Instrumental Variables in Action

Remarks in honor of P.G. Wright's 150th birthday

Joshua D. Angrist

MIT and NBER

October 2011

What is Econometrics Anyway?

- What's the difference between statistics and econometrics?
  - The first is concerned primarily with statistical inference: how to go from sample to population
  - Pretty boring, I'm afraid (perhaps thats why it's mostly statisticians who study it!)
  - By contrast, econometric analysis typically begins with *identification*
- Identification concerns our ability to learn about relationships of interest when the population is fully enumerated
- Identification problems are most interesting when the relationships of interest are causal (say, the effect of price on quantity demanded)
- Identification problems are puzzles; no one algorithm or fundamental theorem solves them all, an insight is required
- The mother of all identification problems is the simultaneous equations model, nowhere expressed more succintly than in **Figure 4** and **Figure 3** of Appendix B (somewhat curiously, in reverse order)

The Miracle of Appendix B

Just how big *is* Appendix B?

- Wright's simple IV estimator is our Swiss Army knife
  - IV is the kernel around which all SEM identification and estimation strategies are built
- IV solves two other identification problems of major theoretical and practical significance
  1. Measurement error in regressors (Wald, 1940; interpreted as IV by Durbin, 1954)
  2. Omitted variables bias (here, I don't have a reference for first use; Goldberger seems a likely progenitor)
- ME and OVB are probably the most common apps today
- These blockbuster problems do not exhaust the power and the glory of IV!
- Bloom (1984), showed how to estimate causal effects in randomized trials with partial compliance
  - This too is IV (Angrist and Imbens, 1991)

Today's Causal Framework

Wright Got It Right: Potential Outcomes and Experiments

The economic concept of elasticity supposes different experiments with prices in the same market at a single instant of time (p. 291)

- PG defined causality in terms of outcomes revealed by a notional experiment. And he understood the identification challenge:

  Obviously such experiments cannot be made. Actual observations must be made at different times and during the period between observations conditions both of supply and demand may change

- Today, we use *potential outcomes* notation to describe what happens under alternative "treatment assignments"
- If the treatment is simply switched off or on, then potential outcomes, Y_{0i} and Y_{1i}, describe what happens under alternative assignments (e.g., policy or individual choices)

Our Constant-Effects Benchmark

- *Instrumental variables*, denoted by Z_i, provide leverage for causal inference when treatment is not randomly assigned
- As in Wright's constant-elasticity analysis of markets, the benchmark IV setup is a linear, constant-effects world

$$\begin{aligned} Y_{0i} &= \alpha + \eta_i \\ Y_{1i} - Y_{0i} &= \rho \\ Y_i &= Y_{0i} + D_i(Y_{1i} - Y_{0i}) = \alpha + \rho D_i + \eta_i \end{aligned}$$

- The difference in means with D_i switched off and on is likely to be a misleading measure of ρ:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \rho + \{E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0]\}$$

  The term in curly brackets is the ... problem-that-must-be-named: *selection bias*, *omitted variables bias*
- Those with health insurance are healthier than those without: causal effect or a difference in Y_{0i}?

Using IV to Eliminate OVB

- WWII vets live longer than non-vets born the same year. Causal effect or selection bias?
- An instrument, Z_i, independent of Y_{0i} and correlated with D_i, solves the OVB problem. From App. B (p. 314):

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} = \frac{Cov(Y_i, Z_i)/V(Z_i)}{Cov(D_i, Z_i)/V(Z_i)} = \frac{RF}{1st}$$

- RF ("reduced form") is the effect of the instrument on the outcome); 1st is the effect of the instrument on the treatment
- Example: draft-lottery estimates of the effects of Vietnam-era service
  - Y_i measures health or earnings
  - D_i indicates Vietnam-era service in a random sample born 1950-52
  - Z_i indicates draft-eligible men as determined by the draft lotteries (randomizing eligibility over birthdays) held 1970-72
- Draft-eligibility RF on 1980s earnings is about $-400; the first stage about .16: Conscription reduced earnings by $2,500 (Angrist, 1990)
  - Assuming $Cov(Y_{0i}, Z_i) = 0$, we have ID

