ExtrapoLATE-ing: External Validity and Overidentification in the LATE framework

Josh Angrist — MIT
Iván Fernández-Val — BU

March 2011

Instrumental Variables and External Validity

- Local Average Treatment Effects (LATE) capture the causal effect of an instrument-induced shift in treatment
- Each LATE is necessarily tied to the instrument that generates it
- Is this surprising or troubling? Could it be otherwise?
  - Draft lottery instruments tell us the effect of being drafted, not necessarily volunteering
  - Compulsory schooling instruments reflect the effect of high school graduation, not an MBA
- The LATE framework highlights questions of external validity
  - Can one instrument identify the average effect induced by another source of variation?
  - Can we go from average effects on compliers to average effects on the entire treated population or an unconditional effect?
- The answer to these questions is usually "no," at least, not without additional assumptions ...

Routes to External Validity

1. Constant effects and distributional assumptions buy you the world
2. Semi/Non-parametric alternatives
   - Try to exploit latent-index selection w/o normality (e.g., Heckman, Tobias, and Vytlacil, 2003; Angrist, 2004)
   - Non-parametric ID using "super-instruments" (Chamberlain 1986, Heckman 1990; Angrist-Imbens 1991)
   - Bayesian functional-form-based extrapolation (Chamberlain, 2010)
3. Multiple instruments give empirical leverage
   - Ichino and Winter-Ebmer (1999) and Oreopoulos (2006) compare IV estimates of the returns to schooling; PO concludes that LATE is close to the effect on non-treated
   - Angrist, Imbens, Graddy (2000) use multiple IVs to assess homogeneity and linearity of demand
   - Ebenstein (2009) compares LATE estimates of fertility effects for Taiwan and the US, argues the former are close to ATE
   - Angrist, Lavy and Schlosser (2010) compare alternative IV estimates of the quantity-quality trade-off; all effects are zero

The Covariate Approach to External Validity

- LATE is the average causal effect on compliers: we can't *name* them but we can still get to know them
- The distribution of complier characteristics is identified and easily estimated (Abadie 2003; An grist and Pischke, 2009)
- We assume here that differences in complier characteristics are what makes a given LATE estimate special; this is *Conditional Effect Ignorability* (CEI)
  - Given CEI, we might explain the difference in LATE across instruments
  - Given CEI, we might explain the difference between LATE and ATE or effects on the treated and nontreated
- CEI can be used for estimation
  - Covariate-specific LATEs can be weighted to produce other effects, such as the effect on the treated, or hypothetical LATE with a different complier population

Overidentification in the Covariate Approach

- Overid tests compare alternative IV estimates and are said to reject exclusion restrictions when the estimates differ
  - Traditional overid testing is a check on internal validity
  - In the LATE framework, by contrast, two internally valid instruments need not estimate the same thing
- We use covariate-specific overidentification tests to see if complier characteristics explain differences in unconditional effects
  - In other words, we use overid stats to test external validity
- In practice, alternative IV estimates of covariate-specific effects might match for some covariate values only
  - We use overid stats to isolate covariate-defined subsamples for which alternative IV estimates seem to match
- Application: kids and labor supply (Angrist and Evans, 1998)
  - Two good instruments, *twins* and *samesex;* very different IV estimates

Framework

- Potential outcomes, Y_0 and Y_1, describe what happens under alternative assignments of treatment, D
- The observed outcome, Y, is

$$Y = Y_0 + (Y_1 - Y_0)D$$

- Potential treatment status is D_1^z when the instrument, Z, is switched on and D_0^z when Z is switched off
- Observed treatment status is

$$D = D_0^z + (D_1^z - D_0^z)Z$$

- Note the instrument-specific notation for potential treatment status
- Covariates are denoted by the vector X

LATE Theorem: Imbens and Angrist (1994)

Assume:

(a) Independence and Exclusion: $(\mathrm{Y}_1, \mathrm{Y}_0, \mathrm{D}_1^z, \mathrm{D}_0^z) \amalg \mathrm{Z}$

(b) First-stage: $\mathbb{E}[\mathrm{D}_1^z - \mathrm{D}_0^z] \neq 0$ *and* $0 < \mathbb{P}[\mathrm{Z} = 1] < 1$

(c) Monotonicity: $\mathrm{D}_1^z \geq \mathrm{D}_0^z$ *a.s., or vice versa*

Then,

$$\frac{\mathbb{E}[\mathrm{Y} \mid \mathrm{Z} = 1] - \mathbb{E}[\mathrm{Y} \mid \mathrm{Z} = 0]}{\mathbb{E}[\mathrm{D} \mid \mathrm{Z} = 1] - \mathbb{E}[\mathrm{D} \mid \mathrm{Z} = 0]} = \mathbb{E}[\mathrm{Y}_1 - \mathrm{Y}_0 \mid \mathrm{D}_1^z > \mathrm{D}_0^z] := \Delta^z$$

- IV using Z as an instrument for the effect of D on Y with no covariates is the (1940) Wald estimator
- Three instrument-defined subgroups:
  - (a) Z-*compliers*; $\mathrm{D}_1^z = 1$ and $\mathrm{D}_0^z = 0$
  - (b) Z-*always-takers*; $\mathrm{D}_1^z = \mathrm{D}_0^z = 1$
  - (c) Z-*never-takers*; $\mathrm{D}_1^z = \mathrm{D}_0^z = 0$
- LATE is the effect of treatment on Z-compliers

Childbearing Effects on Labor Supply

- The potential for omitted variable bias in the relationship between fertility and labor supply seems clear
  - Mothers with weak labor force attachment or low earnings potential may be more likely to have children
- Angrist and Evans (1998) use two IVs to solve this problem
  1. *Twins*: a dummy for a multiple second birth in a sample of mothers with at least two children
  2. *Samesex*: a dummy for same-sex sibling pairs (American parents want a diversified sibling-sex portfolio)
- Wald estimates are reported in **Table 1**
- *Twins* and *samesex* IV estimates both suggest that the birth of a third child has a substantial effect on weeks worked and employment
- But *Twins* IV estimates are about half as large as those using *samesex* . . . Can this be explained by differences in complier characteristics?

