend more on law enforcement (reverse causality)?

Endogeneity

Endogeneity

$$y = \beta_0 + \beta_1 x_1 + u, \quad \mathrm{E}(u) = 0 \tag{3}$$

- Endogeneity

$$\mathrm{cov}(x_1, u) \neq 0 \implies x_1 \text{ endogenous} \tag{4}$$

- Graphically

$$\begin{array}{ccc} \boldsymbol{x} & \longrightarrow & y \\ \uparrow & \nearrow & \\ u & & \end{array} \tag{5}$$

- Consequence:
  Assume $u = \rho x_1 + v$

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + u \\ &= \beta_0 + (\beta_1 + \rho) x_1 + v \end{aligned} \tag{6}$$

$\widehat{\beta}_1 \longrightarrow \rho + \beta_1 \qquad \implies$ OLS estimate $\widehat{\beta}_1$ inconsistent

Possible causes of endogeneity

- Measurement error (error-in-variables)
- Simultaneity
- Omitted variables

Possible causes of endogeneity

Measurement errors (error-in-variables)

Possible causes of endogeneity

Measurement errors (error-in-variables)

$$y_i^* = \beta_0 + \beta_1 x_i^* + \epsilon_i, \quad \mathrm{E}(\epsilon_i | x_i^*) = 0 \tag{7}$$

- If $\{y_i^*, x_i^*\}$ not observable, but observe $\{y_i, x_i\}$

$$y_i = y_i^* + v_i, \qquad x_i = x_i^* + u_i \tag{8}$$

assuming $\mathrm{E}(v_i) = 0,\ \mathrm{E}(v_i \epsilon_i) = 0,\ \ \mathrm{Corr}\big(u_i, x_i^*\big) = 0$

$$\mathrm{E}(u_i) = 0,\ \mathrm{E}(u_i \epsilon_i) = 0,\ \ \mathrm{Corr}\big(v_i, x_i^*\big) = 0$$

Possible causes of endogeneity

Measurement errors (error-in-variables)

- Model with mismeasured variables
  [plugging $y_i = y_i^* - v_i$ and $x_i = x_i^* - u_i$ into (7)]

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad e_i = \epsilon_i - v_i - \beta_1 u_i \tag{9}$$

- Endogeneity: $\mathrm{E}(e_i x_i) \neq 0$
  $\mathrm{E}(e_i x_i) = \mathrm{E}(\epsilon_i x_i) - \mathrm{E}(v_i x_i) - \mathrm{E}(\beta_1 u_i x_i) = -\beta\sigma_u^2$

- We can show that $\mathrm{E}\big(\widehat{\beta}\big)_1 = \beta_1 \dfrac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_u^2} \quad \Longrightarrow \big|\mathrm{E}\big(\widehat{\beta}_1\big)\big| < |\beta_1|$

- Bias toward zero — attenuation bias

Possible causes of endogeneity

Simultaneity

Possible causes of endogeneity

Simultaneity

Model

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 y_2 + \beta_2 x_1 + u && (10)\\
y_2 &= \alpha_0 + \alpha_1 y_1 + \alpha_2 x_2 + v && (11)
\end{aligned}
$$

E.g., demand & supply (y_1=quantity, y_2=price)

- We have $\mathrm{E}(y_2 u) \neq 0$ and $\mathrm{E}(y_1 v) \neq 0$ (even if $\mathrm{Corr}(u, v) = 0$)
  - Solving for y_2, we have $y_2 = \frac{1}{1-\alpha_1\beta_1}(\alpha_0 + \alpha_1\beta_0 + \alpha_1\beta_2 x_1 + \alpha_2 x_2 + \alpha_1 u + v)$
  - Assuming x_1 and x_2 exogenous and $\mathrm{E}(uv) = 0$, $\mathrm{E}(y_2 u) = \alpha_1 \sigma_u^2$

Possible causes of endogeneity

Omitted variables

Possible causes of endogeneity

Omitted variables

$$y = \beta_0 + \beta x + \alpha w + u \tag{12}$$

Assuming $\mathrm{E}(ux) = 0$, $\mathrm{E}(uw) = 0$, $\mathrm{E}(u) = 0$

- If w not observed and $\mathrm{E}(xw) \neq 0$, say $w = \rho x + e$
- Estimable model

$$y = \beta_0 + \beta^+ x + v, \quad v = \alpha e + u \tag{13}$$

- OLS Bias

$$\mathrm{E}\left(\widehat{\beta}^+\right) = \beta + \rho\alpha$$

OLS estimator

Endogenous variable

$$y_i = \beta x_i + u_i, \quad \mathrm{E}(x_i u_i) \neq 0 \tag{14}$$

- OLS estimation of β

$$\begin{aligned} \widehat{\beta}_{OLS} &= \left(\sum x_i y_i\right) / \left(\sum x_i^2\right) & (15) \\ &= \beta + \left(\sum x_i u_i\right) / \left(\sum x_i^2\right) & (16) \end{aligned}$$

- OLS biasness: $\lim_{n \longrightarrow \infty} \widehat{\beta}_{OLS} \neq \beta$

OLS estimator

Other variables — Are they biased?

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \quad \mathrm{E}(x_{1i} u_i) \neq 0, \mathrm{E}(x_{2i} u_i) = 0 \tag{17}$$

- OLS estimation of β_1 and β_2 (x_{2i} exogenous.)

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{OLS} &= (\sum \boldsymbol{x}_i' \boldsymbol{x}_i')^{-1} (\sum \boldsymbol{x}_i' y_i) & (18) \\ &= \boldsymbol{\beta} + (\sum \boldsymbol{x}_i' \boldsymbol{x}_i)^{-1} (\sum \boldsymbol{x}_i' u_i) & (19) \end{aligned}$$

  - $\boldsymbol{x}_i = [1 \quad x_{1i} \quad x_{2i}]'$, $\quad \boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2]'$; assuming $\mathrm{E}(\boldsymbol{x}_i' u_i) = [0 \quad \rho \quad 0]'$

- Bias: $\lim\limits_{n \longrightarrow \infty} \widehat{\beta}_{2OLS} \neq \beta_2$
  (2nd element of last column of $(\sum \boldsymbol{x}_i' \boldsymbol{x}_i)^{-1}$ not zero)

  "Smearing effect"

The IV Approach

- Valid instruments z:
  1. Exogeneity: $\boldsymbol{z}$ uncorrelated with u, $\mathrm{E}(u|\boldsymbol{z}) = 0 \implies \boldsymbol{z}$ exogenous

     I.e., $\boldsymbol{z}$ not in the model of interest (exclusion restriction)

  2. Relevance: $\boldsymbol{z}$ is correlated with endogenous variable x_1 conditional on other regressors $(x_2, \ldots, x_K)$

     I.e., $x_1 = \delta_0 + \delta_2 x_2 + \delta_3 x_3 + \ldots + \delta_K x_K + \boldsymbol{\theta}\boldsymbol{z} + r_1$, where $\boldsymbol{\theta} \neq 0$

Using IV

Just identified model — One IV, one endogenous variable

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \quad \mathrm{E}(x_{1i}u_i) \neq 0, \ \mathrm{E}(x_{2i}u_i) = 0 \tag{20}$$

$$x_{1i} = \delta_0 + \delta_2 x_{2i} + \theta z_{1i} + r_i, \qquad z_{1i} \text{ is instrument} \tag{21}$$

- Plug (21) (i.e., the reduced form) into (20):

$$\begin{aligned} y &= \beta_0 + \beta_1\delta_0 + (\beta_2 + \beta_1\delta_2)x_{2i} + \beta_1\theta z_{1i} + u_i + \beta_1 r_i \\ &= \alpha_0 + \alpha_2 x_{i2} + \lambda z_{1i} + v_i \end{aligned} \tag{22}$$

where $v_i = u_i + \beta_1 r_i, \quad \alpha_2 = \beta_2 + \beta_1\delta_2, \quad \lambda = \beta_1\theta$

