An Adversarial Approach to Identification and Inference

McGill, March 2025

Irene Botosaru (McMaster), Isaac Loh (UNC Wilmington), Chris Muris (McMaster)

Fixed effects in linear and nonlinear panel models

Linear panel models

The linear panel model with fixed effects is one of the core tools of applied economists

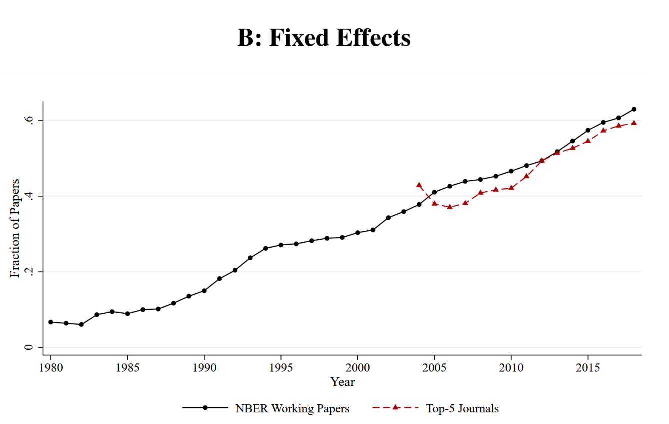


Figure 1: Currie et al., AEA P+P 2020

The linear panel outcome equation:

$$Y_{it} = \alpha_i + X_{it}'\beta + U_{it},\ t = 1, \cdots, T$$

- **fixed effects** (FE) α_i:
  - control for unobserved heterogeneity
  - no restriction on joint distribution $(\alpha_i, X_{i1}, \cdots, X_{iT})$
- OLS of Y_{it} on X_{it} inconsistent for β

The linear panel outcome equation:

$$Y_{it} = \alpha_i + X_{it}'\beta + U_{it},\ t = 1, \cdots, T$$

- **fixed effects** (FE) α_i:
  - control for unobserved heterogeneity
  - no restriction on joint distribution $(\alpha_i, X_{i1}, \cdots, X_{iT})$
- OLS of Y_{it} on X_{it} inconsistent for β

In the linear panel model ($+$ strict exogeneity):

- OLS of $Y_{it} - \overline{Y_i}$ on $X_{it} - \overline{X_i}$ is consistent for β
- β is regression coefficient *and* partial effect
- can estimate (distribution of) $\alpha_i = Y_{it} - X_{it}'\beta - U_{it}$

Nonlinear panel models

Textbook binary choice panel with fixed effects:

$$Y_{it} = 1\left\{\alpha_i + X_{it}'\beta + U_{it} \geq 0\right\},\ t = 1, \cdots, T$$

and where $U_i | X_i \sim F$.

Nonlinear panel models

Textbook binary choice panel with fixed effects:

$$Y_{it} = 1\left\{\alpha_i + X_{it}'\beta + U_{it} \geq 0\right\},\ t = 1, \cdots, T$$

and where $U_i | X_i \sim F$.

If T is fixed, then

- (β, U) does not pin down α_i
  - example: if $X_{it}'\beta + U_{it} = 0$ then any $\alpha_i \geq 0$ is compatible with $Y_{it} = 1$
- (β, F) does not pin down distribution of FEs
- distribution of FEs is **partially identified**

$\Rightarrow$ Partial identification is widespread in nonlinear panels

Consequence 1: partial identification of β

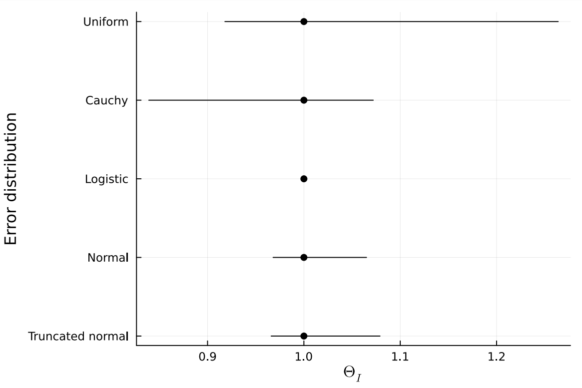


Figure 2: Identified sets in binary choice models, Botosaru, Loh, Muris (2025+)

- partial ID of FE spills over to β
- with exceptions (logit)
- identified sets tend to be small
  - figure: worst case, $X, Y \in \{0, 1\},\ T = 2$
- **today's paper:** characterize identified set for β:
  - in a large class of models
  - with point or partial ID

Consequence 2: partial identification of partial effects

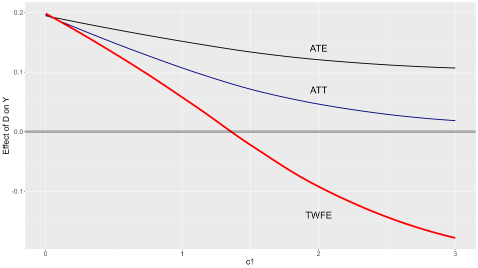
- in applications, focus on **counterfactual choice probabilities**

$$E\left[1\{\alpha_i + x^* \beta + U_{it} \geq 0\} | X_i = x\right]$$

  and differences/derivatives (**partial effects**)
- partial effects depend on FE distribution
- even if β is point identified, partial effects may not be
- estimation and inference are very challenging
- traditional advice: *use random effects or linear models if you want partial effects*
- **today's paper**:
  - ID common parameters **and** PE ...
  - ... in a general class of models.

Does it matter?

1. yes: many models are nonlinear
   - textbook models: binary and (un)ordered choice
   - structural models
2. can't we just do OLS?
   - for textbook cross-sectional models, OLS approximates average partial effects
   - for panels, just do TWFE?



- $Y_t = 1\{\alpha + D_t \times 1 + \lambda_t + U_t \geq 0\}$
- effect of D on Y is positive
- simple DiD ($D_1 = 0, D_2 =$ coin flip)
- standard logistic errors (U_1, U_2)
- fixed effects: $\alpha = -0.5 + c_1 D_2$
- time effects: $\lambda_1 = 0,\ \lambda_2 = 1$

Recap

- **linear** panels with **fixed effects** are central to applied economics
- would like to use fixed effects in nonlinear panel models, too, but:
  - in most models, β not point identified
  - even if β available, cannot get partial effects
- OLS fails due to combination of FE, time effects, and nonlinearity
- partial identification seems unavoidable

This motivates our **adversarial approach to identification**.

This motivates our **adversarial approach to identification**.

Overview:

1. Introduction
2. Model
3. Main result
4. Convexity and semiparametric models
5. Computation via linear programming
6. Parametric models
7. Results for nonlinear panels

Introduction

What? New framework for partial identification + inference

Why?

