Instrumental variable method

Outline

- Assumptions and limitation of instrumental variable method.
- Applications in economics

  -- ways to find a good instrument variable
- Extension of event study

What do we want to estimate?

- The effects of government programs and policies, such as those that subsidize training for disadvantaged workers.
- The effects of individual choices like college attendance.
- The effects of disclosure quality (on cost of capital)

Why do we need instrumental variables ?

Instrumental Variables (IV) estimation is used when your model has **endogenous** x's. That is, we run the following regression Y = α+ $\beta_1 X_1$ + ε but Cov(X_1,ε) $\neq$ 0.

In this case, $\widehat{\beta_1}$ is biased and not asymptotically consistent.

- Omitted variable bias from a variable that is correlated with X but is unobserved, so cannot be included in the regression
- Simultaneous causality bias (endogenous explanatory variables; X causes Y, Y causes X)
- Errors-in-variables bias (X is measured with error)

Instrumental variables regression can eliminate bias from these three sources.

Terminology: Endogeneity and Exogeneity

- An **endogenous** variable is one that is correlated with ε
- An **exogenous** variable is one that is uncorrelated with ε

"Endogenous" literally means "determined within the system," that is, a variable that is jointly determined with , that is, a variable subject to simultaneous causality.

However, this definition is narrow and IV regression can be used to address omitted variable bias and errors-in-variable bias, not just simultaneous causality bias.

Conditions for valid instruments

- Assume that Y is generated according to the following data generating process: $Y = \alpha + \beta_1 X + \varepsilon$ with $Cov(X, \varepsilon) \neq 0$.
- Conditions for valid instrument variable(s) Z:

1. Independence : $Cov(Z,\varepsilon)=0$
2. Exclusion restriction : Z will affect Y only through X
3. relevance : Cov (Z, X)>0 , (The first stage F-statistics from Z >10)
4. Monotonicity

Random Assignment as an instrumental Variable

- Assumptions:

1. Z_i has to be independent with potential outcomes, Y_i and D_i

2. Exclusion: Z_i affects Y only through D

3. E[D | Z_i =1] - E[D | Z_i = 0] > 0

4. No Defiers

- $\boldsymbol{Wald\ estimate} = \beta^{IV}$

$$= \frac{E[Y_i \mid z_i = 1] - E[Y_i \mid z_i = 0]}{E[D_i \mid z_i = 1] - E[D_i \mid z_i = 0]} = \mathbf{E}[Y_i(1) - Y_i(0) \mid \text{complier}]$$

The Validity of Instruments

We can test if Cov(*Z*, *X*) ≠ 0

- Just testing H_0: $\pi_1 = 0$ in the regression: $x = \pi_0 + \pi_1 Z + v$
- Sometimes refer to this regression as the first-stage regression

We can't test if Cov(*Z*, ε) = 0

We have to use common sense and economic theory to decide if it makes sense to assume Cov(*Z*, ε) = 0

- Can't directly test this condition because don't have unbiased estimator for
- OLS estimator of ε is presumed biased and the IV estimator of depends on the validity of Cov(*Z*, ε) = 0 condition

IV Estimation in the Simple Regression Case

Explanation #1:

For $y = \beta_0 + \beta_1 x + \varepsilon$, and given our assumptions

Cov(Z,y) = β_1Cov(Z,x) + Cov(Z, ε), so

β_1 = Cov(z,y) / Cov(z,x)

- Then the IV estimator for β_1 is $\frac{\sum_i (z_i - \bar{z})(y_i - \bar{y})}{\sum_i (z_i - \bar{z})(x_i - \bar{x})}$

- Where Instrument is binary:

$$P \lim \hat{\beta}^{IV} = \frac{E(Y \mid z=1) - E(Y \mid z=0)}{E(x \mid z=1) - E(x \mid z=0)}$$

This is called the **Wald estimator**

IV Estimation in the Simple Regression Case

Explanation #2: Two Stage Least Squares (2SLS)

(1) First isolates the part of X that is uncorrelated with : regress X on Z using OLS

$$X_i = \pi_0 + \pi_1 Z_i + v_i \quad (1)$$

Because Z_i is uncorrelated with ε , $\pi_0 + \pi_1 Z_i$ is uncorrelated with ε_i. We don't know π_0 or π_1 but we have estimated them, so…

Compute the predicted values of X_i ,$\widehat{X_i}$, where ,$\widehat{X}_i = \widehat{\pi}_0 + \widehat{\pi}_1 Z_i$, i=1…n

(2) Replace X_i by $\widehat{X}_i$ in the regression of interest:

regress Y on $\widehat{X}_i$ using OLS:

$$Y_i = \beta_0 + \beta_1 \widehat{X}_i + \varepsilon_i \quad (2)$$

Because $\widehat{X}_i$ is uncorrelated with ε_I in large samples, so the assumption A1 holds.

Thus β_1 can be estimated by OLS using regression (2)

This argument relies on large samples (so that π_0 and π_1 are well estimated using regression (1))

The resulting estimator $\widehat{\beta}_1$ is called the "Two Stage Least Squares" (TSLS or 2SLS) estimator

The General IV Regression Model

- The general IV regression model: notation and jargon

$$Y_i = \beta_0 + \beta_1\widehat{X_{1i}} + \ldots \beta_k\widehat{X_{ki}} + \beta_{k+1}\widehat{W_{1i}} + \ldots\beta_{k+r}\widehat{W_{ri}} + \varepsilon_i$$

Y_i is the dependent variable

$X_{1i}, \ldots, X_{ki}$ are the endogenous regressors (potentially correlated with ε_i)

$W_{1i}, \ldots, W_{ki}$ are the included exogenous variables or included exogenous regressors (uncorrelated with)

$\beta_0, \ldots, \beta_{k+r}$ are the unknown regression coefficients

$Z_{1i}, \ldots, Z_{mi}$ are the instrumental variables (the excluded exogenous variables)

Identification and over-identification

- We need to introduce some new concepts and to extend some old concepts to the general IV regression model:
- Terminology: **identification** and **over-identification**
- TSLS with included exogenous variables o One endogenous regressor o Multiple endogenous regressors
- Assumptions that underlie the normal sampling distribution of TSLS o Instrument validity (relevance and exogeneity) o General IV regression assumptions

Identification

- In general, a parameter is said to be **identified** if different values of the parameter would produce different distributions of the data.
- In IV regression, whether the coefficients are identified depends on the relation between the number of instruments (m) and the number of endogenous regressors (k)
- Intuitively, if there are fewer instruments than endogenous regressors, we can't estimate $\beta_0, \dots, \beta_{k+r}$
- For example, suppose k = 1 but m = 0 (we have no instruments). The coefficients $\beta_1, \dots, \beta_k$ are said to be:
  - **Exactly Identified** if m=k. There are just enough instruments to estimate $\beta_1, \dots, \beta_k$
  - **Overidentified** if m>k. There are more than enough instruments to estimate $\beta_1, \dots, \beta_k$ If so, you can test whether the instruments are valid (a test of the "overidentifying restrictions")
  - **Underidentified** if m<k. There are too few enough instruments to estimate $\beta_1, \dots, \beta_k$ If so, you need to get more instruments!

General IV regression: TSLS with one Endogenous Regressor

- The regression model takes the form,

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 W_{1i} + \ldots + \beta_{1+r} W_{ri} \ \varepsilon_i$$

  Instruments: $Z_{1i}, \ldots, Z_{mi}$

- First stage :

  Regress X_1 on all the exogenous regressors: regress X_1 on $W_1, \ldots, W_r$ and $Z_{1i}, \ldots, Z_{mi}$ using OLS

  Compute predicted values $\widehat{X_1}$, I =1…n.