Using IV

Just identified model — One IV, one endogenous variable

- OLS estimation of (22) and (21) yielding $\widehat{\theta}$ and $\widehat{\lambda}$, $\widehat{\beta}_1 = \widehat{\lambda}/\widehat{\theta}$
- Intuition: $\beta_1 = \dfrac{\partial y_i}{\partial x_{1i}} = \dfrac{\partial y_i}{\partial z_{1i}} \Big/ \dfrac{\partial x_{1i}}{\partial z_{1i}}; \qquad \dfrac{\partial y_i}{\partial z_{1i}} = \lambda, \quad \dfrac{\partial x_i}{\partial z_{1i}} = \theta$
- $$\begin{aligned} \boldsymbol{\beta}_{iv} &= (\textstyle\sum \boldsymbol{z}'_i x_i)^{-1} (\textstyle\sum \boldsymbol{x}'_i y_i) \\ &= \boldsymbol{\beta} + \cancel{(\textstyle\sum \boldsymbol{z}'_i x_i)^{-1} (\textstyle\sum \boldsymbol{z}'_i u_i)}, \qquad \boldsymbol{z}_i = [1 \ x_{2i} \ z_{1i}]' \end{aligned}$$

$\Longrightarrow \mathrm{E}(\boldsymbol{\beta}_{iv}) = \boldsymbol{\beta}$, since $\mathrm{E}\left(\sum \boldsymbol{z}'_i u_i\right) = 0$ by assumption

Using IV

Over identified model — more IVs than endogenous variables

$$y_i \quad = \quad \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \quad \mathrm{E}(x_{1i} u_i) \neq 0, \mathrm{E}(x_{2i} u_i) = 0 \tag{23}$$

- How to use two instruments $\{z_{1i}, z_{2i}\}$ for estimation?
- Define $\boldsymbol{x}_i = [1 \;\; x_{1i} \;\; x_{2i}]'$ and $\boldsymbol{z}_i = [1 \;\; x_{2i} \;\; z_{1i} \;\; z_{2i}]'$
  (i) Projection of $\boldsymbol{x}_i$ in the space of $\boldsymbol{z}_i$: $\widehat{\boldsymbol{x}}_i = \boldsymbol{z}_i \left[(\sum \boldsymbol{z}_i' z_i)^{-1} (\sum \boldsymbol{z}_i' \boldsymbol{x}_i) \right]$

  (ii) Use $\widehat{\boldsymbol{z}}_i$ as instruments: $\beta_{2SLS} = (\sum \widehat{\boldsymbol{x}}_i{}' \boldsymbol{x}_i)^{-1} (\sum \widehat{\boldsymbol{x}}_i{}' y_i)$;

  (iii) Equivalently $\beta_{2SLS} = (\sum \widehat{\boldsymbol{x}}_i{}' \widehat{\boldsymbol{x}}_i)^{-1} (\sum \widehat{\boldsymbol{x}}_i{}' y_i)$ (note: $\widehat{\boldsymbol{x}}_i = [1 \;\; \widehat{x}_{1i} \;\; x_{2i}]'$)
- Two-stage least squares

$$x_{1i} = \delta_0 + \delta_2 x_2 + \theta_1 z_{1i} + \theta_2 z_{2i} + r_i \tag{24}$$

$$y_i = \beta_0 + \beta_1 \widehat{x}_{1i} + \beta_2 x_{2i} + u_i; \; \widehat{x}_{1i} = \widehat{\delta}_0 + \widehat{\delta}_2 x_{2i} + \widehat{\theta}_1 z_{1i} + \widehat{\theta}_2 z_{2i} \tag{25}$$

Using IV

Standard errors of 2SLS estimators

- The standard errors from doing 2SLS by hand are incorrect. Why?
  - Model: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, x_{1i} endogenous
  - Correct residual: $\widehat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{1i} - \widehat{\beta}_2 x_{2i}$
  - Estimate by hand: $\widehat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 \widehat{x}_{1i} - \widehat{\beta}_2 x_{2i}$
- Method extends to multiple endogenous variables—need at least as many instruments as there are endogenous variables (i.e., the *Order Condition*)

Using IV

IV versus OLS estimation

- Model: $y = \beta_0 + \beta_1 x + u$, instrument z
- Homoskedasticity: $\mathrm{E}(u^2|\boldsymbol{x}) = \sigma^2 = \mathrm{Var}(u)$
- Given the asymptotic variance, we can estimate the standard error

$$\widehat{\mathrm{Var}}(\widehat{\beta}_{IV}) = \frac{\widehat{\sigma}^2}{SST_x R^2_{x,z}}$$

  - Standard error in IV case differs from OLS only in the R^2 from regressing x on z
  - Since $R^2 < 1$, IV standard errors are larger
  - But IV is consistent, while OLS is inconsistent, when $\mathrm{Cov}(x, u) \neq 0$
  - The stronger the correlation between z and x, the smaller the IV standard errors

Using IV

Negative R^2

- Standard Formula: $R^2 = 1 - \frac{SSR}{SST}$.
- R^2 from IV estimation can be negative.
- Consider this: Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u_1$
  x_1 endogenous, z_i is instrument

  IV Estimation: $y = \beta_0 + \beta_1 \widehat{x}_1 + \beta_2 x_2 + u_1 + \beta_1 v_2$
  $x_1 = \delta_0 + \delta_1 x_2 + \delta_2 z_1 + v_2$

  — $\mathrm{var}(y) > \mathrm{var}(u_1)$

  — $\mathrm{var}(y) \overset{?}{>} \mathrm{var}(u_1 + \beta_1 v_2)$

Potential Pitfalls with 2SLS

Weak Instruments—Bias

- Model: $y = \beta_0 + \beta x_1 + u$ (instrument: z_1)
  We have:

$$\begin{aligned} \text{plim}\widehat{\beta}_{2sls} &= \beta + \text{cov}(z_1, u) / \text{cov}(z_1, x_1,) \\ &\quad \text{consistent only if } \text{cov}(z_1, u) = 0. \end{aligned}$$

- Consider the case when $\text{cov}(z_1, u) \neq 0$.

$$\text{plim}\widehat{\beta}_{2sls} = \beta + (\sigma_u / \sigma_{x1})[\text{Corr}(z_1, u) / \text{Corr}(z_1, x_1)] \tag{26}$$

  The bias of $\widehat{\beta}_{2sls}$ increases as $\text{Corr}(z_1, x_1)$ approaches zero, especially in finite sample.

Potential Pitfalls with 2SLS

Weak Instrument—Large Standard Error

- Consider: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + u$, x_K is endogenous. Obtain $\widehat{\boldsymbol{\beta}}$ using IV $\boldsymbol{z}$:

$$\text{Avar}(\widehat{\beta}_K) \approx \frac{\widehat{\sigma}^2}{\widehat{\text{SSR}}_K} \qquad \text{Note: Avar}\left(\widehat{\boldsymbol{\beta}}_{2sls}\right) = \widehat{\sigma}^2 \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1} \tag{27}$$

where $\widehat{\text{SSR}}_K$ is the Sum of Squared Residuals of the regression of

$$\widehat{x}_K = \boldsymbol{x}_{K-1}\boldsymbol{\delta} + error; \quad \text{where} \quad \boldsymbol{x}_{K-1} \equiv \{1, x_1, \ldots x_{K-1}\} \tag{28}$$

