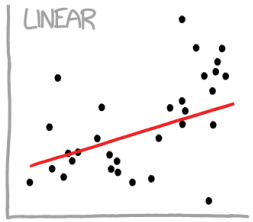
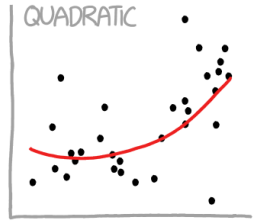
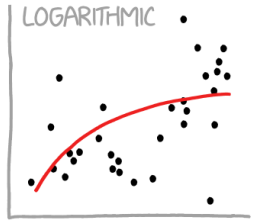
CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND



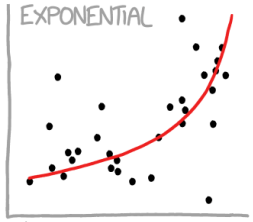
"HEY, I DID A REGRESSION."



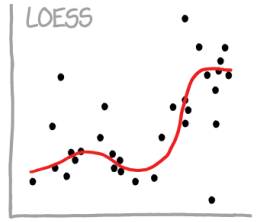
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."



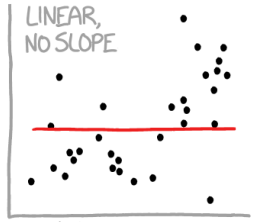
"LOOK, IT'S TAPERING OFF!"



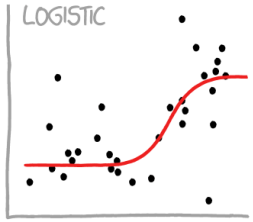
"LOOK, IT'S GROWING UNCONTROLLABLY!"



"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."

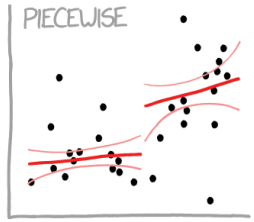


"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."

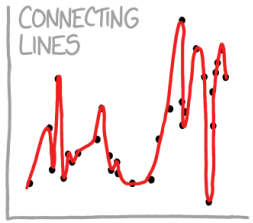


"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."

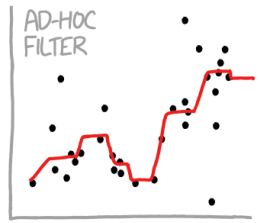
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."



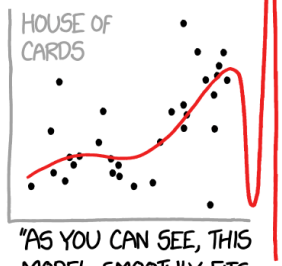
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."



"I CLICKED 'SMOOTH LINES' IN EXCEL."



"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"



"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE– *WAIT NO NO DON'T EXTEND IT AAAAAA!!*"

