Lecture 10: Instrumental Variable

Introduction to Econometrics,Spring 2025

Zhaopeng Qu

Business School,Nanjing University

May 07 2025

Review Previous Lecture of Matching

Matching: Introduction

- Besides OLS, **Matching** is another common method to deal with the **selection bias** in observational studies.
- The idea of matching method is quite simple:
  - construct a reasonable control group by **selecting some samples** in untreated group which have **similar** characteristics with the treated group.
- Assumptions of Matching:
  - **Conditional Independence Assumption**: $Y_i(0), Y_i(1) \perp T_i | X_i$
  - **Common Support Assumption**: $0 < Pr(T_i = 1 | X_i) < 1$

Matching in Practice

- Matching in Xs, Propensity Score Matching and IPW are three common methods in Matching.
- Many details when we use it in practice:
  - **Distance Metric**: Euclidean distance, Mahalanobis distance, Propensity score, etc.
  - **Matching Algorithm**: Nearest Neighbor Matching, Radius Matching, Kernel Matching, etc.
  - **with or without replacement**: one-to-one, one-to-many, many-to-many.
  - ...
- The most part is to make a **balance test** to evaluate the matching quality and **sensitivity analysis** to check the robustness.

Matching v.s OLS

- Essentially, matching is as the same as regression, only different in the weight of estimating the CEF function.
- Why we still need matching? Matching is over regression in some aspects:
  - **Nonparametric**: No need to specify the functional form of the model.
  - **Overlap or Common Support**: Matching explicitly requires the common support assumption, which is not necessary in OLS.
- However, matching still has to rely on **CIA** assumption, which is the same as OLS.
  - Most biases we could suffer in regression, such as **OVB**, **measurement error**, and **simultaneous causality**, will not be avoided even if we use matching.
- We need to use some methods which can deal with these problems,thus Selection-in-unobservables.
  - The first common one is **Instrumental Variable** method.

Instrumental Variable: Introduction

Introduction

- A seemed easy but difficult to answer question:
- *How to estimate the supply or demand curves from the data?*
- **Difficulty**: We can only observe intersections of supply and demand, yielding pairs.
- **Solution**: Wright(1928) use variables that appear in one equation to shift this equation and trace out the other.
- The variables that do the shifting came to be known as **Instrumental Variables** method.

OLS Regression with endogenous variable

- Suppose our model is still a simple OLS regression

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- But now $E[u_i|X_i] \neq 0$, thus violate the **Assumption 1** as we suffer OVB,ME or Simultaneity, then OLS estimator $\hat{\beta}_1$ is *biased and inconsistent*.

Exogeneity and Endogeneity in DAGs

- **DAGs** can help us to understand the relationship between endogeneity and exogeneity.
- **Exogeneity**
- **Endogeneity in confounders**

Instrumental Variable in a intuitive way

- To correct this potential bias, we can use an **instrumental variable**(Z_i) to obtain a consistent estimation of coefficient β.
- Intuitively, we want to split X_i into **two parts**:

- If we can **isolate** the variation in X_i that is **uncorrelated** with u_i,then we can use the exogenous part(Z) to restore a consistent estimate of the causal effect of X_i on Y_i.

Key Assumptions of IV

- An instrumental variable Z_i must satisfy the following two properties:
  1. **Instrumental relevance**:
  2. **Instumental exogeneity**:

Instrument Variable in DAGs

- The IV solution

- Z can **isolate the variation** in X_i that is **uncorrelated with** u_i and restore a consistent estimate of the causal effect of X_i on Y_i.

2SLS Estimator

IV Estimator: Two Steps Least Squares (2SLS)

- The intuition leads us to obtain the 2SLS-IV estimator in following two steps:

1. **First stage**:

- The predicted values of X_i, thus $\hat{X}_i$ only contain variations in X_i that is **uncorrelated** with u_i

IV Estimator: Two Steps Least Squares (2SLS)

2. **Second stage**:

- Because $\hat{X}$ directly comes from Z,thus $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$ where $Cov[Z, u] = 0$,then

$$E[u_i|\hat{X}_i] = 0 \; or \; Cov(\hat{X}_i, u_i) = 0$$

  the **Assumption 1** of OLS can be satisfied, this OLS estimator will be **unbiased and consistent** again.
- Then we can write down the 2SLS estimator $\hat{\beta}_{2SLS}$ in a simple OLS estimator formula:

IV Estimator: Two Steps Least Squares (2SLS)

- Because $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$,then

$$\overline{\hat{X}} = \hat{\pi}_0 + \hat{\pi}_1 \overline{Z}_i$$

- Then, we have

$$\hat{X}_i - \overline{\hat{X}} = \hat{\pi}_1 (Z_i - \bar{Z})$$

- Also because $\hat{\pi}_1$ is the estimating coefficient of Z_i on X_i, then again base on *a simple OLS estimation coefficients formula*,

IV Estimator: Two Steps Least Square (2SLS)

- Then we could obtain

IV Estimator: Two Steps Least Square (2SLS)

- Which gives the 2SLS IV estimator
  - Where s_{ZY} and s_{ZX} are **sample covariances** of Z & Y and Z & Z respectively.
  - The 2SLS estimator of β_1 is the ratio of *the sample covariance between* Z and Y to *the sample covariance between* Z and X.
- If $Z_i = X_i$, which means that X_i itself is *exogenous*, then
- **NOTE**: In this sense, you can see OLS estimator as **a special case** of 2SLS-IV estimator.

IV Estimator: First Stage and Reduced Form

- **First-Stage** regression: regress endogenous variable on IV

- **Reduced-Form** regression: regress outcome variable on IV

- 2SLS estimator can also be seen as a **ratio** of the estimated coefficient in **Reduced Form** to the one in **First Stage**.

Statistical properties of 2SLS estimator

Unbiasedness

- Consider $E[\hat{\beta}_{IV}]$

Unbiasedness

- Because $Cov(X_i, u_i) \neq 0$,then $E[u_i|Z_i, X_i] \neq 0$,then

- Then we have

  - It means that 2SLS estimator is **biased**.
  - In contrast,OLS estimator is *unbiased* even when the sample size is small.

Consistency

- We have a simple regression $Y_i = \beta_0 + \beta_1 X_i + u_i$ and take a covariance of Y_i and Z_i

- Thus if the instrument is valid,

$$\beta_1 = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)}$$

- The population coefficient of 2SLS is the ratio of *the population covariance between* Z and Y to *the population covariance between* Z and X.

Consistency

- As discussed in Section 3.7,the sample covariance is a consistent estimator of the population covariance when the sample size is large, thus
- Then the 2SLS estimator is **consistent**.
- Like the OLS estimator if key assumptions are satisfied,it is the reason why we use it.

Sampling Distribution

- Similar to the expression for the OLS estimator in Equation (4.30,page 183 in S.W), it is easy to show that

- Then as we did in the lecture of statistical inference of OLS regression, it can be derived that

$$\hat{\beta}_{2SLS} \xrightarrow{d} N(\beta, \sigma^2_{\hat{\beta}_{2SLS}})$$

  - Where

$$\sigma^2_{\hat{\beta}_{2SLS}} = \frac{\sigma^2_{\bar{q}}}{[Cov(Z_i, X_i)]^2} = \frac{1}{n}\frac{Var[(Z_i - \mu_Z)u_i]}{Cov[(Z_i, X_i)]^2} \qquad (12.8 \text{ in SW})$$

Statistical Inference

- The standard deviation of $\hat{\beta}_{2SLS}$ can be obtained by estimating the variance and covariance terms appearing in Equation (12.8), thus **the standard error of the 2SLS IV estimator** is

$$SE(\hat{\beta}_{2SLS}) = \sqrt{\frac{\frac{1}{n}\sum(Z_i - \bar{Z})^2 \hat{u}_i^2}{n\left(\frac{1}{n}\sum(Z_i - \bar{Z})(X_i - \bar{X})\right)^2}}$$

- Because $\hat{\beta}_{2SLS}$ is normally distributed in large samples, hypothesis tests about β can be performed by computing **the t-statistic**,and a 95% large-sample **confidence interval** is given by

Standard Errors of 2SLS v.s OLS

- **Recall**: Under the assumption of homoskedasticity, we obtain the variance of a multiple OLS estimator of β_j as

- Where the sample variance of X is $\frac{1}{n-1}\sum(X_i-\bar{X})^2$, which converges to the population variance σ_X^2 as n approaches infinity, thus

$$\frac{1}{n-1}\sum(X_i-\bar{X})^2 \xrightarrow{p} \sigma_X^2$$

- Then we have

Standard Errors of 2SLS v.s OLS

- Likewise, we could prove that the *population variance of the 2SLS estimator* is
    - where ρ_{xz} is the correlation between X and Z.
- The **detailed proof** is somewhat technical, so I have included all the details in *the appendix* of the lecture notes if you are interested in exploring it further.

Standard Errors of 2SLS v.s OLS

- Because $\rho_{xz}^2 \leq 1$,then

- Thus the variance of the 2SLS estimator is **always larger than** that of the OLS estimator if OLS is unbiased.
- In other words, 2SLS is always **less efficient** than OLS.
- This makes intuitive sense because:
  - 2SLS only uses the variation in X that can be explained by Z.
  - This is necessarily less variation than what OLS uses (all variation in X).
  - Less variation leads to less precise estimates (larger standard errors).