Sometimes You Get What You Need

- In today's *design-based* framework, observational data are viewed "as if" from a randomized trial
- *Internal* and *external validity:*
  - A good instrument captures an *internally valid* causal effect: the (average) impact on those "in the experiment"
  - The *external validity* of this effect is it's predictive value in populations other than the one for which the experiment is observed
- The modern framework highlights this distinction, emphasizing the case for internal validity (empirical work from the 1940s-80s often treated the choice of instruments casually)
- Examples
  - Draft-lottery estimates of the effects of Vietnam-era military service
  - Quarter-of-birth estimates of the effects of schooling on earnings
- In these examples, IV captures causal effects for a well-defined subpopulation (a subset of the treated)
- With variable treatment intensity, we get effects over a limited (but knowable) range

Children and Their Parents Labor Supply

Heterogeneous FX at work ...

- Do parents work and earn less as the price of childbearing? Or would those with bigger families have worked less anyway?
- Dependent variables = employment, hours worked, weeks worked, earnings
  - $D_i = 1[kids > 2]$ in families with at least two children
  - Z_i = twins or same-sex sibship at second birth
- With a Bernoulli instrument and no covariates, IV is Wald:

$$\rho = \frac{Cov(Y_i, Z_i)/V(Z_i)}{Cov(D_i, Z_i)/V(Z_i)} = \frac{RF}{1st}$$

$$= \frac{E[\mathrm{Y}_i|\mathrm{Z}_i = 1] - E[\mathrm{Y}_i|\mathrm{Z}_i = 0]}{E[\mathrm{D}_i|\mathrm{Z}_i = 1] - E[\mathrm{D}_i|\mathrm{Z}_i = 0]}$$

- **Results**
- Two good instruments (I love them both equally), two good (but different) estimates!

IV in the Real (heterogeneous) World

The Local Average Treatment Effects (LATE) Framework

- We assume that IV initiates a causal chain: the instrument, Z_i, affects D_i, which in turn affects Y_i.
- To flesh this out, we first define *potential treatment status*, indexed against Z_i
  - D_{1i} is i's treatment status when $Z_i = 1$
  - D_{0i} is i's treatment status when $Z_i = 0$
- The first link in the chain is observed treatment status:

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$$

- The causal effect of Z_i on D_i is $D_{1i} - D_{0i}$

LATE assumptions

Independence The instrument is as good as randomly assigned.

- Independence says that draft lottery numbers are independent of potential outcomes and potential treatments.

Exclusion The instrument affects Y_i only through D_i.

- The exclusion restriction says that draft lottery numbers affect earnings only through veteran status; sex composition affects labor supply only through family size.
- Exclusion takes us from RF causal effects to treatment effects.

Monotonicity $D_{1i} \geq D_{0i}$ for everyone (or vice versa).

- By virtue of monotonicity, an instrument can only push treatment in one direction.
- Wright would probably recognize only the second of these, though the others are implicit in his setup

IV with Heterogeneous Potential Outcomes

THE LATE THEOREM (Imbens and Angrist, 1991)

$$\frac{E[\mathrm{Y}_i|\mathrm{Z}_i=1]-E[\mathrm{Y}_i|\mathrm{Z}_i=0]}{E[\mathrm{D}_i|\mathrm{Z}_i=1]-E[\mathrm{D}_i|\mathrm{Z}_i=0]}=\frac{RF}{1st}=E[\mathrm{Y}_{1i}-\mathrm{Y}_{0i}|\mathrm{D}_{1i}>\mathrm{D}_{0i}]$$

- LATE *compliers* are subjects with $\mathrm{D}_{1i}>\mathrm{D}_{0i}$
- This language comes from randomized trials where Z_i is treatment assigned and D_i is treatment received (more on this soon)
- LATE partitions the world (Angrist, Imbens, and Rubin, 1996):
  - Compliers $D_{1i} > D_{0i}$
  - Always-takers $D_{1i} = D_{0i} = 1$
  - Never-takers $D_{1i} = D_{0i} = 0$
- IV is uninformative for always-takers and never-takers because treatment status for these types is unchanged by the instrument
- Note that {treated} = {always-takers} + {compliers with $\mathrm{Z}_i = 1$}, hence IV does not usually identify average effects on all treated