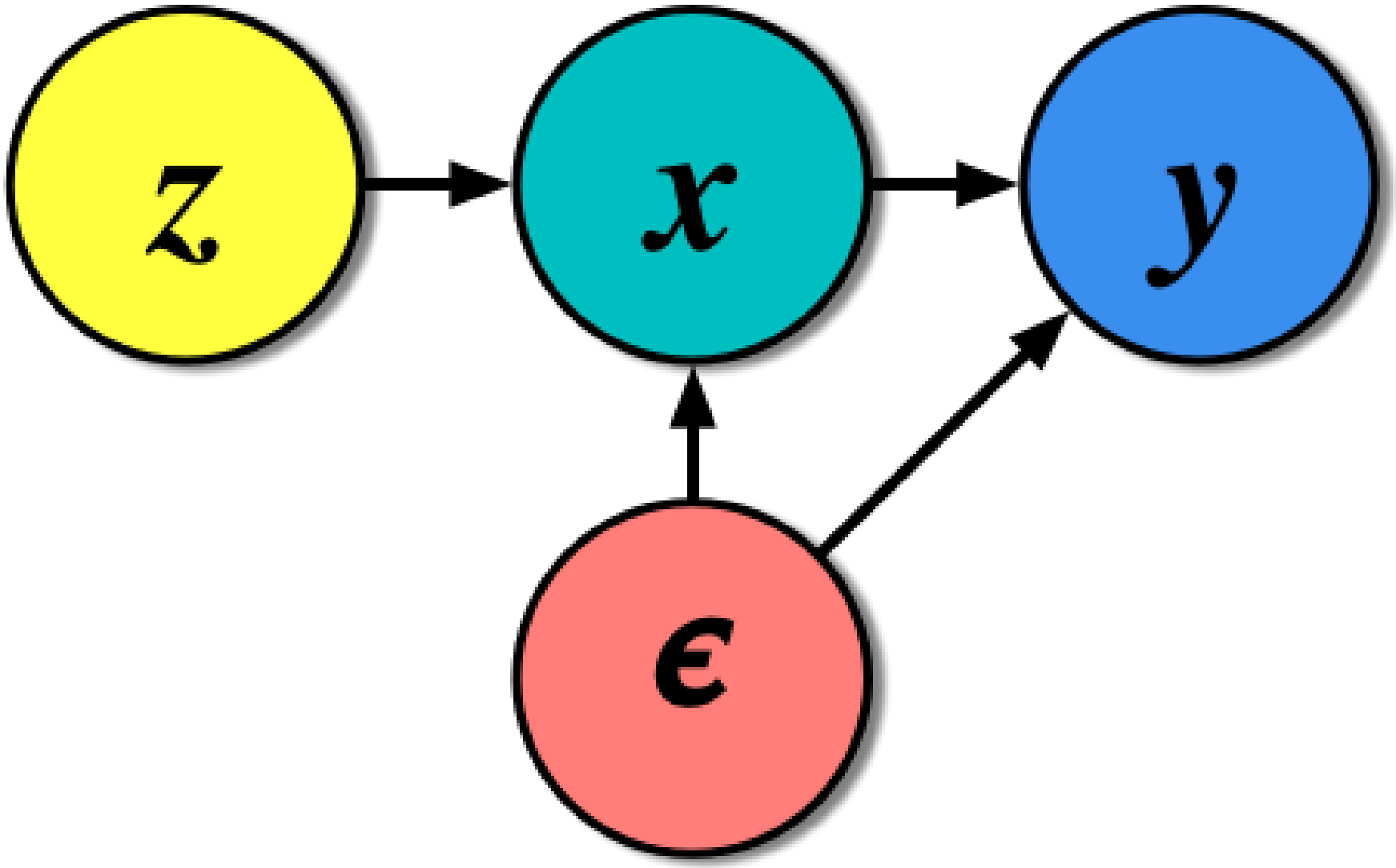
Recitation 8: Instrumental Variables

14.32 Fall 2023

Ian Sapollnik

Nov 20, 2023

What problem does IV solve?



Example: what is the effect of economic conditions on civil conflicts?

- It is generally well known that civil conflict and economic conditions are correlated
- Finding a causal link between the two is more difficult

Example: what is the effect of economic conditions on civil conflicts?

- It is generally well known that civil conflict and economic conditions are correlated
- Finding a causal link between the two is more difficult
  - → Endogeneity: conflict may be causing poor economic growth
  - → OVB: e.g. poor government institutions may cause both poor growth and civil wars
- We can use instrumental variables to address the endogeneity and OVB concerns

An instrument for economic conditions

We are running the regression

$$\underbrace{Y_i}_{\text{civil conflict}} = \beta_0 + \beta_1 \underbrace{X_i}_{\text{economic growth}} + \varepsilon_i$$

and we are looking for an instrument Z_i.

A good instrument for economic conditions should satisfy two conditions:

An instrument for economic conditions

We are running the regression

$$\underbrace{Y_i}_{\text{civil conflict}} = \beta_0 + \beta_1 \underbrace{X_i}_{\text{economic growth}} + \varepsilon_i$$

and we are looking for an instrument Z_i.

A good instrument for economic conditions should satisfy two conditions:

- Relevance: $\text{Cov}(X_i, Z_i) \neq 0$
- Exogeneity: $\text{Cov}(\varepsilon_i, Z_i) = 0$

Any ideas?

IV estimators

We saw two ways to construct IV estimators, given dependent variable Y, endogenous independent variable X and instrument Z.

IV estimators

We saw two ways to construct IV estimators, given dependent variable Y, endogenous independent variable X and instrument Z.

Reduced form divided by first stage:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 Z_i + \varepsilon_i \quad \text{(reduced form)} \\
X_i &= \gamma_0 + \gamma_1 Z_i + \mu_i \quad \text{(first stage)} \\
\hat{\beta}_{IV} &= \frac{\beta_1}{\gamma_1}
\end{aligned}
$$

IV estimators

We saw two ways to construct IV estimators, given dependent variable Y, endogenous independent variable X and instrument Z.

