An Adversarial Approach to Identification and Inference

UT Knoxville, January 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 economist.

Linear panel models

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

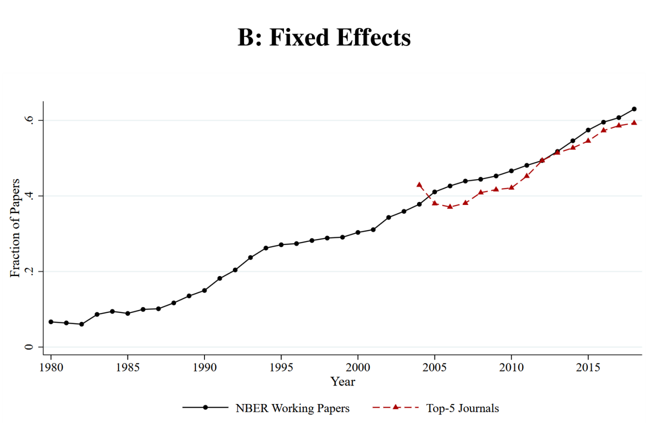


Figure 1: Currie et al., AEA P+P 2020

Outcome equation of linear panel model:

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

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

Outcome equation of linear panel model:

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

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

In 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

Outcome equation of linear panel model:

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

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

In 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

Note:

- 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 $U_i | X_i \sim F$.

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

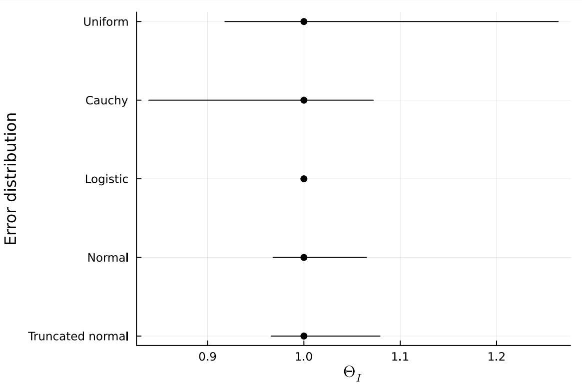


Figure 2: Identified sets in binary choice models, Botosaru, Loh, Muris (2025+)

- partial ID of FE spills over to β
- with exceptions (logit)
- identified sets tend to be small
    - figure: worst case, $X, Y \in \{0, 1\},\ T = 2$

Consequence 1: literature

- huge literature on point ID for specific models
  - Chamberlain (REStud 1980; ECMA 2010); Manski (ECMA 1987)
- small literature on point identification for larger classes
  - Bonhomme (ECMA 2012); Botosaru, Muris, Pendakur (JoE 2023)
- partial identification results for specific models
  - Shi et al. (ECMA 2018); Aristodemou (JoE 2020); Khan et al. (QE 2021); Pakes and Porter (QE 2024); Mbakop (JPE RR)
- **today's paper:** characterize identified set for β:
  - in a large class of models
  - with point or partial ID

I have contributed to this literature:

- static **ordered** choice (Muris, REStat 2017)
- interval-censored models (Abrevaya and Muris, JAE 2020)
- general result + collective households (Botosaru, Muris, Pendakur, JoE 2023)
- **dynamic** ordered choice:
  - Muris, Raposo, Vandoros (REStat 2025+)
  - Honore, Muris, Weidner (QE 2025+)

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 are not
- estimation and inference is very challenging
- traditional advice: *use random effects or linear models if you want partial effects*

Consequence 2: literature

- Wooldridge's grad textbook (2010): "Unfortunately, we cannot estimate the partial effects on the response probabilities …"
- recent work makes progress on this issue:
  - Chernozhukov et al. (ECMA 2013); Honore and Tamer (ECMA 2006)
  - Botosaru and Muris (WP 2017; JoE 2023; JoE 2024)
  - logit models
    - Dobronyi et al. (REStud RR); Davezies et al. (REStud RR); Aguirregabiria and Carro (REStat 2025+); Dano (WP, 2025+); Pakel and Weidner (WP 2025+)
  - literature is fragmented, few solutions and they depend on model/parameter
- **today's paper**:
  - ID common parameters **and** PE …
  - … in a general class of models.