- Applicable to a **large class of econometric models**
- Sharp identification of **structural and counterfactual** parameters
- Computational efficiency via **linear programming**
- **Inference** via sample analogs (paper)
- Break new ground in **nonlinear panels**

How?

- Construct a new **discrepancy function** with maximin formulation
- Leverage:
  - **convexity** of the set of **model probabilities**
  - a **linearity** property of many econometric models

Semiparametric binary choice model

- example: *semiparametric binary choice (SPBC) model* (Manski, 1975, 1985)
- cross-sectional binary choice model with

$$Y = 1\{X'\theta + U \geq 0\}$$

and a median-zero assumption

$$\text{med}(U|X) = 0.$$

- if all regressors are discrete:
  - β partially identified (even with a scale normalization)
  - partial effects are partially identified

- SPBC model has three ingredients:
  - unobserved error term U
  - observed regressors X
  - observed outcome $Y \in \{0, 1\}$
- **inputs** $W = (X, U)$ have **probability measure** γ
  - we know: its marginal distribution with respect to X
  - we know: $P(U \leq 0 | X = x) = 0.5$ for each x
- **outputs** $Z = (X, Y)$ have **probability measure** μ_Z
  - observed one: μ_Z^*
- the focus on γ and μ is key to our analysis

- SPBC model has three properties:
  1. model is a map $(\theta, \gamma) \mapsto \mu_{Z,(\theta,\gamma)}$
  2. at each θ, map from γ to μ is **linear**
  3. set of all γ compatible with "what we know" is **convex**
     - set of all median-zero γ with known X- marginals is convex
- paper: properties 1-3 hold for many econometric models
  - derive results from these basic properties
  - we use convex analysis, functional analysis, and convex functional analysis

Model

Setup and Notation

- Z: observable Borel measurable random variable, support $\mathcal{Z}$
- μ_Z^*: true probability measure of Z
- $\mu_{Z,(\theta,\gamma)}$: model probability for each $\theta \in \Theta$ and $\gamma \in \Gamma_\theta$
  - Θ: parameter space for parameter of interest θ
  - Γ_θ: parameter space for auxiliary parameters γ

Setup and Notation

- Z: observable Borel measurable random variable, support $\mathcal{Z}$
- μ^*_Z: true probability measure of Z
- $\mu_{Z,(\theta,\gamma)}$: model probability for each $\theta \in \Theta$ and $\gamma \in \Gamma_\theta$
    - Θ: parameter space for parameter of interest θ
    - Γ_θ: parameter space for auxiliary parameters γ

In many models: γ is distribution of unobserved heterogeneity.

Model Probabilities

Set of model probabilities

$$\mathcal{M}_\theta \equiv \{\mu_{Z,(\theta,\gamma)} : \gamma \in \Gamma_\theta\}$$

for a fixed θ.

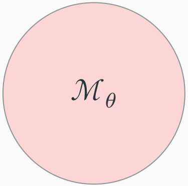


Figure 3: Each point corresponds to a model probability for a given (θ, γ).

Model Probabilities

Set of model probabilities

$$\mathcal{M}_\theta \equiv \{\mu_{Z,(\theta,\gamma)} : \gamma \in \Gamma_\theta\}$$

for a fixed θ.

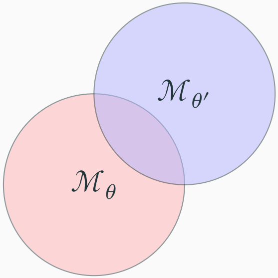


Figure 3: Each point corresponds to a model probability for a given (θ, γ).

Identified Set

Identified set for θ is:

$$\Theta_{\mathrm{I}} \equiv \{\theta \in \Theta : \mu_Z^* \in \overline{\mathcal{M}}_\theta\}$$

where $\overline{\mathcal{M}}_\theta$ is closure of $\mathcal{M}_\theta$.

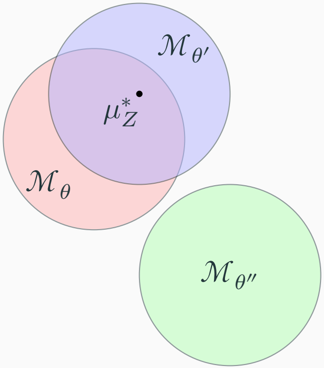


Figure 4: $\{\theta, \theta'\} \subset \Theta_I,\ \theta'' \notin \Theta_I$.

Identified set for θ is:

$$\Theta_{\mathrm{I}} \equiv \{\theta \in \Theta : \mu_Z^* \in \overline{\mathcal{M}}_\theta\}.$$

Notes:

- **Problem:** this definition of Θ_{I} is **not tractable.**
- closure of $\mathcal{M}_\theta$

$$\overline{\mathcal{M}}_\theta \equiv \{m \in \mathcal{P}(\mathcal{Z}) : \forall \varepsilon > 0, \exists \gamma \in \Gamma_\theta \text{ such that } d_{\mathrm{TV}}(m, \mu_{Z,(\theta,\gamma)}) < \varepsilon\}$$

includes limit points that matter for fixed effects models.

Main result

Discrepancy function

- goal: tractability of the identification problem

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$

- $\mathbb{E}_{\mu_Z^*}[\phi]$: what feature ϕ looks like in data

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$

  - $\mathbb{E}_{\mu_Z^*}[\phi]$: what feature ϕ looks like in data
  - $\mathbb{E}_\mu[\phi]$: what ϕ looks like according to model, under (θ, γ)

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

  where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$
  - $\mathbb{E}_{\mu_Z^*}[\phi]$: what feature ϕ looks like in data
  - $\mathbb{E}_\mu[\phi]$: what ϕ looks like according to model, under (θ, γ)
- intuition: view as an **adversarial game**

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \Big(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \Big)$$

  where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$
  - $\mathbb{E}_{\mu_Z^*}[\phi]$: what feature ϕ looks like in data
  - $\mathbb{E}_\mu[\phi]$: what ϕ looks like according to model, under (θ, γ)
- intuition: view as an **adversarial game**
  - critic (sup) chooses feature ϕ to maximize discrepancy

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu^*_Z}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

  where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$
  - $\mathbb{E}_{\mu^*_Z}[\phi]$: what feature ϕ looks like in data
  - $\mathbb{E}_\mu[\phi]$: what ϕ looks like according to model, under (θ, γ)
- intuition: view as an **adversarial game**
  - critic (sup) chooses feature ϕ to maximize discrepancy
  - defender (inf) chooses measure μ ("chooses γ") to minimize discrepancy