Complier Characteristics

- How do *twins* and *samesex* compliers differ?
- Suppose X is an indicator for college graduates. Then,

$$\frac{\mathbb{P}[\mathrm{X}=1 \mid \mathrm{D}_1^z > \mathrm{D}_0^z]}{\mathbb{P}[\mathrm{X}=1]} = \frac{\mathbb{E}[\mathrm{D} \mid \mathrm{Z}=1, \mathrm{X}=1] - \mathbb{E}[\mathrm{D} \mid \mathrm{Z}=0, \mathrm{X}=1]}{\mathbb{E}[\mathrm{D} \mid \mathrm{Z}=1] - \mathbb{E}[\mathrm{D} \mid \mathrm{Z}=0]}$$

- Relative likelihood a complier is college grad is given by the ratio of the first stage for college graduates to the overall first stage
- For general X, the mean of covariates for compliers can be obtained using Abadie's (2003) kappa-weighting scheme:

$$\mathbb{E}[\mathrm{X} \mid \mathrm{D}_1 > \mathrm{D}_0] = \frac{\mathbb{E}[\kappa(\mathrm{Z})\mathrm{X}]}{\mathbb{E}[\kappa(\mathrm{Z})]},$$

where

$$\kappa(\mathrm{Z}) = 1 - \frac{\mathrm{D}(1-\mathrm{Z})}{1-\mathbb{P}[\mathrm{Z}=1]} - \frac{(1-\mathrm{D})\mathrm{Z}}{\mathbb{P}[\mathrm{Z}=1]}$$

- (We can also look at the **distribution of potential outcomes**)

Complier Characteristics (cont.)

- Panel A of **Table 2** reports compliers' characteristics ratios for schooling and age dummies; Kappa weighting is illustrated in Panel B
- Both panels show that *twins* compliers are:
  - More likely than *samesex* compliers to have a young second-born (women w/very young second borns can only have a 3rd via multiples)
  - More educated than *samesex* compliers
- *Samesex* compliers have/are:
  - Older than average second-borns
  - Less educated than average (less likely to be college graduates)
- Age of second born and education are important mediators of the causal effects of childbearing
  - Women with a young second-born at home should be less affected by the birth of a third child
  - Educated women adjust labor supply less when family size increases (they pay for child care)

Conditional Effect Ignorability

Covariates play two roles in our setup:

1. Supporting identification (conditional ignorability of IV)
2. A possible basis for extrapolation

Assumption (Conditional LATE assumptions)

(a) Independence and Exclusion: $(\text{Y}_1, \text{Y}_0, \text{D}_1^z, \text{D}_0^z) \perp\!\!\!\perp \text{Z} \mid \text{X}$ *a.s.*
(b) First stage: $\mathbb{E}[\text{D}_1^z - \text{D}_0^z \mid \text{X}] \neq 0$ *and* $0 < \mathbb{P}[\text{Z} = 1 \mid \text{X}] < 1$ *a.s.*
(c) Monotonicity: $\mathbb{P}[\text{D}_1^z \geq \text{D}_0^z \mid \text{X}] = 1$ *or* $\mathbb{P}[\text{D}_1^z \leq \text{D}_0^z \mid \text{X}] = 1$ *a.s.*

Assumption (Conditional Effect Ignorability for an instrument Z)

$$\mathbb{E}[\text{Y}_1 - \text{Y}_0 \mid \text{D}_1^z, \text{D}_0^z, \text{X}] = \mathbb{E}[\text{Y}_1 - \text{Y}_0 \mid \text{X}] \quad a.s.$$

- A sufficient condition for CEI is

$$\text{Y}_1 = \text{Y}_0 + g(\text{X}) + \nu,$$

where ν is mean-independent of $(\text{D}_1^z, \text{D}_0^z)$ conditional on X

CEI in a latent-index model

- $D = 1[h(X, Z) > \eta]$; errors indep. of instruments and covs
- Potential treatments:

$$D_0^z = 1[h(X, 0) > \eta], \quad D_1^z = 1[h(X, 1) > \eta]$$

- Potential outcomes:

$$\begin{aligned} Y_0 &= g_0(X) + \epsilon_0 \\ Y_1 &= g_1(X) + \epsilon_1 \end{aligned}$$

- Assuming $h(X, 1) \geq h(X, 0)$ a.s., conditional LATE can be written

$$\begin{aligned} \Delta^z(X) &= \mathbb{E}[Y_1 - Y_0 \mid h(X, 1) > \eta > h(X, 0), X] \\ &= g_1(X) - g_0(X) + \mathbb{E}[\epsilon_1 - \epsilon_0 \mid h(X, 1) > \eta > h(X, 0), X] \end{aligned}$$

- The CEI says η and $\epsilon_1 - \epsilon_0$ are independent given X, so

$$\Delta^z(X) = \mathbb{E}[Y_1 - Y_0 \mid X] = g_1(X) - g_0(X)$$

- No Roy Selection!

