Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters

Magne Mogstad	Andres Santos	Alexander Torgovitsky
U. Chicago	UCLA	U. Chicago

November 10, 2017

Basic Question

IV and Heterogeneity

- IV estimand interpretable as LATE (Imbens & Angrist 1994).
- Sometimes LATE has clear policy relevance.
- Other times, different parameters are of interest (external validity).

Basic Question

IV and Heterogeneity

- IV estimand interpretable as LATE (Imbens & Angrist 1994).
- Sometimes LATE has clear policy relevance.
- Other times, different parameters are of interest (external validity).

Our Paper

- Framework for extrapolation in IV model.
- Use insight of marginal treatment effect (Heckman & Vytlacil 2005).
- Allow flexible specifications and computational tractability.

Goal: Allow for different choices of parameters and assumptions.

The Model

Outcome

- Treatment $D \in \{0, 1\}$, potential outcomes (Y_0, Y_1), and actual outcome

$$Y = DY_1 + (1 - D)Y_0.$$

The Model

Outcome

- Treatment $D \in \{0,1\}$, potential outcomes (Y_0, Y_1), and actual outcome

$$Y = DY_1 + (1-D)Y_0.$$

Selection

- For $U \sim U[0,1]$ and (Y_0, Y_1, U) independent of observable instrument Z

$$D = 1\{U \le p(Z)\}$$

The Model

Outcome

- Treatment $D \in \{0, 1\}$, potential outcomes (Y_0, Y_1), and actual outcome

$$Y = DY_1 + (1 - D)Y_0.$$

Selection

- For $U \sim U[0, 1]$ and (Y_0, Y_1, U) independent of observable instrument Z

$$D = 1\{U \leq p(Z)\}$$

Comments

- $U \sim U[0, 1]$ normalization $\Rightarrow p(Z) = P(D = 1|Z)$ (propensity score).
- Instrument monotonicity equivalent to separability (Vytlacil 2002).
- Covariates omitted for simplicity but easily incorporated.

The Model

$$D = 1\{U \leq p(Z)\}$$

$\Rightarrow$ Individuals with smaller unobservable U more likely to receive treatment.

Example: Suppose $Z \in \{0, 1\}$ is binary and that $p(1) > p(0)$, then we have

$$\underbrace{U \in [0, p(0)]}_{\text{always-takers}} \qquad \underbrace{U \in (p(0), p(1)]}_{\text{compliers}} \qquad \underbrace{U \in (p(1), 1]}_{\text{never-takers}}$$

Comments

- Unobservable U determines likelihood of receiving treatment.
- Key Concern: U may not be independent of (Y_0, Y_1) (selection).

The Model

Heckman and Vytlacil (1999, 2005, "HV") define marginal treatment effect

$$\text{MTE}(u) \equiv E[Y_1 - Y_0 | U = u]$$

$\Rightarrow$ summary of unobserved heterogeneity in average treatment effects.

The Model

Heckman and Vytlacil (1999, 2005, "HV") define marginal treatment effect

$$\text{MTE}(u) \equiv E[Y_1 - Y_0 | U = u]$$

$\Rightarrow$ summary of unobserved heterogeneity in average treatment effects.

Key: HV show that many parameters are weighted averages of the MTE

$$\underbrace{\beta^\star}_{\text{parameter}} = E\left[\int_0^1 \text{MTE}(u) \times \underbrace{\omega^\star(u, Z)}_{\text{identified weights}} du\right]$$

The Model

Heckman and Vytlacil (1999, 2005, "HV") define marginal treatment effect

$$\text{MTE}(u) \equiv E[Y_1 - Y_0 | U = u]$$

$\Rightarrow$ summary of unobserved heterogeneity in average treatment effects.

Key: HV show that many parameters are weighted averages of the MTE

$$\underbrace{\beta^\star}_{\text{parameter}} = E[\int_0^1 \text{MTE}(u) \times \underbrace{\omega^\star(u, Z)}_{\text{identified weights}} du]$$

Comments

- $\beta^\star$ may (or may not) be identified.
- Recall $U \sim U[0, 1]$ so that $\beta^\star$ is a weighted average.