Potential Pitfalls with 2SLS

Weak Instrument—Large Standard Error

- Consider: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + u$, x_K is endogenous. Obtain $\widehat{\boldsymbol{\beta}}$ using IV $\boldsymbol{z}$:

$$\text{Avar}(\widehat{\beta}_K) \approx \frac{\widehat{\sigma}^2}{\widehat{\text{SSR}}_K} \qquad \text{Note: Avar}\left(\widehat{\boldsymbol{\beta}}_{2sls}\right) = \widehat{\sigma}^2 \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1} \tag{27}$$

where $\widehat{\text{SSR}}_K$ is the Sum of Squared Residuals of the regression of

$$\widehat{x}_K = \boldsymbol{x}_{K-1}\boldsymbol{\delta} + error; \quad \text{where} \quad \boldsymbol{x}_{K-1} \equiv \{1, x_1, \ldots x_{K-1}\} \tag{28}$$

$$\widehat{\text{SSR}}_K = \widehat{\text{SST}}_K(1 - \widehat{R}_K) \tag{29}$$

Potential Pitfalls with 2SLS

Weak Instrument—Large Standard Error

- Consider: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + u$, x_K is endogenous. Obtain $\widehat{\boldsymbol{\beta}}$ using IV $\boldsymbol{z}$:

$$\text{Avar}(\widehat{\beta}_K) \approx \frac{\widehat{\sigma}^2}{\widehat{\text{SSR}}_K} \qquad \text{Note: Avar}\left(\widehat{\boldsymbol{\beta}}_{2sls}\right) = \widehat{\sigma}^2 \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1} \tag{27}$$

where $\widehat{\text{SSR}}_K$ is the Sum of Squared Residuals of the regression of

$$\widehat{x}_K = \boldsymbol{x}_{K-1}\boldsymbol{\delta} + error; \quad \text{where} \quad \boldsymbol{x}_{K-1} \equiv \{1, x_1, \ldots x_{K-1}\} \tag{28}$$

$$\widehat{\text{SSR}}_K = \widehat{\text{SST}}_K (1 - \widehat{R}_K) \tag{29}$$

where $\widehat{\text{SST}}_K = \sum_{i=1}^{N} (\widehat{x}_{iK} - \overline{\widehat{x}}_K)$, measures $\widehat{x}_{iK}$'s total variation.
$\widehat{R}_K$ = the R^2 from regression (28).
$1 - \widehat{R}_K^2$ measures multicollinearity.

Potential Pitfalls with 2SLS

Weak Instrument—Large Standard Error

- Large Avar($\widehat{\beta}_K$) —when $\widehat{\mathrm{SST}}_K$ is small, or $\widehat{R}_K^2$ close to 1.
  1. Corr($x_{iK}, \boldsymbol{z}$) is small. ($\widehat{\mathrm{SST}}_K$ is the variation of x_K explained by $\boldsymbol{z}$).
  2. Poor instruments: $\widehat{R}_K^2 \longrightarrow 1$ (i.e., $\widehat{x}_K$ mostly a function of $\boldsymbol{x}_{K-1}$ only).

Potential Pitfalls with 2SLS

Weak Instrument—Large Standard Error

- Consider an example
  - Endogenous variable: x_K
  - Instrument: z_1, $\boldsymbol{z} \equiv \{1, x_1, \ldots, x_{K-1}, z_1\}$
  - Regression: $x_K = \delta_0 + \delta_1 x_1 + \ldots + \delta_{K-1} x_K + \theta_1 z_1 + r \implies \widehat{x}_K$
  - Regression: $\widehat{x}_K = \boldsymbol{x}_{K-1} \boldsymbol{\alpha} + q \implies \widehat{R}_K^2$
  - $\theta_1 = 0 \implies \widehat{x}_K$ is function of $\boldsymbol{x}_{K-1}$ only.
    $\implies \widehat{R}_K^2 = 1$

Possible Cures for Weak Instruments

Jackknife Bias Correction

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{J2SLS} &= n\left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\boldsymbol{X})\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'Y \\
&\quad -\textstyle\sum_i^n \frac{n-1}{n}\left[\boldsymbol{X}'_{-i}\boldsymbol{Z}_{-i}(\boldsymbol{Z}'_{-i}\boldsymbol{Z}_{-i})^{-1}(\boldsymbol{Z}'_{-i}\boldsymbol{X}_{-i})\right]^{-1}\boldsymbol{X}'_{-i}\boldsymbol{Z}_{-i}(\boldsymbol{Z}'_{-i}\boldsymbol{Z}_{-i})^{-1}\boldsymbol{Z}'_{-i}Y_{-i} \\
\boldsymbol{X}_{-i} &= \{\boldsymbol{x}_1,\ldots\boldsymbol{x}_{i-1},\boldsymbol{x}_{i+1},\ldots\boldsymbol{x}_{1n}\} \;\; (\boldsymbol{x}_i \text{ omitted}) \\
Y_{-i} &= \{y_1,\ldots y_{i-1},y_{i+1},\ldots y_n\} \;\; (y_i \text{ omitted}) \\
\boldsymbol{Z}_{-i} &= \{\boldsymbol{z}_{11},\ldots\boldsymbol{z}_{1i-1},\boldsymbol{z}_{1i+1},\ldots\boldsymbol{z}_{1n}\} \;\; (\boldsymbol{z}_{1i} \text{ omitted})
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{var}\left(\sqrt{n}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_{J2SLS})\right) &= \frac{\sigma_u^2}{\Theta}+\frac{L}{n}\frac{\sigma_u^2\sigma_v^2+\sigma_{uv}}{\Theta^2}, \\
\Theta &= (\boldsymbol{Z}'\boldsymbol{\pi})'(\boldsymbol{Z}'\boldsymbol{\pi})/n, \\
\boldsymbol{x}_i &= \boldsymbol{z}_i\boldsymbol{\pi}+v_i
\end{aligned}
$$

Hahn, J.; J. Hausman, G. Kuersteiner (2004), "Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations," *Econometrics Journal*, volume 7, pp. 272–306.

Possible Cures for Weak Instruments

k-Class Estimators

$$\widehat{\boldsymbol{\beta}}(k) = [\boldsymbol{Y}'(\boldsymbol{I} - k\boldsymbol{M_Z})\boldsymbol{Y}]^{-1}[\boldsymbol{Y}'(\boldsymbol{I} - k\boldsymbol{M_Z})y], \quad \boldsymbol{M_Z} = \boldsymbol{I} - \boldsymbol{P_Z} \tag{30}$$

1. Limited Information Estimator:
   $k = k_{LIML}$: smallest root of $\det(\boldsymbol{Y}'\boldsymbol{Y} - k\boldsymbol{Y}'\boldsymbol{M_Z}\boldsymbol{Y}) = 0$
2. Fuller's (1977) Modified LIML:
   $k = k_{LIML} - b/(T - K_2),\ b = 1$
3. Bias-Adjusted 2SLS
   $k = k_{B2SLS} = T/(T - K + 2)$

Possible Cures for Weak Instruments

k-Class Estimators

Staiger and Stock (1997), "Instrumental Variables Regression with Weak Instruments," Econometrica