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

  where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$
  - $\mathbb{E}_{\mu_Z^*}[\phi]$: what feature ϕ looks like in data
  - $\mathbb{E}_\mu[\phi]$: what ϕ looks like according to model, under (θ, γ)
- intuition: view as an **adversarial game**
  - critic (sup) chooses feature ϕ to maximize discrepancy
  - defender (inf) chooses measure μ (“chooses γ”) to minimize discrepancy
  - $T(\theta) > 0$: critic finds a feature where model fails to replicate data at θ

Discrepancy function

- goal: tractability of the identification problem
- from the definitions of $\mathcal{M}_\theta$ and Θ_I, we **construct** a discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

  where $\Phi_b(\mathcal{Z})$ is the set of bounded Borel measurable functions from $\mathcal{Z}$ to $[0, 1]$
  - $\mathbb{E}_{\mu_Z^*}[\phi]$: what feature ϕ looks like in data
  - $\mathbb{E}_\mu[\phi]$: what ϕ looks like according to model, under (θ, γ)
- intuition: view as an **adversarial game**
  - critic (sup) chooses feature ϕ to maximize discrepancy
  - defender (inf) chooses measure μ ("chooses γ") to minimize discrepancy
  - $T(\theta) > 0$: critic finds a feature where model fails to replicate data at θ
  - $T(\theta) = 0$: defender can always match all observed features at θ

Discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

is central to the paper.

- for identification: main result
- for computation: $T(\theta)$ can be evaluated using LP
- for inference: results are based on $T_n(\theta)$ (paper)

Discrepancy function

$$T(\theta) \equiv \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu^*_Z}[\phi] - \mathbb{E}_\mu[\phi] \right)$$

is central to the paper.

- for identification: main result
- for computation: $T(\theta)$ can be evaluated using LP
- for inference: results are based on $T_n(\theta)$ (paper)

Main results:

- define $\Theta_{\mathrm{MI}} \equiv \{\theta \in \Theta : T(\theta) = 0\}$
- under mild conditions, $\Theta_{\mathrm{MI}} = \Theta_{\mathrm{I}}$

Main result

Assumption (1)

$\mathcal{Z}$ is a Polish space.

Assumption (2)

For all $\theta \in \Theta$, there exists some σ-finite positive measure $\lambda_\theta \in \mathfrak{B}(\mathcal{Z})$ with respect to which every $\mu \in \mathcal{M}_\theta$ is continuous.

Theorem (1)
Let Assumptions 1 and 2 hold.

For any $\mu_Z^* \in \mathcal{P}(\mathcal{Z})$, $\Theta_{\mathrm{I}} \subseteq \Theta_{\mathrm{MI}}$.

Additionally, let $\overline{\mathcal{M}}_\theta$ be convex for all θ. Then $\Theta_{\mathrm{I}} = \Theta_{\mathrm{MI}}$.

Discussion: convexity

- convexity of $\mathcal{M}_\theta$ is important (else outer set)
- main result says:

$$T(\theta) = 0 \Leftrightarrow \mu_Z^* \in \overline{\text{co}}(\mathcal{M}_\theta)$$

with

$$T(\theta) = \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi]\right)$$

- Same as checking, for each ϕ,

$$\mathbb{E}_{\mu_Z^*}[\phi] \leq \inf_{\mu \in \mathcal{M}_\theta} \mathbb{E}_\mu[\phi]$$

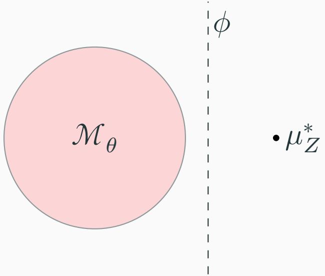


Figure 5: $\theta \notin \Theta_I$, with separating hyperplane

Convexity in semiparametric models

Models with unobserved heterogeneity

- we consider a class of models where:
  - convexity is often satisfied and easy to verify
  - extremal point characterization for computational tractability
- consider model where γ is distribution of unobserved heterogeneity:
  - input variables $W \in \mathcal{W}$, with probability measure $\gamma \in \Gamma_\theta(\mathcal{W})$
  - observed variables $Z \in \mathcal{Z}$
  - map $\psi_\theta : \mathcal{W} \mapsto \mathcal{Z}$ known up to θ
  - pushforward measure $(\psi_\theta)_* : \Gamma_\theta(\mathcal{W}) \to \mathcal{P}(\mathcal{Z})$
  - $\mathcal{M}_\theta = (\psi_\theta)_* \Gamma_\theta(\mathcal{W})$.

- nests semiparametric regression models with $Y = h(X, U; \theta)$:
  - $Y \in \mathcal{Y}$: outcome variables
  - $X \in \mathcal{X}$: observed regressors, instruments, ...
  - $U \in \mathcal{U}$: unobserved variables (error terms, fixed effects, ...)
  - $h : \mathcal{X} \times \mathcal{U} \times \Theta \to \mathcal{Y}$: structural function
  - with (exogeneity and other) restrictions on the distribution of (U, X).
- mapping the notation:
  - inputs $W = (X, U)$ with probability measure γ
  - observed $Z = (X, Y)$
  - $\psi_\theta(x, u) = (x, h(x, u; \theta))$
  - restrictions incorporated in Γ_θ

Corollary 1: Convexity and sharp identification

Assumption (3)
For any θ, there exists a map $\psi_\theta : \mathcal{W} \mapsto \mathcal{Z}$ such that for any $\gamma \in \Gamma_\theta$, the model probability is given by:

$$\mu_{Z,(\theta,\gamma)}(S) = (\psi_\theta)_* \gamma(S), \text{ for any Borel } S \subseteq \mathcal{Z}.$$

Corollary (1)
Let Assumptions 1-3 hold, and assume that $\Gamma_\theta(\mathcal{W})$ is convex. Then, $\mathcal{M}_\theta$ is convex and $\Theta_{\mathrm{I}} = \Theta_{\mathrm{MI}}$.

Many econometric models have convex $\Gamma_\theta(\mathcal{W})$:

1. unrestricted Γ_θ if U are fixed effects
2. linear restrictions on Γ_θ, e.g.:
   - mean or median restrictions on (U, X)
   - θ includes partial effects or other moments of U
   - see maximum score illustration
3. parametric restrictions: stay tuned

Many econometric models have convex $\Gamma_\theta(\mathcal{W})$:

1. unrestricted Γ_θ if U are fixed effects
2. linear restrictions on Γ_θ, e.g.:
    - mean or median restrictions on (U, X)
    - θ includes partial effects or other moments of U
    - see maximum score illustration
3. parametric restrictions: stay tuned

Not covered: independence restrictions involving multiple unobservables.