Example: LATE

$$\beta^\star = E[\int_0^1 \text{MTE}(u) \times \omega^\star(u, Z) du]$$

Example: LATE

$$\beta^\star = E[\int_0^1 \text{MTE}(u) \times \omega^\star(u, Z) du]$$

Example: Suppose $Z \in \{0, 1\}$ is binary and that $p(1) > p(0)$, then we have

$$\underbrace{E[Y_1 - Y_0 | p(0) < U \leq p(1)]}_{\beta^\star \text{ (LATE)}} = \int_0^1 \underbrace{E[Y_1 - Y_0 | U = u]}_{\text{MTE}} \underbrace{\frac{1\{p(0) < u \leq p(1)\}}{p(1) - p(0)}}_{\omega^\star(u)} du$$

Comments

- LATE identified as IV estimand (i.e. identification of MTE not needed).
- Hence, LATE imposes restrictions on possible values of MTE.

Example: ATE

$$\beta^\star = E[\int_0^1 \mathsf{MTE}(u) \times \omega^\star(u, Z) du]$$

Example: ATE

$$\beta^\star = E[\int_0^1 \text{MTE}(u) \times \omega^\star(u, Z) du]$$

Example: Suppose $Z \in \{0, 1\}$ is binary and that $p(1) > p(0)$, then we have

$$\underbrace{E[Y_1 - Y_0]}_{\beta^\star \text{ (ATE)}} = \int_0^1 \underbrace{E[Y_1 - Y_0 | U = u]}_{\text{MTE}} \underbrace{(\times 1)}_{\omega^\star(u)} du$$

Comments

- ATE is not necessarily identified with binary instrument Z.
- But! LATE still provides information on MTE and hence on ATE.

Basic Idea

$$\beta^\star = E[\int_0^1 \text{MTE}(u) \times \omega^\star(u, Z) du] \qquad (\star)$$

Basic Idea

$$\beta^\star = E[\int_0^1 \mathsf{MTE}(u) \times \omega^\star(u, Z) du] \qquad (\star)$$

Step 1

- Find parameters $\beta^\star$ that are separately identified (e.g. LATE).
- Employ relationship ($\star$) to restrict possible values of MTE function.

Basic Idea

$$\beta^\star = E[\int_0^1 \text{MTE}(u) \times \omega^\star(u, Z) du] \qquad (\star)$$

Step 1

- Find parameters $\beta^\star$ that are separately identified (e.g. LATE).
- Employ relationship ($\star$) to restrict possible values of MTE function.

Step 2

- Define the parameter of interest (e.g. ATE).
- Impose desired restriction on MTE (optional).

Basic Idea

$$\beta^\star = E[\int_0^1 \text{MTE}(u) \times \omega^\star(u, Z) du] \qquad (\star)$$

Step 1

- Find parameters $\beta^\star$ that are separately identified (e.g. LATE).
- Employ relationship ($\star$) to restrict possible values of MTE function.

Step 2

- Define the parameter of interest (e.g. ATE).
- Impose desired restriction on MTE (optional).

Step 3

- Conduct inference on possible values of parameter of interest.
- Values must be consistent with MTE restrictions from Steps 1 and 2.

Related Literature

Extrapolation

Heckman & Robb (1985), Heckman, Tobias, & Vytlacil (2003), Manski (2003), Heckman & Vytlacil (2005), Chamberlain (2010), Angrist & Fernandez-Val (2013), Brinch, Mogstad, & Wiswall (2015) ... (and more!)

$\Rightarrow$ We avoid: parametric assumptions, support conditions.
$\Rightarrow$ We allow for: treatment heterogeneity, no closed form solutions.

Related Literature

Extrapolation

Heckman & Robb (1985), Heckman, Tobias, & Vytlacil (2003), Manski (2003), Heckman & Vytlacil (2005), Chamberlain (2010), Angrist & Fernandez-Val (2013), Brinch, Mogstad, & Wiswall (2015) ... (and more!)

