Econ 2148, spring 2019
Instrumental variables II, continuous treatment

Maximilian Kasy

Department of Economics, Harvard University

Recall instrumental variables part I

- Origins of instrumental variables: Systems of linear structural equations
  Strong restriction: Constant causal effects.
- Modern perspective: Potential outcomes, allow for heterogeneity of causal effects
- Binary case:
  1. Keep IV estimand, reinterpret it in more general setting:
     Local Average Treatment Effect (LATE)
  2. Keep object of interest average treatment effect (ATE):
     Partial identification (Bounds)

Agenda instrumental variables part II

- Continuous treatment case:
  1. Restricting heterogeneity in the structural equation: Nonparametric IV (conditional moment equalities)
  2. Restricting heterogeneity in the first stage: Control functions
  3. Linear IV: Continuous version of LATE

Takeaways for this part of class

- We can write linear IV in three numerically equivalent ways:
  1. As ratio $\text{Cov}(Z, Y) / \text{Cov}(Z, X)$.
  2. As regression of Y on first stage predicted values $\widehat{X}$.
  3. As regression of Y on X controlling for the first stage residual V.
- The literature on IV identification with continuous treatment generalizes these ideas to non-linear settings.

Takeaways continued

1. Moment restrictions:
   - Assume one-dimensional additive heterogeneity in structural equation of interest
   - $\Rightarrow$ nonparametric regression of Y on non-parametric prediction $\widehat{X}$.
2. Control functions:
   - Assume one-dimensional heterogeneity in first stage relationship.
   - $\Rightarrow$ X is independent of structural heterogeneity conditional on $V = F_{X|Z}(X|Z)$.
3. Continuous LATE:
   - No restrictions on heterogeneity.
   - Interpret linear IV coefficient as weighted average derivative.

Alternative ways of writing the linear IV estimand

- Linear triangular system:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + U \\ X &= \gamma_0 + \gamma_1 Z + V \end{aligned}$$

- Exogeneity (randomization) conditions:

$$\mathrm{Cov}(Z, U) = 0, \qquad \mathrm{Cov}(Z, V) = 0.$$

- Relevance condition:

$$\mathrm{Cov}(Z, X) = \gamma_1 \,\mathrm{Var}(Z) \neq 0.$$

- Under these conditions,

$$\beta_1 = \frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, X)}.$$

Moment conditions

- Write $\text{Cov}(Z, U) = 0$ as

$$\text{Cov}(Z, Y - \beta_0 - \beta_1 X) = 0$$

- Let $\widehat{X}$ be the predicted value from a first stage regression,

$$\widehat{X} = \gamma_0 + \gamma_1 Z.$$

- Multiply $\text{Cov}(Z, U)$ by γ_1,

$$\text{Cov}(\widehat{X}, Y - \beta_0 - \beta_1 X) = 0,$$

  and note $\text{Cov}(\widehat{X}, X) = \text{Var}(\widehat{X})$, to get

$$\beta_1 = \frac{\text{Cov}(\widehat{X}, Y)}{\text{Var}(\widehat{X})}.$$

- $\Rightarrow$ two-stage least squares!

Conditional moment equalities

- Under the stronger mean independence restriction $E[U|Z] \equiv 0$,

$$
\begin{aligned}
0 &= E[(Y - \beta_0 - \beta_1 X)|Z = z] \\
&= E[Y|Z = z] - \beta_0 - \beta_1 E[X|Z = z]
\end{aligned}
$$

  for all z.
- "Conditional moment equality"
- Suggest 2 stage estimator:
  1. Regress both Y and X (non-parametrically or linearly) on Z.
  2. Then regress $E[Y|Z = z]$ or Y (linearly) on $E[X|Z = z]$.
- $\Rightarrow$ two-stage least squares!