Extremal Point Characterization

- **Extremal point result:** If no restrictions, the inner inf simplifies:

$$T(\theta) = \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \Big(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \Big)$$

$$= \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{w \in \mathcal{W}} \Big(\mathbb{E}_{\mu_Z^*}[\phi] - \phi \circ \psi_\theta(w) \Big)$$

- similar for linear restrictions
- shows tractability of $T(\theta)$ and Θ_{MI}:
  - search over $\mathcal{W}$...
  - ... instead of search over distributions over $\mathcal{W}$

Notation:

- unrestricted: $\Gamma_\theta(\mathcal{W}) = \mathcal{P}(\mathcal{W})$ is the set of all probability measures on $\mathcal{W}$.
- linear restrictions,

$$\mathcal{P}(\mathcal{W})^g \equiv \{\gamma \in \mathcal{P}(\mathcal{W}) : g \in L^1(\gamma) \text{ and } \mathbb{E}_\gamma g(\theta, w) = 0\}$$

with $g : \Theta \times \mathcal{W} \to \mathbb{R}^{d_g}$ a vector of known functions

Assumption (4)
$\mathcal{W}$ is a Polish space.

Theorem
Under Assumptions 1-4:

if $\Gamma_\theta(\mathcal{W}) = \mathcal{P}(\mathcal{W})$, *then* $\theta \in \Theta_{\mathrm{I}}$ *if and only if*

$$\mathbb{E}_{\mu_Z^*}[\phi] \leq \sup_{w \in \mathcal{W}} (\phi \circ \psi_\theta)(w)$$

for all $\phi \in \Phi_b(\mathcal{Z})$

if $\Gamma_\theta(\mathcal{W}) = \mathcal{P}(\mathcal{W})^g$, *then* $\theta \in \Theta_{\mathrm{I}}$ *if and only if*

$$E_{\mu_Z^*}[\phi] \leq \sup_{\{c_j, w_j\}_{j=1}^{d_g+1}} \sum_{j=1}^{d_g+1} c_j \left(\phi \circ \psi_\theta(w_j)\right) \text{ for all } \phi \in \Phi_b(\mathcal{Z})$$

$$\text{subject to } \sum_{j=1}^{d_g+1} c_j g(\theta, w_j) = 0, \; \sum_{j=1}^{d_g+1} c_j = 1, \; c_j \geq 0. \tag{1}$$

Illustration: SPBC

Consider the binary choice model:

$$Y = 1\{\beta_0 + \beta_1 X - U \geq 0\}$$

where:

- $Y \in \{0, 1\}$: binary outcome
- $X \in \{x_1, \cdots, x_K\}$: discrete explanatory variable
- $U \in \mathbb{R}$: scalar error term
- $\beta = (\beta_0, \beta_1)$: unknown regression coefficient

Assumption: $\mathbb{P}(U \leq 0 | X = x_k) = \frac{1}{2}$, for $k = 1, \cdots, K$.

Fits our framework:

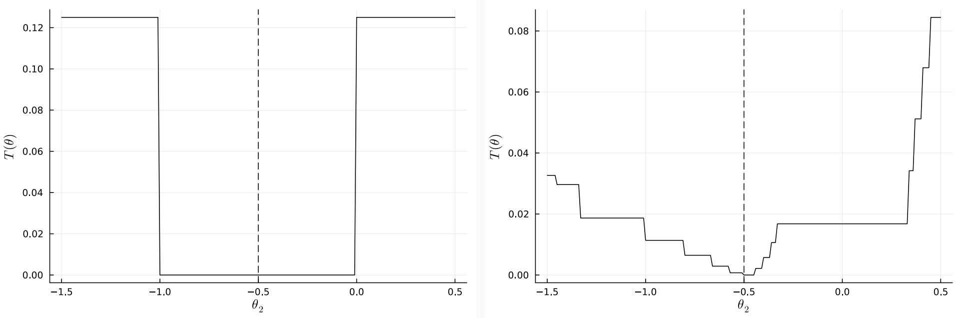
- $Z = (Y, X)$, $W = (X, U)$,
- $\mathcal{Z} = \{0, 1\} \times \{x_1, \cdots, x_K\}$, $\mathcal{W} = \{x_1, \cdots, x_K\} \times \mathbb{R}$
- $\psi_\theta(x, u) = (1\{\beta_0 + \beta_1 x - u \geq 0\}, x)$
- $\Gamma_\theta(\mathcal{W})$: set of distributions of W satisfying the (linear!) median restriction, i.e.

$$E_{\gamma_{U|x}} [\tilde{g}(U, \theta) \mid X = x] = 0,$$

  with

$$\tilde{g}(U, \theta) = 1\{U \leq 0\} - \frac{1}{2}$$

- we provide Corollary 2 for working with $U|X$, continuous X



(a) Design 1 **(b)** Design 4: $\mathcal{X} = \{-3, -2.75, \cdots, 3\}$

Figure 6: $T(\theta)$ for maximum score.

- counterfactual choice probabilities:

$$\tau_k(x^*) \equiv E\left[1\{\beta_0 + x^*\beta_1 - U \geq 0 | X = x_k\}\right]$$

and set

$$\theta = (\beta, \tau_1(x^*), \cdots, \tau_K(x^*))$$

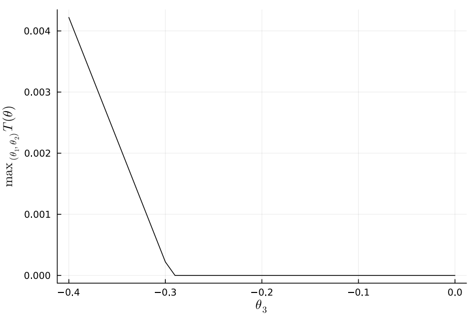
- median-zero restrictions and definition of counterfactual probabilities as

$$E_{\gamma_{U|x}}\left[\tilde{g}(U, \theta) \mid X = x\right] = 0$$

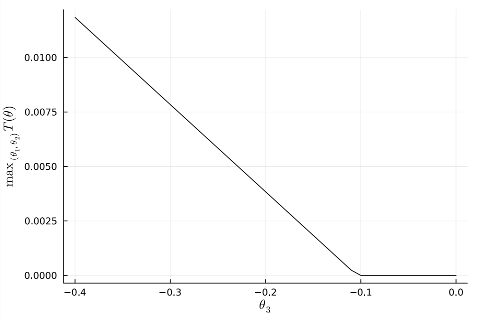
with

$$\tilde{g}(U, \theta) = \begin{bmatrix} 1\{U \leq 0\} - \frac{1}{2} \\ 1\{\beta_0 + x^*\beta_1 - U \geq 0\} - \tau_k(x^*) \end{bmatrix}. \tag{2}$$

