EXTRAPOLATION AND EXTENSIONS

Ivan A. Canay
Northwestern University[1]
ECON 481

1. Brinch, C. N., Mogstad, M., & Wiswall, M. (2017). Beyond LATE with a discrete instrument. Journal of Political Economy, 125(4), 985-1039.
2. Mogstad, M., Santos, A., & Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. Econometrica, 86(5), 1589-1619.

[1]Today's class is based on Alex Torgovitsky's notes. I'd like to thank him for kindly sharing them.

Topics of Part I

- Lec I: Selection on Observables
  1. Potential Outcomes vs Latent Variables
  2. Causal Inference
  3. Selection Bias
  4. Selection on Observables & Selection on Prop. Score
- Lec II: Roy Models and LATE
  1. The role of heterogeneity
  2. Multiple instruments, Covariates, and Abadie's κ
- Lec III: Marginal Treatment Effect
  1. Parameters as functions of MTEs
  2. Policy Relevant Treatment Effects
- Lec IV: Extrapolations
  1. Semi-Parametrics MTEs
  2. Weights for Target Parameters

Outline

1. Parameterizing MTEs
2. Application: QQ in Fertility
3. Extrapolation: Mogstad, Santos, and Torgovitsky
4. Application: Bed Nets

Taking Stock

Summary

- Unobserved heterogeneity widely viewed as prevalent and important
- Fully parametric selection models allow for it, but not deemed credible
- Linear IV models with heterogeneity might not yield a useful parameter
  At best some sort of weighted LATE expression — policy question?
- MTE models provide a conceptual framework for defining parameters
  But many counterfactual parameters will not be point identified

Extrapolation

- We want to extrapolate from those affected by the instrument to others
  Those who would be relevant in our counterfactual question
- Natural way to do this is to add some *attractive* parametric structure
  But want to avoid the fully parametric normal selection model
- Another natural response is to allow for partial identification

No Unobserved Heterogeneity

The easy way out

- All identification issues here are caused by unobserved heterogeneity U dependent with $Y_1 - Y_0$, given X
- Assuming no such heterogeneity in treatment effects:

$$E[Y_1 - Y_0 | U = u, X = x] = E[Y_1 - Y_0 | X = x] \text{ for all } u \in [0,1]$$

- Equivalent to assuming that the MTE is constant as a function of u
- Any conditional-on-x LATE is sufficient to point identify the entire MTE

Sufficient conditions

- This will be true with **constant effects**: $Y_1 - Y_0 | X = x$ deterministic
- More generally, holds if agents choose D with no knowledge of (Y_1, Y_0)
- Notice in principle constant effects still allows $E[Y_d | U, X] \neq E[Y_d | X]$
- So it allows for selection bias, but not **selection on the gain**
- However, most endogeneity stories feature both forms of selection

Parameterizing the MTE

- Brinch et al (2017, "BMW") show how to parameterize the MTE
- For example, with no covariates, suppose we assume that

$$\begin{aligned} m_d(u) &\equiv E[Y_d|U=u] = \alpha_d + \beta_d u \text{ for } d = 0,1 \\ \Rightarrow \quad \text{MTE}(u) &\equiv m_1(u) - m_0(u) = (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)u \end{aligned}$$

- It can be shown that under this assumption

$$\begin{aligned} E[Y|D=1,P=p] &= \alpha_1 + \frac{\beta_1}{2}p \\ E[Y|D=0,P=p] &= \left(\alpha_0 + \frac{\beta_0}{2}\right) + \frac{\beta_0}{2}p \end{aligned}$$

- So regress Y on P among $D=d$ to identify (α_d, β_d) for $d = 0,1$
- Requires two points of support in $P \Rightarrow Z \in \{0,1\}$ suffices
- Implication: Linearity is sufficient to point identify *any* mean contrast

Generalization to Polynomials

- More generally, one could specify

$$m_d(u) \equiv E[Y_d | U = u] = \sum_{k=0}^{K} \beta_{dk} u^k$$

- $E[Y | D = 1, P = p]$ is a Kth degree polynomial in this case
- So we can point identify a polynomial of degree $K = |\text{supp}(Z)| - 1$
- Stated differently, we need an instrument with $|\text{supp}(Z)| \geq K + 1$

- Stratifying on $D = d$ is important, otherwise in the linear case

$$E[Y | P = u] = -\frac{\beta_0}{2} + (\alpha_1 - \alpha_0)u + \frac{1}{2}(\beta_1 + \beta_0)u^2$$

- So you would need a trinary instrument if you didn't stratify
- In the general case, would need $|\text{supp}(Z)| \geq K + 2$