CEI: A Structural Story

- Combine Olsen (1980) and Vytlacil (2002):

$$\mathbb{E}[\epsilon_1 - \epsilon_0 \mid \eta, \mathrm{x}] = \rho(\mathrm{x})\eta \quad \text{and} \quad \eta \mid \mathrm{x}, \mathrm{z} \sim U(0,1)$$

- Write conditional LATE as

$$\begin{aligned} \Delta^z(\mathrm{x}) &= g_1(\mathrm{x}) - g_0(\mathrm{x}) + \rho(\mathrm{x})\mathbb{E}[\eta \mid h(\mathrm{x},1) > \eta > h(\mathrm{x},0), \mathrm{x}] \\ &= g_1(\mathrm{x}) - g_0(\mathrm{x}) + \rho(\mathrm{x})[h(\mathrm{x},1) + h(\mathrm{x},0)]/2 \end{aligned}$$

  For each x, CEI turns on $\rho(\mathrm{x})$
- Treatment decision based on predicted benefits minus costs of childbearing (Imbens and Newey, 2009)

$$1[h(\mathrm{x},\mathrm{z}) > \eta] = 1\{\lambda(\mathrm{x})\mathbb{E}[\mathrm{Y}_1 - \mathrm{Y}_0 \mid \mathrm{x}, \eta] > c(\mathrm{x},\mathrm{z})\}$$

  where $\lambda(\mathrm{x})$ is the weight given to outcome gaps, $c(\mathrm{x},\mathrm{z})$ is the expected cost of having a third child, and η is private information
- $\rho(\mathrm{x})$ is smaller when:
  1. $\lambda(\mathrm{x})$ is small (labor supply consequences are of little import);
  2. η matters little given x (e.g., for women with a young second-born)

LATE-Reweight Theorem

Covariate-specific LATE using Z,

$$\Delta^z(x) := \mathbb{E}[\mathrm{Y}_1 - \mathrm{Y}_0 \mid \mathrm{D}_1^z > \mathrm{D}_0^z, \mathrm{X} = x] \tag{1}$$

Theorem (Reweighting LATE)

Let Z *be an instrument that satisfies CEI and cond. LATE identification and let* $\mathrm{S}^z = \mathrm{S}(\mathrm{D}_0^z, \mathrm{D}_1^z, \mathrm{Z})$ *be an indicator for any subpopulation defined by* Z*. For example, for* Z*-compliers* $\mathrm{S}^z = \mathrm{D}_1^z - \mathrm{D}_0^z$*, for the treated* $\mathrm{S}^z = (1 - \mathrm{Z})\mathrm{D}_0^z + \mathrm{Z}\mathrm{D}_1^z = \mathrm{D}$*, and for the population* $\mathrm{S}^z = 1$*.*
Assuming $\mathbb{E}[|\mathrm{Y}|] < \infty$*,*

$$\mathbb{E}[\mathrm{Y}_1 - \mathrm{Y}_0 \mid \mathrm{S}^z = 1] = \mathbb{E}[\Delta^z(\mathrm{X}) \mid \mathrm{S}^z = 1] = \int \Delta^z(x)\omega_{\mathrm{S}}^z(x) dF_{\mathrm{X}}(x),$$

where

$$\omega_{\mathrm{S}}^z(x) = \mathbb{P}[\mathrm{S}^z = 1 \mid \mathrm{X} = x] / \mathbb{P}[\mathrm{S}^z = 1]$$

and $\int \omega_{\mathrm{S}}^z(x) dF_{\mathrm{X}}(x) = 1$*.*

Reweighting LATE

- This result allows us to construct Δ^z, TOT, TNT, ATE using weights

$$\omega^z_\Delta(x) = \frac{\mathbb{E}[\mathrm{D} \mid \mathrm{Z} = 1, \mathrm{X} = x] - \mathbb{E}[\mathrm{D} \mid \mathrm{Z} = 0, \mathrm{X} = x]}{\mathbb{E}[\mathrm{D} \mid \mathrm{Z} = 1] - \mathbb{E}[\mathrm{D} \mid \mathrm{Z} = 0]}$$

  for Z-compliers;

$$\omega^z_{TOT}(x) = \mathbb{E}[\mathrm{D} \mid \mathrm{X} = x]/\mathbb{E}[\mathrm{D}]$$

  for the treated;

$$\omega^z_{TNT}(x) = \mathbb{E}[1 - \mathrm{D} \mid \mathrm{X} = x]/\mathbb{E}[1 - \mathrm{D}]$$

  for the non-treated; and

$$\omega^z_{ATE}(x) = 1$$

  for the population

From One LATE to Another

- Let $\Delta^w(\mathrm{X})$ be defined as in (**??**) using the instrument W. Then,

$$\begin{aligned}\Delta^z - \Delta^w &= \int [\Delta^z(x) - \Delta^w(x)]\omega_\Delta^z(x) dF_{\mathrm{X}}(x) \\ &\quad + \int \Delta^w(x)[\omega_\Delta^z(x) - \omega_\Delta^w(x)] dF_{\mathrm{X}}(x)\end{aligned}$$

- If Z and W satisfy CEI,

$$\Delta^w(\mathrm{X}) = \Delta^z(\mathrm{X})\text{a.s.}$$

We can then reweight to go from one LATE to another:

Theorem (OLTA)

Let $\mathrm{S}^w = \mathrm{S}(\mathrm{D}_0^w, \mathrm{D}_1^w, \mathrm{W})$ *be an indicator for any subpopulation defined by instrument* W. *If* $\mathbb{E}[|\mathrm{Y}|] < \infty$,

$$\mathbb{E}[\mathrm{Y}_1 - \mathrm{Y}_0 \mid \mathrm{S}^w = 1] = \int \Delta^w(x)\omega_{\mathrm{S}}^w(x) dF_{\mathrm{X}}(x) = \int \Delta^z(x)\omega_{\mathrm{S}}^w(x) dF_{\mathrm{X}}(x)$$

