Instrumental Variables and the Rubin Causal Model

Charlie Gibbons
Political Science 236

October 22, 2008

Outline

Reference

This presentation is based largely on:

Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin. 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association.* 91(434): 444-455.

The problem

Thus far we have only considered treatment status, but now we want to distinguish between treatment assignment and treatment status; the former is prescribed by the researcher and the latter is the actual treatment taken up by the individual.

The problem

Thus far we have only considered treatment status, but now we want to distinguish between treatment assignment and treatment status; the former is prescribed by the researcher and the latter is the actual treatment taken up by the individual.

These two measures would be the same under perfect compliance. But compliance is a choice made by the individual based upon unobservable factors—specifically, we have non-ignorable treatment status.

The problem

Thus far we have only considered treatment status, but now we want to distinguish between treatment assignment and treatment status; the former is prescribed by the researcher and the latter is the actual treatment taken up by the individual.

These two measures would be the same under perfect compliance. But compliance is a choice made by the individual based upon unobservable factors—specifically, we have non-ignorable treatment status.

Recall the definition of ignorability:

- Selection on observables:
  $(Y_i(0), Y_i(1)) \perp T_i | X_i$
- Common support on covariates:
  $0 < \Pr(T = 1 | X = x) < 1 \; \forall \; x \in X$

Framework

Z_i is an instrument taking values 0 or 1
$\mathbf{Z}$ is an N-dimensional vector of treatment assignments to all individuals

$D_i(\mathbf{Z})$ is the treatment status of individual i taking values 0 or 1 (*i.e.*, there is no partial compliance)
$\mathbf{D}(\mathbf{Z})$ is the treatment status of all individuals

If there was perfect compliance, then $\mathbf{D}(\mathbf{Z}) = \mathbf{Z}$.

$Y_i(\mathbf{Z}, \mathbf{D}(\mathbf{Z}))$ is the response of individual i given all treatment assignments and statuses.
$\mathbf{Y}(\mathbf{Z}, \mathbf{D}(\mathbf{Z}))$ is the responses of all individuals

Since compliance isn't perfect, both and $Y_i(\mathbf{Z}, D(\mathbf{Z}))$ $\mathbf{D}_i(\mathbf{Z})$ are potential outcomes.

SUTVA

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA)

Both potential outcomes for individual i are independent of the treatment assignment to other individuals.

SUTVA

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA)

Both potential outcomes for individual i are independent of the treatment assignment to other individuals.

- If an individual has the same value of the instrument in two different treatment assignments, then his treatment status will be the same under both assignments; if $Z_i = Z_i'$, then $D_i(\mathbf{Z}) = D_i(\mathbf{Z}')$.
- If an individual has the same treatment assignment and status under two different treatment assignments, then his response will be the same under both assignments; if $Z_i = Z_i'$ and $D_i(\mathbf{Z}) = D_i(\mathbf{Z}')$, then $Y_i(\mathbf{Z}, \mathbf{D}(\mathbf{Z})) = Y_i(\mathbf{Z}', \mathbf{D}(\mathbf{Z}'))$.

SUTVA

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA)

Both potential outcomes for individual i are independent of the treatment assignment to other individuals.

This assumption permits us to write $D_i(\mathbf{Z}) = D_i(Z_i)$ and $Y_i(\mathbf{Z}, \mathbf{D}(\mathbf{Z})) = Y_i(Z_i, D_i(Z_i))$.

Randomization

Assumption 2: Random assignment of $\mathbf{Z}$

$$\Pr(Z_i = 1) = \Pr(Z_j = 1) \quad \forall\, i, j$$

This assumption is not strictly necessary; the assignment of Z simply needs to be ignorable.

Causal effects

The causal effect of Z_i on D_i is

$$\Delta_{ZD} = D_i(1) - D_i(0).$$

The causal effect of Z_i on Y_i is

$$\Delta_{ZY} = Y_i(1, D_i(1)) - Y_i(0, D_i(0)).$$

These are known as *intention to treat* effects.

Causal effects

Given Assumptions 1 and 2, we can write unbiased estimators for these effects, which are simply the differences in means:

$$\begin{aligned}
\Delta_{ZD} &= \frac{1}{\#\{i\ :\ Z_i = 1\}} \sum_{i\,:\,Z_i=1} D_i \\
&- \frac{1}{\#\{j\ :\ Z_j = 0\}} \sum_{j\,:\,Z_j=0} D_j \\
\Delta_{ZY} &= \frac{1}{\#\{i\ :\ Z_i = 1\}} \sum_{i\,:\,Z_i=1} Y_i \\
&- \frac{1}{\#\{j\ :\ Z_j = 0\}} \sum_{j\,:\,Z_j=0} Y_j
\end{aligned}$$

The IV estimator will be the ratio of these two causal effects, subject to three more assumptions.