Covariates and Separability

Saturated specifications

- Covariates could be fully interacted, e.g. with $X \in \{0,1\}$

$$m_1(u,x) = \alpha_1 + x\gamma_1 + \beta_1 u + ux\delta_1$$
$$\Rightarrow \quad E[Y|D=1, P=p, X=x] = \alpha_1 + \gamma_1 x + \frac{\beta_1}{2}p + \frac{\delta_1}{2}px$$

- Instrument requirement becomes conditional on $X = x$
  Variation in P given $X = 0$ for (α_1, β_1); given $X = 1$ for (γ_1, δ_1)

Separability

- Removing interactions ($\delta_1 = 0$) allows one to combine variation in X:

$$m_1(u,x) = \alpha_1 + x\gamma_1 + \beta_1 u$$
$$\Rightarrow \quad E[Y|D=1, P=p, X=x] = \alpha_1 + x\gamma_1 + \frac{\beta_1}{2}p$$

- Same variation in X and P, but fewer parameters $\Rightarrow$ overidentified

Outline

1. Parameterizing MTEs
2. **Application: QQ in Fertility**
3. Extrapolation: Mogstad, Santos, and Torgovitsky
4. Application: Bed Nets

Quantity-Quality in Fertility

Question

- BMW revisit the causal effect of family size on child outcomes
  Motivated by quantity-quality (QQ) model of Becker and Lewis (1973)
- Could be negative due to (e.g.) resource dilution
- Could be positive due to (e.g.) increased marriage stability

Data and design

- Same as in Black, Devereaux and Salvanes (2005)
- Norwegian administrative data ($N \approx$ 514,000 families)
- Y is the first-born child's years of attained education
- Treatment and instruments are the same as in Angrist and Evans (1998)
- D is an indicator for having more than 2 children (vs. exactly 2)
- Z is either twins or same-sex
- X includes cohort, parent's age at first birth, and parent's schooling

Various IV Estimates

TABLE 3
OLS and IV estimates

<table>
<tr><th></th><th>$P(Z)$ as Instrument (1)</th><th>$Z\_$ as Instrument (2)</th></tr>
<tr><td>IV:</td><td></td><td></td></tr>
<tr><td>Same-sex instrument</td><td>−.208
(.105)</td><td>.174
(.115)</td></tr>
<tr><td>Twins instrument</td><td>−.065
(.060)</td><td>.050
(.062)</td></tr>
<tr><td>Both instruments</td><td>−.015
(.053)</td><td>.076
(.055)</td></tr>
<tr><td>OLS</td><td colspan="2">−.052
(.007)</td></tr>
</table>

Note.—This table reports OLS and IV estimates of the effect of family size on the educational attainment of firstborn children. Column 1 reports linear IV estimates with $P(Z)$ as instrument. We construct $P(Z)$ using the parameter estimates from the logit model with average derivatives reported in table 2. Column 2 reports standard linear IV estimates with $Z\_$ as instrument. We use the same specification for the covariates as reported in table 2. The first row excludes the same sex, first and second children instrument from the second stage, the second row excludes the twins at second parity instrument from the second stage, and the third row excludes both instruments from the second stage. The OLS estimate of the second-stage specification (20) is reported in the fourth row. Standard errors in parentheses are robust to heteroskedasticity.

- Outcome equation has $1, D, X$ without any interactions
- Difference in columns is linear IV vs. using estimated propensity score
- Difference in rows is what is excluded from the outcome equation
- If treatment effects were constant, one would expect stable estimates

Linear MTE Estimates

TABLE 4
Estimates of Linear MTE Model and LATE Based on Same-Sex Instrument

	p = .473	p = .531	Intercept	Slope
	A. Estimates of Linear MTE Model and Its Components			
Linear MTE model:				
$\mu_1 + K_1(P) = E(Y_1 \mid U_D < p)$	12.086	12.131	11.720	$+.775p$
	(.008)	(.007)	(.095)	(.188)
$\mu_0 + K_0(P) = E(Y_0 \mid U_D > p)$	12.462	12.450	12.564	$-.216p$
	(.007)	(.008)	(.091)	(.181)
$\mu_1 + k_1(p) = E(Y_1 \mid U_D = p)$	12.453	12.542	11.720	$+1.550p$
	(.084)	(.105)	(.095)	(.376)
$\mu_0 + k_0(p) = E(Y_0 \mid U_D = p)$	12.576	12.551	12.780	$-.432p$
	(.101)	(.080)	(.272)	(.0362)
$\text{MTE}(p) = E(Y_1 - Y_0 \mid U_D = p)$	−.123	−.008	−1.006	$+1.981p$
	(.129)	(.130)	(.290)	(.529)
	B. LATE from IV and Linear MTE Model			
Instrumental variables:				
$[E(Y \mid \Pr(D) = .531) - E(Y \mid \Pr(D) = .473)]/(.531 - .473)$				−.065
				(.129)
Linear MTE model:				
$\int_{.471}^{.531} \text{MTE}(p) = \text{MTE}[(.531 + .471)/2]$				−.065
				(.129)