Two stage least squares:

$$X_i = \gamma_0 + \gamma_1 Z_i + \mu_i \quad \text{(first stage)}$$

$$Y_i = \beta_0 + \beta_{TSLS} \hat{X}_i + \varepsilon_i \quad \text{(second stage)}$$

Relevance

The relevance condition for an instrumental variable is Cov $(X_i, Z_i) \neq 0$

It is an empirically verifiable condition!

Weak instruments can cause problems: if Cov $(X_i, Z_i) \approx 0$ our standard asymptotic theories fail.

Exogeneity

The exogeneity condition for an instrumental variable is Cov $(\varepsilon_i, Z_i) = 0$.

It is generally **not** an empirically verifiable condition when the number of endogenous regressors is equal to the number of instruments!

It is up to the econometrician to determine whether the exogeneity condition is plausibly satisfied.

Rainfall as an instrument

One candidate for an instrument is rainfall.

In sub-Saharan Africa, the heavily agricultural economy requires rainfall for output (relevance).

Rainfall is uncorrelated(?) with other things that determine civil conflict (exogeneity).

Miguel, Satyanath and Sergenti (2004) do just this to answer the causal question.

First stage

TABLE 2
RAINFALL AND ECONOMIC GROWTH (First-Stage)
Dependent Variable: Economic Growth Rate, t

EXPLANATORY VARIABLE	ORDINARY LEAST SQUARES				
	(1)	(2)	(3)	(4)	(5)
Growth in rainfall, t	.055***	.053***	.049***	.049***	.053***
	(.016)	(.017)	(.017)	(.018)	(.018)
Growth in rainfall, $t-1$	.034**	.032**	.028**	.028*	.037**
	(.013)	(.014)	(.014)	(.014)	(.015)
Growth in rainfall, $t+1$				.001	
				(.019)	
Growth in terms of trade, t					−.002
					(.023)
Log(GDP per capita), 1979		−.011			
		(.007)			
Democracy (Polity IV), $t-1$		.0000			
		(.0007)			
Ethnolinguistic fractionalization		.006			
		(.044)			
Religious fractionalization		.045			
		(.044)			
Oil-exporting country		.007			
		(.019)			
Log(mountainous)		.001			
		(.005)			
Log(national population), $t-1$		−.009			
		(.009)			
Country fixed effects	no	no	yes	yes	yes
Country-specific time trends	no	yes	yes	yes	yes
R^2	.02	.08	.13	.13	.16
Root mean square error	.07	.07	.07	.07	.06
Observations	743	743	743	743	661

NOTE.—Huber robust standard errors are in parentheses. Regression disturbance terms are clustered at the country level. A country-specific year time trend is included in all specifications (coefficient estimates not reported).
* Significantly different from zero at 90 percent confidence.
** Significantly different from zero at 95 percent confidence.
*** Significantly different from zero at 99 percent confidence.

Second stage

TABLE 5
INTERACTIONS BETWEEN ECONOMIC GROWTH AND COUNTRY CHARACTERISTICS
Dependent Variable: Civil Conflict ≥25 Deaths

	IV-2SLS				
EXPLANATORY VARIABLE	(1)	(2)	(3)	(4)	(5)
Economic growth rate, t	−1.20	.92	−9.9	−.99	−1.85
	(1.43)	(2.62)	(22.9)	(1.26)	(1.81)
Economic growth rate, $t-1$	−2.86*	−3.01*	−6.4	−2.37**	−2.97**
	(1.46)	(1.70)	(6.1)	(1.04)	(1.39)
Economic growth rate, t × democracy (Polity IV), $t-1$	.01				
	(.21)				
Economic growth rate, $t-1$ × democracy (Polity IV), $t-1$	−.10				
	(.16)				
Economic growth rate, t × log(per capita income, 1979)		−1.98			
		(2.70)			
Economic growth rate, $t-1$ × log(per capita income, 1979)		.58			
		(1.09)			
Economic growth rate, t × ethnolinguistic fractionalization			12.1		
			(30.1)		
Economic growth rate, $t-1$ × ethnolinguistic fractionalization			5.1		
			(8.1)		
Economic growth rate, t × oil-exporting country				−2.8	
				(6.9)	
Economic growth rate, $t-1$ × oil-exporting country				3.2	
				(3.1)	
Economic growth rate, t × log(mountainous)					.39
					(.83)
Economic growth rate, $t-1$ × log(mountainous)					.23
					(.62)
Country fixed effects	yes	yes	yes	yes	yes
Country-specific time trends	yes	yes	yes	yes	yes
Root mean square error	.33	.34	.41	.32	.32
Observations	743	743	743	743	743

NOTE.—Huber robust standard errors are in parentheses. Regression disturbance terms are clustered at the country level. The instrumental variables are growth in rainfall, t and growth in rainfall, $t-1$ and these two terms interacted with the appropriate explanatory variable. A country-specific year time trend is included in all specifications (coefficient estimates not reported). Similar interaction patterns hold when civil conflict ≥1,000 deaths is the dependent variable and in most OLS specifications (results not shown).