TABLE II

ESTIMATED EFFECTS OF YEARS OF EDUCATION ON LOG WEEKLY EARNINGS IN THE 1980 CENSUS

	I	II	III	IV
A. *Men Born* 1930–39 (n = 329, 509)				
OLS	.0632	.0632	.0632	.0628
(S.E.)	(.0003)	(.0003)	(.0003)	(.0003)
TSLS	.0990	.0806	.0600	.0811
(S.E.)	(.0207)	(.0164)	(.0290)	(.0109)
LIML	.0999	.0838	.0574	.0982
(S.E.)	(.0210)	(.0179)	(.0385)	(.0153)
A-R Confidence Interval	[.052, .153]	[−.003, .179]	[−.441, .490]	[−.015, .240]
TSLS Bonferroni Confidence Interval	[.052, .152]	[.038, .137]	[$-\infty$, $+\infty$]	[.048, .172]
LIML Bonferroni Confidence Interval	[.052, .153]	[.036, .134]	[$-\infty$, $+\infty$]	[.043, .158]
F (first stage)	30.53	4.747	1.613	1.869
{p-value}	{.000}	{.000}	{.021}	{.000}
Durbin test, TSLS	3.087	1.126	0.013	2.853
{p-value}	{0.079}	{.289}	{.910}	{.091}
Basmann test, LIML	2.318	22.45	19.55	161.1
{p-value}	{.314}	{.801}	{.849}	{.800}
B. *Men Born* 1940–49 (n = 486,926)				
OLS	.0520	.0520	.0520	.0516
(S.E.)	(.0003)	(.0003)	(.0003)	(.0003)
TSLS	−.0734	.0393	.0779	.0666
(S.E.)	(.0273)	(.0145)	(.0239)	(.0113)
LIML	−.0902	.0286	.1243	.0878
(S.E.)	(.0301)	(.0197)	(.0420)	(.0178)
A-R Confidence Interval	[∅]	[∅]	[∅]	[.033, .148]
TSLS Bonferroni Confidence Interval	[−.155, −.018]	[−.004, .076]	[.000, .219]	[.027, .150]
LIML Bonferroni Confidence Interval	[−.174, −.028]	[−.023, .079]	[−.009, .290]	[.023, .156]
F (first stage)	26.32	6.849	2.736	1.929
{p-value}	{.000}	{.000}	{.000}	{.000}
Durbin test, TSLS	28.90	0.780	1.188	1.780
{p-value}	{.000}	{.377}	{.276}	{.182}
Basmann test, LIML	9.356	93.29	49.22	200.36
{p-value}	{.009}	{.000}	{.006}	{.110}
Controls				
Race, Standard Metropolitan Statistical Area, Married, Region, Year of Birth Dummies	yes	yes	yes	yes
Age, Age^2	no	no	yes	yes
State of Birth	no	no	no	yes
Instruments				
Quarter of birth	yes	yes	yes	yes
Quarter of birth *(year of birth)	no	yes	yes	yes
Quarter of birth *(state of birth)	no	no	no	yes
# Instruments	3	30	28	178

Testing for Weak IV

One Endogenous Variable

1. Rule of Thumb (Staiger and Stock, 1997)
   F-statistic > 10 in first stage regression

2. Relative Bias of $\boldsymbol{\beta}_{2SLS}$ less than 10% (Stock and Yogo, 2005),
   i.e., $B = \left| \dfrac{\mathrm{E}\left(\widehat{\beta}^{IV} - \beta\right)}{\mathrm{E}\left(\widehat{\beta}^{OLS} - \beta\right)} \right| < 0.1$

Number of Instruments K_2	F-statistic Critical Value at 5%
3	9.08
4	10.27
5	10.83
6	11.12
7	11.29
8	11.39
9	11.46
10	11.49
15	11.51
20	11.45

Testing for endogeneity

Hausman Test

- Relationship between OLS and 2SLS estimators:
  Exogeneity: $\widehat{\beta}_{1,\text{OLS}} = \widehat{\beta}_{1,\text{2SLS}}$, but $\widehat{\beta}_{1,\text{2SLS}}$ is inefficient
  Endogeneity: $\widehat{\beta}_{1,\text{OLS}} \neq \widehat{\beta}_{1,\text{2SLS}}$, and $\widehat{\beta}_{1,\text{OLS}}$ is biased

- Testing for the difference in the OLS and 2SLS estimators of the coefficient of the endogenous regressor: $\widehat{\beta}_{1,\text{2SLS}}$ and $\widehat{\beta}_{1,\text{OLS}}$.

$$\text{var}\left(\widehat{\beta}_{1,\text{2SLS}} - \widehat{\beta}_{1,\text{OLS}}\right) = \text{var}\left(\widehat{\beta}_{1,\text{2SLS}}\right) - \text{var}\left(\widehat{\beta}_{1,\text{OLS}}\right) \tag{31}$$

- Test Statistic:

$$\tau_{\text{Hausman}} = \frac{\left(\widehat{\beta}_{1,\text{2SLS}} - \widehat{\beta}_{1,\text{OLS}}\right)^2}{\text{var}\left(\widehat{\beta}_{1,\text{2SLS}}\right) - \text{var}\left(\widehat{\beta}_{1,\text{OLS}}\right)} \tag{32}$$

NOTE: valid only under homoskedasticity.

Testing for endogeneity

Regression Based Test (control function)

Structural Model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u_1, \quad \mathrm{Corr}(x_1, u_1) \neq 0 \tag{33}$$

Reduced Form Model:

$$x_1 = \pi_0 + \pi_1 x_2 + \pi_2 z_1 + v_2, \quad z_1 \text{ is instrument} \tag{34}$$

Testing for $\mathrm{cov}(u_1, v_2) = 0$, assume $u_1 = \rho_1 v_2 + e_1$

- $\rho_1 = 0$, if and only if x_1 is exogenous

$$y = \alpha_0 + \alpha_1 x_1 + x_2 \alpha_2 + \rho_1 \widehat{v}_2 + e_1 \tag{35}$$

- We can use the OLS t-statistic for testing.
- Even though $\widehat{v}_2$ is a generated regressor, OLS t-statistic is correct under H: $\rho_1 = 0$.

Validity of IV

Overidentifying restriction test

- If there is just one instrument for our endogenous variable, we can't test whether the instrument is uncorrelated with the error
- The model is just identified
- If we have multiple instruments, it is possible to test the overidentifying restrictions—to see if some of the instruments are correlated with the error

Validity of IV

Overidentifying restriction test

- Estimate the structural model using IV and obtain the residuals
- Regress the residuals on all the exogenous variables and obtain the R^2 to form nR^2
- Under the null that all instruments are uncorrelated with the error, LM$\sim \chi^2_q$ where q is the number of extra instruments

Validity of IV

Overidentifying restriction test

- Model:

$$y_1 = \delta_1 z_1 + \alpha_1 y_2 + u_1;$$
$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2 \quad (\textit{overidentified})$$

- Testing Overidentifying Restrictions
  $\mathrm{H}_0: \ \mathrm{cov}(z_2, u_1) = 0$ and $\mathrm{cov}(z_3, u_1) = 0$
  (i.e., validity of exclusion restriction)

- Regression-Based Test:
  1. Obtain $\widehat{u}_1$ by 2SLS using all instruments.
  2. Regress $\widehat{u}_1$ on $\boldsymbol{z}$ (all exogenous variables)
  3. Test Statistic: NR_u^2
  4. Under the H_0: $\mathrm{Corr}(z_2, u_1) = 0$ & $\mathrm{Corr}(z_3, u_1) = 0$: $NR_u^2 \overset{a}{\sim} \chi_2^2$.

Validity of IV

Can we really test the IV's validity

- How does overidentification restriction test work?
  - Compare estimates with different instruments
- Can we justify the validity of z_2 and z_3 solely by overidentification restriction test?
  - No! When the null hypothesis is accepted, all instruments are either equally good or equally bad.
  - If all of the instruments z_2 and z_3 are of similar nature, it is difficult for overidentification restriction test to tell whether they are valid or not.
- Example: in Levitt (1997) all instruments pertain to election cycles.

Validity of IV

Can we really test the IV's validity

- Adding more instruments is good if these additional instruments have explanatory power towards the endogenous variable.
  - This will reduce the standard error.
  - If the additional instruments do no increase the R^2 in the first-stage regression, adding them will increase the bias.
- Must use intuition, theory or institutional settings to argue for the validity of the instruments (internal validity). E.g., IVs arise from reforms/policy changes (i.e., natural experiments).
- Getting similar results using different instruments increases the credibility of the instruments.
- Conclusion: We can never confirm an instrument's validity.

Application of instrumental variables

Return to Education

- Objective: estimate the effect of education (*educ*) on log of wage (denoted log(*wage*)).

$$\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + u \qquad (36)$$

Application of instrumental variables

Return to Education

- Potential instrument 1: *fatheduc* father's education.
  1. *fathereduc* correlated with *educ*
  2. but *fathereduc* may be correlated with omitted factors in u, e.g., one's ability, family background, etc.