Note.—This table displays LATE and linear MTE estimates of family size on the educational attainment of firstborn children. Panel A reports estimates from the linear MTE model with same sex, first and second as the excluded instrument. Panel B reports estimates of LATE from the IV estimator and the linear MTE model, with same sex, first and second as the excluded instrument. We do not include any covariates in the MTE estimation or the IV estimation. Standard errors in parentheses are computed by nonparametric bootstrap with 100 bootstrap replications.

- No covariates in these specifications (reason for large standard errors)
- The linear MTE model replicates the LATE ($-.065$) — see problem set
- Linearity means the MTE is point identified everywhere — credible?
- Provides ATE estimate of -0.02 and ATT estimate of 0.48

Nonparametric MTE w/ Separability

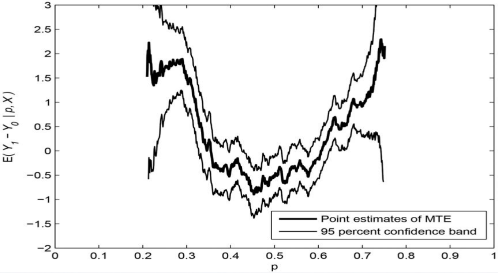


FIG. 5.—This figure displays the MTE estimates based on assumptions 1 and 2. We use same-sex as the excluded instrument. The MTE estimates are evaluated at the mean values of the covariates. We construct $P(Z)$ using the parameter estimates from the logit model specified in the note to table 2. The MTE estimates are based on a double residual regression separately for the treated and nontreated, using a local quadratic regression with uniform kernel and a bandwidth of 0.0615. The 95 percent confidence band (dashed lines) is computed from a nonparametric bootstrap with 100 bootstrap replications. The y-axis measures the value of the MTE in years of schooling, whereas the x-axis represents the unobserved component of parents' net gain from having three or more children rather than two children. A high value of p means that a family is less likely to have three or more children.

- Impose separability to combine $\approx$ 32,000 support points for $p(X, Z)$
- Positive effects (MTE) for both smaller and larger U

  Those least likely to change fertility because of sex composition
- Negative effects (MTE) for those more affected by sex composition

Tests: Unobserved Heterogeneity

TABLE 5
Comparing LATEs across Different Intervals of the Propensity Score

	LATE over Intervals				
	(.22, .27) − (.31, .36)	(.31, .36) − (.40, .45)	(.40, .45) − (.49, .54)	(.49, .54) − (.58, .63)	(.58, .63) − (.67, .72)
Point estimate	1.102	1.011	.046	−.413	−1.006
Standard error	.521	.307	.257	.241	.301
p-value	.034	.001	.859	.087	.001
p-value of joint test			.000		

Note.—This table reports tests of constant MTE of family size on the educational attainment of firstborn children. The MTE estimates are based on assumptions 1 and 2, with same sex, first and second as the excluded instrument (see fig. 5). We construct $P(Z)$ using the parameter estimates from the logit model with average derivatives reported in table 2. We use the same specification for the covariates as reported in table 2. The MTE estimates are based on double residual regression separately for the treated and nontreated, using local quadratic regression with uniform kernel and bandwidth of 0.0615. The LATEs are derived from the MTE estimates by integrating over the indicated intervals. Standard errors are based on nonparametric bootstrap (of both estimation stages) with 100 bootstrap replications.

- Same testing procedure as in CHV, but much more power here
- Unobserved treatment effect heterogeneity is clearly present

Outline

1. Parameterizing MTEs
2. Application: QQ in Fertility
3. **Extrapolation: Mogstad, Santos, and Torgovitsky**
4. Application: Bed Nets

Mogstad, Santos & Torgovitsky (18)

Background motivation

- We want to specify precise parameters that answer specific questions
- The MTE framework provides a good way to do this
- But shows these parameters are often not nonparametrically identified
- 3 options: More assumptions, partial identification, change the question