Control function perspective

- V is the residual of a first stage regression of X on Z.
- Consider a regression of Y on X and V,

$$Y = \delta_0 + \delta_1 X + \delta_2 V + W$$

- Partial regression formula:
  - δ_1 is the coefficient of a regression of $\tilde{Y}$ on $\tilde{X}$ (or of Y on $\tilde{X}$),
  - where $\tilde{Y}$, $\tilde{X}$ are the residuals of regressions on V.
- By construction:

$$\tilde{X} = \gamma_0 + \gamma_1 Z = \widehat{X}$$
$$\tilde{Y} = \beta_0 + \beta_1 \tilde{X} + \tilde{U}$$

- $\mathrm{Cov}(Z, U) = \mathrm{Cov}(Z, V) = 0$ implies $\mathrm{Cov}(\tilde{X}, \tilde{U}) = 0$, and thus

$$\delta_1 = \beta_1.$$

Recap

- Three numerically equivalent estimands:
  1. The slope
  $$\text{Cov}(Z, Y)/\text{Cov}(Z, X).$$
  2. The two-stage least squares slope from the regression
  $$Y = \beta_0 + \beta_1 \widehat{X} + \tilde{U},$$
  where $\tilde{U} = (\beta_1 V + U)$, and $\widehat{X}$ is the first stage predicted value $\widehat{X} = \gamma_0 + \gamma_1 Z$.
  3. The slope of the regression with control
  $$Y = \delta_0 + \delta_1 X + \delta_2 V + W,$$
  where the control function V is given by the first stage residual, $V = X - \gamma_0 - \gamma_1 Z$.

Roadmap

- Nonparametric IV estimators generalize these approaches in different ways, dropping the linearity assumptions:
  1. If heterogeneity in the structural equation is one-dimensional: conditional moment equalities
  2. If heterogeneity in the first stage is one-dimensional: control functions
  3. Without heterogeneity restrictions: continuous versions of the LATE result for the linear IV estimand
- Objects of interest:
  - Average structural function (ASF) $\bar{g}(x) = E[g(x, U)]$.
  - Quantile structural function (QSF) $g_\tau(x)$ defined by $P(g(x, U) < g_\tau(x)) = \tau$.
  - Weighted averages of marginal causal effect, $\int E[\omega_x \cdot g'(x, U)]dx$ for weights ω_x.

Approach I:
Conditional moment restrictions (nonparametric IV)

- Consider the following generalization of the linear model:

$$\begin{aligned} Y &= g(X) + U \\ X &= h(Z, V) \\ Z &\perp (U, V) \end{aligned}$$

- Here the ASF $\bar{g}$ equals g.

Practice problem

- Under these assumptions, write out the conditional expectation $E[Y|Z = z]$ as an integral with respect to $dP(X|Z = z)$.
- Consider the special case where both X and Z have finite support of size n_x and n_z, and rewrite the integral as a matrix multiplication.

Solution

- Using additivity of structural equation, and independence,

$$\begin{aligned} k(z) = E[Y|Z=z] &= E[g(X)|Z=z] + E[U|Z=z] \\ &= E[g(X)|Z=z] \\ &= \int g(x) dP(X=x|Z=z). \end{aligned}$$

- In the finite support case, let
  - $\boldsymbol{k} = (k(z_1), \ldots, k(z_{n_z}))$, $\boldsymbol{g} = (g(x_1), \ldots, g(x_{n_x}))$,
  - and let P be the $n_z \times n_x$ matrix with entries $P(X=x|Z=z)$.
- Then the integral equation can be written as

$$\boldsymbol{k} = P \cdot \boldsymbol{g}.$$

Completeness

- The function $k(z) = E[Y|Z = z]$ and the conditional distribution $P_{X|Z}$ are identified.
- In the finite-support case, the equation $\boldsymbol{k} = P \cdot \boldsymbol{g}$ implies that $\boldsymbol{g}$ is identified if the matrix P has full column rank n_x.
- The analogue of the full rank condition for the continuous case (integral equation) is called "completeness."
- Completeness requires that variation in Z induces enough variation in X, like the "instrument relevance" condition in the linear case.
- Completeness is a feature of the observable distribution $P_{X|Z}$, in contrast to the conditions of exogeneity / exclusion, or restrictions on heterogeneity.