IV in Randomized Trials

The *compliance problem* in RCTs: Not all those randomly assigned to the treatment group are treated; many self-select

- When compliance is voluntary, an *as-treated* analysis is contaminated by selection bias
- *Intention-to-treat* analyses preserve independence but are diluted by non-compliance
- IV solves this problem: Z_i, is a dummy variable indicating random assignment to the treatment group; D_i is a dummy indicating whether treatment was actually received
- There are no always-takers (no controls treated), so here LATE $=$ the average effect on the treated:

$$\frac{E[\mathrm{Y}_i|\mathrm{Z}_i = 1] - E[\mathrm{Y}_i|\mathrm{Z}_i = 0]}{E[\mathrm{D}_i|\mathrm{Z}_i = 1]} = \frac{\text{ITT effect}}{\text{compliance rate}} = E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i}|\mathrm{D}_i = 1]$$

- Direct proof is due to Bloom (1984), though he does not mention (and was probably unaware of) the connection to IV

Bloom Example 1: Training

The Job Training Partnership Act (JTPA) included a large randomized trial to evaluate the effect of training on earnings

- The JTPA *offered* treatment randomly; participation was voluntary
- Roughly 60 percent of those offered training received it
- The IV setup is
  - D_i indicates those who received JTPA services
  - Z_i indicates the random offer of treatment
  - Y_i is earnings in the 30 months since random assignment
- The first-stage here is the compliance rate

$$\begin{aligned} E[\mathrm{D}_i|\mathrm{Z}_i &= 1] - E[\mathrm{D}_i|\mathrm{Z}_i = 0] = .62 - .02 \\ &\cong P[\mathrm{D}_i = 1|\mathrm{Z}_i = 1] \end{aligned}$$

  (about .02 of the control group received JTPA services)
- **Table 4.4.1** Selection bias in OLS (as delivered); ITT (as assigned) is diluted; IV (TOT) is . . . just right!

Bloom Example 2: Battered Wives

What's the best police response to domestic violence? The Minneapolis Domestic Violence Experiment (MDVE; Sherman and Berk, 1984) tries to find out

- Police were randomly assigned to advise, separate, or arrest
- Substantial compliance problems as officers made their own judgements in the field

Table 1: Assigned and Delivered Treatments in Spousal Assault Cases

Assigned Treatment	Delivered Treatment			
		Coddled		
	Arrest	Advise	Separate	Total
Arrest	98.9 (91)	0.0 (0)	1.1 (1)	29.3 (92)
Advise	17.6 (19)	77.8 (84)	4.6 (5)	34.4 (108)
Separate	22.8 (26)	4.4 (5)	72.8 (83)	36.3 (114)
Total	43.4 (136)	28.3 (89)	28.3 (89)	100.0(314)

MDVE First-Stage and Reduced Forms

Table 2: First Stage and Reduced Forms for Model 1

Endogenous Variable is Coddled				
	First-Stage		Reduced Form (ITT)	
	(1)	(2)*	(3)	(4)*
Coddled-assigned	0.786 (0.043)	0.773 (0.043)	0.114 (0.047)	0.108 (0.041)
Weapon		-0.064 (0.045)		-0.004 (0.042)
Chem. Influence		-0.088 (0.040)		0.052 (0.038)
Dep. Var. mean	0.567 (coddled-delivered)		0.178 (failed)	

MDVE OLS and 2SLS

The effect of coddling on the coddled: 2SLS tells us how much less likely those not arrested would have been to re-offend if they had in fact been arrested . . .

Table 3: OLS and 2SLS Estimates for Model 1

Endogenous Variable is Coddled				
	OLS		IV/2SLS	
	(1)	(2)*	(3)	(4)*
Coddled-delivered	0.087 (0.044)	0.070 (0.038)	0.145 (0.060)	0.140 (0.053)
Weapon		0.010 (0.043)		0.005 (0.043)
Chem. Influence		0.057 (0.039)		0.064 (0.039)

Wrapping Up . . .