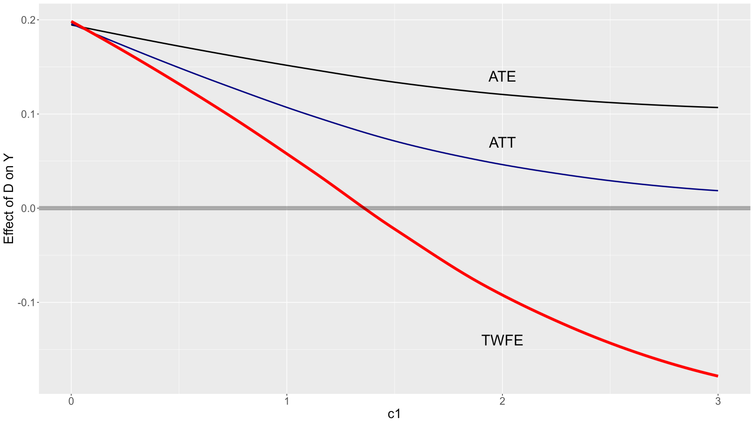
Recap

- **linear** panels with **fixed effects** are central to applied economics
- would like to use FE in nonlinear models, but:
  - in most models, β not point identified
  - even if β point identified, partial effects are not

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?

No.



- Previous slide DGP:
  - binary choice outcomes
  - 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: $\lambda_1 = 0, \ \lambda_2 = 1$,
  - outcome equation:
$$Y_t = 1\{\alpha + D_t \times 1 + \lambda_t + U_t \geq 0\}$$
  - effect of D on Y is positive

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

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

Can we develop an approach that works under partial identification, and that is easy to implement in applied practice?

This is where **adversarial identification** comes in!

This is where **adversarial identification** comes in!

Overview for "Adversarial identification"

1. Introduction
2. Model
3. Main result
4. Computation via linear programming
5. Results for nonlinear panels

Introduction

What? Framework for partial identification + inference

Why?

- Applicable to a **wide range of 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 **discrepancy function** with maximin formulation
- Leverage:
  - **convexity** of the set of **model probabilities**
  - a **linearity** property of most econometric models

Semiparametric binary choice model

To establish notation and terminology:

- running example: *semiparametric binary choice (SPBC) model*
- start from cross-sectional logit/probit model with

$$Y_i = 1\{X_i'\theta + U_i \geq 0\}$$

  with $U_i \perp X_i$ and U_i is standard logistic/normal
- weaken assumptions on (U, X) to $\text{med}(U_i|X_i) = 0$.

Semiparametric binary choice model

To establish notation and terminology:

- running example: *semiparametric binary choice (SPBC) model*
- start from cross-sectional logit/probit model with

$$Y_i = 1\{X_i'\theta + U_i \geq 0\}$$

  with $U_i \perp X_i$ and U_i is standard logistic/normal
- weaken assumptions on (U, X) to $\text{med}(U_i|X_i) = 0$.

Compare: linear model with $U_i \sim \mathcal{N}(0, \sigma^2)$ relaxed to $E(U_i|X_i) = 0$.

- SPBC model is surprisingly hard to analyze...
- ... which is why we don't see it in the wild
- if all regressors are discrete:
    - β partially identified (even with a scale normalization)
    - partial effects partially identified

One of our innovations: how we set up the econometric model.

One of our innovations: how we set up the econometric model.

- 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: $P(U_i \leq 0 | X_i = x) = 0.5$ for each x
- **outputs** $Z = (X, Y)$ have **probability measure** μ_Z
  - observed one: μ_Z^*

One of our innovations: how we set up the econometric model.