Application: Angrist and Krueger(1991)

A classical application of IV

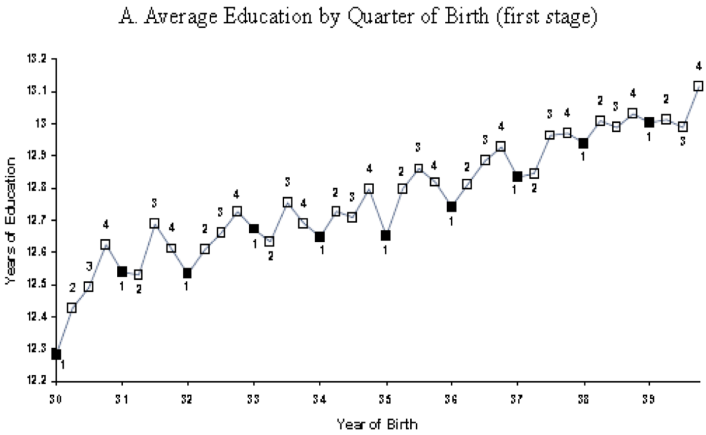
- Angrist, Joshua D. and Alan B. Krueger. 1991. *"Does Compulsory School Attendance Affect Schooling and Earnings?"*,The Quarterly Journal of Economics 106 (4):pp979–1014.
- A well-known fact that an OLS regression to estimate the returns to schooling will suffer OVB bias
- **Question**:
  - How to explain the implication of β_1?
  - Why we cannot obtain an ubiased estimate of β_1?
- To deal with the OVB problem, they used *the quarter of birth* as an instrument for education to estimate the returns to schooling.

Quarter of Birth as IVs

- Why is the 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 by compulsory schooling laws.

Quarter of Birth as IVs

- Is Schooling related to Quarter of Birth?(Assumption 1)



First Stage

- Does quarter of birth affect education?
- Regress education outcomes on quarter of birth dummy variables:

  - where individual i, cohort c, education outcome S, birth quarter Q_j and ϵ_{ijc} is the error term.
- It is the **first stage** regression

First Stage

- It shows that Q_j **does** impact education outcomes such as total years of education and high school graduation.

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

Exogeneity

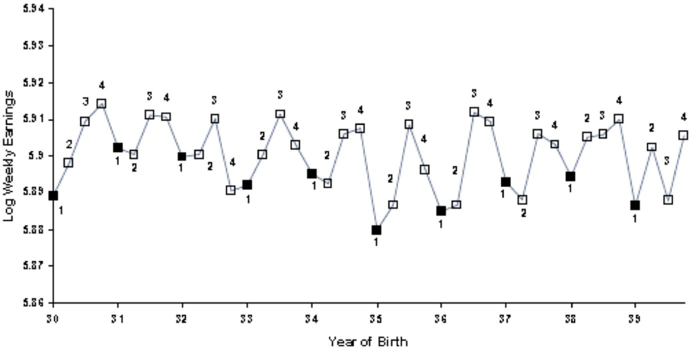
- *Does the birth quarter is exogenous to the wage determination?*
  - It seems that one's birth date should not be related with his/her earnings.
- Moreover,*does the effect of birth quarter on educational outcome fully due to compulsory schooling laws which is exogenous?*

			(0.011)	(0.011)	(0.010)	[0.0017]
College graduate	1930–1939	0.24	−0.005	0.003	0.002	5.0
			(0.002)	(0.002)	(0.002)	[0.0021]
	1940–1949	0.30	−0.003	0.004	0.000	5.0
			(0.002)	(0.002)	(0.002)	[0.0018]
Completed master's degree	1930–1939	0.09	−0.001	0.002	−0.001	1.7
			(0.001)	(0.001)	(0.001)	[0.1599]
	1940–1949	0.11	0.000	0.004	0.001	3.9
			(0.001)	(0.001)	(0.001)	[0.0091]
Completed doctoral degree	1930–1939	0.03	0.002	0.003	0.000	2.9
			(0.001)	(0.001)	(0.001)	[0.0332]
	1940–1949	0.04	−0.002	0.001	−0.001	4.3
			(0.001)	(0.001)	(0.001)	[0.0050]

Reduced form

- Is Earnings related to Quarter of Birth?

B. Average Weekly Wage by Quarter of Birth (reduced form)



Results: OLS v.s 2SLS

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS
Years of education	0.0711 (0.0003)	0.0891 (0.0161)	0.0711 (0.0003)	0.0760 (0.0290)
Race (1 = black)	—	—	—	—
SMSA (1 = center city)	—	—	—	—
Married (1 = married)	—	—	—	—
9 Year-of-birth dummies	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No
Age	—	—	−0.0772 (0.0621)	−0.0801 (0.0645)
Age-squared	—	—	0.0008 (0.0007)	0.0008 (0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]

Results: OLS v.s 2SLS

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS
Years of education	0.0711	0.0891	0.0711	0.0760
	(0.0003)	(0.0161)	(0.0003)	(0.0290)
Race (1 = black)	—	—	—	—
SMSA (1 = center city)	—	—	—	—
Married (1 = married)	—	—	—	—
9 Year-of-birth dummies	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No
Age	—	—	−0.0772	−0.0801
			(0.0621)	(0.0645)
Age-squared	—	—	0.0008	0.0008
			(0.0007)	(0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]

Checking Instrument Validity

The Properties of IVs

- An instrumental variable Z_i must satisfy the following 2 properties:
  1. **Instrumental relevance**: Z_i should be **correlated** with the casual variable of interest, X_i (endogenous variable),thus
  2. **Instumental exogeneity**: Z_i is as good as randomly assigned and Z_i only affect on Y_i through X_i affecting Y_i channel.

Instrument Variable in DAGs

- The IV solution

Assumption #1 Instrument Relevance

OVB in 2SLS

- Recall 2SLS: a simple OLS regression equation is

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Get the predict value from the first stage

$$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$$

- Running the second stage regression

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$$

- So following the OLS formula in large sample, we can obtain

OVB in 2SLS

- A 2SLS version of OVB

Weak Instruments

- **Assumption 1**: Instrument Relevance

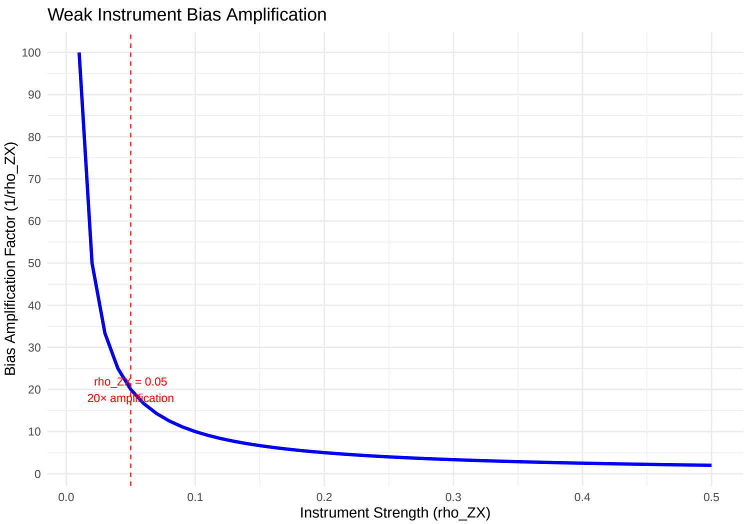
- **Intuition**: the *more the variation in* X is explained by the instruments, thus *the more information is available for use* in IV regression.
- On the contrary, instruments explain little of variation in X are called **Weak Instruments**, thus there is a very weak correlation between X(endogenous variable) and Z(IV).

Weak Instruments

- Because

- In many cases, IV cannot be perfect random and exogenous, thus $Cov(Z, u) \neq 0$ or $\rho_{Zu} \neq 0$.
- Then if $\rho_{ZX} = 0$,thus X and Z is *irrelevant*,the bias will approximate to *infinity*.
- Even the correlation coefficient,ρ_{ZX} is **not ZERO but very small**
- Only if the correlation is large enough, the OVB will approximate to ZERO.

Weak Instruments: Bias Amplification



Weak Instruments: How to test weak instruments ?

- **Reminder**: We should therefore **always** check *whether an instrument is relevant enough*.

- Stock and Yogo(2005) showed that

$$E(\beta_{2SLS}) - \beta \cong \frac{E(\beta_{ols}) - \beta}{E(F) - 1}$$

- $E(F)$ is the expectation of the first stage F-statistics.And if $E(F) = 10$,the bias of 2SLS, relative to the bias of OLS,is approximately $\frac{1}{9}$, which is small enough to be acceptable.
- *A Rule of Thumb*: **F-statistic exceeds** 10,don't need worry about too much.

Angrist and Krueger(1991): Why IV over OLS?

- Despite large samples sizes, the F-statistics for a test of the joint statistical significance of the excluded exogenous variables in the first-stage regression are not over 2.

Coefficient	.063	.083	.063	.081
	(.000)	(.009)	(.000)	(.011)
F (excluded instruments)		2.428		1.869
Partial R^2 (excluded instruments, ×100)		.133		.101
F (overidentification)		.919		.917
Age Control Variables				
Age, Age^2			x	x
9 Year of birth dummies	x	x	x	x
Excluded Instruments				
Quarter of birth		x		x
Quarter of birth × year of birth		x		x
Quarter of birth × state of birth		x		x

Alternative Estimators

- If the instruments are irrelevant, it is not possible to obtain an unbiased estimator of β_1, even in large samples.
- Nevertheless, when instruments are weak, some alternative IV estimators tend to be more centered on the true value of β_1 than 2SLS.
- One such estimator is the limited information maximum likelihood (**LIML**) estimator, which is a maximum likelihood estimator of β_1.
  - If instruments are weak, then the LIML estimator is more centered on the true value than 2SLS.
  - If instruments are strong, then LIML and 2SLS will coincide in large samples.