- IV provides a powerful and flexible framework for causal inference
  - The core identification strategy for simultaneous equations models - problem solved!
  - An elegant solution to the problem of mismeasured regressors
  - A general framework for the elimination of omitted variables bias
  - IV solves the compliance problem in randomized trials and related research designs
- Philip Wright's pathbreaking (in more ways than one!) analysis - the obscure Appendix B - laid the foundation for progress on the problems that define our field
- The best econometric theory then as now, emerges from real empirical questions. PGW's interest in the economic consequences of tariffs left a remarkable legacy indeed
- Like Wright, today's empiricists start with causality and identification, but improve on the empirical work published in the wake of the 1940s re-emergence of IV

Why We Now Do Better

- Wright understood what my empirical colleagues today hold dear, IV empirical work lives or dies on the choice of instruments:

  Success with this method depends on success in discovering factors of the type A and B (p. 314)

- Path-breaking for it's time, Geary's (1949) Cowles-era demand analysis says about the choice of instruments:

  For the instrumental set we have no fewer than 15 series which Stone ([1947], page 11) numbers 3 to 17: they need not be particularized here.

- Indifference to the details of the underlying experiment is typical of the immediate post-Cowles era; this work often used instruments with no real independent information (see, e.g., Christ [1994, p. 55])
- Today, the experiment is where we begin . . . and end

Tables and Figures

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

	1980 PUMS			1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:			Wald estimate using as covariate:			Wald estimate using as covariate:	
Variable	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Twins-2*	*More than 2 children*	*Number of children*
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	−0.0080 (0.0016)	−0.133 (0.026)	−0.104 (0.021)	−0.0053 (0.0015)	−0.084 (0.024)	−0.063 (0.018)	−0.0459 (0.0086)	−0.076 (0.014)	−0.057 (0.011)
Weeks worked	−0.3826 (0.0709)	−6.38 (1.17)	−5.00 (0.92)	−0.3233 (0.0743)	−5.15 (1.17)	−3.87 (0.88)	−1.982 (0.386)	−3.28 (0.63)	−2.45 (0.47)
Hours/week	−0.3110 (0.0602)	−5.18 (1.00)	−4.07 (0.78)	−0.2363 (0.0620)	−3.76 (0.98)	−2.83 (0.73)	−1.979 (0.327)	−3.28 (0.54)	−2.44 (0.40)
Labor income	−132.5 (34.4)	−2208.8 (569.2)	−1732.4 (446.3)	−119.4 (42.4)	−1901.4 (670.3)	−1428.0 (502.6)	−570.8 (186.9)	−946.4 (308.6)	−705.2 (229.8)
ln(*Family income*)	−0.0018 (0.0041)	−0.029 (0.068)	−0.023 (0.054)	−0.0085 (0.0047)	−0.136 (0.074)	−0.102 (0.056)	−0.0341 (0.0223)	−0.057 (0.037)	−0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

TABLE 4.4.1
Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status (OLS)		Comparisons by Assignment Status (ITT)		Instrumental Variable Estimates (IV)	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

Notes: Authors' tabulation of JTPA study data. The table reports OLS, ITT, and IV estimates of the effect of subsidized training on earnings in the JTPA experiment. Columns 1 and 2 show differences in earnings by training status; columns 3 and 4 show differences by random-assignment status. Columns 5 and 6 report the result of using random-assignment status as an instrument for training. The covariates used for columns 2, 4, and 6 are high school or GED, black, Hispanic, married, worked less than 13 weeks in past year, AFDC (for women), plus indicators for the JTPA service strategy recommended, age group, and second follow-up survey. Robust standard errors are shown in parentheses. There are 5,102 men and 6,102 women in the sample.