- 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: $P(U_i \leq 0 | X_i = x) = 0.5$ for each x
- **outputs** $Z = (X, Y)$ have **probability measure** μ_Z
  - observed one: μ_Z^*

Our focus on γ and μ is novel and key to our analysis.

SPBC model has three properties:

1. for each parameter β and for each distribution of inputs γ, it returns a distribution of outputs $\mu_{Z,(\beta,\gamma)}$
   - in other words: model is a map $(\beta, \gamma) \mapsto \mu_{Z,(\beta,\gamma)}$
2. at each β, 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 is convex

SPBC model has three properties:

1. for each parameter β and for each distribution of inputs γ, it returns a distribution of outputs $\mu_{Z,(\beta,\gamma)}$
   - in other words: model is a map $(\beta, \gamma) \mapsto \mu_{Z,(\beta,\gamma)}$
2. at each β, 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 is convex

- paper: properties 1-3 hold for most econometric models
  - derive results from these basic properties
  - we use convex analysis, functional analysis, and convex functional analysis
- this talk: 1-3 hold for pmf version of SPBC model

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.

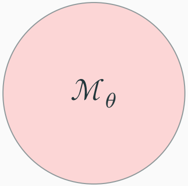


Figure 3: Each point corresponds to a model probability for a given (θ, γ).

Model Probabilities

Set of model probabilities

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

for a fixed θ.

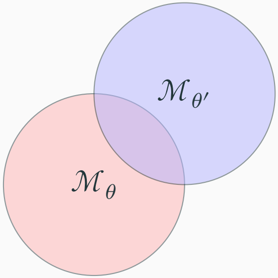


Figure 3: Each point corresponds to a model probability for a given (θ, γ).

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 the closure of $\mathcal{M}_\theta$.

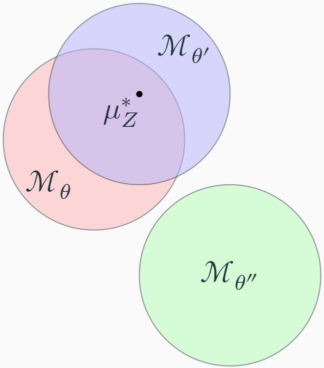


Figure 4: $\{\theta, \theta'\} \subset \Theta_I,\ \theta'' \notin \Theta_I$.

Identified set for θ is:

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

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

Problem: this definition is not tractable.

Main result

Discrepancy function

- Goal: tractability of the identification problem

Discrepancy function

- Goal: tractability of the identification problem
- From 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 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 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 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 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

Discrepancy function

- Goal: tractability of the identification problem
- From 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 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 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,

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

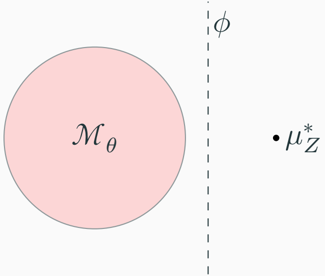


Figure 5: $\theta \notin \Theta_I$, with separating hyperplane

Convexity in econometric models

- paper: applies result, verifying convexity for large classes of econometric models
  - common theme: many econometric models satisfy:
    1. linearity: for each θ, there is a L_θ such that $\mu_{Z,(\theta,\gamma)} = L_\theta \gamma$
    2. convexity: the set of allowed $\gamma \in \Gamma_\theta$ is convex
  - if linearity and convexity,
    - then $\mathcal{M}_\theta$ is convex (Proposition 1)
    - and our theory applies
    - paper: additional results under linearity and convexity (Propositions 1 and 2)
- this talk:
  - demonstrate usefulness of adversarial approach via computation
  - emphasize tractability

Computation: Linear Programming

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
- I will now show that this 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

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

- 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

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

- 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
- 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.