- counterfactual choice probabilities are functionals of γ
- recall: Γ_θ can depend on θ



(a) Design 4



(b) Design 4b: U|X discrete uniform

Figure 7: Identified set for the partial effect in the maximum score model

Computation via LP

Discrepancy function, pmf

- computing the identified set requires evaluating

$$T(\theta) = \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \Big(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi] \Big)$$

- involves optimization over measures μ and functions ϕ
- reduces to a linear program (LP)
- to make things concrete:
  - use pmf instead of probability measures
  - demo using the SPBC model

- discretize support: $\mathcal{Z} = \{z_1, \cdots, z_L\}$, $\mathcal{W} = \{w_1, \cdots, w_M\}$

- discretize support: $\mathcal{Z} = \{z_1, \cdots, z_L\}$, $\mathcal{W} = \{w_1, \cdots, w_M\}$
- represent probability measure μ_Z^* by pmf

$$p_Z^* = \left(p_{Z,l}^*\right) = \left(p_{Z,1}^*, \cdots, p_{Z,L}^*\right)$$

- discretize support: $\mathcal{Z} = \{z_1, \cdots, z_L\}$, $\mathcal{W} = \{w_1, \cdots, w_M\}$
- represent probability measure μ_Z^* by pmf

$$p_Z^* = \left(p_{Z,l}^*\right) = \left(p_{Z,1}^*, \cdots, p_{Z,L}^*\right)$$

- first term in $T(\theta) = \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi]\right)$ is

$$E_{\mu_Z^*}[\phi] = \sum_{l=1}^{L} \phi(z_l) p_{Z,l}^* = \phi' p_Z^*$$

- discretize support: $\mathcal{Z} = \{z_1, \cdots, z_L\}$, $\mathcal{W} = \{w_1, \cdots, w_M\}$
- represent probability measure μ_Z^* by pmf

$$p_Z^* = \left(p_{Z,l}^*\right) = \left(p_{Z,1}^*, \cdots, p_{Z,L}^*\right)$$

- first term in $T(\theta) = \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi]\right)$ is

$$E_{\mu_Z^*}[\phi] = \sum_{l=1}^{L} \phi(z_l) p_{Z,l}^* = \phi' p_Z^*$$

- represent every model probability $\mu_{Z,(\theta,\gamma)}$ by pmf $p_Z^{(\theta,\gamma)}$

- discretize support: $\mathcal{Z} = \{z_1, \cdots, z_L\}$, $\mathcal{W} = \{w_1, \cdots, w_M\}$
- represent probability measure μ_Z^* by pmf

$$p_Z^* = \left(p_{Z,l}^*\right) = \left(p_{Z,1}^*, \cdots, p_{Z,L}^*\right)$$

- first term in $T(\theta) = \sup_{\phi \in \Phi_b(\mathcal{Z})} \inf_{\mu \in \mathcal{M}_\theta} \left(\mathbb{E}_{\mu_Z^*}[\phi] - \mathbb{E}_\mu[\phi]\right)$ is

$$E_{\mu_Z^*}[\phi] = \sum_{l=1}^{L} \phi(z_l) p_{Z,l}^* = \phi' p_Z^*$$

- represent every model probability $\mu_{Z,(\theta,\gamma)}$ by pmf $p_Z^{(\theta,\gamma)}$
- second term, for some (θ, γ), is

$$E_{\mu_{Z,(\theta,\gamma)}}[\phi] = \phi' p_Z^{(\theta,\gamma)}$$

- represent every γ for W by a pmf

$$p_W = \left(p_{W,m}\right) = (p_{W,1}, \cdots, p_{W,M})$$

- represent every γ for W by a pmf

$$p_W = \left(p_{W,m}\right) = (p_{W,1}, \cdots, p_{W,M})$$

- to enforce $\gamma \in \Gamma_\theta$, impose $p_W \geq 0,\ A_\theta p_W = b_\theta$

- represent every γ for W by a pmf

$$p_W = \left(p_{W,m}\right) = (p_{W,1}, \cdots, p_{W,M})$$

- to enforce $\gamma \in \Gamma_\theta$, impose $p_W \geq 0,\ A_\theta p_W = b_\theta$
- there exists a $L \times M$ matrix $\widetilde{C}_\theta$ such that

$$p_Z^{(\theta,\gamma)} = \widetilde{C}_\theta p_W$$

- represent every γ for W by a pmf

$$p_W = (p_{W,m}) = (p_{W,1}, \cdots, p_{W,M})$$

- to enforce $\gamma \in \Gamma_\theta$, impose $p_W \geq 0,\ A_\theta p_W = b_\theta$
- there exists a $L \times M$ matrix $\widetilde{C}_\theta$ such that

$$p_Z^{(\theta,\gamma)} = \widetilde{C}_\theta p_W$$

  - given θ, there is a linear map from pmf of W to a model pmf of Z

- represent every γ for W by a pmf

$$p_W = \left(p_{W,m}\right) = (p_{W,1}, \cdots, p_{W,M})$$

- to enforce $\gamma \in \Gamma_\theta$, impose $p_W \geq 0,\ A_\theta p_W = b_\theta$
- there exists a $L \times M$ matrix $\widetilde{C}_\theta$ such that

$$p_Z^{(\theta,\gamma)} = \widetilde{C}_\theta p_W$$

  - given θ, there is a linear map from pmf of W to a model pmf of Z
- write $T(\theta)$ as

$$T(\theta) = \max_{\phi \in \mathbb{R}^L :\ 0 \leq \phi \leq 1} \quad \min_{p_W \in \mathbb{R}^M :\ p_W \geq 0,\ A_\theta p_W = b_\theta} \quad \underbrace{\phi' p_Z^* - \phi' \widetilde{C}_\theta p_W}_{\phi' C_\theta p_W}$$

Semiparametric binary choice

- SPBC with binary regressor and an error term with 3 points of support,

$$Y \in \{0,1\},\ X \in \{x_1, x_2\} \subset \mathbb{R},\ U \in \{-1,0,1\}$$

  and

$$Y = 1\{\beta_1 + X\beta_2 - U \geq 0\}$$

- define

$$p_Z^{(\theta,\gamma)} = \begin{bmatrix} p_Z^{(\theta,\gamma)}(x_1,1) \\ p_Z^{(\theta,\gamma)}(x_1,0) \\ p_Z^{(\theta,\gamma)}(x_2,1) \\ p_Z^{(\theta,\gamma)}(x_2,0) \end{bmatrix}, \quad p_W = \begin{bmatrix} p_W(x_1,-1) \\ p_W(x_1,0) \\ p_W(x_1,1) \\ p_W(x_2,-1) \\ p_W(x_2,0) \\ p_W(x_2,1) \end{bmatrix}$$