$F > 10$ is NOT everything

- *the Rule of Thumb* of $F > 10$ is not good enough as we thought.
  - Lee et al(2020) show that $F > 10$ does not permit valid inference, but a reliable inference at the 5% level is possible with $F > 143$
- *the Rule of Thumb* of $F > 10$ with the strong assumption of homoskedastic errors, which often leads to a smaller S.E. The assumption is often violated due to heteroskedasticity,serial or spatial auto-correlation, or clustering.
- Two more robust alternatives

Wrap up

- If the correlation between the instruments and the endogenous variable is small, then even the enormous sample sizes do not guarantee that quantitatively important finite sample biases will be eliminated from IV estimates.
- The first assumption of IV method, thus relevance of IV, can be justified by the first stage regression and **F-statistic**.
- Potential Solutions
  - If you have many IVs, some are strong, some are weak. Then discard weak ones.
  - If you only have an weak IV, then find another more stronger IV(easy to say, very hard to do)
  - Using other estimator(LIML) as a supplement to 2SLS estimator.

Assumption #2 Instrument Exogeneity

Instrument Exogeneity

- The idea of instrumental variables regression is that the instrument contains information about variation in X_i that is unrelated to the error term u_i.If the instruments are not exogenous, then TSLS is inconsistent.
- More specifically, it includes two distinct points:
  - **Enough exogenous**: As-good-as-random assignment
  - **Exclusion restriction** :

Instrument Exogeneity in DAGs

- Enough exogenous

- Exclusion restriction

Instrument Exogeneity: Exclusion restriction

- Suppose we could run the following regression

$$Y_i = \beta_0 + \beta_1 X_i + \gamma Z_i + i$$

- Then the exclusion restriction implies that $\gamma = 0$,
- what if $\gamma \neq 0$?

Instrument Exogeneity: Exclusion restriction

- Recall the 2SLS estimator

- The β_1^{2sls} is **inconsistent** estimator, the bias is $\frac{\gamma}{\phi_1}$
- When $\phi_1 \to 0$(weak instrument), the bias will be very large.
- When $\phi_1 \to 0$(weak instrument), a very small violation of the exclusion restriction can lead to a large bias.

Angrist and Krueger(1991): Exclusion restriction

- It should prove that what all of the association between quarter of birth(IV) and both education and earnings should be attributed to **compulsory education law**, which is an exogenous policy, nothing else.
- Angrist and Krueger(1994) provides a supporting evidence that the association between quarter of birth(IV) and educational attained is much weaker for college and above graduates, who are unrestricted by **compulsory education law**.

			(0.011)	(0.011)	(0.010)	[0.0017]
College graduate	1930–1939	0.24	−0.005	0.003	0.002	5.0
			(0.002)	(0.002)	(0.002)	[0.0021]
	1940–1949	0.30	−0.003	0.004	0.000	5.0
			(0.002)	(0.002)	(0.002)	[0.0018]
Completed master's degree	1930–1939	0.09	−0.001	0.002	−0.001	1.7
			(0.001)	(0.001)	(0.001)	[0.1599]
	1940–1949	0.11	0.000	0.004	0.001	3.9
			(0.001)	(0.001)	(0.001)	[0.0091]
Completed doctoral degree	1930–1939	0.03	0.002	0.003	0.000	2.9
			(0.001)	(0.001)	(0.001)	[0.0332]
	1940–1949	0.04	−0.002	0.001	−0.001	4.3
			(0.001)	(0.001)	(0.001)	[0.0050]

AK's Exclusion restriction may be not valid

- Bound and Jeager(2000) find that men born in the 19th century, who were not affected by compulsory schooling laws also display variation in earnings with respect to quarter-of-birth.
- This suggests that quarter-of-birth also influences earnings through other channels rather than solely educational attainment, and that the exclusion restriction of IV is violated.

AK's Exclusion restriction may be not not valid

Table 4. Reduced Form: Quarter of Birth Effects on Imputed Log Weekly Earnings and I (Agriculture) for White Men Educated Prior to Compulsory Schooling Laws

	OLS: Imputed Log Weekly Earnings				Logit: I (Agriculture)			
	Men Born 1840–55		Men Born 1840–75		Men Born 1840–55		Men Born 1840–75	
Qtr of Birth	Age only	Age & Demo.	Age only	Age & Demo.	Age only	Age & Demo.	Age only	Age & Demo.
Jan.–Mar.	–0.019	–0.023	–0.050	–0.038	0.043	0.070	0.127	0.115
	(0.026)	(0.022)	(0.016)	(0.014)	(0.070)	(0.080)	(0.044)	(0.051)
Apr.–June	0.014	–0.006	0.023	0.005	–0.042	0.016	–0.059	–0.011
	(0.026)	(0.023)	(0.017)	(0.015)	(0.071)	(0.082)	(0.044)	(0.051)
July–Sep.	0.049	0.027	0.016	0.021	–0.129	–0.090	–0.046	–0.073
	(0.026)	(0.024)	(0.017)	(0.015)	(0.071)	(0.081)	(0.045)	(0.052)
Oct.–Dec.	–0.044	0.002	0.011	0.012	0.128	0.004	–0.021	–0.031
	(0.027)	(0.024)	(0.017)	(0.015)	(0.073)	(0.082)	(0.045)	(0.052)
Q_3–Q_1	0.068	0.049	0.066	0.058	–0.172	–0.160	–0.173	–0.188
	(0.042)	(0.037)	(0.027)	(0.024)	(0.115)	(0.131)	(0.072)	(0.083)
$\Sigma\|Q_i\|$	0.126	0.058	0.101	0.075	0.342	0.180	0.254	0.229
	(0.060)	(0.078)	(0.039)	(0.029)	(0.164)	(0.161)	(0.087)	(0.101)

Wrap up

- **Question**: *Can we statistically test the assumption that the instruments are exogenous?*
- **Answer**:
- Assessing whether the instruments are exogenous necessarily requires making an expert judgment based on personal knowledge and expert opinion of the application.(“**讲好故事**”)
- And you should provide some solid indirect evidences that exclusion restriction is impossibly violated.
- Several new tests try to loose the exogenous assumption of IV
  - eg. Conley et al.(2012) is to relax the exclusion restriction.
- Reference: *Conley T G, Hansen C B, Rossi P E. Plausibly exogenous[J]. Review of Economics and Statistics, 2012, 94(1): 260-272.*

Overidentification Test

When you have more IVs

- In some case,you can test partially,thus **overidentification test**.
- Terminology: The relationship between the number of instruments(m) and the number of endogenous regressors(k)
- when the coefficients are just identified, you can't do a formal statistical test of the hypothesis that the instruments are in fact exogenous.
- If, however, there are more instruments than endogenous regressors, then there is a statistical tool that can be helpful in this process: the so-called test of *overidentifying restrictions*.

Intuition

- Suppose there are two valid instruments: Z_1 Z_2(you are very lucky.)
- Then you could compute two separate TSLS estimates.
- Intuitively,if these 2 TSLS estimates are very different from each other, then something must be wrong: one or the other (or both) of the instruments must be invalid.
- The **overidentifying restrictions test** makes this comparison in a statistically precise way.

Extension of Multiple OLS Regression

- Our model is a multiple regression

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + \beta_{k+1} W_{1,i} + ... + \beta_{k+r} W_{r,i} + u_i$$

- Where
  - Y_i is the *dependent variable*
  - $X_1, X_2, ...X_k$ are K *endogenous regressors*
  - $W_1, X_2, ...W_r$ are the *additional exogenous variables*
  - we have m instruments,$Z_1, Z_2, ...Z_m$,*instrumental variables*
  - u_i is the regression error term.

M instruments

- We have a set of m instruments, $Z_1, Z_2, ...Z_m$, then run TSLS regression

- Obtain the predict value of $\hat{u}_i{}^{TSLS}$,which should be approximately uncorrelated with instruments $Z_1, Z_2, ...Z_m$.

- Accordingly, if the instruments are in fact **exogenous**, then the coefficients on the instruments in a regression of $\hat{u}_i{}^{TSLS}$ on the instruments and the included exogenous variables should all be *ZERO*.

Overidentification test

- The new regression model of $\hat{u}$ on Z and W

- Let F denote the homoskedasticity-only F-statistic testing the hypothesis that $\delta_0 = ... = \delta_m = 0$
- Then the overidentifying restrictions test statistic is $J = mF$
- Under the null hypothesis that all the instruments are exogenous,

- Where $m - k$ is the "degree of over-identification," that is, the number of instruments minus the number of endogenous regressors.

Application: Demand for Cigarettes

- Smoking poses a serious public health issue with significant externalities.
- One effective policy tool is **taxing cigarettes** heavily enough to reduce consumption among current smokers and discourage potential new smokers.
- Key Question: What tax increase would significantly reduce cigarette consumption?
- For instance, what after-tax price would achieve a 20% reduction in cigarette consumption?
- The answer depends on the **elasticity of demand** for cigarettes.
- **Recall**: Due to supply-demand interactions, cigarette demand elasticity cannot be consistently estimated using simple OLS regression of log quantity on log price.