Compatible Average Treatment Effects (CATE)

- In practice, CEI may be satisfied only for some covariates
- Empirically, this means a good match across instruments for some covariate values; for others, not so good
- Estimates for a population defined by *compatible* covariate-specific effects may have high predictive value . . . for the compatible
- CATE is

$$\Delta_c^{z,w} = \int \Delta^z(x) dF_{\mathrm{X}}(x \mid \Delta^z(x) = \Delta^w(x))$$

- If CEI holds for Z and W for all values of X, CATE is ATE; Otherwise, CATE is ATE for the *compatible subpopulation*
- Rationales for CATE
    - We'd like to report estimates for the largest and most representative subpopulation possible; that might be TOT|CATE
    - We'd like to make predictions for each covariate value, in practice we'll borrow info from compatible instruments and average to get something more precise

Estimation and Inference for Reweighting Estimators

Assumption (Sampling)

$\{\mathrm{R}_i = (\mathrm{Y}_i, \mathrm{D}_i, \mathrm{X}_i, \mathrm{Z}_i, \mathrm{W}_i); i = 1, \ldots, n\}$ are i.i.d. observations from the vector $\mathrm{R} = (\mathrm{Y}, \mathrm{D}, \mathrm{X}, \mathrm{Z}, \mathrm{W})$

Assumption (Discrete covariates)

For a finite set $\mathcal{X}$, $\mathbb{P}[\mathrm{X} \in \mathcal{X}] = 1$

- Random sampling is standard for cross sectional data
- Education and age covariates in the empirical example are discrete
- Generalization to continuous covariates is straightforward under CEI, but more complicated for CATE, so we leave this for another day

The Plug-in Principle

- Our reweighting estimands other than CATE have the form,

$$\theta_{\mathrm{S}} = \mathbb{E}[\Delta^z(\mathrm{X})\omega^j_{\mathrm{S}}(\mathrm{X})],\ \omega^j_{\mathrm{S}}(x) = \mathbb{P}[\mathrm{S}^j = 1 \mid \mathrm{X} = x]/\mathbb{P}[\mathrm{S}^j = 1],$$

  for $j \in \{z, w\}$
- Replace expectations $\mathbb{E}$ and probabilities $\mathbb{P}$ by empirical analogs, $\mathbb{E}_n$ and $\mathbb{P}_n$:

$$\hat{\theta}_{\mathrm{S}} = \mathbb{E}_n[\hat{\Delta}^z(\mathrm{X}_i)\hat{\omega}^j_{\mathrm{S}}(\mathrm{X}_i)],\ \hat{\omega}^j_{\mathrm{S}}(x) = \mathbb{P}_n[\mathrm{S}^j_i = 1 \mid \mathrm{X} = x]/\mathbb{P}_n[\mathrm{S}^j_i = 1]$$

- $\hat{\Delta}^z(x)$ can be Wald with instrument Z in cell x, or 2SLS (or GMM) using both Z and W as instruments
- For LATE, S^j_i is not observable, so we use

$$\hat{\omega}^j_{\Delta}(x) = \frac{\mathbb{E}_n[\mathrm{D}_i \mid \mathrm{J} = 1, \mathrm{X} = x] - \mathbb{E}_n[\mathrm{D}_i \mid \mathrm{J} = 0, \mathrm{X} = x]}{\mathbb{E}_n[\mathrm{D}_i \mid \mathrm{J} = 1] - \mathbb{E}_n[\mathrm{D}_i \mid \mathrm{J} = 0]},\ \mathrm{J} \in \{\mathrm{Z}, \mathrm{W}\}$$

- Consistency of $\hat{\theta}_{\mathrm{S}}$ follows from LLN and the Slutsky theorem

Inference under CEI

- Estimators are smooth functions of IV/GMM estimators and therefore asymptotically normal under standard regularity conditions
- Two routes to inference: delta method to derive the asymptotic variance, or resampling methods
- The bootstrap saves us from having to evaluate complicated analytical formulas
- Consistency of bootstrap for reweighting estimators follows from Hall and Horowitz (1996), Hahn (1996), Brown and Newey (2002) theorems for GMM, and delta method for bootstrap (see, e.g., Theorem 23.5 in van der Vaart, 1998)
- How to bootstrap in practice? We go with the empirical likelihood (EL) bootstrap as described in Brown and Newey (1992)
  - EL bootstrap works well for moment restriction models because it uses an efficient estimator of the DGP

EL Bootstrap for Reweighting Estimators

- EL bootstrap resamples from the empirical likelihood distribution (ELD) that imposes CEI (compatibility) in all cells, instead of the empirical distribution
- Let the compatibility conditions be $\mathbb{E}[g(\mathrm{R}, \theta)] = 0$. The ELD $(\hat{\pi}_1, ..., \hat{\pi}_n)$ is the solution to

$$\max_{\pi_1,\ldots,\pi_n} \sum_{i=1}^{n} \ln(\pi_i), \text{ s.t. } \sum_{i=1}^{n} \pi_i g(\mathrm{R}_i, \hat{\theta}) = 0, \sum_{i=1}^{n} \pi_i = 1, \pi_i \geq 0,$$

  where $\hat{\theta}$ is the EL or another consistent estimator of θ
- The EL bootstrap resamples from the data with probabilities $\hat{\pi}_i$ instead of $1/n$

Theorem (Bootstrap consistency)