- Second stage :

  Regress Y_i on $\widehat{X_1}$, $W_1, \ldots, W_r$ using OLS

  The coefficients from this second stage regression are the TSLS estimators, but the standard errors are again wrong

General IV regression: TSLS with Multiple Endogenous Regressors

- The regression model takes the form,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots \beta_k X_{ki} + \beta_{k+1} W_{1i} + \ldots \beta_{k+r} W_{ri} + \varepsilon_i$$

  Instruments: $Z_{1i}, \ldots, Z_{mi}$

- Now there are k first stage regressions:

Regress X_1 on all the exogenous regressors: regress X_1 on $W_1, \ldots, W_r$ and $Z_{1i}, \ldots, Z_{mi}$ using OLS. Compute predicted values $\widehat{X_{1i}}$, i= 1…n

Regress X_2 on all the exogenous regressors: regress X_1 on $W_1, \ldots, W_r$ and $Z_{1i}, \ldots, Z_{mi}$ using OLS. Compute predicted values $\widehat{X_{2i}}$, i= 1…n

Repeat for all X's, obtaining $\widehat{X_{1i}} \ldots \widehat{X_{ki}}$

- Second stage:

Regress Y_i on $\widehat{X_{1i}} \ldots \widehat{X_{ki}}, W_1, \ldots, W_r$ using OLS

The coefficients from this second stage regression are the 2SLS estimators (but the standard errors are wrong)

Example 1: Using season of birth

DOES COMPULSORY SCHOOL ATTENDANCE AFFECT SCHOOLING AND EARNINGS ?*

BY Joshua D. Angrist and Alan B. Krueger (1991), *J.L.E*

Ideas

Students who are born early in the calendar year are typically older when they enter school than children born late in the year.

(6.45 v.s. 6.07)

Older students can drop out after completing less schooling than the younger students.

Season of birth generates exogenous variation in education that can be used to estimate the impact of compulsory schooling on education and earnings.

First Look of the Data

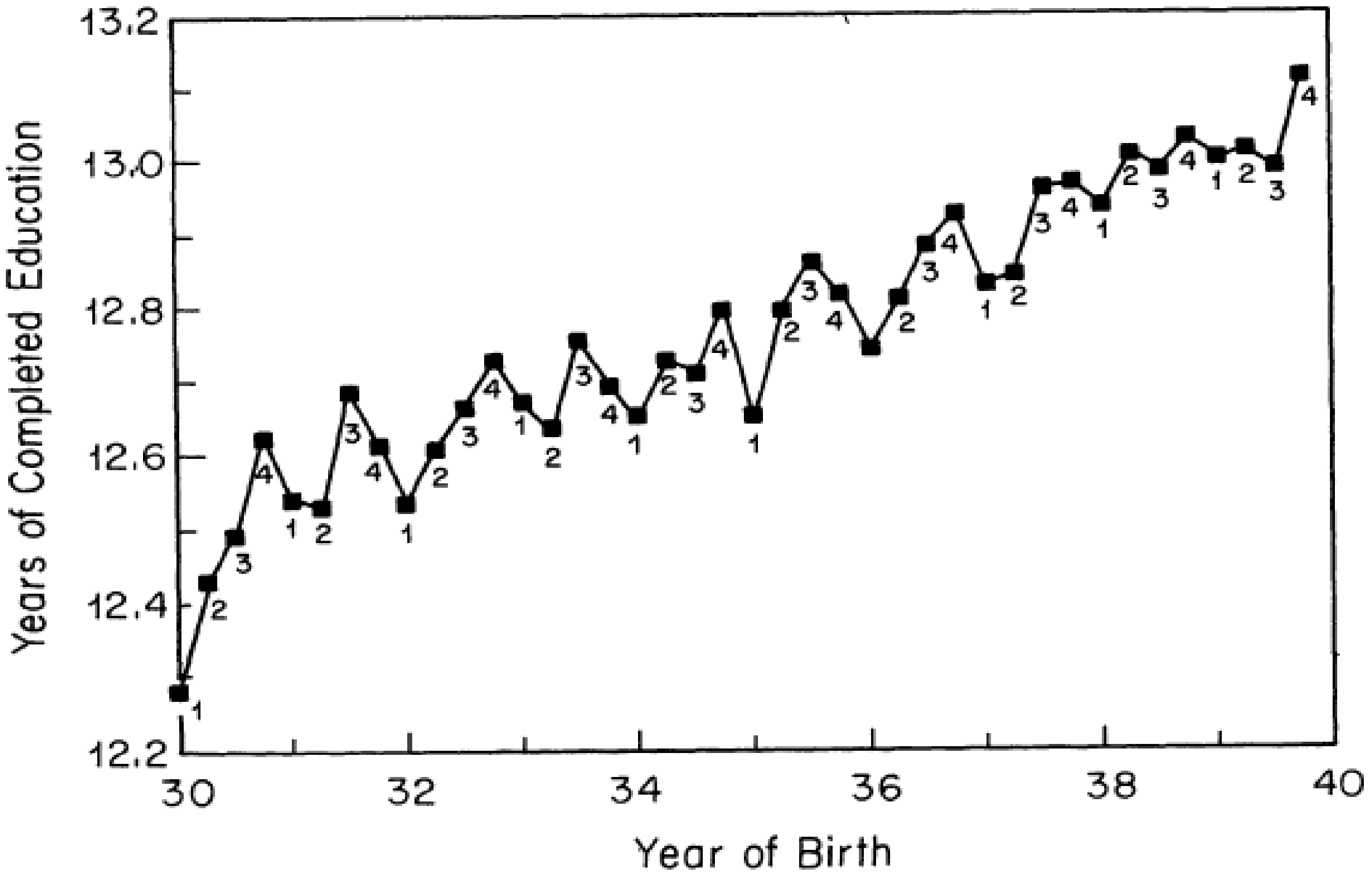

FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Using Census data conducted in 1960 April

- Students who were born between January and March of 1944 were age sixteen when the 1960 Census was conducted (Census Day is April 1)
- Those who were born between April and December of 1944 were not yet age sixteen.
- Students born in January-March were able to drop out of school in the states that had an age sixteen compulsory attendance law, but were not able to legally drop out of school in states that had an age seventeen or age eighteen compulsory attendance law.
- Students born in April-December of 1944 were not able to legally withdraw from school under either regime.