Application: Demand for Cigarettes

- We use TSLS with annual data from the 48 contiguous U.S. states in 1995 to estimate cigarette demand elasticity.
- Our instrumental variable, $SalesTax_i$, represents the portion of cigarette tax from general sales tax, measured in dollars per pack.
- Cigarette consumption, $Q_i^{cigarettes}$, is measured as packs sold per capita in each state.
- The price, $P_i^{cigarettes}$, represents the average real price per pack including all taxes.

Application: Demand for Cigarettes

- We analyze changes in quantity and price over a 10-year period.
- Dependent variable: Change in log cigarette consumption

$$\Delta ln(Q_i^{cigarettes}) = ln(Q_{i,1995}^{cigarettes}) - ln(Q_{i,1985}^{cigarettes})$$

- Independent variable: Change in log cigarette price

$$\Delta ln(P_i^{cigarettes}) = ln(P_{i,1995}^{cigarettes}) - ln(P_{i,1985}^{cigarettes})$$

- Control variable: Change in log income

$$\Delta ln(Inc_i) = ln(Inc_{i,1995}) - ln(Inc_{i,1985})$$

- *Question: why we use change in log quantity and price?* you will learn it in the lectures of Panel Data.

Application: Demand for Cigarettes

- Two instruments which should be not correlated with the error term but correlated with the endogenous variable, the change in the price of cigarettes.
  1. the change in the sales tax over 10 years,

$$\Delta SalesTax_i = SalesTax_{i,1995} - SalesTax_{i,1985}$$

  2. the change in the cigarette-specific tax over 10 years

$$\Delta CigTax_i = CigTax_{i,1995} - CigTax_{i,1985}$$

Demand for Cigarettes: First stage and 2SLS regressions

TABLE 12.1 Two Stage Least Squares Estimates of the Demand for Cigarettes Using Panel Data for 48 U.S. States

Dependent variable: $\ln(Q_{i,1995}^{cigarettes}) - \ln(Q_{i,1985}^{cigarettes})$

Regressor	(1)	(2)	(3)
$\ln(P_{i,1995}^{cigarettes}) - \ln(P_{i,1985}^{cigarettes})$	−0.94** (0.21)	−1.34** (0.23)	−1.20** (0.20)
$\ln(Inc_{i,1995}) - \ln(Inc_{i,1985})$	0.53 (0.34)	0.43 (0.30)	0.46 (0.31)
Intercept	−0.12 (0.07)	−0.02 (0.07)	−0.05 (0.06)
Instrumental variable(s)	Sales tax	Cigarette-specific tax	Both sales tax and cigarette-specific tax
First-stage F-statistic	33.70	107.20	88.60
Overidentifying restrictions J-test and p-value	—	—	4.93 (0.026)

These regressions were estimated using data for 48 U.S. states (48 observations on the 10-year differences). The data are described in Appendix 12.1. The J-test of overidentifying restrictions is described in Key Concept 12.6 (its p-value is given in parentheses), and the first-stage F-statistic is described in Key Concept 12.5. Individual coefficients are statistically significant at the *5% significance level or **1% significance level.

Demand for Cigarettes: Over-identifying J-test

- Over-identifying J-test **reject** the null hypothesis that both the instruments are exogenous at the 5% significant level($p - value = 0.026$)

TABLE 12.1 Two Stage Least Squares Estimates of the Demand for Cigarettes Using Panel Data for 48 U.S. States

Dependent variable: $\ln(Q_{i,1995}^{cigarettes}) - \ln(Q_{i,1985}^{cigarettes})$

Regressor	(1)	(2)	(3)
$\ln(P_{i,1995}^{cigarettes}) - \ln(P_{i,1985}^{cigarettes})$	−0.94** (0.21)	−1.34** (0.23)	−1.20** (0.20)
$\ln(Inc_{i,1995}) - \ln(Inc_{i,1985})$	0.53 (0.34)	0.43 (0.30)	0.46 (0.31)
Intercept	−0.12 (0.07)	−0.02 (0.07)	−0.05 (0.06)
Instrumental variable(s)	Sales tax	Cigarette-specific tax	Both sales tax and cigarette-specific tax
First-stage F-statistic	33.70	107.20	88.60
Overidentifying restrictions J-test and p-value	—	—	4.93 (0.026)

These regressions were estimated using data for 48 U.S. states (48 observations on the 10-year differences). The data are described in Appendix 12.1. The J-test of overidentifying restrictions is described in Key Concept 12.6 (its p-value is given in parentheses), and the first-stage F-statistic is described in Key Concept 12.5. Individual coefficients are statistically significant at the *5% significance level or **1% significance level.

Application: Demand for Cigarettes

- The J-statistic rejection says that *at least one of the instruments* is **endogenous**, you have to choose one of them to be the instrument.
- There are *three* logical possibilities
  - The *sales tax* is exogenous but the *cigarette-specific tax* is not, in which case the column (1) regression is reliable;
  - The *cigarette-specific tax* is exogenous but the *sales tax* is not, so the column (2) regression is reliable;
  - or neither tax is exogenous, so neither regression is reliable. The statistical evidence cannot tell us which possibility is correct, so we must use our judgement.
- Therefore, how to choose in the case?

Application: Demand for Cigarettes

- The exogeneity of the *general sales tax* may be **stronger** than that for the cigarette-specific tax.
- Because the political process can link changes in the cigarette-specific tax to changes in the cigarette market and smoking policy.In other words, the cigarette-specific tax is more likely endogenous.
    - e.g. If smoking decreases in a state because it falls out of fashion,there will be fewer smokers and a weakened lobby against cigarette specific tax increases, which in turn could lead to higher cigarette-specific taxes.
- Therefore, the result using the *general sales tax* as an instrument is more **reliable**.
- Conclusion: The estimate of -0.94 indicates that cigarette consumption is somewhat **elastic**.
    - *An increase in price of* 1% *leads to a decrease in consumption of* 0.94%*.*

The Nature of IV: identification of Heterogeneous Causal Effects

Heterogeneous Populations

- So far, all models we learned have to be satisfied by a strong latent hypothesis: **No Heterogeneity**.
- It means that if the sample could be divide by m heterogeneous groups, then we assume that the estimate coefficient β_j for the jth independent variable, X_j,are the same among all groups(M) of the sample. Thus
- If the population is heterogeneous, then the i^{th} individual now has his or her own causal effect, β_{1i}.
- Taking the heterogeneous effect model can help us to understand further where the identification comes from when we use IV.

Angrist(1990)

Question: Earnings and Veteran

"Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records".The American Economic Review,Vol. 80, No. 3 (Jun., 1990), pp. 313-336

- **Topic**: How does **veteran** status effect on **earnings** for Americans.
- **Question**: *What is the difficulty of identification?*
  - Normall, joinning the military is a selection process, which means that the earnings is not exogenous.
- **Methods**: IV, the lottery outcome as an instrument for veteran status

Background

- In the 1960s and 70s young men in the US were at risk of being drafted for military service in Vietnam.
- Fairness concerns led to the institution of a draft lottery in 1970 that was used to determine priority for conscription.
- In each year from 1970 to 1972, random sequence numbers were randomly assigned to each birth date in cohorts of 19-year-olds.
  - Men with lottery numbers below a cutoff were eligible for the draft
  - Men with lottery numbers above the cutoff were not.

Lottery as an IV

- The instrument(Z_i) is thus defined as follows:
- The econometrician observes treatment status(D_i) as follows:

IV's Relevance and Exogenous

- While the lottery didn't completely determine veteran status, it certainly mattered: **relevance**.
- The lottery outcome was **random** and seems reasonable to suppose that its only effect was on veteran status: **exogenous**.

Heterogeneous Effects

- It classifies the population according to assignment(Z) an treatment(D) into **four** groups

	Z=1	
	D=0	D=1
D=0		

Heterogeneous Effects

- It classifies the population according to assignment(Z) an treatment(D) into **four** groups

		Z=1	
		D=0	D=1
Z=0	D=0		
	D=1		

Heterogeneous Effects

- It classifies the population according to assignment(Z) an treatment(D) into **four** groups

		Z=1	
		D=0	D=1
Z=0	D=0		
	D=1		

Heterogeneous Effects

- It classifies the population according to assignment(Z) an treatment(D) into **four** groups

		Z=1	
		D=0	D=1
Z=0	D=0		
	D=1		

Local Average Treatment Effect(LATE)

- Because the IV relevance means that the variations of IV can explain some variations of endogenous variable(D).
- And first and second stages regression means that only the variations of Z is used to restore the true value of β.
- In other words, IV estimate the D effect on Y on based on the "behavior-changers" under the instrument, who is only the sub-population: compliers.
- Angrist and Imbens(1994) called it as Local Average Treatment Effect(LATE), thus the treatment effect on those that change their behaviors under the instrument.

Local Average Treatment Effect(LATE)

- Two basic assumptions for IV estimation
  - relevance: $Cov(Z_i, D_i) neq = 0$
  - exogenous: $Cov(Z_i, u_i) = 0$
- The third assumption: **Monotonicity**

  - It ensures that the responders all go in one direction
  - It excludes the the unreasonable defiers in the sample.

IV with Heterogeneous Causal Effects: Generalization

- If we assume effects are the *same* for all three groups, then the constant-effects IV model is still valid.
- But if the population is *heterogeneous*, then *the LATE could be not ATE*.
- Let us assume that i^{th} individual now has his or her own causal effect, β_{1i},then the population regression equation can be written

$$Y_i = \beta_{0i} + \beta_{1i} X_i + u_i \tag{13.9}$$

- β_{1i} now is a **random variable** that,just like u_i, reflects unobserved variation across individuals.
- The average causal effect is the population mean value of the causal effect,$E(\beta_{1i})$ which is the *expected causal effect* of a randomly selected member of the population.