Let Z and W be two instruments that satisfy cond. LATE and CEI Assumptions. If $E[\mathrm{Y}^2] < \infty$, the empirical likelihood bootstrap is consistent for the asymptotic distribution of $\hat{\theta}_{\mathrm{S}}$

But Weight . . . estimation and inference for CATE

- We want to condition on $\{x : \Delta^z(x) = \Delta^w(x)\}$, but in finite samples we never have $\hat{\Delta}^z(x) = \hat{\Delta}^w(x)$
- We can use covariate-specific overidentification tests to find compatible values of x
  - The cov-specific overid test, $J(x)$, measures $\|\hat{\Delta}^z(x) - \hat{\Delta}^w(x)\|$
  - Under the null, $J(x)$ is $\chi^2(1) = N(0,1)^2$
  - So we weight cell estimates as follows

$$\hat{\Delta}_c^{z,w} = \mathbb{E}_n[\hat{\Delta}^{z,w}(\mathrm{x}_i)\hat{\omega}_c(\mathrm{x}_i)];\ \hat{\omega}_c(x) \propto \exp\left[-\frac{J(x)}{a_n(x)}\right],$$

  where $\hat{\Delta}^{z,w}(x)$ is a covariate-specific GMM estimator using z and w, and $a_n(x) \to \infty$ and $a_n(x) = o(n)$; the application sets $a_n(x) = \frac{\ln n_x}{k}$ (n_x = cell size; k = number of cells)

- $\hat{\Delta}_c^{z,w}$ is consistent for CATE because: (a) $plim \frac{J(x)}{\ln n_x} = \infty$ where CEI fails; (b) $plim \frac{J(x)}{\ln n_x} = 0$ otherwise
- We can view the weights as a moment selection device in the spirit of Andrews (1999)

CATE Inference Issues

- Data-driven selection of an estimator is problematic (like pre-testing); see, e.g., Leeb and Potscher (2009)
- Convergence of finite-sample distribution to the limiting distribution is non-uniform in the DGP and potentially slow
- This is most relevant when compatibility is not empirically clear-cut (like having a marginally weak instrument, or near-unit root)
- EL versus an unrestricted nonparametric bootstrap
  - Virtue of EL is potentially a vice for CATE: EL imposes compatibility but CATE allows CEI to fail for some cells
  - We therefore check EL against a nonparametric bootstrap that does not impose CEI
- Subsampling may improve on the bootstrap in pretesting scenarios (see, e.g., Andrews and Guggenberger, 2009)
  - Our application may not be well-suited to subsampling because some cells are small

Estimates

- Our empirical work focuses on a 12-cell representation of second child age and mother's education
  - Three 2nd-born age groups: ≤ 4, $(4, 8]$, and > 8 (all mothers 2+)
  - Four schooling groups: HS dropout, HS graduate, some college, and college graduate
- These covariates are strongly related to compliance probabilities
  - Multiple births have larger first-stage effects on women with a young second-born and for women with some college and college degrees
  - Sex-composition compliers are relatively unlikely to be college graduates and have older second-borns
- Labor supply effects are also likely to vary with these covariates
  - The birth of a third child has little effect on the work behavior of a woman with a young second-born if she is at home anyway
  - Relatively educated women with high wages should be affected less by the birth of a child than less-educated women

Estimates (cont.)

- **Table 3** reports the conditional estimates $\hat{\Delta}^z(x)$ and $\hat{\Delta}^{z,w}(x)$, and the weighting functions $\hat{\omega}^z_{\mathrm{s}}(x)$ and $\hat{\omega}^z_{CATE}(x)$ for each cell
    - Women with a third child differ from those without
    - As we saw in table 2, *Twins* and *samesex* compliers have different characteristics and differ from the random sample
    - The overid rejects CEI in only two cells for weeks worked (only one for employment), though covariate-specific IV estimates are fairly noisy
- **Table 4** compares IV, GMM, and LATE-reweighted estimates
    - Reweighting covariate-specific *twins* estimates using *samesex* weights - and vice versa - brings the estimates together
    - ATE, TOT, and to a lesser extent TNT are poorly matched (twins LATE *is* TNT)
    - Downweighting a few bad cells, **CATE matches well** (compare -3.80 and -3.66 for weeks and $-.099$ and $-.095$ for employment)

The Compatible Subpopulation

- For given over-identified model, the CATE weighting function (and subpopulation) potentially depend on the dependent variable
- **Figure 1** describes the compatible subpopulation for two versions of CATE, the first using weeks worked, the second using employment
- The first version shows $\hat{\omega}_c(x)$; the second plots $\hat{\omega}_c(x)dF_{\mathrm{X}}(x)$
- Relative to the covariate histogram, *CATE* subpopulations emphasize:
  - Women with young second-borns
  - More educated mothers among those with older second-borns
- CATE weights are something like those for LATE(twins) and hence the population of non-treated
  - (not quite the same since TNT did not match all that well, but both emphasize women with younger second-borns and more educated)
- We knew that twins identify TNT, but this result shows samesex results - appropriately reweighted - are almost equally general

Summary and Directions for Further Work

- We develop a covariate-based approach to external validity for IV estimates
- By assuming differences in complier characteristics are what make IV estimates special, we can construct estimates for new subpopulations from covariate-specific LATEs
- When CEI fails for some cells, as seems likely in practice, we use the traditional over-identification test statistic to find a population for which a given pair of IV estimates are compatible
- CATE achieves external validity by favoring cells with complier-independent effects (external validity is for these cells only: no free lunch)
- Implicit pretesting makes inference for compatible effects econometrically challenging
- The development of robust and convenient inference procedures for CATE is a natural direction for further work