TABLE II

PERCENTAGE OF AGE GROUP ENROLLED IN SCHOOL BY BIRTHDAY AND LEGAL DROPOUT AGE[a]

Date of birth	School-leaving age: 16 (1)	School-leaving age: 17 or 18 (2)	Column (1) − (2)
	Type of state law[b]		
	Percent enrolled April 1, 1960		
1. Jan 1–Mar 31, 1944	87.6	91.0	−3.4
(age 16)	(0.6)	(0.9)	(1.1)
2. Apr 1–Dec 31, 1944	92.1	91.6	0.5
(age 15)	(0.3)	(0.5)	(0.6)
3. Within-state diff.	−4.5	−0.6	−4.0
(row 1 − row 2)	(0.7)	(1.0)	(1.2)
	Percent enrolled April 1, 1970		
4. Jan 1–Mar 31, 1954	94.2	95.8	−1.6
(age 16)	(0.3)	(0.5)	(0.6)
5. Apr 1–Dec 31, 1954	96.1	95.7	0.4
(age 15)	(0.1)	(0.3)	(0.3)
6. Within-state diff.	−1.9	0.1	−2.0
(row 1 − row 2)	(0.3)	(0.6)	(0.6)
	Percent enrolled April 1, 1980		
7. Jan 1–Mar 31, 1964	95.0	96.2	−1.2
(age 16)	(0.1)	(0.2)	(0.2)
8. Apr 1–Dec 31, 1964	97.0	97.7	−0.7
(age 15)	(0.1)	(0.1)	(0.1)
9. Within-state diff.	−2.0	−1.5	0.5
(row 1 − row 2)	(0.1)	(0.2)	(0.3)

Estimation of return of schooling

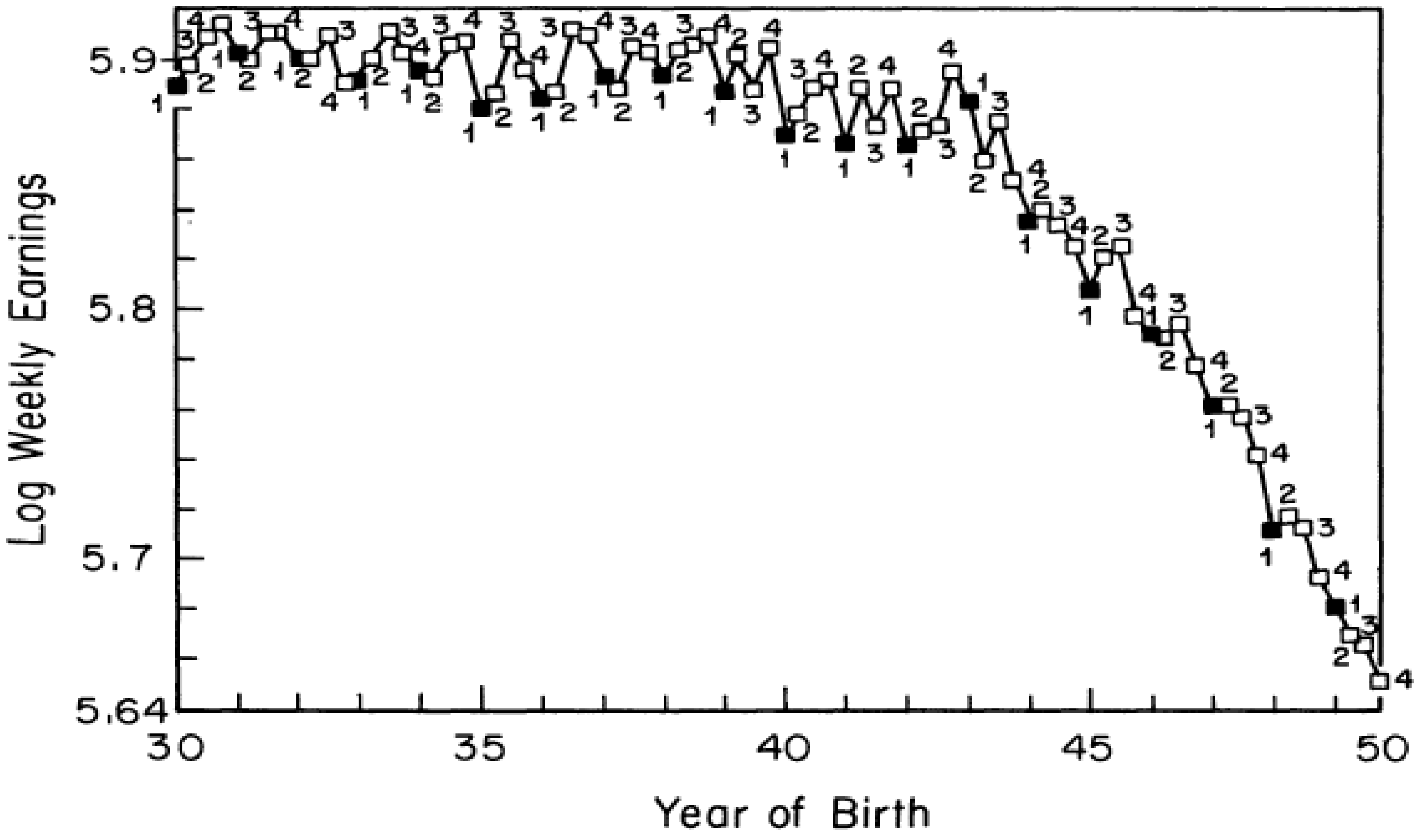

FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

TSLS

- First stage and second stage

(1) $E_i = X_i\pi + \sum_C Y_{iC}\delta_i + \sum_C \sum_j Y_{iC}Q_{ij}\,\theta_{iC} + \varepsilon_i$

(2) $\ln W_i = X_i\beta + \sum_C Y_{iC}\vartheta_i + \rho E_i + \mu_i$

E is the education of the ith individual

X is a vector of covariates

Q is a dummy variable indicating whether the individual was born in quarter j (j = 1,2,3)

Y is a dummy variable indicating whether the individual was born in year c (c =1,. .., 10)

W is the weekly wage.

The coefficient **ρ** is the return to education.

TABLE V

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0711	0.0891	0.0711	0.0760	0.0632	0.0806	0.0632	0.0600
	(0.0003)	(0.0161)	(0.0003)	(0.0290)	(0.0003)	(0.0164)	(0.0003)	(0.0299)
Race (1 = black)	—	—	—	—	−0.2575	−0.2302	−0.2575	−0.2626
					(0.0040)	(0.0261)	(0.0040)	(0.0458)
SMSA (1 = center city)	—	—	—	—	0.1763	0.1581	0.1763	0.1797
					(0.0029)	(0.0174)	(0.0029)	(0.0305)
Married (1 = married)	—	—	—	—	0.2479	0.2440	0.2479	0.2486
					(0.0032)	(0.0049)	(0.0032)	(0.0073)
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	—	—	−0.0772	−0.0801	—	—	−0.0760	−0.0741
			(0.0621)	(0.0645)			(0.0604)	(0.0626)
Age-squared	—	—	0.0008	0.0008	—	—	0.0008	0.0007
			(0.0007)	(0.0007)			(0.0007)	(0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]	—	22.5 [29]	—	19.6 [27]

Interpretation

- Can we call the estimated effect “average treatment effect”?

- The estimated effects based on the IV are Local Average Treatment Effects (LATE)

-- In this study, the results are based on those who are about to dropout.

Table 4.4.2: Probabilities of compliance in instrumental variables studies

Source (1)	Endogenous Variable (D) (2)	Instrument (Z) (3)	Sample (4)	$P[D=1]$ (5)	1st Stage, $P[D_1 > D_0]$ (6)	$P[Z=1]$ (7)	$P[D_1 > D_0 \mid D=1]$ (8)	$P[D_1 > D_0 \mid D=0]$ (9)
Angrist (1990)	Veteran Status	Draft eligibility	White men born in 1950	0.267	0.159	0.534	0.318	0.101
			Non-white men born in 1950	0.163	0.060	0.534	0.197	0.033
Angrist and Evans (1998)	More than 2 children	Twins at second birth	Married women aged 21-35 with two or more children in 1980	0.381	0.603	0.008	0.013	0.966
		First two children are of the same sex	Married women aged 21-35 with two or more children in 1980	0.381	0.060	0.506	0.080	0.048
Angrist and Krueger (1991)	High school graduate	Third or fourth quarter birth	Men born between 1930 and 1939	0.770	0.016	0.509	0.011	0.034
Acemoglu and Angrist (2000)	High school graduate	State requires 11 or more years of school attendance	White men aged 40-49	0.617	0.037	0.300	0.018	0.068

Notes: The table shows an analysis of the absolute and relative size of the complier population for a number of instrumental variables. The first-stage, reported in column 6, gives the absolute size of the complier group. Columns 8 and 9 show the size of the complier population relative to the treated and untreated populations.