Application of instrumental variables

Return to Education

- Potential instrument for *educ*: *first_quarter* (i.e., whether born in the first quarter of ther year). See Angrist, Joshua D. and Alan B. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *QJE*, 106(4), 979–1014.

$$\begin{aligned} educ &= \delta_0 + \delta_1 + \delta_2 gender + \delta_3 exper \\ &\quad + \delta_4 exper^2 + \theta_1 first\_quarter + r \end{aligned} \tag{37}$$

1. *first_quarter* correlated with educ because of compulsory education (i.e., child under a certain age must be in school)

2. *first_quarter* is unlikely to be correlated with other omitted factors affecting educ

Application of instrumental variables

Return to Education

TABLE IV

OLS and TSLS Estimates of the Return to Education for Men Born 1920–1929: 1970 Census[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0802	0.0769	0.0802	0.1310	0.0701	0.0669	0.0701	0.1007
	(0.0004)	(0.0150)	(0.0004)	(0.0334)	(0.0004)	(0.0151)	(0.0004)	(0.0334)
Race (1 = black)	—	—	—	—	0.2980	−0.3055	−0.2980	−0.2271
					(0.0043)	(0.0353)	(0.0043)	(0.0776)
SMSA (1 = center city)	—	—	—	—	0.1343	0.1362	0.1343	0.1163
					(0.0026)	(0.0092)	(0.0026)	(0.0198)
Married (1 = married)	—	—	—	—	0.2928	0.2941	0.2928	0.2804
					(0.0037)	(0.0072)	(0.0037)	(0.0141)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region of residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	0.1446	0.1409	—	—	0.1162	0.1170
			(0.0676)	(0.0704)			(0.0652)	(0.0662)
Age-squared	—	—	−0.0015	−0.0014	—	—	−0.0013	−0.0012
			(0.0007)	(0.0008)			(0.0007)	(0.0007)
χ^2 [dof]	—	36.0 [29]	—	25.6 [27]	—	34.2 [29]	—	28.8 [27]

a. Standard errors are in parentheses. Sample size is 247,199. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the State, County, and Neighborhoods 1 percent samples of the 1970 Census (15 percent form). The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

Application of instrumental variables

Return to Education

- Potential instrument for educ: *born*55 (i.e., aged 12 or younger in 1968—affected by Taiwan's compulsory 9-year education, enacted in 1968). This instrument arises from **Natural Experiment**. See Spohr, Chris A. (2003) "Formal schooling and workforce participation in a rapidly developing economy: evidence from "compulsory" junior high school in Taiwan" *JDE*, 70(2), 291–327.

$$educ = \delta_0 + \delta_1 age + \delta_2 age^2 + \theta_1 born1955 + r \tag{38}$$

1. *born*1955 correlated with educ because children born after 1956 have at least 9 years of education.

2. *born*1955 is likely to be uncorrelated with other omitted factors affecting educ

Application of instrumental variables

Return to Education

Table 6
OLS and 2SLS estimates of returns to schooling for all positive income respondents

	Males			Females		
	OLS	IV using cohort deviations		OLS	IV using cohort deviations	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is log(annual earnings from any type of work)						
EDYRS	0.054	0.025	0.058	0.095	0.164	0.167
	(0.000)	(0.022)	(0.026)	(0.001)	(0.052)	(0.052)
Control for propensity score, p.s. squared:	No	No	Yes	No	No	Yes

Regressions control for the full set of age-specific dummies (dropping age 36) and transformed survey year dummies (dropping 1979–1980, see Appendix B); a homogeneous cohort trend for GDP growth and inflation; controls for cohort size, squared-size.
Standard errors displayed in parentheses are robust to heteroscedasticity.

Application of instrumental variables

Children and Labor Supply

- Objective: Estimate the effect of number of children on labor supply

$$\log(hours) = \alpha\, children + \boldsymbol{x}\boldsymbol{\beta} + u \tag{39}$$

Potential Instrument: gender mix of the first two children, i.e, *gender_mix* (whether the first two child are of the same sex).

$$children = \boldsymbol{x}\boldsymbol{\delta} + \theta_1 gender\_mix + r \tag{40}$$

1. *children* and *gender_mix* are correlated because people prefer to have children of mixed genders ($\theta_1 > 0$).
2. *gender_mix* is exogenous, not decided by the parents.
3. See Joshua D. Angrist and William N. Evans (1998) "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *AER*, 88(3), 450–477.

Application of instrumental variables

Children and Labor Supply

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:									
Worked for pay	−0.176 (0.002)	−0.120 (0.025)	−0.113 (0.025) [0.013]	−0.167 (0.002)	−0.120 (0.028)	−0.113 (0.028) [0.013]	−0.008 (0.001)	0.004 (0.009)	0.001 (0.008) [0.013]
Weeks worked	−8.97 (0.07)	−5.66 (1.11)	−5.37 (1.10) [0.017]	−8.05 (0.09)	−5.40 (1.20)	−5.16 (1.20) [0.071]	−0.82 (0.04)	0.59 (0.60)	0.45 (0.59) [0.030]
Hours/week	−6.66 (0.06)	−4.59 (0.95)	−4.37 (0.94) [0.030]	−6.02 (0.08)	−4.83 (1.02)	−4.61 (1.01) [0.049]	0.25 (0.05)	0.56 (0.70)	0.50 (0.69) [0.71]
Labor income	−3768.2 (35.4)	−1960.5 (541.5)	−1870.4 (538.5) [0.126]	−3165.7 (42.0)	−1344.8 (569.2)	−1321.2 (565.9) [0.703]	−1505.5 (103.5)	−1248.1 (1397.8)	−1382.3 (1388.9) (0.549)
ln(*Family income*)	−0.126 (0.004)	−0.038 (0.064)	−0.045 (0.064) [0.319]	−0.132 (0.004)	−0.051 (0.056)	−0.053 (0.056) [0.743]	—	—	—
ln(*Non-wife income*)	—	—	—	−0.053 (0.005)	0.023 (0.066)	0.016 (0.066) [0.297]	—	—	—

Notes: The table reports estimates of the coefficient on the *More than 2 children* variable in equations (4) and (6) in the text. Other covariates in the models are *Age, Age at first birth,* plus indicators for *Boy 1st, Boy 2nd, Black, Hispanic,* and *Other race.* The variable *Boy 2nd* is excluded from equation (6). The *p*-value for the test of overidentifying restrictions associated with equation (6) is shown in brackets. Standard errors are reported in parentheses.

Application of instrumental variables

Police and Crime

- Objective: Estimate the effect of law enforcement on crime rate

$$\text{crime}_{it} = \alpha \ \text{police}_{it} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \eta_i + \epsilon_{it} \tag{41}$$

  Identification Problem: Cities with more crimes needs more police officers–reverse causality (endogeneity).

- Strategy 1: Use electoral cycles as instruments [Levitt, S. (1997) "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime" *AER*, 87, 270–290; and comment by McCrary, Justin, (2001, *American Economic Review*)].
  - First Stage:

$$\begin{aligned} \Delta \ \text{police}_{it} &= \text{Mayoral Election}_{it}\gamma_1 + \text{Gubernatorial Election}_{it}\gamma_2 \\ &\quad + \boldsymbol{x}_{it}\boldsymbol{\theta} + \delta_i + \phi_t + u_{it} \end{aligned} \tag{42}$$

  - Second Stage:

$$\Delta \ \text{crime}_{it} = \Delta \ \widehat{\text{police}}_{it}\alpha + \boldsymbol{x}_{it}\boldsymbol{\beta} + \eta_i + \kappa_t + u_{it} \tag{43}$$

Application of instrumental variables

Police and Crime

- Strategy 2: Use public sector employees as instruments (Levitt, S. (2002) “Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Reply” *AER*, 92, 1244–1250.)
  - First Stage:

$$\text{police}_{it} = \text{ firefighter}_{it}\gamma_1 + \text{ municipal workers}_{it}\gamma_2 + \boldsymbol{x}_{it}\boldsymbol{\theta} + \delta_i + \phi_t + u_{it}$$