Tables and Figures

Table 1: Wald estimates of the effects of family size on labor supply

Dependent Variable	Mean	OLS	Twins First Stage	Twins Wald Estimates	Samesex First Stage	Samesex Wald Estimates	Both 2SLS Estimates
		(1)	(2)	(3)	(4)	(5)	(6)
Weeks Worked	20.83	-8.98	0.603	-3.28	0.060	-6.36	-3.97
		(0.072)	(0.008)	(0.634)	(0.002)	(1.18)	(0.558)
	Overid: $\chi^2(1)$ (p-value)	-	-	-	-	-	5.3(.02)
Employment	0.565	-0.176		-0.076		-0.132	-0.088
		(0.002)		(0.014)		(0.026)	(0.012)
	Overid: $\chi^2(1)$ (p-value)	-	-	-	-	-	3.5(.06)

Note: The table reports OLS, Wald, and 2SLS estimates of the effects of a third birth on labor supply using twins and sex composition instruments. Data are from the Angrist and Evans (1998) extract including women aged 21-35 with at least two children in the 1980 census. OLS models include controls for mother's age, age at first birth, ages of the first two children, and dummies for race. The first stage is the same for all dependent variables.

Table 2: Complier characteristics ratios for twins and sex composition instruments

	Mean	Twins		Samesex	
	$E[x_{1i}]$	$E[x_{1i} \mid D_{1i} > D_{0i}]$	$E[x_{1i} \mid D_{1i} > D_{0i}]/E[x_{1i}]$	$E[x_{1i} \mid D_{1i} > D_{0i}]$	$E[x_{1i} \mid D_{1i} > D_{0i}]/E[x_{1i}]$
Variable	(1)	(2)	(3)	(4)	(5)
			A. Bernoulli		
Age of second child less than or equal to 16 quarters	0.343	0.449	1.31	0.194	0.565
High school graduate	0.488	0.498	1.02	0.515	1.06
Some College	0.202	0.212	1.05	0.212	1.05
College graduate	0.132	0.151	1.14	0.092	0.702
			B. Discrete, ordered		
Age of second child	26.37	22.04	.835	28.55	1.083
Mother's schooling	12.13	12.43	1.025	12.09	0.997

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. The values in columns 2 and 4 in Panel B. represent Abadie's (2003) kappa-weighted means. Data are from the 1980 census 5 percent sample including mothers aged 21-35 with at least two children, as in Angrist and Evans (1998). The sample size is 394,840 for all columns.

Table 3: LATE decompositions (n = 394,840)

Covariates		Covariates pmfs			IV estimates and weighting functions						
					Twins		Samesex		Twins & Samesex		
Age	Education	P(X)	P(X|D=1)/P(X)	P(X|D=0)/P(X)	$\Delta^{z}(X)$	$\omega_{\Delta}^{z}(X)$	$\Delta^{w}(X)$	$\omega_{\Delta}^{w}(X)$	$\Delta^{z,w}(X)$	J-pvalue	$\omega_{c}(X)$
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
					A. Weeks worked						
[0, 4]	HS drop	0.06	0.65	1.24	-4.63	1.23	-1.26	0.77	-4.35	0.55	1.45
					(1.84)		(5.49)		(1.60)		
	HS grad	0.15	0.46	1.36	-4.24	1.36	-2.66	0.59	-4.17	0.76	1.99
					(1.15)		(5.00)		(1.04)		
	Some Col	0.06	0.46	1.36	-3.79	1.36	-4.93	0.66	-3.86	0.87	2.14
					(1.69)		(6.86)		(1.59)		
	Col grad	0.05	0.41	1.40	-5.36	1.40	-2.21	0.54	-5.24	0.75	1.95
					(1.99)		(9.77)		(1.84)		
(4, 8]	HS drop	0.07	1.29	0.81	-4.43	0.81	-9.35	1.20	-6.29	0.25	0.47
					(2.65)		(3.36)		(2.06)		
	HS grad	0.17	0.93	1.05	-3.07	1.05	-5.59	1.38	-3.94	0.33	0.79
					(1.52)		(2.09)		(1.23)		
	Some Col	0.07	0.91	1.06	-1.02	1.06	-7.39	1.13	-2.59	0.16	0.21
					(2.22)		(3.91)		(1.93)		
	Col grad	0.04	0.89	1.08	-1.52	1.08	-2.88	0.94	-1.72	0.83	2.09
					(2.63)		(5.97)		(2.32)		
(8+]	HS drop	0.10	1.79	0.47	0.29	0.47	-12.04	0.80	-5.53	0.06	0.04
					(4.47)		(4.74)		(3.24)		
	HS grad	0.17	1.40	0.73	-2.41	0.73	-9.76	1.30	-6.15	0.02	0.01
					(2.22)		(2.25)		(1.60)		
	Some Col	0.06	1.33	0.78	-4.40	0.78	-4.72	1.10	-4.52	0.96	2.20
					(3.55)		(4.54)		(2.82)		
	Col grad	0.02	1.15	0.90	6.78	0.90	17.90	1.12	9.48	0.28	0.44
					(4.99)		(9.01)		(4.17)		

Notes: Standard errors for estimates in parentheses. J-weights are the exponential of minus the overidentification test statistic, normalized to have mean one. The p-value for the joint J-statistic for all the covariate values is 0.25 for weeks and 0.29 for LFP.

