# Introduction to Instrumental Variable Strategy

## Methods Workshop

Youngjoon Lee
(2nd Year PhD student / University of Maryland)

March 22, 2021


Notes: Kee Hyun Park shared this video with me on February 19, 2021. Original file is available at https://twitter.com/i/status/1362636561813303297

# Structure of My Presentation (45 min)

1. When We Need An Instrumental Variable Strategy
2. When We Can Use An Instrumental Variable Strategy
3. How to Interpret Tables
4. Useful R Codes
5. Quiz  

# Structure of the Methods Workshop

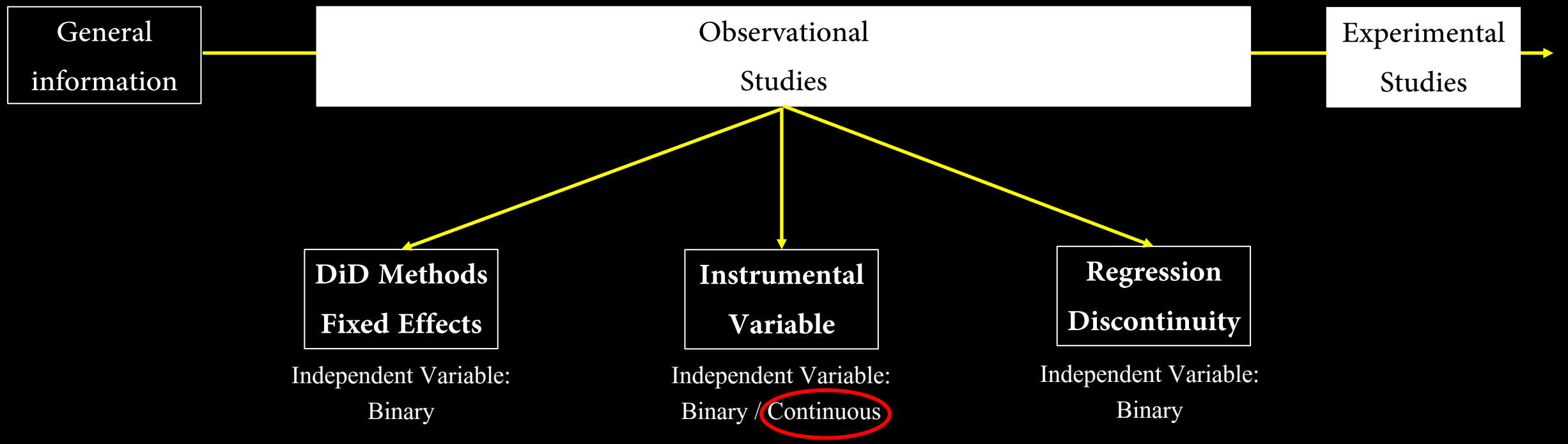
Evan (Feb 15)

Kee Hyun
Xiaonan (Mar 8)

Youngjoon

Chase (Apr 5)

Alauna, Autumn, William (Apr 19)



# When We Need An Instrumental Variable Strategy

1. Reverse Causality (e.g., Dissent-Repression Nexus)

2. Measurement Error (e.g., # of officials punished for corruption = corruption level?)

3. Omitted Variables

# When We Need An Instrumental Variable Strategy

Dissent-Repression Nexus  Dissent ↑ → Repression ↑ & Repression ↑ → Dissent ↓

Our Hypothesis

*First-Stage Regression*

$$Strike_{i,t} = \tau + \omega Casualties_{i,t-1} + \eta\delta_{i,t} + \varepsilon_{i,t}$$

*Second-Stage Regression*

$$Repress_{i,t} = \alpha + \beta Strike_{i,t} + \eta\delta_{i,t} + \varepsilon_{i,t}$$

TSLS Model

# When We Need An Instrumental Variable Strategy

**First Quiz**  **10$** *First Come First Serve: Message to Chat*



Model 1 (No TSLS / Simple OLS)

$$Repress_{i,t} = \alpha + \beta_1 Strike_{i,t} + \eta\delta_{i,t} + \varepsilon_{i,t}$$

Model 2 (Yes TSLS)

$$Repress_{i,t} = \alpha + \beta_2 Strike_{i,t} + \eta\delta_{i,t} + \varepsilon_{i,t}$$

*Statement: $\beta_2$ should be bigger than $\beta_1$.*

*Is this statement correct?*

*1) Yes*
*2) No*

# When We Can Use An Instrumental Variable Strategy

Assumption: the number of instrumental variable = the number of endogenous regressor.