Ill posed inverse problem

- Even if completeness holds, estimation in the continuous case is complicated by the “ill posed inverse” problem.
- Consider the discrete case. The vector $\boldsymbol{g}$ is identified from

$$\boldsymbol{g} = (P'P)^{-1}P'\boldsymbol{k}$$

- Suppose that $P'P$ has eigenvalues close to zero. Then $\boldsymbol{g}$ is very sensitive to minor changes in $P'\boldsymbol{k}$.

- Continuous analog: notation

$$\tilde{k}(z) = E[Y|Z = z]f_Z(z)$$
$$(\boldsymbol{P}g)(z) = \int g(x)f_{X,Z}(x,z)dx$$
$$(\boldsymbol{P}'k)(x) = \int k(z)f_{X,Z}(x,z)dz$$
$$\boldsymbol{T} = \boldsymbol{P}' \circ \boldsymbol{P}$$

- Thus the moment conditions can be rewritten as $\tilde{k} = \boldsymbol{P}g$ or $\boldsymbol{P}'\tilde{k} = \boldsymbol{T}g$,
- Therefore

$$g = \boldsymbol{T}^{-1}\boldsymbol{P}'\tilde{k},$$

if the inverse of $\boldsymbol{T}$ exists – which is equivalent to completeness.

- $\boldsymbol{T}$ is a linear, self-adjoint ($\approx$ symmetric) positive definite operator on L^2.
- Functional analysis:
  If $\int\int f_{X,Z}(x,z)^2 fxdz \leq \infty$, then 0 is the unique accumulation point of the eigenvalues of $\boldsymbol{T}$,
- and the eigenvectors form an orthonormal basis of L^2.
- Implication: g is *not* a continuous function of $\boldsymbol{P}'\tilde{k}$ in L^2.
- Minor estimation errors for $\tilde{k}$ can translate into arbitrarily large estimation errors for g.
- Takeaway: Estimation needs to use regularization, convergence rates are slow.

Estimation using series

- Implementation is surprisingly simple.
- Use series approximation $g(x) \approx \sum_{j=1}^{k} \beta_j \phi_j(x)$.
- Then we get

$$E[\phi_{j'}(Z)Y] \approx \sum_{j=1}^{k} \beta_j E[\phi_{j'}(Z)\phi_j(X)]$$

- and thus

$$\beta \approx (E[\phi_{j'}(Z)\phi_j(X)])_{j,j'}^{-1}(E[\phi_{j'}(Z)Y])_{j'}.$$

- Sample analog: Two stage least squares, where the regressors $\phi_j(X)$ are instrumented by the instruments $\phi_{j'}(Z)$.

Additive one-dimensional hetereogeneity is crucial for conditional moment equality

- Consider the following non-additive example:

$$Y = X^2 \cdot U$$
$$X = Z + V$$
$$(U, V) \sim N\left(0, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

- Average structural function:

$$\bar{g}(x) = E[x^2 \cdot U] = 0.$$

- Conditional moment equality is solved by $\tilde{g}(x) = x$:

$$\begin{aligned} E[Y - \tilde{g}(X)|Z = z] &= E[(Z + V)^2 U|Z = z] - z \\ &= 2zE[VU] + E[V^2 U] - z = 0. \end{aligned}$$

Non-additive heterogeneity

- Consider now the slightly more general model

$$Y = g(X, U)$$
$$X = h(Z, V)$$
$$Z \perp (U, V)$$

- where $\dim(U) = 1$ and g is strictly monotonic in U.
- We can assume w.l.o.g. $U \sim Uniform([0, 1])$.
- Here the QSF $g_\tau(x)$ equals $g(x, \tau)$.

Practice problem

- Under these assumptions, show that the conditional probability $P(Y \leq g(X, \tau)|Z = z)$ equals τ.
- Propose an estimator for $g(\cdot, \tau)$.