Table 3: LATE decompositions (n = 394,840)

Covariates		Covariates pmfs			IV estimates and weighting functions						
					Twins		Samesex		Twins & Samesex		
Age	Education	P(X)	P(X|D=1)/ P(X)	P(X|D=0)/ P(X)	$\Delta^{z}(X)$	$\omega_{\Delta}^{z}(X)$	$\Delta^{w}(X)$	$\omega_{\Delta}^{w}(X)$	$\Delta^{z,w}(X)$	J-pvalue	$\omega_{c}(X)$
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
					B. Employment						
[0, 4]	HS drop	0.06	0.65	1.24	-0.154	1.23	-0.035	0.77	-0.143	0.43	1.15
					(0.048)		(0.143)		(0.043)		
	HS grad	0.15	0.46	1.36	-0.081	1.36	-0.123	0.59	-0.083	0.73	2.14
					(0.027)		(0.117)		(0.026)		
	Some Col	0.06	0.46	1.36	-0.089	1.36	-0.035	0.66	-0.086	0.74	2.15
					(0.038)		(0.157)		(0.037)		
	Col grad	0.05	0.41	1.40	-0.130	1.40	-0.023	0.54	-0.125	0.65	1.90
					(0.047)		(0.231)		(0.046)		
(4, 8]	HS drop	0.07	1.29	0.81	-0.140	0.81	-0.150	1.20	-0.144	0.92	2.42
					(0.064)		(0.082)		(0.051)		
	HS grad	0.17	0.93	1.05	-0.081	1.05	-0.156	1.38	-0.108	0.19	0.39
					(0.034)		(0.046)		(0.028)		
	Some Col	0.07	0.91	1.06	-0.034	1.06	-0.161	1.13	-0.066	0.19	0.34
					(0.048)		(0.084)		(0.042)		
	Col grad	0.04	0.89	1.08	0.075	1.08	-0.202	0.94	0.028	0.06	0.03
					(0.061)		(0.134)		(0.054)		
(8+]	HS drop	0.10	1.79	0.47	0.047	0.47	-0.188	0.80	-0.064	0.11	0.13
					(0.101)		(0.106)		(0.073)		
	HS grad	0.17	1.40	0.73	-0.066	0.73	-0.167	1.30	-0.117	0.13	0.20
					(0.046)		(0.047)		(0.033)		
	Some Col	0.06	1.33	0.78	-0.110	0.78	-0.023	1.10	-0.075	0.47	1.32
					(0.071)		(0.092)		(0.058)		
	Col grad	0.02	1.15	0.90	0.034	0.90	0.224	1.12	0.082	0.33	0.68
					(0.098)		(0.171)		(0.081)		

Notes: Standard errors for estimates in parentheses. J-weights are the exponential of minus the overidentification test statistic, normalized to have mean one. The p-value for the joint J-statistic for all the covariate values is 0.25 for weeks and 0.29 for LFP.

Table 4: Reweighting LATE (n = 394,840)

Effect	Conditional LATE $\Delta(X)$	Weighting function $\omega(X)$	Weeks worked: Estimate [1]	Weeks worked: |t| for diff [2]	Employment: Estimate [3]	Employment: |t| for diff [4]
LATE twins	twins	twins	-3.15		-0.075	
			(0.62)	1.32	(0.014)	0.93
	samesex		-2.71		-0.068	
			(0.81)		(0.018)	
LATE samesex	samesex	samesex	-6.30		-0.131	
			(1.15)	1.44	(0.026)	0.77
	twins		-5.08		-0.115	
			(1.58)		(0.037)	
ATE	twins	1	-2.84		-0.067	
			(0.76)	1.99	(0.017)	1.58
	samesex		-5.88		-0.123	
			(1.35)		(0.031)	
	twins, samesex		-4.36		-0.092	
			(0.60)		(0.014)	

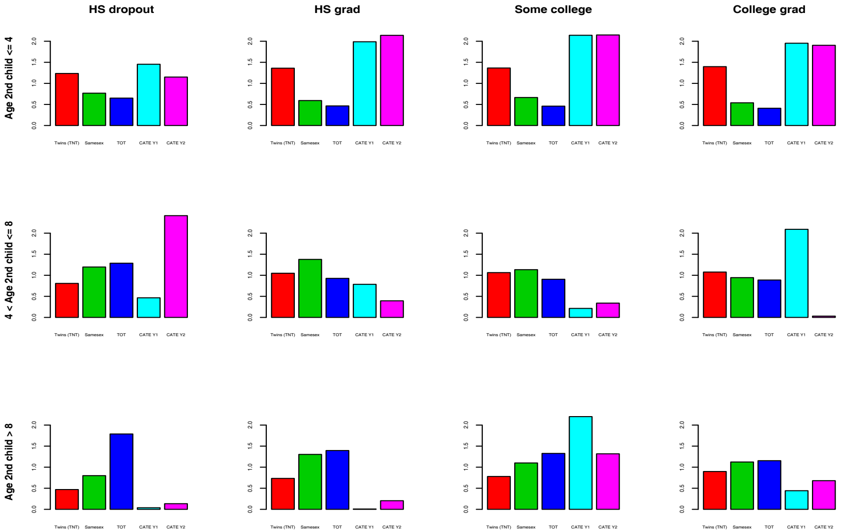
Notes: Standard errors for estimates in parentheses. T-statistics are for the difference between samesex and twins estimates. Standard errors and t-statistics obtained by Brown and Newey (2002) GMM bootstrap with 1,000 repetions. In brackets, we report standard errors and t-statistics obtained by nonparametric bootstrap with 1,000 repetitions.