Motivation relative to BMW

- More assumptions, but functional forms dictated by instrument support
  This reflects a framework that *requires* point identification
- Instead, we would like to take the assumptions as the primitive
- Possible for these to be informative even without point identification
- Would also like to treat the parameter of interest as central
  Instead of deriving it from the MTE as a second step

Overview

Motivation

- Assume that the analyst has a **target parameter**, $\beta^\star$
- They choose $\beta^\star$ on the outset to answer a specific question
- Forward engineering instead of backward engineering
- Suppose $\beta^\star(m)$ can be written as function of $m \equiv (m_0, m_1)$

  $m_d \equiv E[Y_d | U = u]$ is a **marginal treatment response** (MTR) function

Main question

- There are two constraints on the MTR functions m:
  1. Must be observationally equivalent (consistent with data)
  2. Must satisfy our assumptions ($m \in \mathcal{M}$, the parameter space)
- These constraints restrict m to be in some set $\mathcal{M}^\star$
- So they also restrict $\beta^\star(m)$ to be in some set $\mathcal{B}^\star$

  How do we determine $\mathcal{B}^\star$?

Form for the Target Parameter

Form

- Suppose $\beta^\star(m)$ is a single number for each pair $m \equiv (m_0, m_1)$
- Means we are looking for $\mathcal{B}^\star \subseteq \mathbf{R}$
- Assume that $\beta^\star$ can be written as

$$\beta^\star(m_0, m_1) \equiv \sum_{d \in \{0,1\}} E\left[\int_0^1 \underbrace{m_d(u, X)}_{\text{MTR for } Y_d} \overbrace{\omega_d^\star(u, X, Z)}^{\text{identified weights}} \, du\right]$$

- The weights can (usually) be derived from the choice of $\beta^\star$

Scope

- Covers basically any mean contrast — recall ATE, ATT, LATE, PRTE ...
- HV and MST provide extensive tables cataloguing $\omega_0^\star, \omega_1^\star$ forms ...
- In many (but not all) cases the weights are symmetric: $\omega_0^\star = -\omega_1^\star$

Weights for Target Parameters

Target Parameter	Expression	Weights $\omega_0^\star(u,x,z)$	Weights $\omega_1^\star(u,x,z)$
Average Untreated Outcome	$\mathbb{E}[Y_0]$	1	0
Average Treated Outcome	$\mathbb{E}[Y_1]$	0	1
Average Treatment Effect (ATE)	$\mathbb{E}[Y_1 - Y_0]$	-1	1
ATE given $X = \bar{x}$ where $\mathbb{P}[X = \bar{x}] > 0$	$\mathbb{E}[Y_1 - Y_0 \mid X = \bar{x}]$	$-\omega_1^\star(u,x,z)$	$\frac{\mathbb{1}[x = \bar{x}]}{\mathbb{P}[X = \bar{x}]}$
Average Treatment on the Treated (ATT)	$\mathbb{E}[Y_1 - Y_0 \mid D = 1]$	$-\omega_1^\star(u,x,z)$	$\frac{\mathbb{1}[u \le p(x,z)]}{\mathbb{P}[D = 1]}$
Average Treatment on the Untreated (ATU)	$\mathbb{E}[Y_1 - Y_0 \mid D = 0]$	$-\omega_1^\star(u,x,z)$	$\frac{\mathbb{1}[u > p(x,z)]}{\mathbb{P}[D = 0]}$
Local Average Treatment Effect (LATE) for $z_0 \to z_1$ given $X = x$, where $p(x,z_1) > p(x,z_0)$	$\mathbb{E}[Y_1 - Y_0 \mid p(x,z_0) < U \le p(x,z_1), X = x]$	$-\omega_1^\star(u,x,z)$	$\frac{\mathbb{1}[p(x,z_0) < u \le p(x,z_1)]}{p(x,z_1) - p(x,z_0)}$