⇒ We avoid: parametric assumptions, support conditions.
⇒ We allow for: treatment heterogeneity, no closed form solutions.

Inference

Imbens & Manski (2004), Romano & Shaikh (2008), Beresteanu & Molinari (2008), Bontemps, Magnac & Maurin (2012), Bugni, Canay, & Shi (2015), Kaido, Molinari, & Stoye (2016), Chernozhukov, Newey, & Santos (2016) ...

⇒ Our problem: specific convex programming problem.
⇒ Allows for: uniformly valid inference under weak assumptions.

1. General Framework
2. Numerical Illustration
3. Inference: Basic Outline
4. Efficacy of Price Subsidies for Bed Nets

Notation

Assumption M: (Y, D, Z) are generated according to the model

$$Y = DY_1 + (1 - D)Y_0 \qquad (Y_0, Y_1, U) \perp Z$$

$$D = 1\{U \leq p(Z)\} \qquad U \sim U[0, 1]$$

Note: Covariates ignored for notational simplicity but simple to add.

Notation

Assumption M: (Y, D, Z) are generated according to the model

$$Y = DY_1 + (1 - D)Y_0 \qquad (Y_0, Y_1, U) \perp Z$$

$$D = 1\{U \leq p(Z)\} \qquad U \sim U[0, 1]$$

Note: Covariates ignored for notational simplicity but simple to add.

Marginal Treatment Response (MTR)

$$m_d(u) \equiv E[Y_d | U = u] \text{ for } d \in \{0, 1\}$$

where $m = (m_0, m_1) \in \mathcal{M}$ for some known set $\mathcal{M}$ (prior assumptions)

Note: By definition, MTE$(u) = m_1(u) - m_0(u)$ (but + flexibility with m_d).

Target Parameter

$$\beta^\star = E[\int_0^1 m_0(u)\omega_0^\star(u,Z)du] + E[\int_0^1 m_1(u)\omega_1^\star(u,Z)du]$$

where $\omega_d^\star$ are known or identified weighting functions for $d \in \{0,1\}$.

Target Parameter

$$\beta^\star = E[\int_0^1 m_0(u)\omega_0^\star(u,Z)du] + E[\int_0^1 m_1(u)\omega_1^\star(u,Z)du]$$

where $\omega_d^\star$ are known or identified weighting functions for $d \in \{0,1\}$.

Example 1: Average Treatment Effect (ATE)

$$E[Y_1 - Y_0] = \int_0^1 \underbrace{E[Y_0|U=u]}_{m_0(u)} \times \underbrace{(-1)}_{\omega_0^\star(u,Z)} du + \int_0^1 \underbrace{E[Y_1|U=u]}_{m_1(u)} \times \underbrace{(+1)}_{\omega_1^\star(u,Z)} du$$

Target Parameter

$$\beta^\star = E[\int_0^1 m_0(u)\omega_0^\star(u,Z)du] + E[\int_0^1 m_1(u)\omega_1^\star(u,Z)du]$$

where $\omega_d^\star$ are known or identified weighting functions for $d \in \{0,1\}$.

Example 2: Average Treatment on Treated (ATT)

$$E[Y_1 - Y_0|D=1]$$
$$= E[\int_0^1 m_0(u)\underbrace{(-\frac{1\{u \le p(Z)\}}{P(D=1)})}_{\omega_0^\star(u,Z)}du] + E[\int_0^1 m_1(u)\underbrace{(\frac{1\{u \le p(Z)\}}{P(D=1)})}_{\omega_1^\star(u,Z)}du]$$

What We Know

Problem: Target parameter identified when MTR $m = (m_0, m_1)$ identified ...

... but the MTR functions $m = (m_0, m_1)$ may not be identified ...

However: We do have some information about $m = (m_0, m_1)$ through LATE

What We Know

Problem: Target parameter identified when MTR $m = (m_0, m_1)$ identified ...

... but the MTR functions $m = (m_0, m_1)$ may not be identified ...