OLS with Heterogeneous Causal Effects

- If there is heterogeneity in the causal effect and if X_i is **randomly assigned**, then the OLS with heterogeneous estimator is still a **consistent** estimator of the average causal effect.

- Thus, if X_i is randomly assigned, $\hat{\beta}_1$ is a *consistent* estimator of the average causal effect $E(\beta_{1i})$.

IV Regression with Heterogeneous Causal Effects

- Specifically, suppose that X_i is related to Z_i by the linear model

$$X_i = \pi_{0i} + \pi_{1i} Z_i + v_i$$

- where the coefficients π_{0i} and π_{1i} vary from one individual to the next.
- And it is the first-stage equation of TSLS with the modification of heterogeneous effect of Z on X.
- Then we can prove it that when there is population heterogeneity in the treatment effect and in the influence of the instrument on the receipt of treatment, the IV estimator will have the following formula

- *Please prove it by yourself (refers to S.W. Appendix 13.2) or see the Appendix.*

IV Regression with Heterogeneous Causal Effects

- Then

- It is a **weighted** average of the individual causal effects β_{1i}, The weights are $\frac{\pi_{1i}}{E(\pi_{1i})}$, which measure the **relative degree** to which the instrument influences whether the i_{th} individual receives treatment,
- In other words,TSLS estimator is a consistent estimator of a *weighted average of the individual causal effects*, where the individuals who receive the *most weight* are those for *whom the instrument is most influential*.

IV Regression with Heterogeneous Causal Effects

- Three special cases:
  - The treatment effect is the same for all individuals.

$$\beta_{1i} = \beta_1$$

  - The instrument affects each individual equally.

$$\pi_{1i} = \pi_1$$

  - The heterogeneity in the treatment effect and heterogeneity in the effect of the instrument are uncorrelated.

$$Cov(\beta_{1i}\pi_{1i}) = 0$$

IV Regression with Heterogeneous Causal Effects

- LATE equals to the ATE: all three cases we have

$$\frac{E(\beta_{1i}\pi_{1i})}{E(\pi_{1i})} = E(\beta_{1i}) = \beta_1$$

- Aside from these three special cases, in general the local average treatment effect **differs** from the average treatment effect.

IV Regression with Heterogeneous Effects

- Different instruments can identify different parameters because they estimate the impact on different populations.
- The difference arises because each researcher is implicitly estimating a different weighted average of the individual causal effects in the population.
- Recall: **J-test of over-identifying restrictions** can reject if the two instruments estimate different local average treatment effects,even if both instruments are valid. In general neither estimator is a consistent estimator of the average causal effect.

Wrap Up

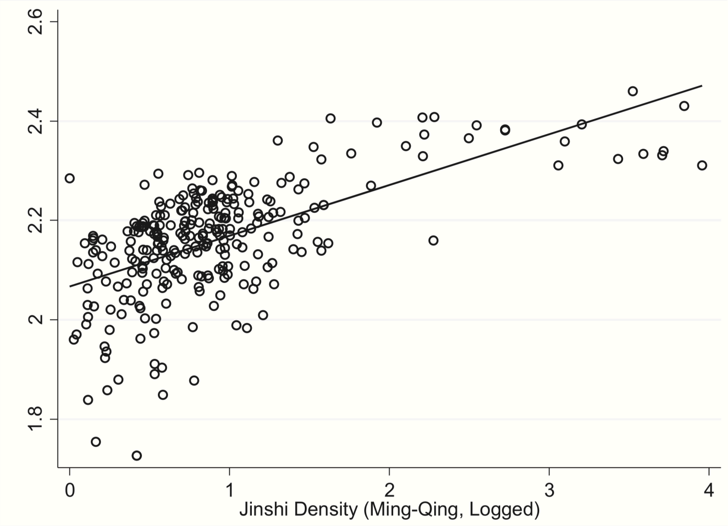
- The IV paradigm provides a powerful and flexible framework for causal inference.
- An alternative to random assignment with a strong claim on internal validity.
- The LATE framework highlights questions of external validity
  - Can one instrument identify the average effect induced by another source of variation?
  - Can we go from average effects on compliers to average effects on the entire treated population or an unconditional effect?
- The answer to these questions is usually: **NO**, at least without additional assumptions.

A Good Example: Long live Keju(“科举万岁”)

Chen, Kung and MA(2020)

- Ting Chen, James Kai-sing Kung(龚启圣) and Chicheng Ma(2020), *"Long Live Keju! The Persistent Effects of China's Imperial Examination System"*, *The Economic Journal*, 130 (October), 2030–2064.
- Topic: Long term persistence of human capital:the effect of **Keju**
- Dependent Variable: average schooling years in 2010
- Independent Variable: the density of **jinshi** in the Ming-Qing dynasties
- Data: 272 prefectures in *jinshi*.

Chen, Kung and MA(2020)



Chen, Kung and MA(2020)

- The effect of Keju on human capital at present
- Run regression

$$lnY_i = \alpha + \beta ln(Keju_i) + \gamma_1 X_i^c + \gamma_2 X_i^h + u_i$$

- Y_i: 2010 年 i 地区 (地级市或" 府") 的平均受教育年限。
- $Keju_i$: 明清时期 i 地区获得进士的人数。
- X_i^c: 控制变量（当代），包括经济繁荣程度 (夜间灯光)；地理因素：该地区到海选距离、地形（免于遭受自然灾害）。
- X_i^c: 控制变量（历史）：历史经济繁荣程度、基础教育设施、社会和政治影响力等等

Chen, Kung and MA(2020): OLS

Table 3. Impact of *Jinshi* Density on Contemporary Human Capital: OLS Estimates

	Average Years of Schooling in 2010 (logged)					
	(1)	(2)	(3)	(4)	(5)	(6)
Jinshi Density (logged)	0.092***	0.065***	0.070***	0.067***	0.058***	
	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	
Jinshi Density (logged, excludes migrant)						0.053***
						(0.019)
						[0.016]
Economic Prosperity						
Population Density (logged)			-0.049***	-0.051***	-0.053***	-0.049***
			(0.016)	(0.016)	(0.015)	(0.015)
			[0.013]	[0.013]	[0.012]	[0.013]
Urbanization Rate			0.062	0.093	0.051	0.234
			(0.163)	(0.156)	(0.164)	(0.180)
			[0.167]	[0.162]	[0.173]	[0.169]
Commercial Center			-0.012	-0.014	-0.020	-0.026*
			(0.014)	(0.014)	(0.013)	(0.014)
			[0.011]	[0.011]	[0.011]	[0.012]
Agricultural Suitability			-0.005	-0.005	-0.003	-0.004
			(0.014)	(0.014)	(0.014)	(0.014)
			[0.009]	[0.009]	[0.009]	[0.009]

Chen, Kung and MA(2020): Potential Bias

- **OVB**: that are simultaneously associated with both historical “jinshi” density and years of schooling today.
- For instance, prefectures that had produced more “jinshi” may be associated with unobserved (natural or genetic) endowments.

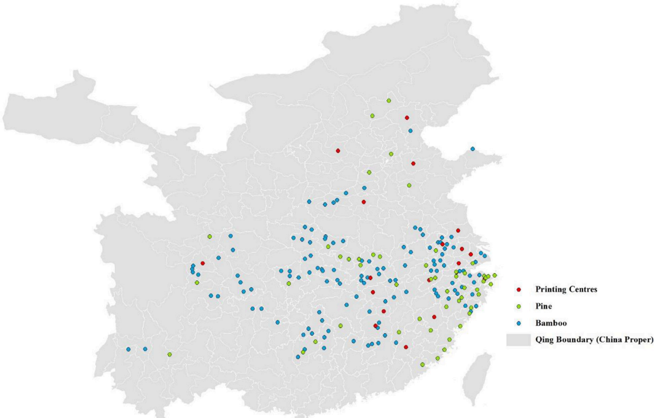
Chen, Kung and MA(2020): Instrumental Variable

- IV: Distance to the Printing Ingredients (Pine and Bamboo) as the Instrumental Variable of "Keju"
- A logic chain:

$$\begin{aligned} \textit{More jinshi} &\Longleftarrow \textit{more references books} \\ &\Longleftarrow \textit{more print in print centers} \\ &\Longleftarrow \textit{centers closer nearby some ingredients} \end{aligned}$$

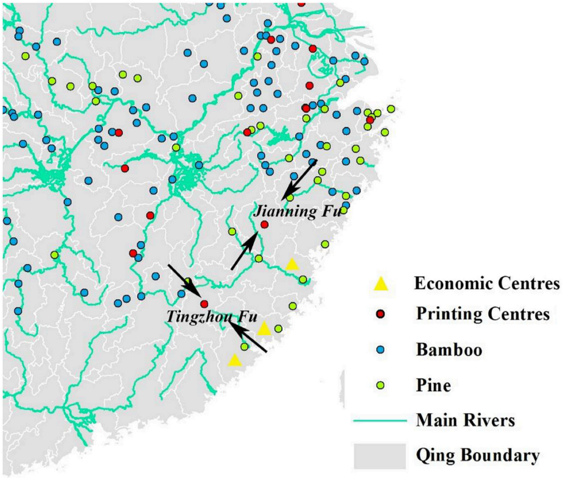
Chen, Kung and MA(2020): Instrumental Variable