  - Second Stage:

$$\Delta\text{crime}_{it} = \widehat{\text{police}}_{it-1}\alpha + \boldsymbol{x}_{it}\boldsymbol{\beta} + \eta_i + \kappa_t + u_{it} \tag{44}$$

Application of instrumental variables

Police and Crime

280 THE AMERICAN ECONOMIC REVIEW JUNE 1997

TABLE 3—ESTIMATES OF THE ELASTICITY OF VIOLENT CRIME RATES WITH RESPECT TO SWORN POLICE OFFICERS

Variable	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) LIML
ln Sworn officers per capita	0.28 (0.05)	−0.27 (0.06)	−1.39 (0.55)	−0.90 (0.40)	−0.65 (0.25)	−1.16 (0.38)
State unemployment rate	−0.65 (0.40)	−0.25 (0.31)	−0.00 (0.36)	−0.19 (0.33)	−0.13 (0.32)	−0.02 (0.33)
ln Public welfare spending per capita	−0.03 (0.02)	−0.03 (0.02)	−0.03 (0.02)	−0.03 (0.02)	−0.02 (0.02)	−0.03 (0.02)
ln Education spending per capita	0.04 (0.07)	0.06 (0.06)	0.02 (0.07)	0.03 (0.07)	0.05 (0.06)	0.03 (0.06)
Percent ages 15–24 in SMSA	1.43 (1.00)	−2.61 (3.71)	−1.47 (4.12)	−2.55 (3.88)	−2.02 (3.76)	−1.50 (3.86)
Percent black	0.010 (0.003)	−0.017 (0.011)	−0.034 (0.015)	−0.025 (0.013)	−0.022 (0.012)	−0.031 (0.013)
Percent female-headed households	0.003 (0.006)	0.007 (0.023)	0.040 (0.030)	0.023 (0.027)	0.018 (0.025)	0.033 (0.027)
Data differenced?	No	Yes	Yes	Yes	Yes	Yes
Instruments:	None	None	Elections	Election * city-size interactions	Election * region interactions	Election * region interactions
P-value of cross-crime restriction on police elasticity	<0.01	<0.01	0.09	0.13	0.33	0.28

Application of instrumental variables

Police and Crime

VOL. 87 NO. 3 LEVITT: USING ELECTORAL CYCLES IN POLICE HIRING 281

TABLE 4—ESTIMATES OF THE ELASTICITY OF PROPERTY CRIME RATES WITH RESPECT TO SWORN POLICE OFFICERS

Variable	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) LIML
ln Sworn officers per capita	0.21 (0.05)	−0.23 (0.09)	−0.38 (0.83)	−0.05 (0.59)	−0.24 (0.36)	−0.34 (0.54)
State unemployment rate	1.40 (0.46)	0.99 (0.46)	1.04 (0.55)	0.94 (0.50)	1.01 (0.47)	1.04 (0.49)
ln Public welfare spending per capita	0.01 (0.03)	−0.02 (0.03)	−0.02 (0.04)	−0.02 (0.04)	−0.02 (0.03)	−0.02 (0.03)
ln Education spending per capita	0.51 (0.08)	0.01 (0.09)	0.01 (0.11)	0.02 (0.10)	0.01 (0.09)	0.01 (0.10)
Percent ages 15–24 in SMSA	1.43 (1.00)	1.15 (4.92)	−1.47 (4.12)	−2.55 (3.88)	−2.02 (3.76)	2.69 (5.16)
Percent black	−0.002 (0.003)	−0.019 (0.014)	−0.029 (0.018)	−0.021 (0.016)	−0.022 (0.015)	−0.027 (0.016)
Percent female-headed households	0.007 (0.006)	0.013 (0.030)	0.025 (0.039)	0.012 (0.034)	0.016 (0.031)	0.022 (0.034)
Data differenced?	No	Yes	Yes	Yes	Yes	Yes
Instruments:	None	None	Elections	Election * city-size interactions	Election * region interactions	Election * region interactions
P-value of cross-crime restriction on police elasticity	<0.01	0.18	0.91	0.92	0.96	0.93

Application of instrumental variables

Police and Crime

1248 THE AMERICAN ECONOMIC REVIEW SEPTEMBER 2002

TABLE 3—THE IMPACT OF POLICE ON CRIME

	Violent crime			Property crime		
Variable	OLS	OLS	IV	OLS	OLS	IV
ln(Police per capita)$_{t-1}$	0.562	−0.076	−0.435	0.113	−0.218	−0.501
	(0.056)	(0.061)	(0.231)	(0.038)	(0.052)	(0.235)
ln(State prisoners per capita)$_{t-1}$	0.250	−0.131	−0.171	0.189	−0.273	−0.305
	(0.039)	(0.036)	(0.044)	(0.030)	(0.028)	(0.037)
Unemployment rate	3.573	−0.741	−0.480	1.283	1.023	1.231
	(0.473)	(0.365)	(0.404)	(0.312)	(0.274)	(0.326)
State income per capita (×10,000)	0.050	−0.003	0.003	0.010	0.005	0.009
	(0.005)	(0.006)	(0.007)	(0.003)	(0.004)	(0.006)
Effective abortion rate (×100)	−0.214	−0.150	−0.141	−0.184	−0.118	−0.111
	(0.045)	(0.023)	(0.025)	(0.020)	(0.021)	(0.024)
ln(City population)	0.072	0.203	0.178	−0.064	−0.333	−0.355
	(0.012)	(0.063)	(0.067)	(0.006)	(0.063)	(0.066)
Percentage black	0.627	0.233	0.398	−0.136	0.411	0.517
	(0.074)	(0.334)	(0.345)	(0.057)	(0.271)	(0.291)
City-fixed effects and year dummies included?	only year dummies	yes	yes	only year dummies	yes	yes
R^2:	0.601	0.930	—	0.238	0.819	—
Number of observations:	2,005	2,005	2,005	2,032	2,032	2,032

Notes: The dependent variable is listed at the top of each column. The police and prison variables are once-lagged. Heteroscedasticity-robust standard errors are in parentheses. Columns (1), (2), (4), and (5) are estimated using weighted least squares, with city populations as weights. Columns (3) and (6) are instrumental variables estimates using the once-lagged, logged number of firefighters as an instrument for the once-lagged number of police. With the exception of columns (1) and (4), other regressions include both city-fixed effects and year dummies. See the notes to Table 1 for further description of the variables and the data sources.

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

- Objective: causal effect of education on trusting

$$trusting = \beta edu + \boldsymbol{\alpha x} + u$$

- Importance:
  - Trust enhances growth (Guido Tabellini, 2008 *JEEA*; Algan and Cahuc, 2013 *Annual Review of Economics*)
  - What determines trusting attitude? Only correlational studies (Alesina and La Ferrara, 2002 *Journal of Public Economics*).
- Endogeneity:
  - u: parents' trusting attitude, community's average trusting (good for an individual's socioeconomic outcomes) $\implies$ $\text{Corr}(u, \ edu) > 0$
  - Measurement error (education)

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

- Method: Use compulsory education reforms in 14 European countries to provide exogenous variation in years of schooling.