Applications:
How to find good instrument variables?

- Institution/culture:

  -- Angrist return to school

  -- culture : ghost month paper
- Weather: rainfalls

  Economic Shocks and Civil Conflicts

  Network size of migrants and labor market outcomes
- Event shocks
- Policy influences (DID+IV):

  1. Duflo school constructions

  2. Faculty quality and student performance

IV applications: Economic Shocks and Civil Conflict (2004; JPE)

Economic Shocks and Civil Conflict: An Instrumental Variables Approach

Edward Miguel

University of California, Berkeley and National Bureau of Economic Research

Shanker Satyanath and Ernest Sergenti

New York University

Main story

- Economic condition shall be negatively correlated with civil conflict
- A lot of political science research suggest so, but ignore the endogeneity problem
- Omitted variable and simultaneity
- And here comes rainfall, and all is bright!

41 African countries 81-99: 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.

Reduce form

TABLE 3
Rainfall and Civil Conflict (Reduced-Form)

Explanatory Variable	Dependent Variable: Civil Conflict ≥25 Deaths (OLS) (1)	Dependent Variable: Civil Conflict ≥1,000 Deaths (OLS) (2)
Growth in rainfall, t	−.024	−.062**
	(.043)	(.030)
Growth in rainfall, $t-1$	−.122**	−.069**
	(.052)	(.032)
Country fixed effects	yes	yes
Country-specific time trends	yes	yes
R^2	.71	.70
Root mean square error	.25	.22
Observations	743	743

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.

OLS & 2SLS

TABLE 4
ECONOMIC GROWTH AND CIVIL CONFLICT

EXPLANATORY VARIABLE	Dependent Variable: Civil Conflict ≥25 Deaths						Dependent Variable: Civil Conflict ≥1,000 Deaths
	Probit (1)	OLS (2)	OLS (3)	OLS (4)	IV-2SLS (5)	IV-2SLS (6)	IV-2SLS (7)
Economic growth rate, t	−.37	−.33	−.21	−.21	−.41	−1.13	−1.48*
	(.26)	(.26)	(.20)	(.16)	(1.48)	(1.40)	(.82)
Economic growth rate, $t-1$	−.14	−.08	.01	.07	−2.25**	−2.55**	−.77
	(.23)	(.24)	(.20)	(.16)	(1.07)	(1.10)	(.70)
Log(GDP per capita), 1979	−.067	−.041	.085		.053		
	(.061)	(.050)	(.084)		(.098)		
Democracy (Polity IV), $t-1$	.001	.001	.003		.004		
	(.005)	(.005)	(.006)		(.006)		
Ethnolinguistic fractionalization	.24	.23	.51		.51		
	(.26)	(.27)	(.40)		(.39)		
Religious fractionalization	−.29	−.24	.10		.22		
	(.26)	(.24)	(.42)		(.44)		
Oil-exporting country	.02	.05	−.16		−.10		
	(.21)	(.21)	(.20)		(.22)		
Log(mountainous)	.077**	.076*	.057		.060		
	(.041)	(.039)	(.060)		(.058)		
Log(national population), $t-1$	.080	.068	.182*		.159*		
	(.051)	(.051)	(.086)		(.093)		
Country fixed effects	no	no	no	yes	no	yes	yes
Country-specific time trends	no	no	yes	yes	yes	yes	yes
R^2	...	.13	.53	.71	...	...	...
Root mean square error	...	.42	.31	.25	.36	.32	.24
Observations	743	743	743	743	743	743	743

NOTE.—Huber robust standard errors are in parentheses. Regression disturbance terms are clustered at the country level. Regression 1 presents marginal probit effects, evaluated at explanatory variable mean values. The instrumental variables for economic growth in regressions 5–7 are growth in rainfall, t and growth in rainfall, $t-1$. A country-specific year time trend is included in all specifications (coefficient estimates not reported), except for regressions 1 and 2, where a single linear time trend is included.

* Significantly different from zero at 90 percent confidence.
** Significantly different from zero at 95 percent confidence.
*** Significantly different from zero at 99 percent confidence.

exclusion restriction

- Exclusion restriction hold or not:

Is rainfall affecting conflict only through economic condition? Does it work through other channels? Or even itself?

- Any way to show that the exclusion restriction does not hold?

Dams help increase agriculture Production(Duflo)

FIGURE 1: District Maps of India

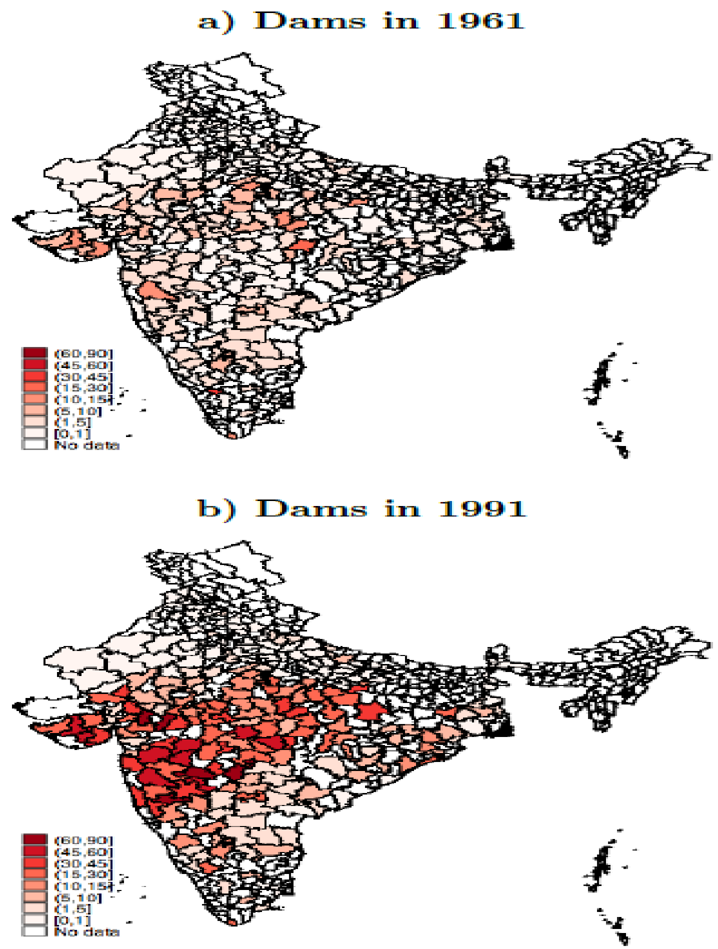

Table 2
IV, First Stage, and Reduced Form Results

	IV	First Stage (OLS)			Reduced Form (OLS)		
Dep. Variable:	Conflict	Agricultural Wage			Conflict		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rain Shock		0.345*** (0.093)	0.057 (0.120)	0.501*** (0.180)	-0.076*** (0.025)	-0.070** (0.034)	-0.078** (0.037)
Agricultural Wage	-0.243** (0.111)						
District & Year Fixed Effects	x	x	x	x	x	x	x
Sample	full	full	dam-fed	rain-fed	full	dam-fed	rain-fed
Number of Districts	142	142	92	56	142	92	56
Number of Obs.	2908	2908	1854	1054	2908	1854	1054

NETWORKS IN THE MODERN ECONOMY: MEXICAN MIGRANTS IN THE U. S. LABOR MARKET

By Kaivan Munshi

Research Questions

- The value of social capital/non-market institution
- These institutions emerge in response to market failure, harnessing social ties to avoid information, enforcement, and coordination problems
- Identify the effects of job networks among Mexican migrants in the U. S. labor market.