Exclusion restriction

Assumption 3: Exclusion restriction

The instrument can only effect the outcome by altering treatment status.

Exclusion restriction

Assumption 3: Exclusion restriction

The instrument can only effect the outcome by altering treatment status.

$$\mathbf{Y}(\mathbf{Z}, \mathbf{D}) = \mathbf{Y}(\mathbf{Z}', \mathbf{D}) \quad \forall\, \mathbf{Z}, \mathbf{Z}', \mathbf{D}$$

That is, the effect of Z on Y must be solely through the effect of Z on D and a change in the value of the instrument does not effect the outcome unless it changes treatment status.

Exclusion restriction

Assumption 3: Exclusion restriction

The instrument can only effect the outcome by altering treatment status.

$$\mathbf{Y}(\mathbf{Z}, \mathbf{D}) = \mathbf{Y}(\mathbf{Z}', \mathbf{D}) \quad \forall \, \mathbf{Z}, \mathbf{Z}', \mathbf{D}$$

That is, the effect of Z on Y must be solely through the effect of Z on D and a change in the value of the instrument does not effect the outcome unless it changes treatment status.

The combination of the exclusion restriction, ignorable treatment effects, and the linearity assumption imply that Z_i is uncorrelated with both error terms conditional on the observables: $\mathbb{E}[Z_i \nu_i | X] = \mathbb{E}[Z_i \epsilon_i | X] = 0$.

Exclusion restriction

Assumption 3: Exclusion restriction

The instrument can only effect the outcome by altering treatment status.

This lets us write $\mathbf{Y}(\mathbf{Z}, \mathbf{D})$ as $\mathbf{Y}(\mathbf{D})$.

Assumption 1 lets us write $Y_i(\mathbf{D})$ as $Y_i(D_i)$.

Hence, the causal effect of D on Y is $Y_i(1) - Y_i(0)$, the standard result of the Rubin causal model framework.

Inclusion restriction

Assumption 4: Inclusion restriction

The instrument has a non-zero average effect on treatment status.

$$\mathbb{E}[D_i(1) - D_i(0)|X] \neq 0$$

This can also be expressed as $\text{Cov}(Z, X) \neq 0$.

Inclusion restriction

Assumption 4: Inclusion restriction

The instrument has a non-zero average effect on treatment status.

$$\mathbb{E}[D_i(1) - D_i(0)|X] \neq 0$$

This can also be expressed as $\text{Cov}(Z, X) \neq 0$.

The inclusion restriction considers the effect of the assignment mechanism on treatment status, while the exclusion restriction considers the effect of the assignment mechanism on the outcome.

Monotonicity

Assumption 5: Monotonicity

The instrument effects all individuals' treatment statuses in the same direction.

$$D_i(1) \geq D_i(0) \text{ or } D_i(1) \leq D_i(0) \quad \forall\, i$$

Combined with Assumption 4, this equality must be strict for some individual i.

Definition of an instrument

Now we can define an instrument precisely.

Definition of an instrument

Now we can define an instrument precisely.

A covariate Z is an instrument for the causal effect of D on Y if:

- SUTVA holds
- It is randomly (or ignorably) assigned
- It satisfies the exclusion restriction
- Its average effect on D is non-zero (*i.e.*, it satisfies the inclusion restriction)
- It satisfies the monotonicity assumption

Respondent types

		$D_i(0)$	
		0	1
$D_i(1)$	0	Never-taker	Defier
	1	Complier	Always-taker

Respondent types

		$D_i(0)$	
		0	1
$D_i(1)$	0	Never-taker	Defier
	1	Complier	Always-taker

The behavior of never-takers and always-takers does not change in response to a change in the instrument Z and treatment status is always the same. Because only one status is ever observed, there can be no causal effect of a change in status for these groups.

Respondent types

		$D_i(0)$	
		0	1
$D_i(1)$	0	Never-taker	Defier
	1	Complier	Always-taker

The behavior of never-takers and always-takers does not change in response to a change in the instrument Z and treatment status is always the same. Because only one status is ever observed, there can be no causal effect of a change in status for these groups.

The monotonicity assumption rules out defiers.

Respondent types

		$D_i(0)$	
		0	1
$D_i(1)$	0	Never-taker	Defier
	1	Complier	Always-taker

The behavior of never-takers and always-takers does not change in response to a change in the instrument Z and treatment status is always the same. Because only one status is ever observed, there can be no causal effect of a change in status for these groups.

The monotonicity assumption rules out defiers.

Thus, the effect of Z on Y is identified solely by this effect for compliers.