- Data: *European Social Survey* (2002, 2004, 2006, 2008, 2010, 2012).
  - Generalized trust.
    "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" (0–10)
  - Perceived fairness of other people.
    "Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair?" (0–10)
  - Perceived helpfulness of other people.
    "Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves?" (0–10)
  - Trust in various institutions.
    Country's parliament, Legal system, Police, Politicians, Political parties, European Parliament, United Nations (0–10)

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

— Countries and reforms

Country (Region)	Reform year	Affected cohort	Change
Austria	1962	1947	$8 \rightarrow 9$
Belgium	1983	1969	$8 \rightarrow 12$
Denmark	1971	1957	$7 \rightarrow 9$
Finland (Uusimaa)	1977	1966	$6 \rightarrow 9$
Finland (Etela-Suomi)	1976	1965	$6 \rightarrow 9$
Finland (Ita-Suomi)	1974	1963	$6 \rightarrow 9$
Finland (Vali-Suomi)	1973	1962	$6 \rightarrow 9$
Finland (Pohjois-Suomi)	1972	1961	$6 \rightarrow 9$
France	1967	1953	$8 \rightarrow 10$

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

— Countries and reforms

Country (Region)	Reform year	Affected cohort	Change
Germany (Schleswig-Holstein)	1956	1941	$8 \rightarrow 9$
Germany (Hamburg)	1949	1934	$8 \rightarrow 9$
Germany (Niedersachsen)	1962	1947	$8 \rightarrow 9$
Germany (Bremen)	1958	1943	$8 \rightarrow 9$
Germany (Nordrhein-Westphalia)	1967	1953	$8 \rightarrow 9$
Germany (Hessen)	1967	1953	$8 \rightarrow 9$
Germany (Rheinland-Pfalz)	1967	1953	$8 \rightarrow 9$
Germany (Baden-Wurttemberg)	1967	1953	$8 \rightarrow 9$
Germany (Bayern)	1969	1955	$8 \rightarrow 9$
Germany (Saarland)	1964	1949	$8 \rightarrow 9$

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

— Countries and reforms

Country (Region)	Reform year	Affected cohort	Change
Greece	1975	1963	$6 \rightarrow 9$
Ireland	1972	1958	$8 \rightarrow 9$
Italy	1963	1949	$5 \rightarrow 8$
Netherlands	1975	1959	$9 \rightarrow 10$
Portugal	1964	1956	$4 \rightarrow 6$
Spain	1970	1957	$6 \rightarrow 8$
Sweden	1962	1950	$8 \rightarrow 9$
U.K. (Scotland)	1976	1961	$10 \rightarrow 11$
U.K. (Engld/Wales/N Ire.)	1972	1957	$10 \rightarrow 11$

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

Variable	N	Mean	S.D.
Generalized trust	31153	5.269	2.383
Perceived fairness	31112	5.792	2.217
Perceived helpfulness	31129	5.044	2.254
Years of schooling	30906	12.608	4.322

Application of instrumental variables

***Education & trust* (Kan & Lai, 2014)**

— Possible instruments

1. Reform dummy:

$$Z_1 = \begin{cases} 1 & \text{individuals affected,} \\ 0 & \text{individuals not affected.} \end{cases}$$

2. Years of compulsory education: Z_2

3. % change in years of compulsory education:

$$Z_3 = \begin{cases} \dfrac{\Delta\text{Years of compul. edu.}}{\text{Pre-reform years of compul. edu.}} & \text{(post-reform)} \\ 0 & \text{(pre-reform)} \end{cases}$$

Application of instrumental variables

Education & trust **(Kan & Lai, 2014)**

— How good are the instruments

1. Reform dummy:
   - $Z_1 : F - \text{statistics} = 7.328$
2. Years of compulsory education
   - $Z_2 : F - \text{statistics} = 11.768$
3. % change in years of compulsory education:
   - $Z_2 : F - \text{statistics} = 21.723$

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

- Stata command: `ivreg2`

```
ivreg2 dep_var (endog_var = iv) other_var
```

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

- Stata codes

```
clear all
set more off
set mem 500m
set matsize 2000
use ESS, clear
global controls "age age2 d_female d_minority d_edu_fa d_edu_mo"
xi: ivreg2 ppltrst (eduyrs = dycomp) $controls i.year*i.cntry2    /
               i.yrbrn [pw = dweight], cluster(cntry2 yrbrn) first
```

- Note:

— IV=rdycomp; dep var = ppltrst;
controls: \$controls + year + birth year + country + country×year

— s.d. clustered at country & birth year (within cluster correlation)

— Code: https://www.dropbox.com/s/754butq49jhe7el/ESS.do
Data: https://www.dropbox.com/s/6vozfopju2ln3w9/ESS.dta

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

- Stata output

```
First-stage regressions
-----------------------

First-stage regression of eduyrs:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on cntry2 and yrbrn

Number of clusters (cntry2) =       28                Number of obs =    30862
Number of clusters (yrbrn) =        49                F(197,    27) =     7.80
                                                      Prob > F      =   0.0000
Total (centered) SS     =  577586.5523                Centered R2   =   0.2504
Total (uncentered) SS   =  5419589.162                Uncentered R2 =   0.9201
Residual SS             =  432940.6057                Root MSE      =    3.758

-------------------------------------------------------------------------------
             |               Robust
      eduyrs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
        .           .          .           .       .           .          .
        .           .          .           .       .           .          .
        .           .          .           .       .           .          .

     rdycomp |   1.249514   .2680889     4.66   0.000     .6994406    1.799587
        .           .          .           .       .           .          .
-------------------------------------------------------------------------------
                                           .
                                           .
                                           .
```

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

- Stata output

```
IV (2SLS) estimation
--------------------
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on cntry2 and yrbrn

Number of clusters (cntry2) =        28                Number of obs =    30862
Number of clusters (yrbrn) =         49                F(197,    27) =     6.76
                                                       Prob > F      =   0.0000
Total (centered) SS     =   175112.951                 Centered R2   =   0.1223
Total (uncentered) SS   =  1026386.982                 Uncentered R2 =   0.8502
Residual SS             =  153703.2822                 Root MSE      =    2.232
------------------------------------------------------------------------------
             |               Robust
    ppltrst  |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      eduyrs |   .2570367   .0805986     3.19   0.001     .0990664     .415007
           .          .          .          .       .            .           .
           .          .          .          .       .            .           .
           .          .          .          .       .            .           .
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):              4.198
                                                    Chi-sq(1) P-val =   0.0405
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):               26.815
                         (Kleibergen-Paap rk Wald F statistic):         21.723
Stock-Yogo weak ID test critical values: 10% maximal IV size             16.38
                                         15% maximal IV size              8.96
                                         20% maximal IV size              6.66
                                         25% maximal IV size              5.53
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
```

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

— Results: 7-year window around the year of birth of the first affected cohorts. First affected cohort omitted.

Table 1: First-stage

	trusting	fairness	helpfulness
% change in years of compulsory educucation	1.250*** (0.268)	1.235*** (0.268)	1.249*** (0.270)
F-statistic	21.723	21.249	21.448
Observations	30862	30823	30837

Table 2: Second-stage

	trusting		fairness		helpfulness	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Yesrs of schooling	0.093*** (0.007)	0.257** (0.081)	0.068*** (0.007)	0.262*** (0.087)	0.037*** (0.006)	0.109*** (0.044)

Application of instrumental variables

Education & trust (Kan & Lai, 2014)

— Results: 7-year window around the year of birth of the first affected cohorts. First affected cohort omitted.

Table 3: Second-stage

	Trust in institutions						
	Country's Parliament	Legal System	Police	Politicians	Political Parties	European Parliament	United Nations
2SLS	-0.176**	-0.150	-0.081	-0.103	-0.001	-0.250**	-0.240
	(0.084)	(0.121)	(0.087)	(0.111)	(0.119)	(0.115)	(0.137)
OLS	0.089***	0.068***	0.024**	0.061***	0.051***	0.067***	0.060***
	(0.011)	(0.013)	(0.010)	(0.005)	(0.006)	(0.008)	(0.009)

IV and Heterogeneous Treatment Effects

Heterogeneous effects — A different interpretation

- Interested in effect of D_i on Y_i, D_i: binary
- Potential outcomes: Y_{i1} or $Y_i(1)$ if $D_i = 1$ (not observed)
  Y_{i0} or $Y_i(0)$ if $D_i = 0$
- Observed outcome:

$$\begin{aligned} Y_i &= Y_{0i} + D_i(Y_{1i} - Y_{01}) \\ &= \alpha_{0i} + \rho_i D_i + \eta_i \text{ (i.e., allows heterogeneous effect)} \end{aligned}$$