Difficulties to over come

- How to measure network strength?
- How to identify the effects of social network on job outcomes?
- Can we use OLS specification? What's the problems?
- Can we use fixed effects model? Any problems?

Empirical Strategy

- Including individual fixed effects in the employment and occupation regressions

(While fixed effects control for the individual's unobserved ability, network size could also respond to unobserved shocks in the U. S. labor market.)

- Using rainfall in the origin-community as an instrument for the size of the network at the destination

Data

- Mexican Migration Project (MMP)

-- individuals belonging to multiple origin-communities in Mexico, over a long period of time.

-- excluding communities with no variation in employment over time

- 24 communities in seven states:

Jalisco, Guanajuato, San Luis Potosi (SLP), Michoacan, Zacatecas, Nayarit, and Colima.

- uses data on migration patterns and labor market outcomes, based on a sample of individuals belonging to multiple origin-communities in Mexico, over a long period of time.

Some Descriptive Statistics : Migration Pattern

TABLE III
INDIVIDUAL MIGRATION PATTERNS

Origin state:	Full sample	Jalisco	Guanajuato	SLP	Michoacan	Zacatecas	Nayarit	Colima
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: migration and employment								
% migrants	11.47	10.51	6.79	11.90	12.37	17.71	6.92	4.64
	(8.28)	(7.65)	(3.69)	(7.58)	(4.20)	(10.18)	(1.33)	(1.10)
% new migrants	5.17	5.49	3.09	4.31	5.87	7.12	4.18	2.82
	(3.78)	(3.97)	(2.00)	(3.25)	(2.77)	(4.61)	(1.46)	(0.86)
% established migrants	6.31	5.02	3.70	7.59	6.50	10.58	2.75	1.83
	(5.19)	(4.27)	(2.10)	(4.57)	(1.93)	(6.57)	(0.64)	(0.51)
% employment in the United States	95.66	96.38	96.35	92.95	97.40	95.66	92.39	95.83
% employment in Mexico	86.48	90.48	87.07	82.01	90.17	84.23	82.50	88.01
Panel B: individual migration patterns over the sample period								
Avg. number of trips	1.35	1.45	1.22	1.30	1.29	1.38	1.34	1.21
	(0.69)	(0.77)	(0.54)	(0.56)	(0.59)	(0.73)	(0.69)	(0.63)
Avg. duration at destination (years)	3.57	3.36	2.59	4.16	3.10	4.08	3.18	2.98
	(3.57)	(3.53)	(2.41)	(3.47)	(2.92)	(4.04)	(3.35)	(3.09)
% with 1 trip	74.50	69.23	84.21	74.26	75.25	72.92	76.00	85.71
% with 2 trips	17.85	19.66	9.87	22.77	21.78	18.29	16.00	10.71
% with 3 trips	5.83	8.12	5.92	1.98	0.99	6.94	6.00	3.57
% with 4 trips	1.55	2.99	0.00	0.99	1.98	1.16	2.00	0.00
% with 5 trips	0.27	0.00	0.00	0.00	0.00	0.69	0.00	0.00
Number of observations	1098	234	152	101	101	432	50	28

Estimation

- Reduced form (推力跟拉力)
- New migrants v.s. established migrants
- First Stage
- IV results

  -- Employment

  -- Occupation choices

TABLE IV

Reduced-Form Regressions: Fine Partition of Rainfall Lags

Dependent variable:	Employment at the destination	Employment at the origin
	(1)	(2)
rain (t)	−0.003	**0.027**
	(0.013)	(0.009)
rain (t − 1)	−0.007	**0.027**
	(0.015)	(0.009)
rain (t − 2)	−0.016	**0.035**
	(0.014)	(0.009)
rain (t − 3)	−0.027	**0.024**
	(0.016)	(0.009)
rain (t − 4)	**−0.033**	0.008
	(0.014)	(0.008)
rain (t − 5)	**−0.032**	0.008
	(0.013)	(0.008)
rain (t − 6)	**−0.032**	0.009
	(0.013)	(0.010)
Individual fixed effects	Yes	Yes
Year dummies	Yes	Yes
R^2	0.705	0.812
Box-Pearson Q statistic	0.042	2.813
Number of observations	4,546	41,120

TABLE V

REDUCED-FORM AND FIRST-STAGE REGRESSIONS: COARSE PARTITION OF RAINFALL LAGS

Dependent variable:	Reduced-form					First-stage	
	Employment at the destination				Employment at the origin	New migrants	Established migrants
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Recent-past rainfall	−0.028	−0.049	−0.023	−0.047	0.085	−0.091	0.005
	(0.027)	(0.035)	(0.072)	(0.040)	(0.018)	(0.037)	(0.020)
Distant-past rainfall	−0.125	−0.092	−0.226	−0.129	0.046	0.103	−0.106
	(0.035)	(0.027)	(0.108)	(0.044)	(0.021)	(0.033)	(0.023)
Individual fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.705	0.705	0.647	0.038	0.812	0.768	0.940
Q statistic	0.041	0.041	0.036	0.660	2.813	0.010	0.316
Number of observations	4546	4546	1732	4546	41,120	4546	4546

OLS and Instrumental Variable Regressions

	Employment at the destination										
	OLS	IV									
		Basic specifications				Robustness to individual characteristics			Robustness to sample lengths		
Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
New migrants	−0.032	0.397	0.522	0.093	0.626	0.623	0.394	0.424	0.511	0.377	0.251
	(0.070)	(0.315)	(0.376)	(0.537)	(0.501)	(0.353)	(0.306)	(0.326)	(0.321)	(0.400)	(0.356)
Established migrants	0.670	1.554	1.474	2.073	1.745	2.021	1.321	1.565	1.699	1.304	1.058
	(0.154)	(0.551)	(0.545)	(1.069)	(0.661)	(0.594)	(0.534)	(0.656)	(0.526)	(0.578)	(0.491)
Individual fixed effects	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.707	—	—	—	—	—	—	—	—	—	—
Q statistic	0.042	0.041	0.041	0.036	0.660	0.110	0.0005	0.015	0.049	0.022	0.001
Number of observations	4546	4546	4546	1732	4546	4710	3371	4067	5214	3894	3614

Using Surprising events as exogenous variation

- Using events as instrumental variable

  Example: The effects of news reporting on US government donation

- Extension of Event studies

  Example: Party policy platform and stock market returns

NEWS DROUGHTS, NEWS FLOODS, AND U. S. DISASTER RELIEF (THOMAS EISENSEE AND DAVID STROMBERG)

1. This paper studies the influence of mass media on U. S. government response to approximately 5,000 natural disasters occurring between 1968 and 2002.

2. This paper show that U. S. relief depends on whether the disaster occurs at the same time as other newsworthy events.

3. This paper argue that the only plausible explanation of this is that relief decisions are driven by news coverage of disasters and that the other newsworthy material crowds out this news coverage.