Weights for Policy Relevant TEs

Target Parameter	Expression	$\omega_1^\star(u,x,z) = -\omega_0^\star(u,x,z)$
Generalized LATE for $U \in [\underline{u}, \overline{u}]$	$\mathbb{E}[Y_1 - Y_0 \mid U \in [\underline{u}, \overline{u}]]$	$\dfrac{\mathbb{1}[u \in [\underline{u}, \overline{u}]]}{\overline{u} - \underline{u}}$
Policy Relevant Treatment Effect (PRTE) for policy (p^{a_1}, Z^{a_1}) relative to policy (p^{a_0}, Z^{a_0})	$\dfrac{\mathbb{E}[Y^{a_1}] - \mathbb{E}[Y^{a_0}]}{\mathbb{E}[D^{a_1}] - \mathbb{E}[D^{a_0}]}$	$\dfrac{\mathbb{P}[u \le p^{a_1}(x, Z^{a_1}) \mid X = x] - \mathbb{P}[u \le p^{a_0}(x, Z^{a_0}) \mid X = x]}{\mathbb{E}[p^{a_1}(X, Z^{a_1})] - \mathbb{E}[p^{a_0}(X, Z^{a_0})]}$
Additive PRTE with magnitude α	PRTE with $Z^\star = Z$ and $p^\star(x,z) = p(x,z) + \alpha$	$\dfrac{\mathbb{1}[u \le p(x,z) + \alpha] - \mathbb{1}[u \le p(x,z)]}{\alpha}$
Proportional PRTE with magnitude α	PRTE with $Z^\star = Z$ and $p^\star(x,z) = (1+\alpha)p(x,z)$	$\dfrac{\mathbb{1}[u \le (1+\alpha)p(x,z)] - \mathbb{1}[u \le p(x,z)]}{\alpha\,\mathbb{E}[p(X,Z)]}$
PRTE for an additive α shift of the j^{th} component of Z	PRTE with $Z^\star = Z + \alpha e_j$ and $p^\star(x,z) = p(x,z)$	$\dfrac{\mathbb{1}[u \le p(x, z + \alpha e_j)] - \mathbb{1}[u \le p(x,z)]}{\mathbb{E}[p(X, Z + \alpha e_j)] - \mathbb{E}[p(X,Z)]}$

Observational Equivalence

IV–like estimands

- Moments of the distribution of (Y, D, X, Z) are similar to $\beta^\star$:
  Each m generates a different value of a given moment
- Let s be a function of (D, X, Z) and let $S \equiv s(D, X, Z)$
- Define an **IV–like estimand** as $\beta_s \equiv E[YS]$

- The mapping between m and the β_s it *would* generate is

$$\Gamma_s(m) \equiv \sum_{d \in \{0,1\}} E\left[\int_0^1 \underbrace{m_d(u, X)}_{\text{MTR for } Y_d} \overbrace{\omega_{ds}(u, X, Z)}^{\text{identified weights}} \, du\right]$$

where

$$\omega_{0s}(u, x, z) \equiv s(0, x, z) I\{u \geq p(x, z)\}$$
$$\omega_{1s}(u, x, z) \equiv s(1, x, z) I\{u \leq p(x, z)\}$$

- Same structure as the target parameter, $\beta^\star$, but different weights
- Derivation was part of deriving IV estimand weights (may write supplement)

Examples of IV–Like Estimands

Estimand	β_s	$s(d,x,z)$	Notes
Wald (z_0 to z_1)	$\frac{\mathbb{E}[Y\|Z=z_1]-\mathbb{E}[Y\|Z=z_0]}{\mathbb{E}[D\|Z=z_1]-\mathbb{E}[D\|Z=z_0]}$	$\frac{\frac{\mathbb{1}[z=z_1]}{\mathbb{P}[Z=z_1]}-\frac{\mathbb{1}[z=z]}{\mathbb{P}[Z=z_0]}}{\mathbb{E}[D\|Z=z_1]-\mathbb{E}[D\|Z=z_0]}$	$\mathbb{P}[Z=z_j]\neq 0, j=0,1$ and $\mathbb{E}[D\|Z=z_1]$ $\neq \mathbb{E}[D\|Z=z_0]$
IV slope	$\frac{\text{Cov}(Y,Z)}{\text{Cov}(D,Z)}$	$\frac{z-\mathbb{E}[Z]}{\text{Cov}(D,Z)}$	Z scalar
IV (jth component)	$e_j'\,\mathbb{E}[\widetilde{Z}\widetilde{X}']^{-1}\,\mathbb{E}[\widetilde{Z}Y]$	$e_j'\,\mathbb{E}[\widetilde{Z}\widetilde{X}']^{-1}\widetilde{z}$	$\widetilde{X}\equiv[1,D,X']'$ $\widetilde{Z}\equiv[1,Z,X']'$ Z scalar e_j the jth unit vector
TSLS (jth component)	$e_j'\left(\Pi\,\mathbb{E}[\widetilde{Z}\widetilde{X}']\right)^{-1}\left(\Pi\,\mathbb{E}[\widetilde{Z}Y]\right)$	$e_j'(\Pi\,\mathbb{E}[\widetilde{Z}\widetilde{X}'])^{-1}\Pi\widetilde{Z}$	$\Pi\equiv\mathbb{E}[\widetilde{X}\widetilde{Z}']\,\mathbb{E}[\widetilde{Z}\widetilde{Z}']^{-1}$ Z vector
OLS slope	$\frac{\text{Cov}(Y,D)}{\text{Var}(D)}$	$\frac{d-\mathbb{E}[D]}{\text{Var}(D)}$	—
OLS (jth component)	$e_j'\,\mathbb{E}[\widetilde{X}\widetilde{X}']^{-1}\,\mathbb{E}[\widetilde{X}Y]$	$e_j'\,\mathbb{E}[\widetilde{X}\widetilde{X}']^{-1}\widetilde{x}$	$\widetilde{X}\equiv[1,D,X']'$ e_j the jth unit vector