However: We do have some information about $m = (m_0, m_1)$ through LATE

Example: Suppose $Z \in \{0, 1\}$ is binary and that $p(1) > p(0)$, then we have

$$\begin{aligned}\frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} &= E[Y_1 - Y_0 | p(0) < U \leq p(1)] \\ &= \int_0^1 \{m_1(u) - m_0(u)\} \frac{1\{p(0) < u \leq p(1)\}}{p(1) - p(0)} du\end{aligned}$$

Key: LATE imposes a linear restriction on MTR ... are there others?

MTR Restrictions

Proposition: If Assumption M holds and $E[s^2(D,Z)] < \infty$, then it follows

$$E[Ys(D,Z)] = E[\int_0^1 m_0(u)\omega_{0s}(u,Z)du] + E[\int_0^1 m_1(u)\omega_{1s}(u,Z)du]$$

where $\omega_{0s}(u,Z) \equiv s(0,Z)1\{u > p(Z)\}$
and $\omega_{1s}(u,Z) \equiv s(1,Z)1\{u \leq p(Z)\}$.

Comments

- LATE corresponds to $s(D,Z) = (Z - E[Z])/\text{Cov}(D,Z)$.
- Additional choices of $s(D,Z) \Rightarrow$ more restrictions on MTR.
- Sufficiently many $s(D,Z) \Rightarrow$ reflect all information in conditional means.

Using MTR Restrictions

What We Want

$$\beta^\star = E[\int_0^1 m_0(u)\omega_0^\star(u, Z)du] + E[\int_0^1 m_1(u)\omega_1^\star(u, Z)du]$$

Note: $\beta^\star$ is linear in m – i.e. $\beta^\star = \Gamma^\star(m)$ where $\Gamma^\star : \mathcal{M} \to \mathbf{R}$ equals

$$\Gamma^\star(m) = E[\int_0^1 m_0(u)\omega_0^\star(u, Z)du] + E[\int_0^1 m_1(u)\omega_1^\star(u, Z)du]$$

Using MTR Restrictions

What We Want

$$\beta^\star = E[\int_0^1 m_0(u)\omega_0^\star(u,Z)du] + E[\int_0^1 m_1(u)\omega_1^\star(u,Z)du]$$

Note: $\beta^\star$ is linear in m – i.e. $\beta^\star = \Gamma^\star(m)$ where $\Gamma^\star : \mathcal{M} \to \mathbf{R}$ equals

$$\Gamma^\star(m) = E[\int_0^1 m_0(u)\omega_0^\star(u,Z)du] + E[\int_0^1 m_1(u)\omega_1^\star(u,Z)du]$$

What We Know

$$E[Ys(D,Z)] = E[\int_0^1 m_0(u)\omega_{0s}(u,Z)du] + E[\int_0^1 m_1(u)\omega_{1s}(u,Z)du]$$

Note: Linear restrictions – i.e. $E[Ys(D,Y)] = \Gamma_s(m)$ where $\Gamma_s(m)$ equals

$$\Gamma_s(m) = E[\int_0^1 m_0(u)\omega_{0s}(u,Z)du] + E[\int_0^1 m_1(u)\omega_{1s}(u,Z)du]$$

Using MTR Restrictions

Proposition: Let $\mathcal{M}$ be convex, $\mathcal{S}$ a set of functions of (D, Z), and define

$$\underline{\beta}^\star \equiv \inf_{m \in \mathcal{M}} \Gamma^\star(m) \text{ s.t. } E[Ys(D,Z)] = \Gamma_s(m) \text{ for all } s \in \mathcal{S}$$

$$\bar{\beta}^\star \equiv \sup_{m \in \mathcal{M}} \Gamma^\star(m) \text{ s.t. } E[Ys(D,Z)] = \Gamma_s(m) \text{ for all } s \in \mathcal{S}$$

Then closure of feasible (s.t. linear constraints) values of $\beta^\star$ equals $[\underline{\beta}^\star, \bar{\beta}^\star]$.