TABLE II

SUMMARY STATISTICS FOR DISASTERS

Disaster type	Number of disasters	Share of disasters	Killed per disaster	Affected per disaster	Share receiving OFDA relief
Flood	1,675	0.32	170	1,724,851	0.22
Storm	1,175	0.23	646	601,490	0.17
Epidemic	737	0.14	249	27,528	0.12
Earthquake	559	0.11	1,522	173,015	0.21
Drought	326	0.06	18,657	5,740,623	0.30
Landslide	310	0.06	84	38,789	0.06
Fire	129	0.02	19	69,552	0.13
Cold wave	114	0.02	103	46,656	0.01
Volcano	102	0.02	853	39,008	0.27
Infestation	47	0.01	na	1,100	0.68
Food shortage	38	0.01	4,293	734,630	0.13
Total	*5,212*	*1.00*	*590*	*1,166,505*	*0.19*

News stories on disasters, by days from the disaster

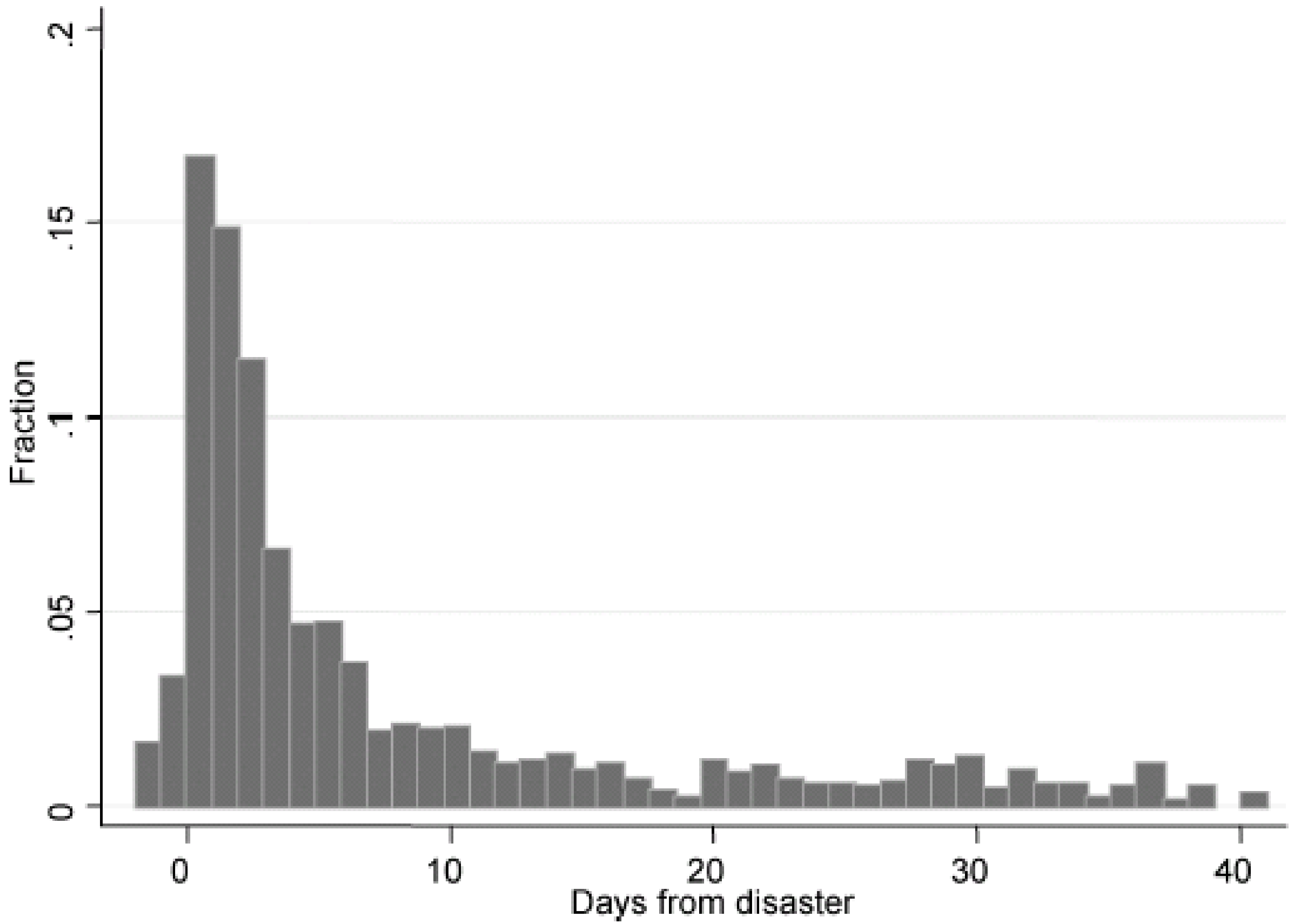

FIGURE I

Daily Number News Stories about Olympic Games, 1992

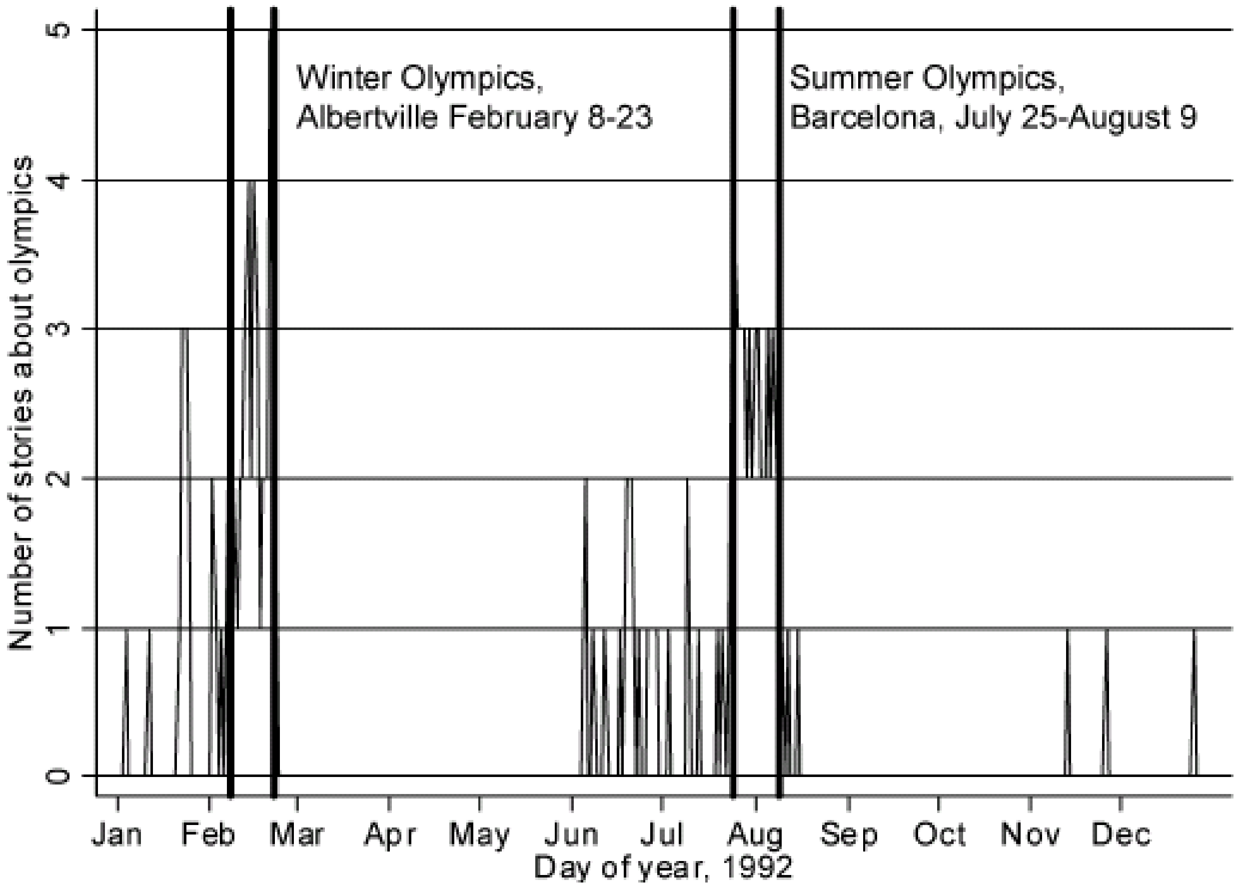

FIGURE II

Daily news pressure (minutes), by day

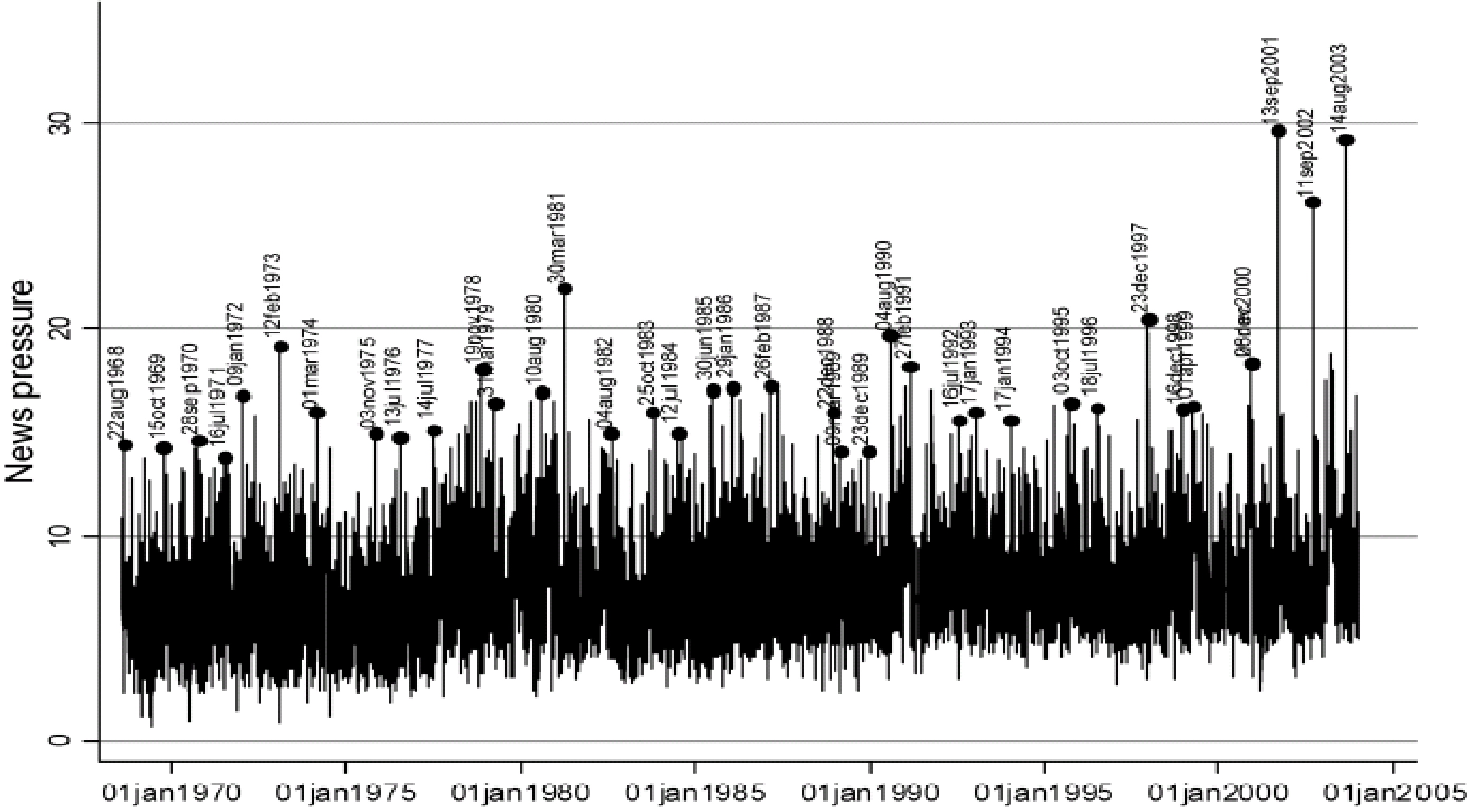

FIGURE III

TABLE III

Two Largest *Daily News Pressure* Dates and Main Story, by Year

Year	Date	Main news story
2003	Aug 14	New York City Blackout
	Mar 22	Invasion of Iraq: Day 3