- next slide: $p_Z^{(\theta,\gamma)} = \widetilde{C}_\theta p_W$

$$\begin{bmatrix} p_Z^{(\theta,\gamma)}(x_1,1) \\ p_Z^{(\theta,\gamma)}(x_1,0) \\ p_Z^{(\theta,\gamma)}(x_2,1) \\ p_Z^{(\theta,\gamma)}(x_2,0) \end{bmatrix} = \begin{bmatrix} 1\{\widetilde{x}_1'\theta+1\geq 0\} & 1\{\widetilde{x}_1'\theta\geq 0\} & 1\{\widetilde{x}_1'\theta-1\geq 0\} & 0 & 0 & 0 \\ 1\{\widetilde{x}_1'\theta+1< 0\} & 1\{\widetilde{x}_1'\theta< 0\} & 1\{\widetilde{x}_1'\theta-1< 0\} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1\{\widetilde{x}_2'\theta+1\geq 0\} & 1\{\widetilde{x}_2'\theta\geq 0\} & 1\{\widetilde{x}_2'\theta-1\geq 0\} \\ 0 & 0 & 0 & 1\{\widetilde{x}_2'\theta+1< 0\} & 1\{\widetilde{x}_2'\theta< 0\} & 1\{\widetilde{x}_2'\theta-1< 0\} \end{bmatrix} \begin{bmatrix} p_W(x_1,-1) \\ p_W(x_1,0) \\ p_W(x_1,1) \\ p_W(x_2,-1) \\ p_W(x_2,0) \\ p_W(x_2,1) \end{bmatrix}$$

where $\tilde{x} = (1, x)$.

- $p_Z^{(\theta,\gamma)} = \widetilde{C}_\theta p_W$
- $\widetilde{C}_\theta$ is a known matrix

- model restrictions are $A_\theta p_W = b_\theta$, with:

$$A_\theta = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}, \; b_\theta = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

- first constraint: p_W is a probability vector, $\sum_{m=1}^{M} p_{W,m} = 1$
- constraints 2 and 3 ensure that the median is zero: $\sum_{u<0} P(X = x, U = u) = \sum_{u>0} P(X = x, U = u)$.

Duality

Inner minimization problem is LP with coefficients $\phi' C_\theta$ on decision variables p_W.

$$\begin{aligned} \text{min} \quad & \phi' C_\theta p_W \\ \text{subject to} \quad & A_\theta p_W = b_\theta, \\ & p_W \geq 0, \end{aligned}$$

Duality

Inner minimization problem is LP with coefficients $\phi' C_\theta$ on decision variables p_W.

$$\begin{aligned} \text{min} \quad & \phi' C_\theta p_W \\ \text{subject to} \quad & A_\theta p_W = b_\theta, \\ & p_W \geq 0, \end{aligned}$$

Its dual is:

$$\begin{aligned} \text{max} \quad & \lambda' b_\theta \\ \text{subject to} \quad & \lambda' A_\theta \leq \phi' C_\theta, \end{aligned}$$

with λ the dual variables for constraints in primal.

Strong duality holds, so can replace inner minimization by its dual.

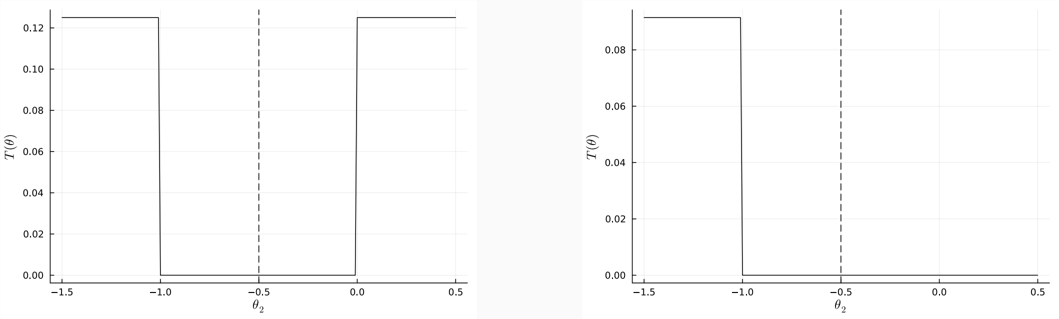
Substituting dual into $T(\theta)$ means we now maximize over ϕ (as before) and λ (dual):

$$T(\theta) = \begin{cases} \text{max}_{\lambda,\phi} & \lambda' b_\theta \\ \text{subject to} & A'_\theta \lambda \leq C'_\theta \phi, \\ & 0 \leq \phi \leq 1. \end{cases} \tag{3}$$

This is a LP: we have achieved tractability for $T(\theta)$ and therefore Θ_I.

Takeaway:

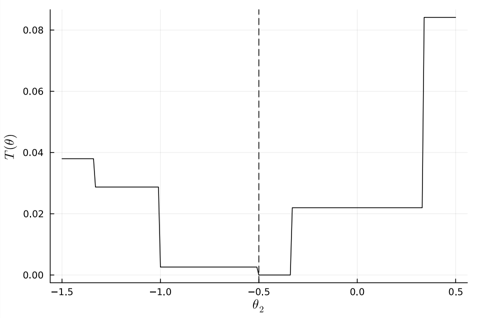
- to determine if $\theta \in \Theta_I$, solve an LP for $T(\theta)$ and check $T(\theta) \leq 0$.
- negligible computation time, even for very large (L, M) (stay tuned)
- writing code for a specific model is trivial. LP solver only needs:
    1. supports $\mathcal{Z}$, $\mathcal{W}$
    2. true parameter values (θ^*, p_W^*) to compute p_Z^*
    3. the matrix $\widetilde{C}_\theta$
    4. restrictions (A_θ, b_θ)



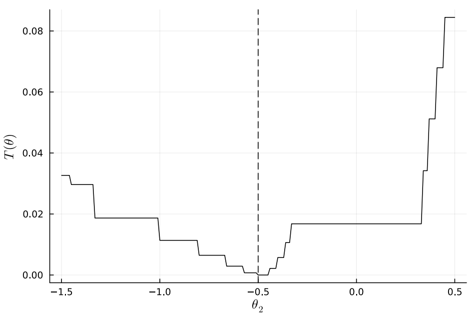
(a) Design 1

(b) Design 2: $\mathcal{U} = \{-5, -4.9, \cdots, 5\}$

Figure 8: $T(\theta)$ for maximum score.