Respondent types

The IV estimator can be written as

$$\frac{\mathbb{E}[Y_i(1, D_i(1))|X] - \mathbb{E}[Y_i(0, D_i(0))|X]}{\mathbb{E}[D_i(1) - D_i(0)|X]}$$
$$= \mathbb{E}[Y_i(1, D_i(1)) - Y_i(0, D_i(0))|D_i(1) - D_i(0) = 1, X]$$

This is called the *local average treatment effect*, where "local" refers to the fact that the estimate is the average treatment effect for the group of compliers only.

Respondent types

The IV estimator can be written as

$$\frac{\mathbb{E}[Y_i(1, D_i(1))|X] - \mathbb{E}[Y_i(0, D_i(0))|X]}{\mathbb{E}[D_i(1) - D_i(0)|X]}$$
$$= \mathbb{E}[Y_i(1, D_i(1)) - Y_i(0, D_i(0))|D_i(1) - D_i(0) = 1, X]$$

This is called the *local average treatment effect*, where "local" refers to the fact that the estimate is the average treatment effect for the group of compliers only.

Note that this group is *unobservable*.

Estimation

The following two equations define the instrumental variables framework:

$$
\begin{aligned}
X &= Z\gamma + \nu \\
Y &= X\beta + \epsilon
\end{aligned}
$$

Note that the Z matrix contains the instruments and the non-endogenous X regressors. The instruments do not appear in the outcome regression.

Estimation

The simplest case is when there are the same number of endogenous regressors as there are instruments (the just=identified case). Then we can solve the problem as follows:

$$
\begin{aligned}
Y &= X\hat{\beta} + \epsilon \\
Z'Y &= Z'X\hat{\beta} + Z'\epsilon \\
Z'Y &= Z'X\hat{\beta} \\
\hat{\beta} &= (Z'X)^{-1}Z'Y
\end{aligned}
$$

Estimation

If there are more instruments than there are endogenous regressors (the over-identified case), then we can follow a two-stage least squares process:

1. Calculate $\hat{\gamma} = (Z'Z)^{-1}Z'X$.
2. Calculate $\hat{X} = Z\hat{\gamma} = Z(Z'Z)^{-1}Z'X$.
3. Use $\hat{X}$ instead of X in the outcome equation: $Y = \hat{X}\beta = Z(Z'Z)^{-1}Z'X\beta$.
4. The estimate of β becomes:

$$
\begin{aligned}
\hat{\beta} &= \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y
\end{aligned}
$$

Estimation

If there are more instruments than there are endogenous regressors (the over-identified case), then we can follow a two-stage least squares process:

1. Calculate $\hat{\gamma} = (Z'Z)^{-1}Z'X$.
2. Calculate $\hat{X} = Z\hat{\gamma} = Z(Z'Z)^{-1}Z'X$.
3. Use $\hat{X}$ instead of X in the outcome equation: $Y = \hat{X}\beta = Z(Z'Z)^{-1}Z'X\beta$.
4. The estimate of β becomes:

$$\begin{aligned}
\hat{\beta} &= \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y
\end{aligned}$$

Estimation

If there are more instruments than there are endogenous regressors (the over-identified case), then we can follow a two-stage least squares process:

1. Calculate $\hat{\gamma} = (Z'Z)^{-1}Z'X$.
2. Calculate $\hat{X} = Z\hat{\gamma} = Z(Z'Z)^{-1}Z'X$.
3. Use $\hat{X}$ instead of X in the outcome equation: $Y = \hat{X}\beta = Z(Z'Z)^{-1}Z'X\beta$.
4. The estimate of β becomes:

$$
\begin{aligned}
\hat{\beta} &= \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y
\end{aligned}
$$

Estimation

If there are more instruments than there are endogenous regressors (the over-identified case), then we can follow a two-stage least squares process:

1. Calculate $\hat{\gamma} = (Z'Z)^{-1}Z'X$.
2. Calculate $\hat{X} = Z\hat{\gamma} = Z(Z'Z)^{-1}Z'X$.
3. Use $\hat{X}$ instead of X in the outcome equation: $Y = \hat{X}\beta = Z(Z'Z)^{-1}Z'X\beta$.
4. The estimate of β becomes:

$$\begin{aligned}
\hat{\beta} &= \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y
\end{aligned}$$

Estimation

If there are more instruments than there are endogenous regressors (the over-identified case), then we can follow a two-stage least squares process:

1. Calculate $\hat{\gamma} = (Z'Z)^{-1}Z'X$.
2. Calculate $\hat{X} = Z\hat{\gamma} = Z(Z'Z)^{-1}Z'X$.
3. Use $\hat{X}$ instead of X in the outcome equation: $Y = \hat{X}\beta = Z(Z'Z)^{-1}Z'X\beta$.
4. The estimate of β becomes:

$$
\begin{aligned}
\hat{\beta} &= \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'Y
\end{aligned}
$$

Estimation

Instrumental variables use only part of the variability in the endogenous regressor—the part that is correlated with the instrument and uncorrelated with omitted variables—to identify the effect of that regressor.