Comments

- Convex optimization problem $\Rightarrow$ simple estimators (when feasible).
- Constraints may be unfeasible $\Rightarrow$ model is misspecified.
- For appropriate $\mathcal{S}$ can exhaust information in conditional means.
  ... but still may not correspond to identified set (unless Y binary).

1 General Framework

2 Numerical Illustration

3 Inference: Basic Outline

4 Efficacy of Price Subsidies for Bed Nets

Basic Design

Data Generating Process

- Instrument takes three values $Z \in \{0, 1, 2\}$.
- Outcome Y is binary $Y \in \{0, 1\}$.
- Propensity score $p(0) = 0.35$, $p(1) = 0.6$, and $p(2) = 0.7$.

Parameter of Interest

$$\text{LATE}(0.35, 0.9) \equiv E[Y_1 - Y_0 | U \in (0.35, 0.9]] = 0.046$$

Comments

- Three LATEs nonparametrically identified.
- Parameter of interest measures sensitivity to expanding complier group.
- MTR functions not identified (unless $\mathcal{M}$ is restricted).

Basic Information

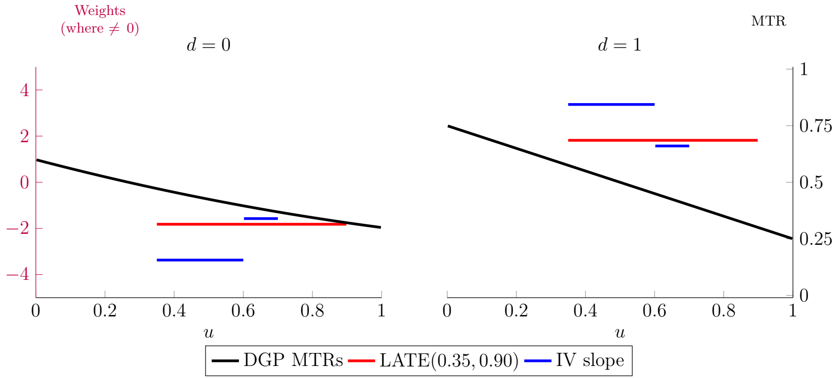

Figure: MTRs Used in the Data Generating Process (DGP)