(a) Design 3: $\mathcal{X} = \{-3, -2, \cdots, 3\}$



(b) Design 4: $\mathcal{X} = \{-3, -2.75, \cdots, 3\}$

Figure 9: $T(\theta)$ for maximum score.

Computational efficiency:

Design	θ_2	θ_3	K_u	K_x
1	0.0024	0.0016	3	2
2	0.0033	0.0023	101	2
3	0.0083	0.0077	101	7
4	0.0522	0.0536	101	25

Table 1: Time, in seconds, for one evaluation of $T(\theta)$.

Competing methods in partial identification: Design 4 picture would take days.

Models with parametric restrictions

Key idea: Parametric error distributions

- What if some components of latent variables follow known distributions?
- Consider model with:

$$Y = h(X, \alpha, V; \beta), \quad (X, \alpha) \sim \gamma_{X,\alpha}, \quad V|X, \alpha \sim F_{V|X,\alpha;\beta}$$

- $\gamma_{X,\alpha}$ is unknown (e.g., fixed effects)
- $F_{V|X,\alpha;\beta}$ is known up to parameter β

Key idea: Parametric error distributions

- What if some components of latent variables follow known distributions?
- Consider model with:

$$Y = h(X, \alpha, V; \beta), \quad (X, \alpha) \sim \gamma_{X,\alpha}, \quad V|X, \alpha \sim F_{V|X,\alpha;\beta}$$

- $\gamma_{X,\alpha}$ is unknown (e.g., fixed effects)
- $F_{V|X,\alpha;\beta}$ is known up to parameter β
- **Main insight**: Similar framework applies!
  - Set of model probabilities still has convex structure
  - Linear structure is maintained, but in a different form

Linear integral operator

- For parametric error models, the linear mapping changes form:
  - Before: pushforward measure $(\psi_\theta)_*$
  - Now: linear integral operator $\mathcal{L}_\beta$

$$\mathcal{L}_\beta[\gamma_{X,\alpha}](S) \equiv \int_{\mathcal{X}\times\mathcal{A}} \int_{\mathcal{Y}} \mathbf{1}_S(y,x) f_{Y|X,\alpha}(y \mid x, a; \beta)\, \mathrm{d}\lambda_y(y)\, \mathrm{d}\gamma_{X,\alpha}(x,a)$$

- Still a linear map: $\mu_{Z,(\beta,\gamma_{X,\alpha})} = \mathcal{L}_\beta[\gamma_{X,\alpha}]$
- Model probabilities $\mathcal{M}_\beta = \{\mathcal{L}_\beta[\gamma_{X,\alpha}] : \gamma_{X,\alpha} \in \mathcal{P}(\mathcal{X} \times \mathcal{A})\}$

Identification and computation

- Our general theory still applies:
  - Θ_I still characterized by zeros of $T(\theta)$
  - $\mathcal{M}_\beta$ is convex (map is linear, domain is convex)
  - Extremal point representation available
- Linear programming approach extends naturally:
  - Dimensionality of LP depends on supports of X and α
  - Not on support of V (integrated out parametrically)
  - Structure of LP remains the same
- Will see this applied to parametric binary choice panel models

Binary choice with fixed effects

Panel Models

- Focus on most challenging flavour:
  - nonlinear / discrete choice
  - fixed effects
  - short-T setting (fixed number of time periods)
- Strict and **sequential** exogeneity
- Get both structural and **counterfactual** parameters

Binary choice, parametric

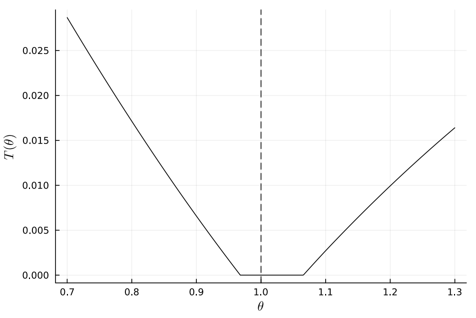
Textbook binary choice panel with fixed effects:

$$Y_{it} = 1\left\{\alpha_i + X_{it}'\beta + U_{it} \geq 0\right\},\ t = 1, \cdots, T$$

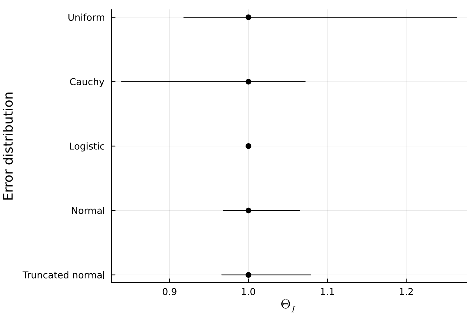
and $U_i | X_i \sim F$.

- $T = 2$, $X_1 = 0,\ X_2 = 1$
- $\alpha \in \{-5, -4.9, \cdots, 4.8, 4.9, 5.0\}$ with $P(\alpha = a) \propto \exp(-a^2/2)$.

- on my laptop, it takes 0.0036 seconds to compute $T(\theta)$
- logit: β_0 is point-identified
- probit: $\Theta_I = [0.968, 1.065]$



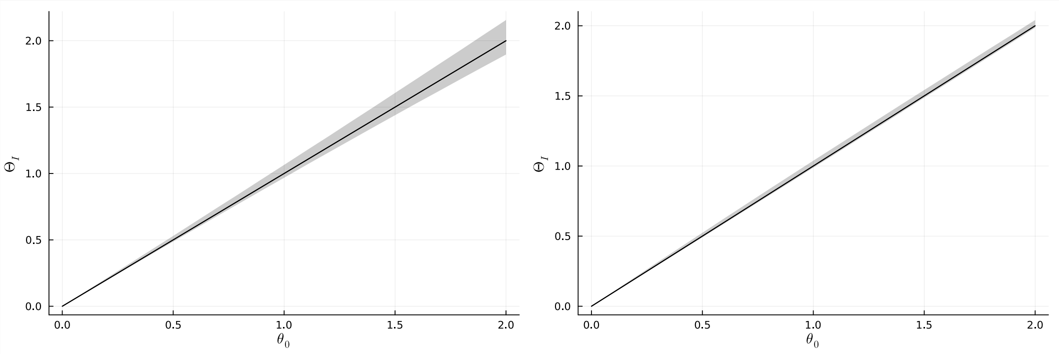
(a) $T(\theta)$ for the probit model.



(b) Θ_I for various error distributions.

Figure 10: Identified sets for the static binary choice model with $X = (0, 1)$, $\beta_0 = 1$.