Solution

- Conditional probability:

$$\begin{aligned} P(Y \leq g(X,\tau)|Z=z) &= P(g(X,U) \leq g(X,\tau)|Z=z) \\ &= P(U \leq \tau|Z=z) \\ &= P(U \leq \tau) = \tau \end{aligned}$$

- This implies

$$g(\cdot,\tau) \in \underset{g(\cdot)}{\text{argmin}}\ E\left[(E[\mathbf{1}(Y \leq g(X))|Z] - \tau)^2\right].$$

- This suggests a series minimum distance estimator:

$$\widehat{g}(\cdot) = \underset{g:g(x)=\sum \beta_j \phi_j(x)}{\text{argmin}} \sum_i \left(\widehat{E}[\mathbf{1}(Y \leq g(X))|Z=Z_i] - \tau\right)^2,$$

with $\widehat{E}$ given in turn by series regression.

One-dimensional hetereogeneity is crucial for conditional quantile restriction

- Consider the following example where heterogeneity U is multidimensional:

$$\begin{aligned} Y &= U_1 X + U_2 \\ X &= Z + V \\ (U_1, U_2, V) &\sim N(0, \Sigma) \end{aligned}$$

- Without proof: In this case, for generic Σ,

$$P(Y \leq g_\tau(X) | Z = z) \neq \tau,$$

where g_τ is the quantile structural function.

Approach II: Control functions

- Consider now the alternative model

$$
\begin{aligned}
Y &= g(X, U) \\
X &= h(Z, V) \\
Z &\perp (U, V)
\end{aligned}
$$

- where $\dim(V) = 1$ and h is strictly monotonic in V.
- We can assume w.l.o.g. $V \sim Uniform([0, 1])$.

Practice problem

- Write V as a function of X and Z.
- Show that
$$X \perp U | V.$$
- Derive an expression for $E[Y|X, V]$.
- Write the average structural function (ASF) $E[g(x, U)]$ in terms of observable distributions.
- Propose an estimator for the ASF.

Solution

- V as a function of X and Z: Let $x = h(z, v)$. Then

$$\begin{aligned} F_{X|Z}(x|z) &= P(h(Z, V) \leq x | Z = z) \\ &= P(h(z, V) \leq h(z, v)) \\ &= P(V \leq v) = v, \end{aligned}$$

  and thus $V = F_{X|Z}(X|Z)$.
- Conditional independence: Write $X \perp U | V$ as

$$h(Z, V) \perp U | V = v,$$

  which follows immediately from $Z \perp (U, V)$.

Solution continued

- Conditional expectation:

$$\begin{aligned} E[Y|X=x,V=v] &= E[g(x,U)|X=x,V=v] \\ &= E[g(x,U)|V=v] \end{aligned}$$

- Since $V \sim Uniform([0,1])$ by assumption, the law of iterated expectations gives

$$E[g(x,U)] = E[E[g(x,U)|V]] = \int_0^1 E[Y|X=x,V=v]dv.$$

Possible estimator

- Estimate $F_{X|Z}$ using kernel regression:

$$\widehat{F}_{X|Z}(x|z) = \sum_i K(Z_i - z)\mathbf{1}(X_i \leq x) \Big/ \sum_i K(Z_i - z)$$

  for some kernel function K.

- Impute V_i:

$$\widehat{V}_i = \widehat{F}_{X|Z}(X_i|Z_i).$$

- Flexibly regress Y_i on X_i and $\widehat{V}_i$.
- Integrate predicted values for x, v over uniform distribution for v.

One-dimensional hetereogeneity in the first stage is crucial for control function

- Consider the following example where heterogeneity V is multidimensional:

$$Y = X + U$$
$$X = V_1 Z + V_2$$
$$(U, V_1, V_2) \sim N(\mu, \Sigma)$$

- Average structural function:

$$g(x) = E[x + U] = x.$$

- Control function $\tilde{V} = F_{X|Z}(X|Z)$.
- Conditional independence $U \perp X|\tilde{V}$ is violated, since $U \perp Z|\tilde{V}$ does not hold:

$$E[U|Z, \tilde{V}] = \mu_U + \Phi^{-1}(\tilde{V}) \frac{\Sigma_{V_2,U} + Z\Sigma_{V_q,U}}{\sqrt{\Sigma_{V_2,V_2} + 2Z\Sigma_{V_1,V_2} + Z^2\Sigma_{V_1,V_1}}}$$