FINDING BOUNDS THROUGH OPTIMIZATION

- Suppose we pick some $s \in \mathcal{S}$ and get IV–like estimands $\{\beta_s : s \in \mathcal{S}\}$
- Then an upper bound on $\beta^\star$ can be found by solving:

$$\overline{\beta}^\star \equiv \sup_m \beta^\star(m) \quad \text{s.t.} \quad \underbrace{m \in \mathcal{M}}_{\text{assumptions}} \quad \text{and} \quad \underbrace{\Gamma_s(m) = \beta_s \; \forall s \in \mathcal{S}}_{\text{observationally equivalent}}$$

- Lower bound $\underline{\beta}^\star$ by replacing "sup" with "inf"
- If we can find these, then we can conclude $\mathcal{B}^\star \subseteq [\underline{\beta}^\star, \overline{\beta}^\star]$

QUESTIONS

- How do we do this in practice? Can these problems be made feasible?
  Turns out the answer is yes if $\mathcal{M}$ has a particular structure
- Are these the best bounds possible? This will depend on $\mathcal{S}$
- Are these bounds useful?
  Depends on $\beta^\star$, $\mathcal{M}$ and the data in a natural and intuitive way

Feasible Computation

Linear programming

- Both $\beta^\star$ and Γ_s are *linear functions* of m
- If $\mathcal{M}$ is polyhedral, the optimization problems are linear programs
- (Finite) linear programs can be solved quickly and reliably
- However, m is a function — how do we optimize over a function?

Linear basis

- Assume that every $m \equiv (m_0, m_1) \in \mathcal{M}$ has the following form:

$$m_d(u,x) = \sum_{k=1}^{K_d} \theta_{dk} b_{dk}(u,x) \quad \text{known basis } b_{dk}, \text{ unknown } \theta_{dk}$$

- Now the optimization problems are finite and linear in θ since (e.g.)

$$\beta^\star(\overbrace{m_0, m_1}^{\text{(now } \theta\text{)}}) \equiv \sum_{d \in \{0,1\}} \sum_{k=1}^{K_d} \theta_{dk} E\left[\int_0^1 b_{dk}(u,X)\omega_d^\star(u,X,Z)\,du\right]$$

Choosing a Linear Basis

- Main concern is $\mathcal{M}$ being polyhedral after adding assumptions
- The following two choices of bases are flexible in this regard

Bernstein polynomials (BPs)

- BPs are just polynomials in a different basis:

$$b_k^K(u) \equiv \binom{K}{k} u^k (1-u)^{K-k} \quad \text{for } k = 0, 1, \ldots, K$$

- They are less collinear than ordinary ("power basis") polynomials
- Bounded, monotone, concave can be ensured by linear constraints on θ

Constant splines (CSs)

- Indicator functions: $b_k(u) = I\{c_{k-1}(u), c_k(u)\}$ — **knots** c_k
- MST show CSs can *exactly* replicate nonparametric bounds
- Idea is to choose the knots correctly (e.g. propensity score values)
- Also easy to constrain to be bounded, monotone using linear constraints

Computationally Tractable Assumptions

Boundedness

- Generally need $Y \in [\underline{y}, \overline{y}]$ to get nontrivial bounds on means
- Natural in economics, although sometimes see resistance to this
- For BPs/CSs these are box (bound) constraints on θ_{dk}

Monotonicity

- Imposes an assumption about the direction of selection
- $m_0(u)$ decreasing — positive selection bias
- Distinct from $(m_1 - m_0)(u)$ decreasing — positive selection on gains
- For BPs/CSs these are sets of inequality constraints on θ_{dk}

Separability

- $m_d(u, x) = m_d^U(u) + m_d^X(x)$ — same meaning as before
- $b_{dk}(u, x) = b_{dk}^U(u) + b_{dk}^X(x)$ — impose $\theta_{dk} = 0$ for interaction terms