Just a Start

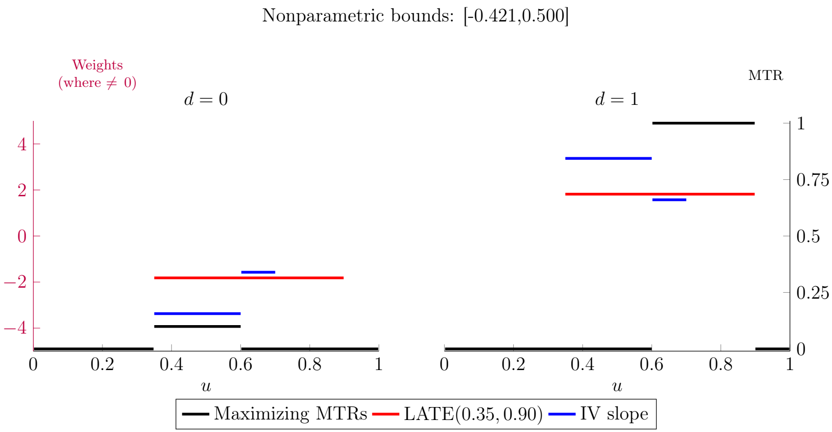

Figure: Maximizing MTRs When Using Only the IV Slope Coefficient

Adding Information

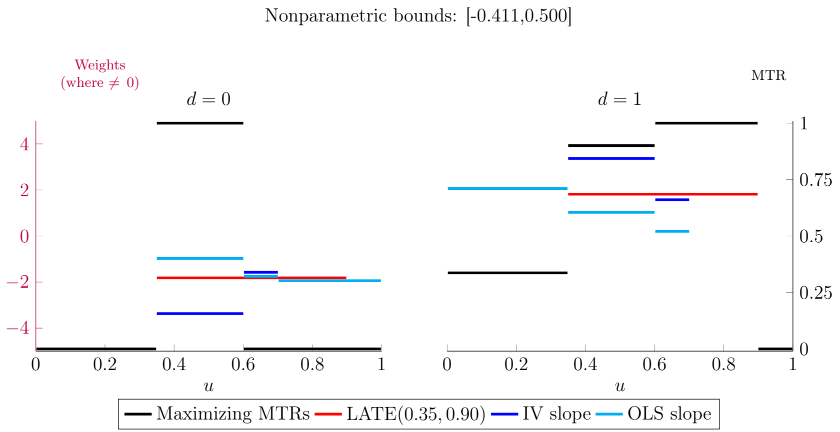

Figure: Maximizing MTRs When Using Both the IV and OLS Slope Coefficients

Adding All Information

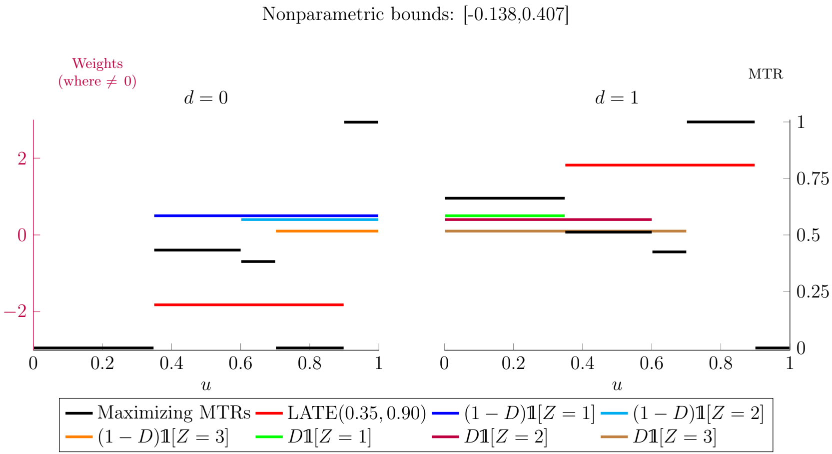

Figure: Maximizing MTRs When Using All IV–like Estimands (Sharp Bounds)