1. **Strong Instrument**

*First-Stage Regression* $Strike_{i,t} = \tau + \omega Casualties_{i,t-1} + \eta\delta_{i,t} + \varepsilon_{i,t}$

*Testable: F-Statistic: bigger than 10.00*

2. **Valid Instrument (Exclusion Restriction)**

$Strike_{i,t} = \tau + \omega Casualties_{i,t-1} + \eta\delta_{i,t} + \varepsilon_{i,t}$

$Repress_{i,t} = \alpha + \beta Strike_{i,t} + \eta\delta_{i,t} + \varepsilon_{i,t}$

*Not Testable*

3. Exogenous Instrument *Not Testable*

# Examples of Instrumental Variable Strategies (TSLS)

## 1. Ritter and Conrad (2016): Africa / US

Rainfall → Count Number of Protest → Count Number of Protests Repressed by Governments

## 2. Guriev, Melnikov, and Zhuravskaya (2020): Global

Lightning Strikes → Mobile Usage Expansion → Government Trust (Yes or No)

## 3. Rogowski, Gerring, Maguire, and Cojocaru (2021): US

County Vote Shares for the Congressional Candidates from the Incumbent President's Party
→ Number of Post Offices → Economic Outputs from Agriculture and Manufacturing Sectors.

# Examples of Instrumental Variable Strategies (TSLS)

## Second Quiz  10$ *First Come First Serve: Message to Chat*

***Which of the following statement (s) is (are) correct?***

(a) Whether the selected instrument is strong is not testable.

(b) Control variables need to be included both in the first and second stage regressions.

(c) Rogowski et al. (2021)'s instrumental variable is in face exogenous.

1. (a)  2. (b)  3. (c)  4. (a) (c)  5. (b) (c)

# How to Interpret Tables

2. The same control variables should be included in both 1st/2nd stage regressions

1. Please locate 1) the instrumental variable, 2) independent variable, and 3) dependent variable!

3. You may choose *whether to* report the results using the reduce form.

4. You should report the F-statistic value.

**Table A.13:** Proximal Effects of Postal Infrastructure in U.S. Counties, 1850-1900 (Instrumental variables)

| | First-stage | Farm value | Manufacturing output | Manufacturing capital |
|---|---|---|---|---|
| **2SLS** | | | | |
| Vote for president's party in Congress | 0.375* | | | |
| | (0.024) | | | |
| Post offices (ln) | | 0.362* | 1.579* | 1.836* |
| | | (0.075) | (0.233) | (0.237) |
| Population (ln) | 0.172* | 0.577* | 0.465* | 0.411* |
| | (0.014) | (0.032) | (0.066) | (0.064) |
| Foreign-born (%, ln) | 0.104 | 0.042 | 1.173 | 1.427 |
| | (0.219) | (0.447) | (0.963) | (0.945) |
| N (observations) | 8,850 | 8,850 | 8,675 | 8,675 |
| N (counties) | 2,258 | 2,258 | 2,226 | 2,226 |
| County fixed effects | ✓ | ✓ | ✓ | ✓ |
| Year fixed effects | ✓ | ✓ | ✓ | ✓ |
| **Reduced form** | | | | |
| Vote for president's party in Congress | | 0.136* | 0.590* | 0.686* |
| | | (0.027) | (0.081) | (0.080) |
| Population (ln) | | 0.639* | 0.730* | 0.720* |
| | | (0.032) | (0.058) | (0.055) |
| Foreign-born (%, ln) | | 0.080 | 1.488 | 1.793 |
| | | (0.473) | (0.981) | (0.950) |
| N (observations) | | 9,186 | 9,004 | 9,004 |
| N (counties) | | 2,594 | 2,555 | 2,555 |
| County fixed effects | | ✓ | ✓ | ✓ |
| Year fixed effects | | ✓ | ✓ | ✓ |

Entries are regression coefficients and standard errors, clustered by county. The dependent variables are the logged values of the outcomes listed at the top of each column. First-stage $F$-statistic = 245.63. * indicates $p < 0.05$ (two-tailed tests).