2002	Sep 11	9/11 Commemoration
	Oct 24	Sniper Shooting in Washington: Arrest of Suspects
2001	Sep 13	9/11 Attack on America: Day 3
	Sep 12	9/11 Attack on America: Day 2
2000	Nov 26	Gore vs. Bush: Florida Recount—Certification by Katherine Harris
	Dec 8	Gore vs. Bush: Florida Recount—Supreme Court Ruling
1999	Apr 1	Kosovo Crisis: U. S. Soldiers Captured
	Jul 18	Crash of Plane Carrying John F. Kennedy, Junior
1998	Dec 16	U. S. Missile Attack on Iraq
	Dec 18	Clinton Impeachment
1997	Dec 23	Oklahoma City Bombing: Trial
	Aug 31	Princess Diana's Death
1996	Jul 18	TWA Flight 800 Explosion
	Jul 27	Olympic Games Bombing in Atlanta
1995	Oct 3	O.J. Simpson Trial: The Verdict
	Apr 22	Oklahoma City Bombing
1994	Jan 17	California Earthquake
	Jun 18	O.J. Simpson Arrested
1993	Jan 17	U. S. Missile Attack on Iraq
	Apr 20	Waco, Texas: Cult Standoff Ends in Fire
1992	Jul 16	Perot Quits 1992 Presidential Campaign
	May 1	Los Angeles Riots
1991	Feb 27	Gulf War: President Bush Declares Kuwait Liberated
	Jan 17	Gulf War: Operation Dessert Storm Launched