Adding Shape Restrictions

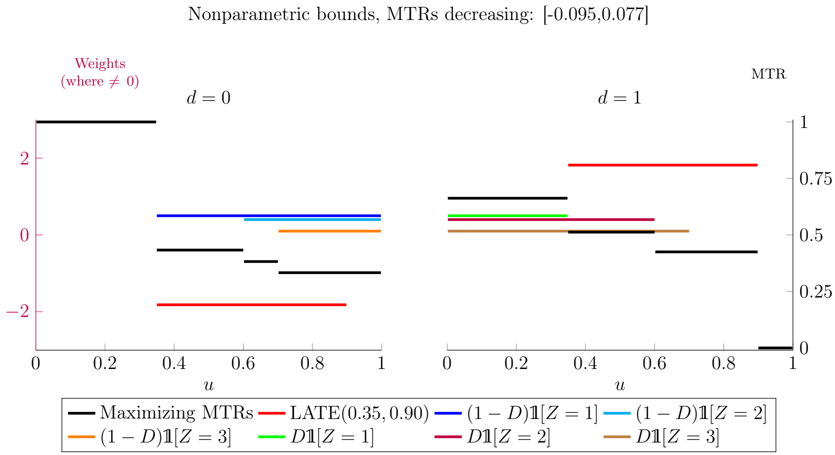

Figure: Maximizing MTRs When Restricted to be Decreasing

Adding Smoothness

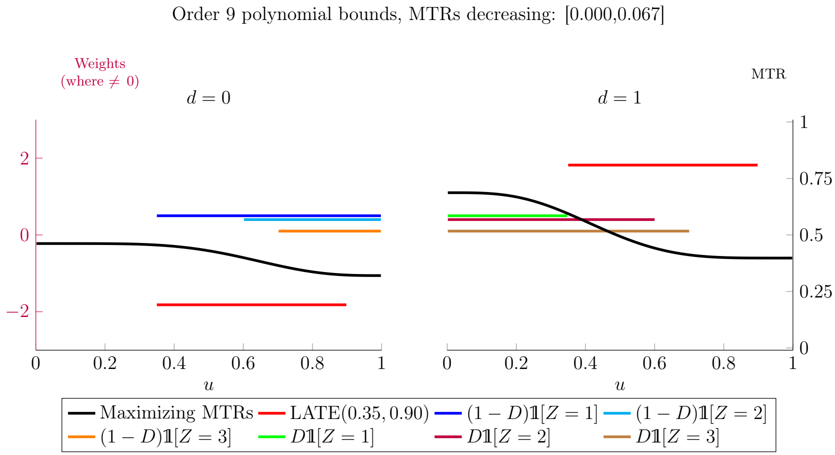

Figure: Maximizing MTRs When Further Restricted to be a 10th Order Polynomial

Summary

Available Information

- Valuable information in the data beyond identified LATEs.
- Parameters can be informative without being interesting (e.g. OLS).

Shape Restrictions

- When credible, they can substantially improve bounds.
- Value in parametric and nonparametric restrictions.

Additional Comments

- Computational approach allows flexibility without analytical solution.
- Different information for different parameters (e.g. LATE$(0.35, \bar{u})$).

1 General Framework

2 Numerical Illustration

3 Inference: Basic Outline

4 Efficacy of Price Subsidies for Bed Nets

Setup

Goal: Build confidence regions and/or conduct specification tests.

Setup

Goal: Build confidence regions and/or conduct specification tests.

Formally: Given i.i.d. sample $\{Y_i, Z_i, D_i\}_{i=1}^n$ with $(Y, Z, D) \sim P \in \mathbf{P}$, we test

$$H_0 : P \in \mathbf{P}_0 \qquad H_1 : P \in \mathbf{P} \setminus \mathbf{P}_0$$

where for some linear map Γ_P and element β_P we define $\mathbf{P}_0$ to equal

$$\mathbf{P}_0 \equiv \{P \in \mathbf{P} : \Gamma_P(m) = \beta_P \text{ for some } m \in \mathcal{M}\}$$

Allows For

- Confidence regions for target parameter.
- Specification tests for different maintained assumptions

Test Statistic

$$\mathbf{P}_0 \equiv \{P \in \mathbf{P} : \Gamma_P(m) = \beta_P \text{ for some } m \in \mathcal{M}\}$$

Test Statistic

$$\mathbf{P}_0 \equiv \{P \in \mathbf{P} : \Gamma_P(m) = \beta_P \text{ for some } m \in \mathcal{M}\}$$

Test Statistic

$$T_n \equiv \inf_{m \in \mathcal{M}} \sqrt{n} \|\hat{\beta} - \hat{\Gamma}(m)\|$$

Comments

- Theory allows for use of sieve $\mathcal{M}_n$ (if needed for computation).
- Linear minimum distance problem, complications arise from $\mathcal{M}$.
- "Irregular" behavior $\Rightarrow$ Bootstrap failure (Fang and Santos (2016)).

Inference Results

Test Statistic

$$T_n \equiv \inf_{m \in \mathcal{M}} \sqrt{n} \|\hat{\beta} - \hat{\Gamma}(m)\|$$

Main Result Propose Bootstrap critical values and establish size control.

Contributions

- Size control is uniform in large class of distributions $\mathbf{P}$.
- Allow shape restrictions, and parametric/nonparametric specifications.
- Employ with finite or infinite number of moment restrictions.
- More generally applicable to linear programming problems.
- Bootstrap statistic obtained as a bilinear optimization problem.

1 General Framework

2 Numerical Illustration

3 Inference: Basic Outline

4 Efficacy of Price Subsidies for Bed Nets

Background

The Data

- Randomized control experiment in Kenya by Dupas (2014).
- Households randomly assigned a price (out of 17) for antimalaria net.
- Total of 1200 households in six villages.
- Follow up check to see if malaria net was in use.