Source: Rogowski et al. 2021. "Public Infrastructure and Economic Development: Evidence from Postal Systems." *American Journal of Political Science* Published Online (March 3, 2021).

# How to Interpret Tables


I can't find the instrumental variable!

= Reports using a reduced form!

2. F-Statistic value

**TABLE 1 MNC Activity and Corruption**

| Model | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| MNC Activity (Factor Score) | 0.49*** | 0.46*** | 0.51*** | 0.48*** | 0.45*** | 0.29*** | 0.22** |
| | (0.10) | (0.11) | (0.10) | (0.11) | (0.11) | (0.10) | (0.09) |
| GDP per Capita (log) | −0.03 | −0.10 | −0.58 | −0.42 | −0.83** | −0.23 | −0.54*** |
| | (0.41) | (0.39) | (0.39) | (0.43) | (0.39) | (0.33) | (0.18) |
| GDP | −0.98 | −0.38 | 2.22 | −1.26 | 1.66 | −0.47 | 4.62*** |
| | (3.60) | (3.19) | (3.25) | (3.55) | (3.05) | (2.27) | (1.44) |
| Population | 1.19 | 0.88 | 0.96 | 1.01 | 0.74 | 0.60 | −0.63 |
| | (0.72) | (0.68) | (0.59) | (0.67) | (0.57) | (0.46) | (0.23) |
| Government Expenditure (% of GDP, log) | 1.08*** | 1.01*** | 1.27*** | 0.80** | 0.94*** | 0.26 | 0.22 |
| | (0.35) | (0.35) | (0.33) | (0.32) | (0.34) | (0.30) | (0.19) |
| Public Employees | 13.62*** | 12.90*** | 15.75*** | 9.59 | 10.93** | 13.10*** | 0.70 |
| | (4.78) | (4.62) | (3.80) | (5.29) | (4.73) | (4.19) | (2.79) |
| Schooling | −0.36 | −0.39 | −0.47 | 0.28 | 0.15 | −0.11 | −0.08 |
| | (0.29) | (0.28) | (0.26) | (0.46) | (0.42) | (0.37) | (0.22) |
| Relative Wages | 0.41 | 0.37 | 0.42 | −0.04 | 0.03 | −0.22 | 0.44*** |
| | (0.29) | (0.28) | (0.28) | (0.28) | (0.31) | (0.27) | (0.16) |
| Gender | 1.48 | 0.91 | 3.73 | 2.35 | 4.62 | 4.63 | −0.12 |
| | (3.37) | (3.55) | (3.40) | (3.82) | (3.90) | (3.44) | (2.25) |
| Time | 0.35 | 0.36 | 0.42 | 0.19 | 0.22 | 0.31 | 0.20 |
| | (0.26) | (0.26) | (0.26) | (0.41) | (0.42) | (0.37) | (0.19) |
| Bureaucratic Integration | | 0.13 | | | 0.06 | 0.02 | −0.08 |
| | | (0.08) | | | (0.10) | (0.07) | (0.04) |
| Four Municipalities (Dummy Variable) | | | 0.85*** | | 0.67** | 0.40 | 0.84*** |
| | | | (0.25) | | (0.29) | (0.26) | (0.14) |
| Trust in Courts | | | | 0.13 | 0.08 | 0.24 | −0.11 |
| | | | | (0.23) | (0.23) | (0.20) | (0.13) |
| Constant | 3.12 | 3.48 | 5.49** | 3.91 | 5.52 | −0.24 | 2.32 |
| | (3.06) | (2.91) | (2.77) | (3.27) | (2.93) | (2.45) | (1.54) |
| N | 61 | 61 | 61 | 55 | 55 | 55 | 54 |
| $R^2$ | 0.38 | 0.40 | 0.46 | 0.36 | 0.45 | 0.52 | 0.64 |
| F-Statistic (Excluded Instrument) | 45.92 | 46.47 | 45.28 | 38.89 | 38.23 | 38.23 | 38.03 |
| Prob > F | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 2SLS | 2SLS | 2SLS | 2SLS | 2SLS | 2SLS | 2SLS |

*Note:* Robust standard errors are in parentheses. The dependent variable (DV) is recovered corrupt funds per filed case in Models 1–5, recovered corrupt funds per capita in Model 6, and senior cadres disciplined per 10,000 public employees in Model 7.
**significant at 5%; ***significant at 1%.