- probit model, $T \in \{2, 3\}$



(a) $T = 2$ **(b)** $T = 3$

Figure 11: Identified sets for regression coefficient in static binary choice probit.

- $T = 4$: point-identified if you don't have a microscope

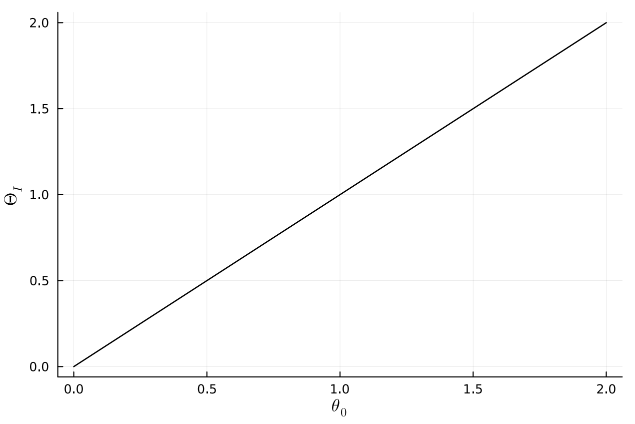
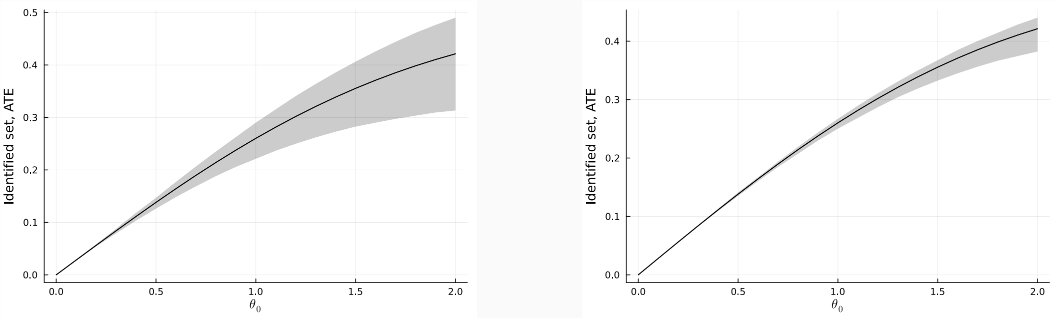


Figure 12: Identified sets for regression coefficient in static binary choice probit, $T = 4$

- average treatment effect of moving x from 0 to 1, i.e.

$$\text{ATE}(0,1;\beta) = E[H(\alpha+\beta) - H(\alpha)].$$



(a) $T = 2$ **(b)** $T = 3$

Figure 13: Identified sets for ATE in static binary choice probit.

Binary choice, strict exogeneity

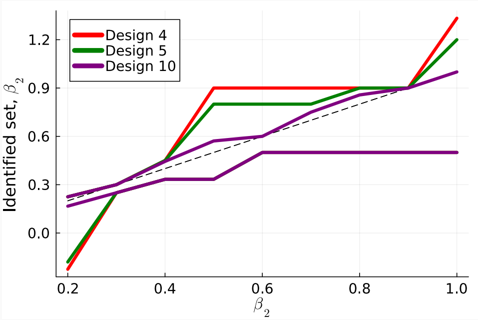
SPBC with fixed effects:

$$Y_t = 1\{X_t'\beta + \alpha + V_t \geq 0\}, \quad t = 1, 2$$

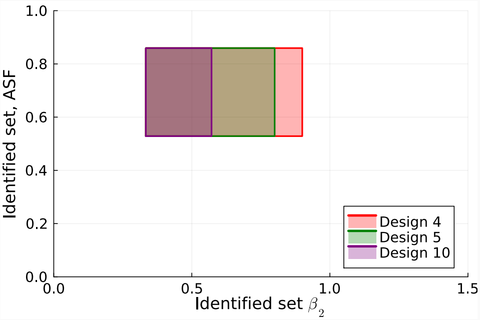
- outcomes $Y_t \in \{0, 1\}$ and regressors $X_t \in \mathcal{X}_t$
- fixed effect $\alpha \in \mathbb{R}$ and error terms $V_t \in \mathbb{R}$
- assume strict stationarity

$$V_1 | \alpha, X_1, X_2 \overset{d}{=} V_2 | \alpha, X_1, X_2$$

- literature
  - β: Manski (1985); Khan et al. (2023); Gao and Wang (2024); Mbakop (2024)
  - partial effects:
    - Botosaru and Muris (2024)
    - parametric: Aguirregabiria and Carro (2021), Davezies et al. (2024); Dobronyi et al. (2021); Pakel and Weidner (2024); Dano (2024)



(a) DGP1: Identified set for β_2.



(b) DGP1: Identified set for β_2 and ASF.

We are the first to obtain results in panel (b).

Sequential exogeneity

- predetermined regressors:

$$V_1|\alpha, X_1 \stackrel{d}{=} V_2|\alpha, X_1, X_2$$

- $\Gamma_\theta(\mathcal{W})$ **not convex** because of $V_2 \perp X_2$
- Theorem 3: $\mathcal{M}_\theta$ **is convex**
- computation via a variant of our LP
- literature:
  - parametric: Arellano and Carrasco (2003); Bonhomme et al. (2023); Chamberlain (2023); Pigini and Bartolucci (2022)
  - ???

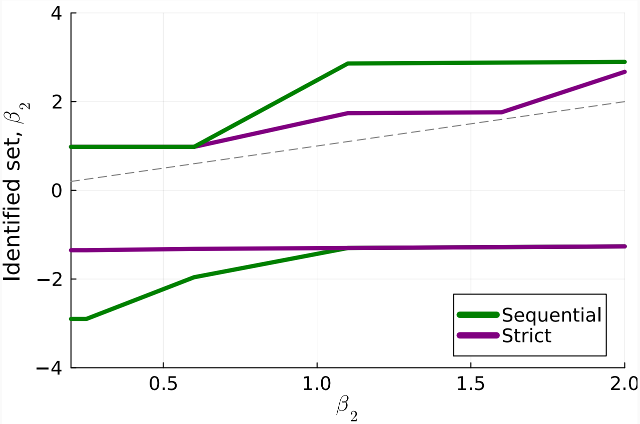


Figure 15: DGP2: Identified set for β_2.

Conclusion

Summary

1. general framework for (partial) identification
2. novel discrepancy function yields tractable, sharp ID
3. works for general class of models, for structural and counterfactual parameters
4. break new ground in nonlinear panels

Paper on arXiv: “An Adversarial Approach to Identification”