* Significantly different from zero at 90 percent confidence.

** Significantly different from zero at 95 percent confidence.

How good is this instrument?

- Sarsons (2014) challenges the exclusion restriction in rainfall instruments
- Shows that in India, rain shocks predict riot incidence even in locations that are downstream from dams
  - → These locations should not be sensitive to the amount of rainfall!
- Other channels through which rainfall might reduce conflict include migration and infrastructure/road damage

Multiple instruments and endogenous regressors

We can extend the analysis of instrumental variables to a setting with multiple endogenous regressors and instruments.

If we have m instruments and k endogenous regressors, we require $m \geq k$.

Estimate through two stage least squares:

- Regress each endogenous variable on the full set of instruments
- Obtain the predicted values from each regression
- Regress our outcome variable on the set of predicted values

F-stat test on first stage

In the single instrument + single endogenous regressor case, we tested for relevance by just looking at the regression coefficient in the first stage.

F-stat test on first stage

In the single instrument + single endogenous regressor case, we tested for relevance by just looking at the regression coefficient in the first stage.

With multiple instruments we can do the natural thing and look at the F-stat for the joint significance of our instruments.

We use a different critical value for this test: comparing the F-stat with 10.

- We may actually require a much higher critical value (see Lee, McCrary, Moreira and Porter AER 2022).

J-test for exogeneity

In the single instrument + single endogenous regressor case, we could not directly test for exogeneity, instead relying on economic insight.

If our model is over-identified ($m > k$) we can conduct a J-test for exogeneity. (see lecture notes for full details)

Basic idea is that if all instruments are valid, IV estimates should be the same regardless of which instrument we use. We can test by regressing the appropriate residuals on our instruments.

- Important: the null in this test is exogeneity

Practice questions

Suppose we have an outcome variable Y, an endogenous regressor X, a valid instrument (for X) Z and controls $W_1, W_2, \ldots, W_j$.

Question 1: If IV can solve omitted variable bias, why do we need to include controls in our TSLS/IV estimates?

Practice questions

Suppose we have an outcome variable Y, an endogenous regressor X, a valid instrument (for X) Z and controls $W_1, W_2, \ldots, W_j$.

Question 1: If IV can solve omitted variable bias, why do we need to include controls in our TSLS/IV estimates?

Answer: Remember that the exogeneity condition is Cov $(Z, e) = 0$. The exogeneity of your instrument might be *conditional on controls*!

- Example on board

Practice questions

Suppose we have an outcome variable Y, an endogenous regressor X, a valid instrument (for X) Z and controls $W_1, W_2, \ldots, W_j$.

Question 2: Suppose our instrument Z is *unconditionally* exogenous. What will the differences in our estimates and our standard errors be if we include vs. exclude controls?

Practice questions

Suppose we have an outcome variable Y, an endogenous regressor X, a valid instrument (for X) Z and controls $W_1, W_2, \ldots, W_j$.

Question 2: Suppose our instrument Z is *unconditionally* exogenous. What will the differences in our estimates and our standard errors be if we include vs. exclude controls?

Answer: our estimates of the causal parameter should be the same, but the regression with controls will have smaller standard errors (ie improved precision).

Practice questions

Suppose we have an outcome variable Y, an endogenous regressor X, a valid instrument (for X) Z and controls $W_1, W_2, \ldots, W_j$.

Question 3: If Z is a valid instrument for X, is X a valid instrument for Z?

Practice questions

Suppose we have an outcome variable Y, an endogenous regressor X, a valid instrument (for X) Z and controls $W_1, W_2, \ldots, W_j$.
Question 3: If Z is a valid instrument for X, is X a valid instrument for Z?

Answer: No! Remember that require that an instrument affects our outcome variable only through its effect on our endogenous regressor. X was endogenous to begin with in this set-up, so it cannot be an instrument for Z since it has a direct relationship with Y.