- Average Treatment Effect: $ATE = \mathrm{E}(Y_{1i} - Y_{0i})$
- When D_i is randomly assigned (i.e., experimental setting)

$$\begin{aligned} ATE &= \mathrm{E}[Y_i | D_i = 1] - \mathrm{E}[Y_i | D_i = 0] \\ &= \mathrm{E}[Y_{1i} | D_i = 1] - \mathrm{E}[Y_{0i} | D_i = 0] \\ &= \mathrm{E}[Y_{1i} - Y_{01} | D_i = 1] + \cancel{\left\{\mathrm{E}[Y_{0i} | D_i = 1] - \mathrm{E}[Y_{0i} | D_i = 0]\right\}} \\ &= \mathrm{E}[Y_{1i} - Y_{01}] \end{aligned}$$

IV and Heterogeneous Treatment Effects

Heterogeneous effects — A different interpretation

- D_i not random:
  - Endogenous D_i:
    $\left\{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]\right\} \neq 0$, $E[Y_{1i}-Y_{0i}|D_i=1] \neq E[Y_{1i}-Y_{0i}]$
  - Could use instrumental variable.
  - Consider binary instrument: Z_i
- Potential Outcomes of D_i conditional on Z_i

$$
\begin{aligned}
D_i &= \begin{cases} D_i(1) & \text{if } Z_i = 1 \\ D_i(0) & \text{if } Z_i = 0 \end{cases} \\
&= D_{01} + (D_{1i} - D_{01}) Z_i \\
&= \pi_0 + \pi_{1i} Z_i + \xi_i \quad \text{(heterogeneous effect of } Z_i \text{ on } D_i \text{)}
\end{aligned}
$$

$$
Y_i = Y_i(d,z) \text{ if } D_i = d, Z_i = z
$$

IV and Heterogeneous Treatment Effects

Assumptions

- Random assignment:

$$Z_i \perp\!\!\!\perp \Big\{ Y_i(0,0),\ Y_i(0,1),\ Y_i(1,0),\ Y_i(1,1),\ D_i(0),\ D_i(1) \Big\} \quad (45)$$

$$\text{i.e.,}\quad \begin{aligned} \mathrm{E}[Y_i|Z_i=1] &= \mathrm{E}[Y_i(0,1)+(Y_i(1,1)-Y_i(0,1))D_{1i}] \\ &= \mathrm{E}[Y_{0i}+(Y_{1i}-Y_{0i})D_{1i}] \end{aligned}$$

- Exclusion restriction: $Y_i(d,z) = Y_i(d,z') \ \ \forall\ d,\ z,\ z'$

  i.e., Y_i, conditional on D_i, is not a function of Z_i

- Relevance: $\mathrm{E}[D_{1i} - D_{0i}] \neq 0$

- Monotonicity: (No-Defiers) $D_{1i} - D_{0i} \geq 0 \ \forall\ i$

IV and Heterogeneous Treatment Effects

The LATE Estimator

- Local Average Treatment Effect

$$LATE = \frac{\mathrm{E}[Y_i|Z_i=1]-\mathrm{E}[Y_i|Z_i=0]}{\mathrm{E}[D_i|Z_i=1]-\mathrm{E}[D_i|Z_i=0]} = \mathrm{E}[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}]$$

  - Same as 2SLS estimator. Different interpretation.
  - average causal effect on the affected group (compliers).
  - not informative about effects on always takers or never takers.
  - treatment effects differ with different instruments.

- Some details:
  - $$\begin{aligned} \mathrm{E}[Y_i|Z_i=1] &= \mathrm{E}[Y_i(0,1)+(Y_i(1,1)-Y_{0i}(0,1))D_{1i}] \\ &= \mathrm{E}[Y_{0i}+(Y_{1i}-Y_{0i})D_{1i}] \\ \mathrm{E}[Y_i|Z_i=0] &= \mathrm{E}[Y_{0i}+(Y_{1i}-Y_{0i})D_{0i}] \end{aligned}$$
  - $\mathrm{E}[Y_i|Z_i=1]-\mathrm{E}[Y_i|Z_i=0] = \mathrm{E}[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}]\mathrm{Prob}(D_{1i}>D_{01})$
    $\mathrm{E}[D_i|Z_i=1]-\mathrm{E}[D_i|Z_i=0] = \mathrm{E}[D_{1i}-D_{0i}] = \mathrm{Prob}(D_{1i}>D_{01})$
  - Note: $D_{1i}-D_{0i}=0$ or 1 and $\mathrm{E}[D_{1i}-D_{0i}] = \mathrm{Prob}(D_{1i}-D_{0i}>0)$

IV and Heterogeneous Treatment Effects

Why do we need monotonicity?

- Without monotonicity:

$$\begin{aligned} E[Y_i|Z_i=1]-E[Y_i|Z_i=0] &= E[Y_{1i}-Y_{0i}(D_{1i}-D_{0i})] \\ &= E[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}]\text{Prob}(D_{1i}>D_{01}) \\ &\quad +E[Y_{1i}-Y_{0i}|D_{1i}<D_{0i}]\text{Prob}(D_{1i}<D_{01}) \end{aligned}$$

- Unlike in the constant effect case where
$E[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}]\text{Prob}(D_{1i}>D_{01}) =$
$E[Y_{1i}-Y_{0i}|D_{1i}<D_{0i}]\text{Prob}(D_{1i}<D_{01})$
i.e., $E[Y_i|Z_i=1]-E[Y_i|Z_i=0]=E[Y_{1i}-Y_{0i}]\times E[D_{1i}-D_{0i}]$

IV and Heterogeneous Treatment Effects

Some LATE examples

- Example 1: Serving in the military and mortality

Outcome Y	Treatment D	Instrument Z
Mortality	Military Service	High/low draft lottery number, likely to be drafted if low

Draft Lottery: each date of birth in the cohorts at risk to be drafted is assigned a random sequence number (RSN, 1–365). The US Selective Service call menfor induction by SRN up to a ceiling. [See Angrist, Imbens, Rubin (*JASA*, 1996)]

$$\widehat{E(D|Z=1)} = 0.3527, \quad \widehat{E(D|Z=0)} = 0.1934,$$
$$\widehat{E(Y|Z=1)} = 0.0204, \quad \widehat{E(Y|Z=0)} = 0.0195$$

$$\widehat{LATE} = \frac{0.0204 - 0.0195}{0.3527 - 0.1934} = 0.0056$$

IV and Heterogeneous Treatment Effects

Some *LATE* examples

- Example 2: Return to education

Outcome Y	Treatment D	Instrument Z
Weekly earnings (log)	12 or more years of schooling	Quarter of birth (whether first quarter or not)

Originally analyzed by Angrist and Krueger (1991, *QJE*) and re-analyzed by Imbens and Rubin (1997, *REStudies*).

$$\widehat{\mathrm{E}(D|Z=1)} = 0.782, \quad \widehat{\mathrm{E}(D|Z=0)} = 0.762,$$
$$\widehat{\mathrm{E}(Y|Z=1)} = 5.905, \quad \widehat{\mathrm{E}(Y|Z=0)} = 5.892$$

$$\widehat{LATE} = \frac{5.905 - 5.892}{0.782 - 0.762} = 0.651 \text{ (1980 Census:men born 1930-38)}$$

IV and Heterogeneous Treatment Effects

Some LATE examples

- Example 3: Serving in the military and earnings

Outcome Y	Treatment D	Instrument Z
Weekly earnings	Serving in the Military	High/low draft lottery number, likely to be drafted if low
Originally analyzed by Angrist (1990, *AER*).		

$$\widehat{\mathrm{E}(D|Z=1)} = 0.1594, \quad \widehat{\mathrm{Prob}(D|Z=0)} = 0.1362,$$
$$\widehat{\mathrm{E}(Y|Z=1) - \mathrm{E}(Y|Z=0)} = -434.8$$

$$\widehat{LATE} = \frac{-434.8}{0.1594 - 0.1362} = -2,195.8 \quad \text{(for men of the 1950 cohort)}$$