Table 4.4.2
Probabilities of compliance in instrumental variables studies

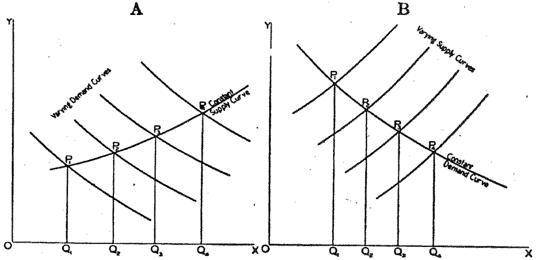
Endogenous Variable (D) (2)	Instrument (Z) (3)	Sample (4)	$P[D=1]$ (5)	First Stage, $P[D_1 > D_0]$ (6)	$P[Z=1]$ (7)	Compliance Probabilities: $P[D_1 > D_0 \mid D=1]$ (8)	Compliance Probabilities: $P[D_1 > D_0 \mid D=0]$ (9)
Veteran status	Draft eligibility	White men born in 1950	.267	.159	.534	.318	.101
		Non-white men born in 1950	.163	.060	.534	.197	.033
More than two children	Twins at second birth	Married women aged 21–35 with two or more children in 1980	.381	.603	.008	.013	.966
	First two children are same sex		.381	.060	.506	.080	.048
High school graduate	Third- or fourth-quarter birth	Men born between 1930 and 1939	.770	.016	.509	.011	.034
High school graduate	State requires 11 or more years of school attendance	White men aged 40–49	.617	.037	.300	.018	.068

TABLE 4.4.3
Complier characteristics ratios for twins and sex composition instruments

		Twins at Second Birth		First Two Children Are Same Sex	
Variable	$P[x_{1i}=1]$ (1)	$P[x_{1i}=1 \mid D_{1i} > D_{0i}]$ (2)	$P[x_{1i}=1 \mid D_{1i} > D_{0i}]/P[x_{1i}=1]$ (3)	$P[x_{1i}=1 \mid D_{1i} > D_{0i}]$ (4)	$P[x_{1i}=1 \mid D_{1i} > D_{0i}]/P[x_{1i}=1]$ (5)
Age 30 or older at first birth	.0029	.004	1.39	.0023	.995
Black or hispanic	.125	.103	.822	.102	.814
High school graduate	.822	.861	1.048	.815	.998
College graduate	.132	.151	1.14	.0904	.704

Notes: The table reports an analysis of complier characteristics for twins and sex composition instruments. The ratios in columns 3 and 5 give the relative likelihood that compliers have the characteristic indicated at left. Data are from the 1980 census 5 percent sample, including married mothers aged 21–35 with at least two children, as in Angrist and Evans (1998). The sample size is 254,654 for all columns.

FIGURE 3. PRICE-OUTPUT DATA REVEAL—
(A) SUPPLY CURVE
(B) DEMAND CURVE

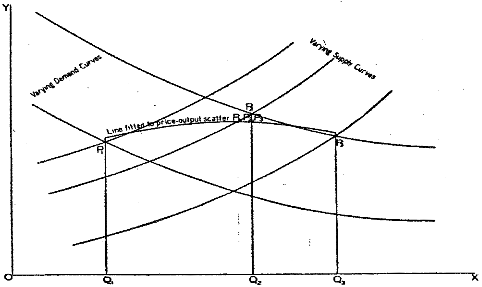


cally by a bodily shifting of the demand curve to the right. Supply conditions are said to move toward lower costs when a greater quantity will be forthcoming at the same price. This also will be shown graphically by a bodily shifting of the supply curve to the right. A weakening of demand or a movement toward higher costs will be shown by a shifting of the curves to the left.

If it can be shown that during a period of time covered by two or more observations either curve remains fixed while the other moves to right or left, price-output data will reveal points on the curve that remains fixed. This should be obvious from the analysis given above but may be illustrated by a diagram. (Figure 3.)

If both supply and demand conditions change, price-output data yield no direct information as to either curve. (Figure 4.)

FIGURE 4. PRICE-OUTPUT DATA FAIL TO REVEAL EITHER SUPPLY OR DEMAND CURVE.



Unfortunately for our problem, the case represented by Figure 4 is the more common, and even if either curve does remain fixed during the period covered by the observations there is no certain way of knowing this fact in advance.[5]