Choosing IV–Like Estimands $\mathcal{S}$

Sharpness

- The bounds *necessarily* get smaller with more $\mathcal{S}$
- MST show the smallest bounds are achieved by making $\mathcal{S}$ "rich enough"
- For identification, the only drawback of increasing $\mathcal{S}$ is computational
- In statistical inference the issue is more complicated (no answer yet)

Reproducing common estimands

- Any s included in $\mathcal{S}$ must be consistent with the derived bounds
- For example, suppose one includes a $z' \to z$ Wald estimand in $\mathcal{S}$
  Then the bounds also reproduce $\text{LATE}^{z}_{z'}$ — as in BMW
- Procedure allows for extrapolation, but does not sacrifice internal validity
- Natural approach might be to include common estimands in $\mathcal{S}$
  "Doesn't hurt to look" attitude — should please all camps (?)

Numerical Illustration: MST

Treatment and instrument

- Motivated by the empirical application in MST (discussed later)
- $D \in \{0,1\}$ is purchasing an anti-malarial bed net
- $Z \in \{1,2,3,4\}$ is a randomly assigned price subsidy for purchase
- Marginal distribution $P[Z = z] = \frac{1}{4}$, propensity score $p(z)$ given by

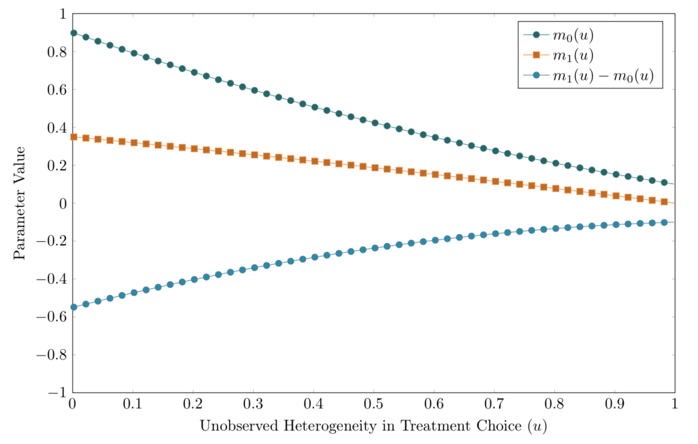
$$\underbrace{p(1) = .12}_{\text{least generous}} \qquad p(2) = .29 \qquad p(3) = .48 \qquad \underbrace{p(4) = .78}_{\text{most generous}}$$

- Roughly the type of variation we have in the data

Outcome

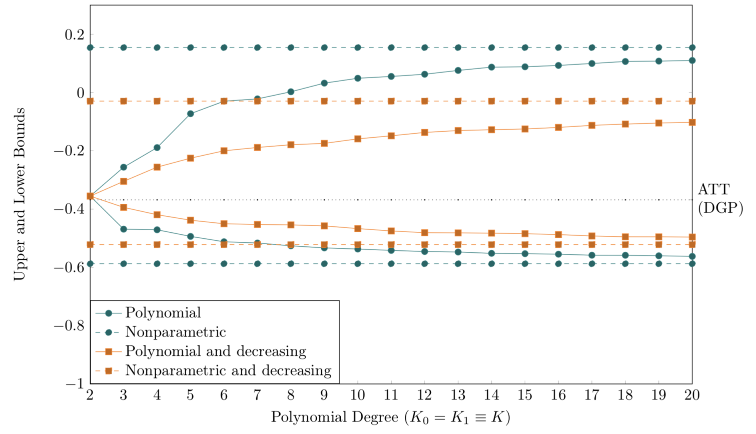
- Suppose $Y \in \{0,1\}$ is being infected by malaria
- D is endogenous if individuals know their propensity to contract malaria
  For example, they live by a lot of mosquitoes, or have poor immunity

MTR & MTE Functions in Simulation



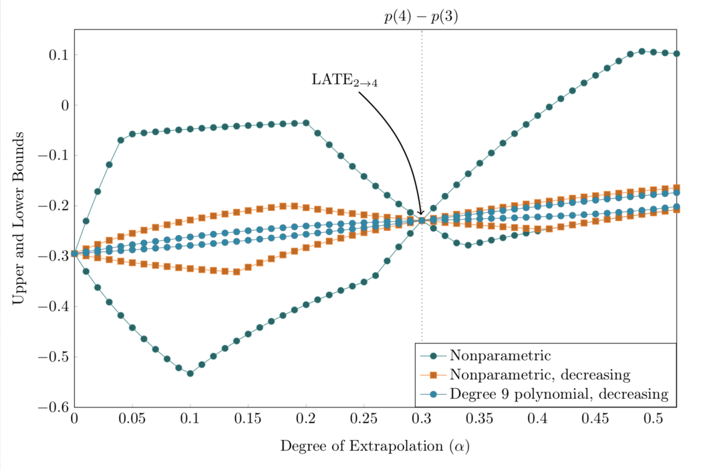
- MTRs and MTE non-constant — selection bias and selection on gain
- Those less likely to buy a net are more likely to get malaria anyway
- Those more likely to buy a net are more likely to gain *more* from it