1. Please locate 1) the instrumental variable, 2) independent variable, and 3) dependent variable!

Source: Zhu. 2017. "MNCs, Rents, and Corruption: Evidence from China." *American Journal of Political Science* 61 (1).

# How to Interpret Tables

**TABLE 2. The Effect of Mobilized Dissent on State Repression in U.S. State-Days**

| | 1 OLS (No Instrument) | 2 IV Regression (Basic Model) | 3 IV Regression (Matched Model) |
|---|---|---|---|
| | | **Second Stage: The Effect of Dissent on Repression** | |
| Mobilized Dissent | 0.353* | 0.397* | 0.459* |
| | (0.011) | (0.088) | (0.118) |
| Urbanization | −0.000* | −0.000 | −0.000 |
| | (0.000) | (0.000) | (0.000) |
| Constant | 0.006* | 0.008 | 0.012 |
| | (0.001) | (0.005) | (0.007) |
| | | **First Stage: Instrumenting Mobilized Dissent** | |
| Rainfall (ln) | — | 0.001* | 0.001* |
| | | (.000) | (0.000) |
| Annual Rainfall | — | −0.000* | −0.000 |
| | | (0.000) | (0.000) |
| Urbanization | — | 0.001* | 0.001* |
| | | (0.000) | (0.000) |
| Constant | — | −0.057* | −0.059* |
| | | (0.001) | (0.001) |
| | | **Model Statistics** | |
| *N* | 700,435 | 699,610 | 703,622 |
| *F* Test of Excluded Instruments | — | 23.39 (0.000) | 13.86 (0.000) |
| Cragg-Donald Wald *F* Statistic | — | 26.33 | 13.86 |
| Sargan-Hansen *J* Statistic ($\chi^2$ *p* value) | — | 4.200 (0.040) | 0.263 (0.608) |

*Notes:* $^*p < 0.05$ in two-tailed tests with robust standard errors reported beneath coefficients in parentheses. Parentheses on instrument statistics report their respective *p* values.

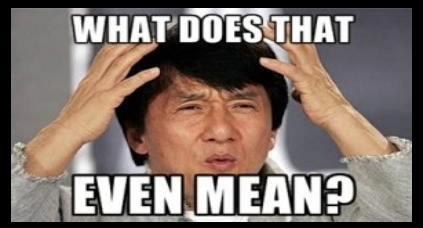
Source: Ritter and Conrad. 2016. "Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression." *American Political Science Review* 110 (1).

1. Please locate 1) the instrumental variable, 2) independent variable, and 3) dependent variable!

2. Urbanization variable was included in both 1st/2nd stage regressions.

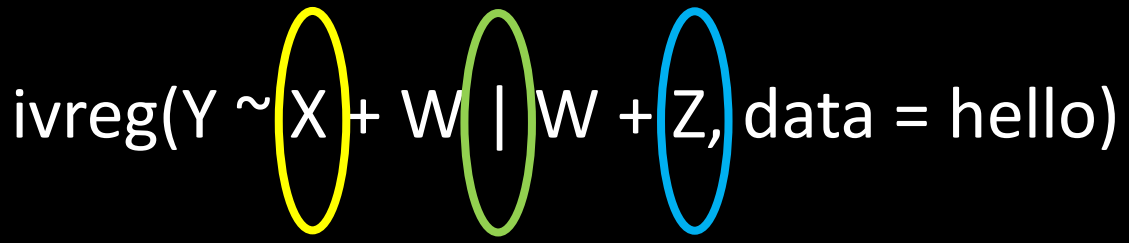
3. F-Statistic value

4. What is this????




***$H_0$: The instrumental variables can be "jointly" used.***

# Useful R Codes




Y: Dependent Variable

X: Endogenous Independent Variable

Z: Instrumental Variable

W: Control Variables

# What We Have Learned Today

1. Assumptions Matter.

   - Strong Instrument  Testable

   - Valid Instrument  Not Testable

   - Exogenous Instrument  Not Testable

2. Please Compare TSLS Results with OLS Results.

3. Please Note that You Can Adopt More Than One Instrument to Instrument One Independent Variable.

# What We Need to Learn in the Future

If I am given one additional opportunity..

## Advanced Instrumental Variable Strategy

Table of Contents

# Thank you

I hereby note that Kee Hyun Park and Xiaonan Wang’s suggestions were very helpful.