News Pressure (minutes) during 405 Days, March 15, 2001–Apr 23, 2002, by Day

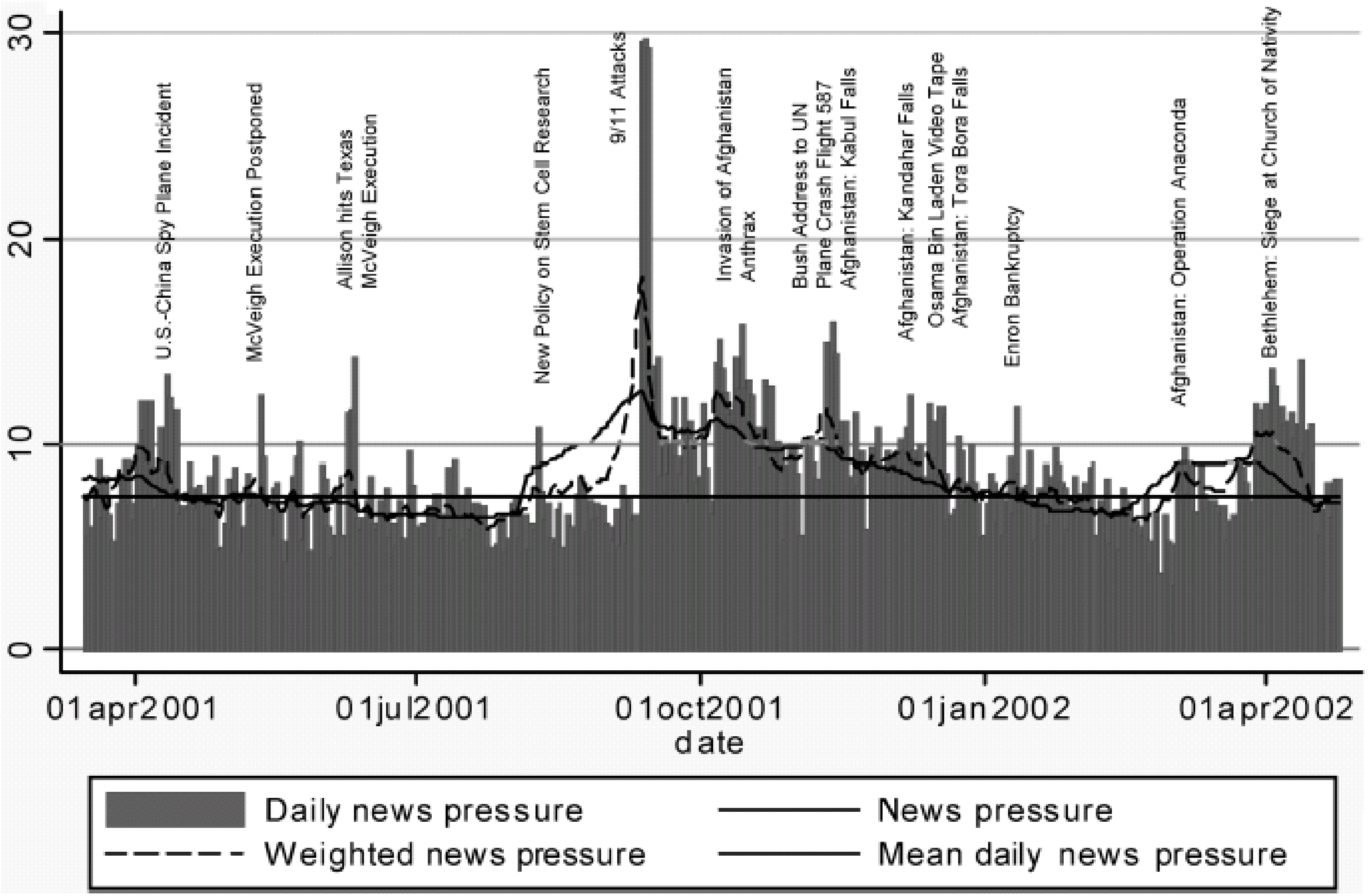

FIGURE IV

Empirical Design

$$(1) \quad relief_i^* = \alpha_1\, news_i + \alpha'\theta_i + \epsilon_i$$

$$(2) \quad relief_i = \begin{cases} 1 & \text{if } relief_i^* > 0 \\ 0 & \text{if } relief_i^* \leq 0 \end{cases}$$

$$(3) \quad news_i^* = \beta_1 news\ pressure_i + \beta_2 Olympics_i + \beta'\theta_i + \omega_i$$

$$\beta_1 < 0 \text{ and } \beta_2 < 0$$

$$(4) \quad news_i = \begin{cases} 1 & \text{if } news_i^* > 0 \\ 0 & \text{if } news_i^* \leq 0 \end{cases}$$

Key Variables

$relief_i^*$: *reliefworthiness* for disaster *i* *describes* the benefits of providing relief from the decision maker's perspective.

$news_i$: *indicates that the disaster was* covered when *news*$_i$ =1 [was not covered when *news*$_i$ =0] in the news.

θ_i : contains disaster specific variables, such as *killed and affected, and fixed* effects for disaster type, country, year, etc.

Key Variables (continued)

$relief_i$: *is the event that OFDA provided* [did not provide] disaster relief to disaster *i when relief*$_i$ *= 1* [*relief*$_i$ *= 0*] .

$news_i^*$: newsworthiness describes the benefits of covering disaster *i from the TV network's* perspective.

Linear probability OLS regressions.

TABLE IV

EFFECT OF THE PRESSURE FOR NEWS TIME ON DISASTER *NEWS* AND *RELIEF*

	Dependent variable: *News*				Dependent variable: *Relief*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
News Pressure	−0.0162	−0.0163	−0.0177	−0.0142	−0.0117	−0.0119	−0.0094	−0.0078
	(0.0041)***	(0.0041)***	(0.0057)***	(0.0037)***	(0.0045)***	(0.0045)***	(0.0058)	(0.0040)**
Olympics	−0.1078	−0.1079	−0.0871	−0.111	−0.1231	−0.1232	−0.1071	−0.1098
	(0.0470)**	(0.0470)**	(−0.0628)	(0.0413)***	(0.0521)**	(0.0521)**	(0.0763)	(0.0479)**
World Series	−0.1133				−0.1324			
	(−0.1065)				(0.1031)			
log Killed			0.0605				0.0582	
			(0.0040)***				(0.0044)***	
log Affected			0.0123			0.0376		
			(0.0024)***				(0.0024)***	
Imputed log Killed				0.0491				0.0442
				(0.0034)***				(0.0037)***
Imputed log Affected				0.0151				0.0394
				(0.0020)***				(0.0020)***
Observations	5,212	5,212	2,926	5,212	5,212	5,212	2,926	5,212
R-squared	0.1799	0.1797	0.3624	0.2875	0.1991	0.1989	0.4115	0.3726

OLS regressions with the instruments *news pressure and Olympics as dependent variables*

TABLE V

CORRELATIONS BETWEEN INSTRUMENTS AND THE SEVERITY OF DISASTERS

	Dependent variable	
	News pressure	*Olympics*
log Killed	−0.0082	0.0003
	(0.0113)	(0.0010)
log *Affected*	0.0005	−0.0006
	(0.0068)	(0.0006)
p-value: F-test of joint insignificance	0.75	0.62
Observations	5,212	5,212
R-squared	0.3110	0.2035

TABLE VI
DEPENDENT VARIABLE: *Relief*

	OLS					IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
News	0.2886	0.158	0.1309	0.2323	0.2611	0.8237	0.6341	0.6769
	(0.0200)***	(0.0232)***	(0.0178)***	(0.0328)***	(0.0569)***	(0.2528)***	(0.3341)*	(0.2554)***
News*abs(Pr(news)-0.5)				−0.4922	−0.302			
				(0.1059)***	(0.0840)***			
abs(Pr(news)-0.5)				0.5374	0.2959			
				(0.0943)***	(0.0831)***			
log Killed		0.0486					0.0198	
		(0.0046)***					(0.0208)	
log Affected		0.0358					0.0299	
		(0.0024)***					(0.0048)***	
Imputed log Killed			0.0378	0.0546	0.0307			0.0109
			(0.0038)***	(0.0049)***	(0.0046)***			(0.0132)
Imputed log Affected			0.0375	0.0445	0.0345			0.0292
			(0.0020)***	(0.0023)***	(0.0026)***			(0.0045)***
F-stat, instruments, 1st stage						11.0	6.1	11.1
Over-id restrictions, χ^2_{df}(p-value)						$0.51_1(0.47)$		$0.64_1(0.42)$
Observations	5,212	2,926	5,212	5,212	5,027	5,212	2,926	5,212
R-squared	0.2443	0.4225	0.3800	0.3860				