- Only 19 printing centres were distributed across the 278 prefectures, and that these 19 centres accounted for 80% of the 13,050 texts published during that period (Zhang and Han, 2006)



Chen, Kung and MA(2020): Instrumental Variable

- Jianning Fu(建寧府) and Tingzhou FU(汀州府)



Chen, Kung and MA(2020): First Stage

Table 4. River Distance to Pine and Bamboo Locations, Printing Centers and *Jinshi* Density

	Jinshi Density (logged)		Printing Center		Printed Books (logged)		*Jinshi* Density (logged)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Printed Books (logged)	0.179***	0.170***						
	(0.031)	(0.036)						
River Distance to Pine/Bamboo			-0.017***	-0.017***	-0.092***	-0.084***	-0.102***	-0.099***
			(0.004)	(0.004)	(0.029)	(0.029)	(0.011)	(0.012)
Baseline Control Variables	No	Yes	No	Yes	No	Yes	No	Yes
Provincial Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	274	274	274	274	274	274	274	274
Adj. R-squared	0.323	0.332	0.132	0.131	0.449	0.463	0.526	0.528

More Books, More Jinshis

Closer to Pine/Bamboo, More Center, then More Books

Closer to Pine/ Bamboo, More Jinshis

Notes: All results are OLS estimates. Baseline controls include agricultural suitability, distance to coast, and terrain ruggedness. Robust standard errors adjusted for clustering at the province level are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

Chen, Kung and MA(2020): Reduced-form and 2SLS

Table 7. Impact of *Keju* on Contemporary Human Capital: Instrumented Results

	Reduced-form			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Jinshi Density (logged)				0.104***	0.080***	0.082***
				(0.008)	(0.013)	(0.013)
Distance to Major Navigable Rivers			0.008			0.008
			(0.006)			(0.006)
				First Stage		
River Distance to Bamboo/Pine	-0.011***	-0.006***	-0.006***	-0.011***	-0.006***	-0.006***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
First Stage F-stat				78.04	58.07	57.76
First Stage Partial R-squared				0.392	0.282	0.282
Baseline + Additional Controls	No	Yes	Yes	No	Yes	Yes
Provincial Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	272	272	272	272	272	272
Adj. R-squared	0.531	0.732	0.735	0.65	0.751	0.752
Cragg-Donald Wald F-statistic				129.156	72.314	72.354

Notes: Baseline controls include nighttime lights in 2010, agricultural suitability, distance to coast, and terrain ruggedness. Additional controls are commercial center, population density, urbanization rate, Confucian academies, private book collections, strength of clan and political elites. Robust standard errors adjusted for clustering at the province level are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10%, respectively.

Chen, Kung and MA(2020): Exclusion Restrictions

- The locations of bamboo and pine geographic distributions were exogenously given. Historians find little if any evidence of planting pine and bamboo intentionally for the purpose of commercial printing.
- But they may be correlated with other omitted variables—most notably **economic prosperity**, which may be correlated with years of schooling today.

Table A3. Exclusion Restrictions

Panel A	Commercial Centers in Ming-Qing	Tea Centers in Ming-Qing	Silk Centers in Ming-Qing	Population Density in Ming (logged)	Population Density in Qing (logged)	Population Density in 1953 (logged)	Urbanization Rate in Ming-Qing	Urbanization Rate in 1920
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
River Distance to Pine/Bamboo	-0.006	-0.007	-0.008	0.007	-0.020	-0.021	-0.001	-0.020
(logged)	(0.005)	(0.005)	(0.007)	(0.045)	(0.019)	(0.017)	(0.001)	(0.024)
Observations	274	274	274	274	274	269	274	274
Adjusted R-squared	0.309	0.216	0.153	0.534	0.624	0.540	0.664	0.296

Practical Guides of Using IV

Pratical tips of using IV

Explaination and Check Revelence

1. Explain your identification strategy very clearly.
   - start with the ideal experiment; why is your setting different? Why is your regressor endogenous?
   - explain theoretically why there should be a first stage and what coefficient we should expect.
   - explain why the instrument is as good as randomly assigned.
   - explain theoretically why the exclusion restriction holds in your setting.
2. Show and discuss the first stage regression
   - start with a raw correlation, graph is the better way if possible.
   - Always report the first stage and think about whether it makes sense(signs and magnitudes)
   - Always report the F-statistic on the excluded instruments to avoid weak IV.

Check Exogeneity

3. Check exogeneity including exclusion restriction
    - Show that the instrument does not predict pre-treatment characteristics.
    - Run and examine the reduced form(regression of dependent variable on instruments) and look at the coefficients, t-statistics and F-statistics for excluded instruments.
    - Exclusion restriction can not be tested directly but a *Falsification test* can help.
    - Consider using the plausibly exogenous bounding procedure by Conley et al. (2012)
4. If you have multiple instruments, report over-identification tests.
    - Pick your best single instrument and report just-identified estimates using this one only because just-identified IV is relatively unlikely to be subject to a weak bias.
    - Worry if it is substantially different from what you get using multiple instruments.
    - Check over-identified 2SLS estimates with LIML. LIML is less than precise than 2SIS but also less biased.

Discuss the Results and Limitations in detail

5. Provide a substantive explanation for observed difference between 2SLS and OLS
   - How bid is the difference? What does this tell you?
   - Is the coefficient bigger when theory of endogeneity suggests it should be smaller? If so, why?
   - Measurement Error or heterogeneous effects?
6. What LATE is being estimated?
   - Whose behavior is affected by the instrument?
   - Is this the LATE you would want? Is it a quantify of theoretical interest?
   - Would other LATEs possible yield different estimates?

How to Evaluate IV paper in a simple way?

1. Relevant: The first stage regression
   - Does the author report the first stage regression?
   - Does the instrument perform well in the first stage?
   - Testable: rule of thumb: first stage $F > 10$
2. Exclusion restriction:
   - Is the instrument exogenous enough?(the random assignment is the best)
   - Would you expect a direct effect of Z on Y
   - Not directly testable: Except when equation is overidentified.

Where Do Valid Instruments Come From?

Where do we find an IV?

- Generally Speaking
  - “可遇不可求”
- Two main approaches
  1. Economic Theory/Logics
  2. Exogenous Source of Variation in X(natural experiments)

Where do we find an IV?

1. Institutional Background

- Angrist(1990)-draft lottery: Vietnam veterans were randomly designated based on birth day used to estimate the wage impact of a shorter work experience.
- Acemoglu, Johnson, and Robinson(2001): the dead rate of some diseases in some areas to estimate the impact of institutions to economic growth.
- Li and Zhang(2007),Liu(2012)- "One Child policy"

Where do we find an IV?

2. Natural conditions(geography,weather,disaster)

- the Rainfall,Hurricane,Earthquake,Tsunami...
- the number of Rivers: Hoxby(2000)
- the distance to print ingredients: Chen,Gong and Ma(2020)

Where do we find an IV?

3. Economic theory and Economic logic

- study the alcohol consumption and income relationship. alcohol price change by government's tax in a local market may be as a instrument of alcohol consumption.
- Angrist & Evans(1998): have same sex or different sex children used to estimate the impact of an additional birth on women labor supply.

Where do we find an IV? Some classic examples

- Example 1: Does putting criminals in jail reduce crime?
- Run a regression of crime rates(d.v.) on incarceration rates(id.v) by using annual data at a suitable level of jurisdiction(states) and covariates (economic conditions)
- *Simultaneous causality bias*: crime rates goes up, more prisoners and more prisoners,reduced crime.
- IV: it must affect the incarceration rate but be unrelated to any of the unobserved factors that determine the crime rate.
- Levitt (1996) suggested that *lawsuits aimed at reducing prison overcrowding* could serve as an instrumental variable.
- Result: The estimated effect was three times larger than the effect estimated using OLS.

Where do we find an IV?: Some classic examples

- Example 2: Does cutting class sizes increase test scores?
- *Omitted Variable bias*: such as parental interest in learning, learning opportunities outside the classroom, quality of the teachers and school facilities.
- IV: correlated with class size (relevance) but uncorrelated with the omitted determinants of test performance.
- Hoxby (2000) suggested biology. Because of random fluctuations in timings of births, the size of the incoming kindergarten class varies from one year to the next.
- But potential enrollment also fluctuates because parents with young children choose to move into an improving school district and out of one in trouble. She used the deviation of potential enrollment from its long-term trend as her instrument.
- Result: the effect on test scores of class size is small.

Where do we find an IV? Some new techniques

4. Bartick or Shift-Share IV

- Consider a **local** labor market regression like the following:

$$Y_c = \beta_0 + \beta_1 X_c + \varepsilon_c$$

  - Where X_c is the shock to location c such as exposures to foreign import competition.
  - Think: how to define and measure the **exposures to foreign import competition**?
- **Concern**: shock may be correlated with error term.
- **Solution**: find a feasible IV for X_c

A Shift-Share Instrumental Variable(SSIV)

- **Solution**: the SSIV is a **weighted sum** of a common set of shocks,with **weights** reflecting heterogeneous exposure shares.

$$Z_c = \sum_{k=1}^{K} s_{ck} g_k$$

  - where s_{ck} is the *exposure* to sector(**share**) k in city c.
  - g_k is the *exogenous shock* to (**shift**) sector k on the country level.

Autor et al(2013): The China Shocks

- The impact of rising Chinese imports on manufacturing employment in U.S. local labor market(locations denoted by c)

$$\Delta Y_{ct} = \beta_0 + \beta_1 E_{ct} + \varepsilon_{ct}$$

where ΔY_{ct} is the change in employment rate in city c at time t.