Approach III: Continuous LATE

- Consider the model without restrictions on heterogeneity:

$$
\begin{aligned}
Y &= g(X, U) \\
X &= h(Z, V) \\
Z &\perp (U, V)
\end{aligned}
$$

- Assume first that $X \in \mathbb{R}$, $Z \in \{0, 1\}$.
- Potential outcome notation:

$$X^z = h(z, V).$$

- Assume $X^0 \leq X^1$ (for non-negative weights).

LATE for binary instrument

- Linear IV slope: As in part I of class,

$$\beta := \frac{\mathrm{Cov}(Z,Y)}{\mathrm{Cov}(Z,X)} = \frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}.$$

- Denominator:

$$E[X|Z=1]-E[X|Z=0] = E[X^1 - X^0].$$

- Numerator:

$$\begin{aligned} E[Y|Z=1]-E[Y|Z=0] &= E[g(X^1,U)-g(X^0,U)] \\ &= E\left[\int_{X^0}^{X^1} g'(x,U)dx\right] \\ &= \int_{-\infty}^{\infty} E[g'(x,U)\mathbf{1}(X^0 \le x \le X^1)]dx \end{aligned}$$

- Taking rations yields:

$$\beta = \int_{-\infty}^{\infty} E[g'(x,U)\cdot\omega]dx$$

where

$$\omega = \frac{\mathbf{1}(X^0 \leq x \leq X^1)}{\int_{-\infty}^{\infty} E[\mathbf{1}(X^0 \leq x \leq X^1)dx}.$$

- $\Rightarrow$ Linear IV gives a weighted average of the slopes (marginal causal effects) $g'(x,U)$.

General instrument

- Now drop restriction that $Z \in \{0, 1\}$, but assume that $X \geq 0$.
- Then

$$\begin{aligned} Y &= g(h(Z,V),U) \\ &= g(0,U) + \int_0^\infty g'(x,U)\mathbf{1}(x \leq h(Z,V))dx. \end{aligned}$$

- Thus

$$\begin{aligned} \mathrm{Cov}(Z,Y) &= E\left[(Z - E[Z]) \cdot \int_0^\infty g'(x,U)\mathbf{1}(x \leq h(Z,V))dx\right] \\ &= \int_0^\infty E[g'(x,U) \cdot \varpi]dx \end{aligned}$$

where

$$\varpi(x) = E[\mathbf{1}(x \leq h(Z,V)) \cdot (Z - E[Z])|V].$$

- If h is increasing in Z, then $\varpi \geq 0$.
- Taking ratios as before yields

$$\beta = \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,X)} = \int_0^\infty E[g'(x,U)\cdot\omega]dx$$

where

$$\omega = \frac{\varpi(x)}{\int_0^\infty E[\varpi(x)]dx}.$$

- As before, linear IV is a weighted average of marginal causal effects $g'(x,U)$.

References

- Nonparametric IV:

Newey, W. K. and Powell, J. L. (2003). Instrumental Variable Estimation of Nonparametric Models. Econometrica, *71(5):1565–1578.*

Horowitz, J. L. (2011). Applied Nonparametric Instrumental Variables Estimation. Econometrica, *79(2):347–394.*

Chernozhukov, V., Imbens, G. W., and Newey, W. K. (2007). Instrumental variable estimation of nonseparable models. Journal of Econometrics, *139(1):4–14.*

Hahn, J. and Ridder, G. (2011). Conditional moment restrictions and triangular simultaneous equations. The Review of Economics and Statistics, *93(2):683–689.*

- Control functions:

Imbens, G. W. and Newey, W. K. (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. Econometrica, *77:1481–1512.*

Kasy, M. (2011). Identification in triangular systems using control functions. Econometric Theory, *27(03):663–671.*

- Continuous LATE:

Angrist, J. D., Graddy, K., and Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. The Review of Economic Studies, *67(3):499–527.*