Policy Concerns

- Subsidising inframarginal consumers that would have purchased.
- Unwilling to purchase may be unwilling to use (nonmonetary cost).
- Higher price may exclude poor or credit constrained individuals.

Background

The Setup

- Y indicator for whether net is in use.
- Z randomly assigned price.
- D indicator for whether net was purchased.

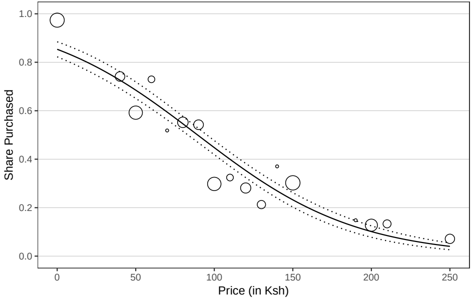

Figure: Impact of Price on the Household's Purchase of Bed Net (logit regression)

Policy Examination

First Target Parameter

- Obtain the average treatment effect.
- Interpretable as comparing no availability of net with free nets.

Second Target Parameter

- Obtain LATE from no net to propensity score at Ksh 150 (avg price).
- Interpretable as the effect of introducing the net into the market.
- Propensity score at 150 estimated via logit prediction.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Information Specification														
Intercept	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Linear in $p(Z)$	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OLS						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$1(Z \leq 50)$											✓	✓	✓	✓	✓
$1(Z \leq 150)$											✓	✓	✓	✓	✓
Panel A.	**Population Average Treatment Effect**														
K (polynomial order)	2	6	10	20	NP	2	6	10	20	NP	2	6	10	20	NP
Bounds															
Lower	.6521	.4646	.3857	.3275	.2533	.6521	.4956	.4700	.4537	.3954	$\emptyset$	.6365	.5602	.5269	.4487
Upper	.6772	.7269	.7362	.7445	.7515	.6521	.7269	.7362	.7445	.7515	$\emptyset$	.7104	.7178	.7229	.7253
90% Confidence Interval															
Lower	.5486	.3761	.2995	.2421		.4282	.4032	.3511	.3204		.5206	.4130	.3652	.3260	
Upper	.7462	.8019	.8102	.8139		.7516	.8093	.8179	.8209		.7491	.7910	.7941	.7978	
Panel B.	**PRTE at Free Provision versus a Price of 150 Ksh**														
K (polynomial order)	2	6	10	20	NP	2	6	10	20	NP	2	6	10	20	NP
Bounds															
Lower	.6600	.5881	.5626	.5444	.4817	.6600	.5881	.5626	.5444	.4856	$\emptyset$	.6758	.6506	.6214	.5573
Upper	.7049	.8140	.8469	.8817	.9732	.6600	.7085	.7172	.7275	.7941	$\emptyset$	.6895	.6988	.7140	.7492
90% Confidence Interval															
Lower	.5417	.5005	.4695	.4479		.3890	.3472	.3414	.3320		.5079	.4755	.4584	.4281	
Upper	.7686	.9161	.9519	.9746		.7732	.9263	.9616	.9838		.7713	.9093	.9291	.9511	
	Specifications of the IV-like Estimands														
Intercept			$s(d,z)=1$					$s(d,z)=1$					$s(d,z)=1$		
Linear in $p(Z)$			$s(d,z)=p(z)$					$s(d,z)=p(z)$					$s(d,z)=p(z)$		
OLS								$s(d,z)=d$					$s(d,z)=d$		
$1(Z \leq 50)$													$s(d,z)=1(z \leq 50)$		
$1(Z \leq 150)$													$s(d,z)=1(z \leq 150)$		

Conclusion

Summary

- General method for inference on a specified target parameter.
- Does not require continuous/large support instruments (though help).
- Allows for specification testing.
- Computation is fast and reliable – linear/bilinear programming.
- Uniformly valid inference in linear programs.

In Progress

- Extensive Monte Carlo experiments.
- Developing R package (with Bradley Setzler).
- Additional guidance on bandwidth selection.