- With import exposure defined by:

$$E_{ct} = \sum_{k=1}^{k} E_{ckt} G_{kt}^{US}$$

Where the E_{ckt} are the start of the period shares of employment in location c in each industry k at time t(or $t-1$), and G_{kt}^{US}US is a normalized measure of growth in imports from China to the U.S in each industry k at period t.

- However, the import exposure here is possibly endogenous to the local employment.

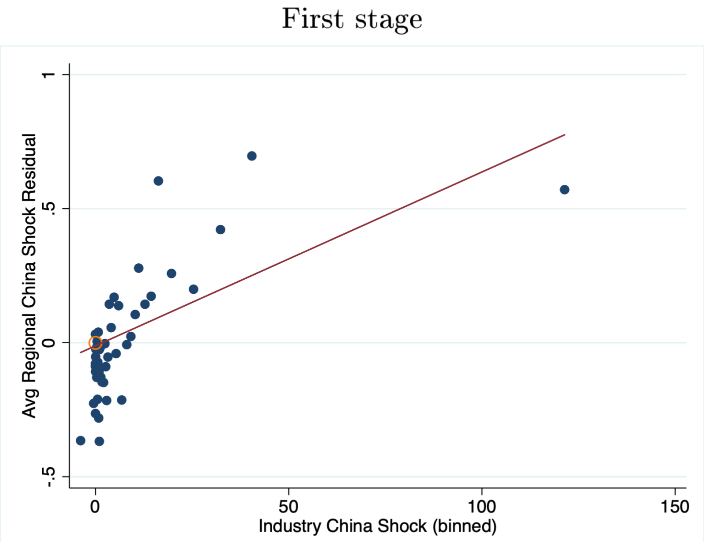
Autor et al(2013): The China Shocks

- The Bartik instrument

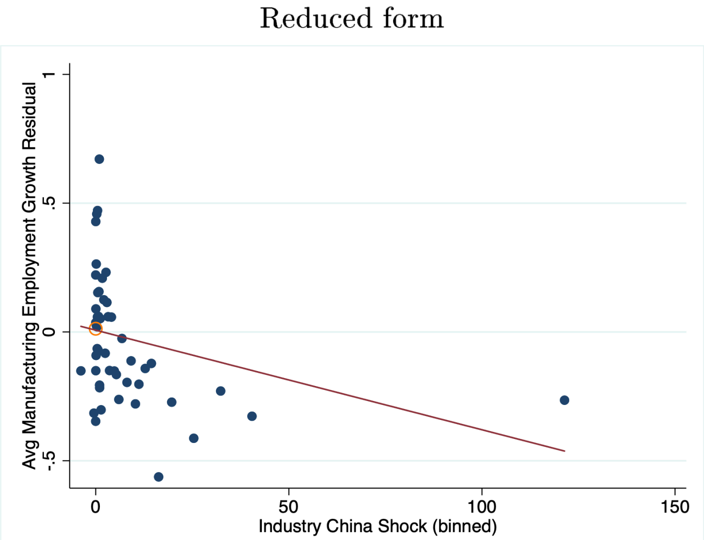
$$B_{ct} = \sum_{k=1}^{K} E_{ck,t-1} G_{kt}^{OtherCountry}$$

- Where the E now are lagged ("initial") shares of employment in location i in industry j, and $G_{jt}^{OtherCountry}$ is growth of imports from China to other high-income countries.
- That is, the predicted exposure of a location to Chinese imports is a weighted average of how much China is exporting in general of different products (the "shift"), with weights that come from the initial industry composition in a location (the "shares").

Autor et al(2013): The China Shocks



Autor et al(2013): The China Shocks



Summary and Appendix

Summary

- IVs have become *less and less popular* in recent years.
  - very difficult to find an IV that fulfills the exclusion restriction.
  - The LATE is often not the desired policy parameter.
  - IV has **unfavourable small sample properties**.
- Many classic IVs have been shown to be invalid.
  - Quarter of birth is correlated with SES.
  - Twin births (IV for family size) are also related to SES.
- In what settings are IVs used these days?
  - In randomized experiments with imperfect compliance.
  - In fuzzy regression discontinuity designs.
  - As a complementary identification strategy, along with FE estimation and DID.
  - Bartick IV

Appendix 1: 2SLS Variance Formula

2SLS Variance Formula in large samples

- Starting with the expression for $\hat{\beta}_{2SLS}$:

$$\hat{\beta}_{2SLS} = \beta_1 + \frac{\frac{1}{n}\sum u_i(Z_i - \bar{Z})}{\frac{1}{n}\sum(X_i - \bar{X})(Z_i - \bar{Z})}$$

- Since β_1 is a constant, it does not affect the variance.Then the variance of the 2SLS estimator is:

$$Var(\hat{\beta}_{2SLS}) = Var\left(\frac{\frac{1}{n}\sum u_i(Z_i - \bar{Z})}{\frac{1}{n}\sum(X_i - \bar{X})(Z_i - \bar{Z})}\right)$$

2SLS Variance Formula in large samples: denominator

- Given Continuous Mapping Theorem, we separate the variance into the numerator and denominator in the large sample limit.
- For large samples, the **denominator** converges to:

$$\frac{1}{n}\sum(X_i - \bar{X})(Z_i - \bar{Z}) \xrightarrow{p} Cov(X, Z) = \rho_{XZ}\sigma_X\sigma_Z$$

  - Where ρ_{XZ} is the correlation between X and Z.
- Then the variance of the denominator in the large sample converges to:

$$Var\left(\frac{1}{n}\sum(X_i - \bar{X})(Z_i - \bar{Z})\right) \xrightarrow{p} \frac{\rho_{XZ}^2\sigma_X^2\sigma_Z^2}{n}$$

2SLS Variance Formula in large samples: numerator

- For the numerator's variance:

$$Var\left(\frac{1}{n}\sum u_i(Z_i - \bar{Z})\right) = \frac{1}{n^2} Var\left(\sum u_i(Z_i - \bar{Z})\right)$$

- Because both u_i and Z_i are i.i.d. and $E(u_i|Z_i) = 0$, we have:

$$Cov\left(u_i(Z_i - \bar{Z}), u_j(Z_j - \bar{Z})\right) = 0$$

- Then we could rewrite the numerator's variance as:

$$\frac{1}{n^2} Var\left(\sum u_i(Z_i - \bar{Z})\right) = \frac{1}{n^2}\left(\sum Var(u_i(Z_i - \bar{Z}))\right)$$

2SLS Variance Formula in large samples: numerator

Math Review: The law of total variance

$$Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$$

- Then we have:

$$\begin{aligned}
Var\left(u_i(Z_i - \bar{Z})\right) &= E\left[Var\left(u_i(Z_i - \bar{Z})|Z\right)\right] + Var\left(E\left[u_i(Z_i - \bar{Z})|Z\right]\right) \\
&= E\left[(Z_i - \bar{Z})^2 Var\left(u_i|Z\right)\right] + Var\left((Z_i - \bar{Z})E[u_i|Z]\right) \\
&= E\left[(Z_i - \bar{Z})^2 \sigma_u^2\right] \ \because Var(u_i|Z) = \sigma_u^2 \ and \ E[u_i|Z] = 0 \\
&= Var(Z_i - \bar{Z})\sigma_u^2 \ \because E(Z_i - \bar{Z}) = 0
\end{aligned}$$

2SLS Variance Formula in large samples: numerator

- Then we have:

$$\begin{aligned}\frac{1}{n^2}\left(\sum Var(u_i(Z_i-\bar{Z}))\right) &= \frac{1}{n^2}\sum Var(Z_i-\bar{Z})\sigma_u^2\\ &= \frac{1}{n^2}\sigma_u^2\cdot\sum Var(Z_i-\bar{Z})\\ &= \frac{\sigma_u^2}{n^2}\sum Var(Z_i-\bar{Z})\end{aligned}$$

- Now we need to determine $\sum Var(Z_i-\bar{Z})$.
- We will show that:

$$\sum_{i=1}^{n}\mathrm{Var}(Z_i-\bar{Z}) \xrightarrow{p} n\sigma_Z^2$$

2SLS Variance Formula in large samples: numerator

- We would like to divide the sum of the variance into two parts: with Z_i and without Z_i.
- Then first we rewrite $Z_i - \bar{Z}$ into two parts:

$$Z_i - \bar{Z} = Z_i - \frac{1}{n}\sum_{j=1}^{n} Z_j = Z_i - \frac{Z_i}{n} - \frac{1}{n}\sum_{j \neq i} Z_j = \frac{n-1}{n} Z_i - \frac{1}{n}\sum_{j \neq i} Z_j$$

- Then the variance of $Z_i - \bar{Z}$ is:

$$\mathrm{Var}(Z_i - \bar{Z}) = \mathrm{Var}\left(\frac{n-1}{n} Z_i - \frac{1}{n}\sum_{j \neq i} Z_j\right)$$

2SLS Variance Formula in large samples: numerator

- Then

$$\begin{aligned}
\mathrm{Var}\left(\frac{n-1}{n}Z_i - \frac{1}{n}\sum_{j\neq i} Z_j\right) &= \mathrm{Var}\left(\frac{n-1}{n}Z_i\right) + \mathrm{Var}\left(\frac{1}{n}\sum_{j\neq i} Z_j\right) \because Z_i \perp \sum_{j\neq i} Z_j \\
&= \frac{(n-1)^2}{n^2}\sigma_Z^2 + \frac{1}{n^2}\mathrm{Var}\left(\sum_{j\neq i} Z_j\right) \\
&= \frac{(n-1)^2}{n^2}\sigma_Z^2 + \frac{1}{n^2}(n-1)\sigma_Z^2 \\
&= \sigma_Z^2\left[\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right] \\
&= \sigma_Z^2\frac{n-1}{n}
\end{aligned}$$