Estimation

Instrumental variables use only part of the variability in the endogenous regressor—the part that is correlated with the instrument and uncorrelated with omitted variables—to identify the effect of that regressor.

IV estimates are built around estimates of $\hat{X}$, rather than a fixed, "true" value X. As a result, IV estimates have finite sample bias. This bias increases with the number of instruments employed, but decreases when these instruments are highly correlated with the endogenous regressors and as the sample size increases. The estimates are consistent, however.

Concerns

Recall the assumptions of IV:

- SUTVA holds
- Z is randomly (or ignorably) assigned
- The exclusion restriction holds
- The average effect of Z on D is non-zero (*i.e.*, it satisfies the inclusion restriction)
- Z satisfies the monotonicity assumption

Concerns

Recall the assumptions of IV:

- SUTVA holds
  Always needed for causal analysis
- Z is randomly (or ignorably) assigned
- The exclusion restriction holds
- The average effect of Z on D is non-zero (*i.e.*, it satisfies the inclusion restriction)
- Z satisfies the monotonicity assumption

Concerns

Recall the assumptions of IV:

- SUTVA holds
  Always needed for causal analysis
- Z is randomly (or ignorably) assigned
- The exclusion restriction holds
- The average effect of Z on D is non-zero (*i.e.*, it satisfies the inclusion restriction)
  This is testable: first-stage F-test
  If the effect is small, "weak" instruments
  Imbens-Rosenbaum standard errors robust to weak instruments
- Z satisfies the monotonicity assumption

Concerns

- Z is randomly (or ignorably) assigned
- The exclusion restriction holds
- Z satisfies the monotonicity assumption

Are unique to IV and are *untestable*.

Concerns

- Z is randomly (or ignorably) assigned
- The exclusion restriction holds
- Z satisfies the monotonicity assumption

Are unique to IV and are *untestable*.

If we know how the instrument was assigned (*i.e.*, randomly), we may be able to justify the ignorability assumption.

Concerns

- Z is randomly (or ignorably) assigned
- The exclusion restriction holds
- Z satisfies the monotonicity assumption

Are unique to IV and are *untestable*.

If we know how the instrument was assigned (*i.e.*, randomly), we may be able to justify the ignorability assumption.

We tell stories, but we need a lot of knowledge to believe these assumptions hold and there is no way to empirically justify these beliefs.

IV v. Matching

Matching assumption: Treatment status is independent of outcomes given a set of covariates.

IV v. Matching

Matching assumption: Treatment status is independent of outcomes given a set of covariates.

IV assumption: Treatment status is *not* independent of unobservables, but treatment assignment is.

IV v. Matching

Matching assumption: Treatment status is independent of outcomes given a set of covariates.

IV assumption: Treatment status is *not* independent of unobservables, but treatment assignment is.

Matching assumption: All unobserved factors effecting outcomes and treatment status are perfectly correlated with the observables to obtain the independence assumption of selection on observables.

IV v. Matching

Matching assumption: Treatment status is independent of outcomes given a set of covariates.

IV assumption: Treatment status is *not* independent of unobservables, but treatment assignment is.

Matching assumption: All unobserved factors effecting outcomes and treatment status are perfectly correlated with the observables to obtain the independence assumption of selection on observables.

IV assumption: Part (though not all) of treatment status must be explained by exogenous instruments; unobserved factors effecting outcomes or treatment must be uncorrelated with the instrument (conditional on the observable covariates).

IV v. Matching

Matching: Balance on covariates between treatment and control groups can be examined.

IV v. Matching

Matching: Balance on covariates between treatment and control groups can be examined.

IV: Exclusion restriction is *not* testable.

IV v. Matching

Matching: Balance on covariates between treatment and control groups can be examined.

IV: Exclusion restriction is *not* testable.

Matching: Treatment effect estimated for an identifiable group.

IV v. Matching

Matching: Balance on covariates between treatment and control groups can be examined.

IV: Exclusion restriction is *not* testable.

Matching: Treatment effect estimated for an identifiable group.

IV: Treatment effect estimated for an unknown, instrument-dependent group.

IV v. Matching

Matching: Balance on covariates between treatment and control groups can be examined.

IV: Exclusion restriction is *not* testable.

Matching: Treatment effect estimated for an identifiable group.

IV: Treatment effect estimated for an unknown, instrument-dependent group.

Matching: Non-parametric.

IV v. Matching

Matching: Balance on covariates between treatment and control groups can be examined.

IV: Exclusion restriction is *not* testable.

Matching: Treatment effect estimated for an identifiable group.

IV: Treatment effect estimated for an unknown, instrument-dependent group.

Matching: Non-parametric.

Linearity assumed (though other functional form assumptions are possible).