Bounds on the ATT ($\neq$ Polynomials)



- Polynomials vs nonparametric, as well as decreasing vs unrestricted
- Polynomial bounds converge to the nonparametric bounds
- Shape restrictions can have a big impact (look at $K = 6$)

The Degree of Extrapolation



- The ATT requires substantial extrapolation, hence bounds are fairly wide
- Contrast to an **extrapolated LATE:** $E[Y_1 - Y_0 | U \in (p(2), p(3) + \alpha]]$
  Width of bounds are a function of extrapolation and assumptions

Outline

1. Parameterizing MTEs
2. Application: QQ in Fertility
3. Extrapolation: Mogstad, Santos, and Torgovitsky
4. **Application: Bed Nets**

Bed Nets (MST 2017 WP)

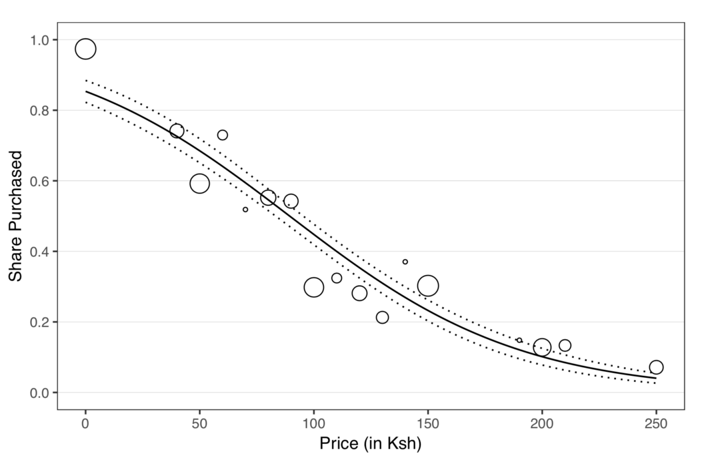
Motivation

- Motivated by Dupas (2014) — same variables as in illustration
- Dupas randomly assigned price subsidies for purchasing the net
- Difference is $Y \in \{0, 1\}$ is whether the net is actually being used
- How to promote (cost-effective) use of preventative health products?
- Encourage wide use without over-subsidizing inframarginal consumers?
- Olyset nets are a new type, so $Y_0 = 0$ (shape restriction)
- Unfortunately no data on interesting economic or health outcomes

Data

- Randomly assigned prices to 1200 households in 6 Kenyan villages
- Prices varied from 0 to 250 Ksh (about $3.80, twice daily wage)
- Overall 17 different prices offered, but only 4 or 5 to each village
- Everything will be conditioned on village (then average parameters)

Propensity Score (First Stage)



Estimate demand (p-score) with a logit (villages combined in graph)

Different Subsidy Regimes

	(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	

- PRTE of free provision (with full-take up) vs. no provision
- Common parameter of interest ($\equiv$ ATE) — but maybe not so interesting
- Bounds depend on $\mathcal{S}$ and K — can be empty (misspecification)

Different Subsidy Regimes

	(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	∅	.6365	.5602	.5269	.4487
Upper	.6772	.7269	.7362	.7445	.7515	.6521	.7269	.7362	.7445	.7515	∅	.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	∅	.6758	.6506	.6214	.5573
Upper	.7049	.8140	.8469	.8817	.9732	.6600	.7085	.7172	.7275	.7941	∅	.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	

- PRTE of free provision vs. market price one year later (150 Ksh)
- Demand at 150 Ksh is predicted from the logit estimate
- This is a LATE, but not nonparametrically point id'd (for every village)

Different Subsidy Regimes

	(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	∅	.6365	.5602	.5269	.4487
Upper	.6772	.7269	.7362	.7445	.7515	.6521	.7269	.7362	.7445	.7515	∅	.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	∅	.6758	.6506	.6214	.5573
Upper	.7049	.8140	.8469	.8817	.9732	.6600	.7085	.7172	.7275	.7941	∅	.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	

- Statistical inference is a challenge for partial identification approaches
- We know these confidence intervals are conservative (excessively wide)
- There is rapid progress here, and work continues …

QUESTIONS?