Table 4: Reweighting LATE (n = 394,840)

Effect	Conditional LATE Δ(X)	Weighting function ω(X)	Weeks worked: Estimate [1]	Weeks worked: |t| for diff [2]	Employment: Estimate [3]	Employment: |t| for diff [4]
TOT	twins	P(X|D=1)/P(X)	-2.38		-0.056	
			(1.07)	2.85	(0.024)	2.14
	samesex		-7.08		-0.136	
			(1.28)		(0.029)	
	twins, samesex		-4.64		-0.092	
			(0.79)		(0.018)	
TNT	twins	P(X|D=0)/P(X)	-3.15		-0.075	
			(0.62)	1.15	(0.014)	1.02
	samesex		-5.08		-0.115	
			(1.58)		(0.037)	
	twins, samesex		-4.18		-0.092	
			(0.53)		(0.012)	
CATE	twins	exp[-12*J(X)/n(X)]	-3.80		-0.099	
			(0.80)	0.18	(0.018)	0.17
			[0.64]	[0.09]	[0.015]	[0.08]
	samesex		-3.66		-0.095	
			(1.01)		(0.023)	
			[0.96]		[0.021]	
	twins, samesex		-4.00		-0.101	
			(0.77)		(0.017)	
			[0.62]		[0.014]	

Notes: Standard errors for estimates in parentheses. T-statistics are for the difference between samesex and twins estimates. Standard errors and t-statistics obtained by Brown and Newey (2002) GMM bootstrap with 1,000 repetions. In brackets, we report standard errors and t-statistics obtained by nonparametric bootstrap with 1,000 repetitions.

Weighting functions: $\hat{\omega}_{\mathrm{S}}(x)$



Weighting functions × Histogram of X: $\hat{\omega}_{\mathrm{S}}(x) \times \mathbb{P}_n(x)$

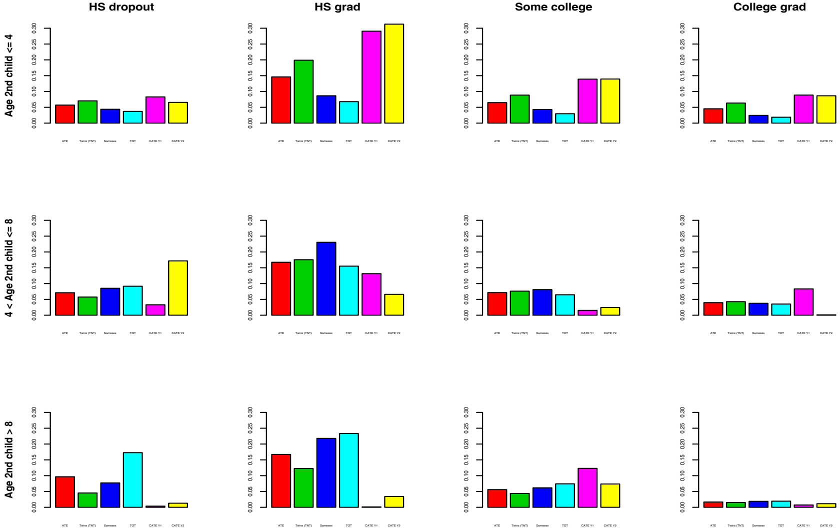


Table 5: Average outcomes by compliance type and covariate value

Covariates		Same sex				Multi second			
		Y_0		Y_1		Y_0		Y_1	
Age	Education	NT	C	C	AT	NT	C	C	AT
		A. Weeks worked							
All	All	23.28	23.20	17.15	17.35	-	23.24	19.96	17.22
[0, 4]	HS drop	14.36	13.91	7.60	7.93	-	14.14	9.51	7.74
	HS grad	18.28	18.16	11.21	10.73	-	18.22	13.98	10.83
	Some Col	20.55	20.42	12.78	12.68	-	20.49	16.69	12.54
	Col grad	19.93	19.79	13.71	13.58	-	19.86	14.50	13.60
(4, 8]	HS drop	19.72	19.16	10.98	11.69	-	19.46	15.03	11.32
	HS grad	24.70	24.25	14.78	14.43	-	24.49	21.43	14.49
	Some Col	27.09	26.77	16.39	16.28	-	26.94	25.92	16.17
	Col grad	25.77	25.16	15.55	15.40	-	25.48	23.96	15.31
(8+]	HS drop	25.72	25.75	17.66	17.94	-	25.73	26.03	17.78
	HS grad	31.58	31.24	22.50	22.92	-	31.43	29.02	22.68
	Some Col	33.67	33.32	24.87	24.61	-	33.51	29.11	24.70
	Col grad	34.39	31.96	25.06	23.42	-	33.27	40.05	24.00
		B. Employment							
All	All	0.61	0.61	0.49	0.49	-	0.61	0.54	0.49
[0, 4]	HS drop	0.45	0.44	0.31	0.31	-	0.45	0.30	0.31
	HS grad	0.51	0.51	0.37	0.37	-	0.51	0.43	0.37
	Some Col	0.57	0.56	0.42	0.41	-	0.56	0.48	0.41
	Col grad	0.57	0.57	0.42	0.44	-	0.57	0.44	0.43
(4, 8]	HS drop	0.56	0.55	0.37	0.38	-	0.55	0.41	0.37
	HS grad	0.64	0.63	0.44	0.43	-	0.63	0.55	0.43
	Some Col	0.69	0.68	0.48	0.47	-	0.69	0.65	0.47
	Col grad	0.70	0.70	0.48	0.48	-	0.70	0.78	0.48
(8+]	HS drop	0.65	0.65	0.50	0.50	-	0.65	0.69	0.50
	HS grad	0.75	0.75	0.59	0.60	-	0.75	0.69	0.60
	Some Col	0.81	0.79	0.65	0.65	-	0.80	0.69	0.65
	Col grad	0.86	0.83	0.68	0.65	-	0.84	0.88	0.66

Data: 1980 Census. NT denotes never takers, C denotes compliers, AT denotes always takers. There are no twins never-takers