- It means that the summation of the variance of $Z_i - \bar{Z}$ is:

$$\sum_{i=1}^{n} \mathrm{Var}(Z_i - \bar{Z}) = n \cdot \sigma_Z^2 \cdot \frac{n-1}{n} = (n-1)\sigma_Z^2 \xrightarrow{p} n\sigma_Z^2$$

2SLS Variance Formula in large samples: numerator

- Now we obtain the numerator's variance:

$$
\begin{aligned}
Var\left(\frac{1}{n}\sum u_i(Z_i - \bar{Z})\right) &= \frac{\sigma_u^2}{n^2}\sum Var(Z_i - \bar{Z}) \\
&= \frac{\sigma_u^2}{n^2} \cdot n\sigma_Z^2 \\
&= \frac{\sigma_u^2\sigma_Z^2}{n}
\end{aligned}
$$

2SLS Variance Formula in large samples: Combining results

- Now we combine the results:

$$
\begin{aligned}
Var(\hat{\beta}_{2SLS}) &\xrightarrow{p} \frac{\frac{\sigma_u^2 \sigma_Z^2}{n}}{(\rho_{XZ}\sigma_X\sigma_Z)^2} \\
&= \frac{\sigma_u^2 \sigma_Z^2}{n \cdot \rho_{XZ}^2 \sigma_X^2 \sigma_Z^2} \\
&= \frac{\sigma_u^2}{n\sigma_X^2 \rho_{XZ}^2}
\end{aligned}
$$

- This is the variance of the 2SLS estimator.

Appendix 2: IV and Heterogeneous Causal Effects

IV and Heterogeneous Causal Effects

- Let us prove that when there is population heterogeneity in both the treatment effect and the influence of the instrument on treatment receipt, the IV estimator converges to the following formula:

$$\hat{\beta}_{2SLS} \xrightarrow{p} \frac{Cov(ZY)}{Cov(ZX)} = \frac{E(\beta_{1i}\pi_{1i})}{E(\pi_{1i})}$$

IV Regression with Heterogeneous Causal Effects

- At first

$$Cov(ZX) = E[(Z - \mu_Z)(X - \mu_X)]$$

IV Regression with Heterogeneous Causal Effects

- At first

$$
\begin{aligned}
Cov(ZX) &= E[(Z-\mu_Z)(X-\mu_X)] \\
&= E[(Z-\mu_Z)X]
\end{aligned}
$$

IV Regression with Heterogeneous Causal Effects

- At first

$$
\begin{aligned}
Cov(ZX) &= E[(Z-\mu_Z)(X-\mu_X)] \\
&= E[(Z-\mu_Z)X] \\
&= E[(Z_i-\mu_Z)\left(\pi_{0i}+\pi_{1i}Z_i+v_i\right)]
\end{aligned}
$$

- At first

$$
\begin{aligned}
Cov(ZX) &= E[(Z-\mu_Z)(X-\mu_X)] \\
&= E[(Z-\mu_Z)X] \\
&= E[(Z_i-\mu_Z)\left(\pi_{0i}+\pi_{1i}Z_i+v_i\right)] \\
&= E(\pi_{0i})E\left(Z_i-\mu_Z\right)+E(\pi_{1i})E\left[Z_i\left(Z_i-\mu_Z\right)\right]+\operatorname{cov}\left(Z_i,v_i\right)
\end{aligned}
$$

IV Regression with Heterogeneous Causal Effects

- At first

$$\begin{aligned}
Cov(ZX) &= E[(Z - \mu_Z)(X - \mu_X)] \\
&= E[(Z - \mu_Z)X] \\
&= E[(Z_i - \mu_Z)\left(\pi_{0i} + \pi_{1i}Z_i + v_i\right)] \\
&= E(\pi_{0i})E\left(Z_i - \mu_Z\right) + E(\pi_{1i})E\left[Z_i\left(Z_i - \mu_Z\right)\right] + \operatorname{cov}\left(Z_i, v_i\right) \\
&= 0 + E(\pi_{1i})E\left[Z_i\left(Z_i - \mu_Z\right)\right] + 0
\end{aligned}$$

IV Regression with Heterogeneous Causal Effects

- At first

$$
\begin{aligned}
Cov(ZX) &= E[(Z-\mu_Z)(X-\mu_X)] \\
&= E[(Z-\mu_Z)X] \\
&= E[(Z_i-\mu_Z)\left(\pi_{0i}+\pi_{1i}Z_i+v_i\right)] \\
&= E(\pi_{0i})E\left(Z_i-\mu_Z\right)+E(\pi_{1i})E\left[Z_i\left(Z_i-\mu_Z\right)\right]+\operatorname{cov}\left(Z_i,v_i\right) \\
&= 0+E(\pi_{1i})E\left[Z_i\left(Z_i-\mu_Z\right)\right]+0 \\
&= Var(Z)E(\pi_{1i})
\end{aligned}
$$

IV Regression with Heterogeneous Causal Effects

- Second,

$$Y_i = \beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i}Z_i + v_i\right) + u_i$$

- Then

$$Cov(ZY) = E[(Z - \mu_Z)(X - \mu_X)]$$

IV Regression with Heterogeneous Causal Effects

- Second,

$$Y_i = \beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i}Z_i + v_i\right) + u_i$$

- Then

$$\begin{aligned} Cov(ZY) &= E[(Z-\mu_Z)(X-\mu_X)] \\ &= E[(Z-\mu_Z)Y] \end{aligned}$$

IV Regression with Heterogeneous Causal Effects

- Second,

$$Y_i = \beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i$$

- Then

$$\begin{aligned} Cov(ZY) &= E[(Z - \mu_Z)(X - \mu_X)] \\ &= E[(Z - \mu_Z)Y] \\ &= E[(Z_i - \mu_Z)\left(\beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i\right)] \end{aligned}$$

IV Regression with Heterogeneous Causal Effects

- Second,

$$Y_i = \beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i}Z_i + v_i\right) + u_i$$

- Then

$$\begin{aligned}
Cov(ZY) &= E[(Z - \mu_Z)(X - \mu_X)] \\
&= E[(Z - \mu_Z)Y] \\
&= E[(Z_i - \mu_Z)\left(\beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i}Z_i + v_i\right) + u_i\right)] \\
&= E(\beta_{0i})E\left(Z_i - \mu_Z\right) + Cov(Z, \beta_{1i}\pi_{0i})
\end{aligned}$$

IV Regression with Heterogeneous Causal Effects

- Second,

$$Y_i = \beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i$$

- Then

$$\begin{aligned}
Cov(ZY) &= E[(Z - \mu_Z)(X - \mu_X)] \\
&= E[(Z - \mu_Z)Y] \\
&= E[(Z_i - \mu_Z)\left(\beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i\right)] \\
&= E(\beta_{0i})E\left(Z_i - \mu_Z\right) + Cov(Z, \beta_{1i}\pi_{0i}) \\
&+ E[\beta_{1i}\pi_{1i}Z_i(Z_i - \mu_Z)] + E[\beta_{1i}v_i(Z_i - \mu_Z)] + \text{cov}\left(Z_i, u_i\right)
\end{aligned}$$

IV Regression with Heterogeneous Causal Effects

- Second,

$$Y_i = \beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i$$

- Then

$$\begin{aligned}
Cov(ZY) &= E[(Z - \mu_Z)(X - \mu_X)] \\
&= E[(Z - \mu_Z)Y] \\
&= E[(Z_i - \mu_Z)\left(\beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i\right)] \\
&= E(\beta_{0i}) E\left(Z_i - \mu_Z\right) + Cov(Z, \beta_{1i}\pi_{0i}) \\
&+ E[\beta_{1i}\pi_{1i} Z_i (Z_i - \mu_Z)] + E[\beta_{1i} v_i (Z_i - \mu_Z)] + \mathrm{cov}\left(Z_i, u_i\right) \\
&= 0 + 0 + E(\beta_{1i}\pi_{1i}) E\left[Z_i\left(Z_i - \mu_Z\right)\right] + 0 + 0
\end{aligned}$$

IV Regression with Heterogeneous Causal Effects

- Second,

$$Y_i = \beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i$$

- Then

$$\begin{aligned}
Cov(ZY) &= E[(Z-\mu_Z)(X-\mu_X)] \\
&= E[(Z-\mu_Z)Y] \\
&= E[(Z_i-\mu_Z)\left(\beta_{0i} + \beta_{1i}\left(\pi_{0i} + \pi_{1i} Z_i + v_i\right) + u_i\right)] \\
&= E(\beta_{0i})E\left(Z_i-\mu_Z\right) + Cov(Z, \beta_{1i}\pi_{0i}) \\
&+ E[\beta_{1i}\pi_{1i}Z_i(Z_i-\mu_Z)] + E[\beta_{1i}v_i(Z_i-\mu_Z)] + \text{cov}\left(Z_i, u_i\right) \\
&= 0 + 0 + E(\beta_{1i}\pi_{1i})E\left[Z_i\left(Z_i-\mu_Z\right)\right] + 0 + 0 \\
&= Var(Z)E(\beta_{1i}\pi_{1i})
\end{aligned}$$