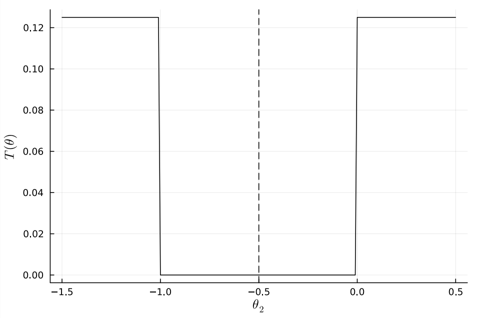
Substituting dual into $T(\theta)$ means we now maximize over ϕ (as before) and λ (dual):

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

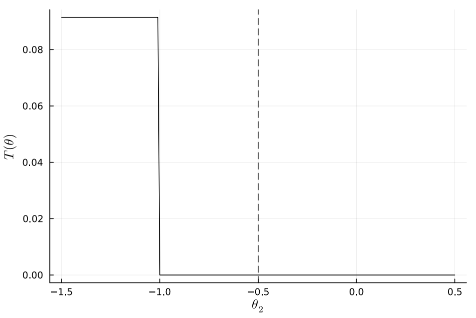
This is a LP: we have achieved tractability for $T(\theta)$ and therefore Θ_I.

Takeaway: to determine if $\theta \in \Theta_I$:

- solve an LP for $T(\theta)$
- checking $T(\theta) \leq 0$.
- 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);
    3. the matrix $\widetilde{C}_\theta$;
    4. restrictions (A_θ, b_θ);

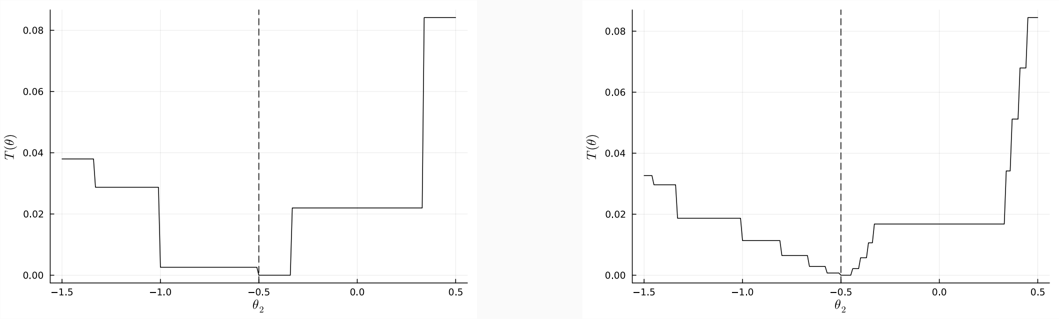


(a) Design 1



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

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



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

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

Figure 7: $T(\theta)$ 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.

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

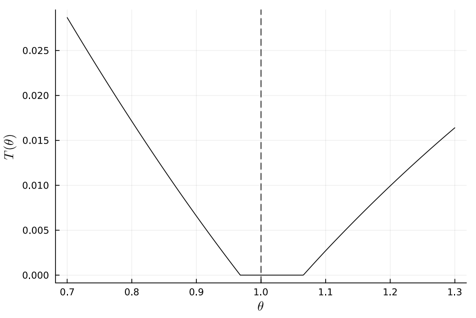
Textbook binary choice panel with fixed effects:

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

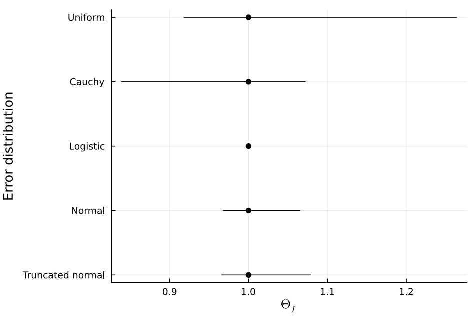
and $U_i | X_i \sim F$.

- $T = 2,\ X_1 = 0,\ X_2 = 1$
- $\alpha \in \{-5, -4.9, \cdots, 4.8, 4.9, 5.0\}$ 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: $\Theta_I = [0.968, 1.065]$



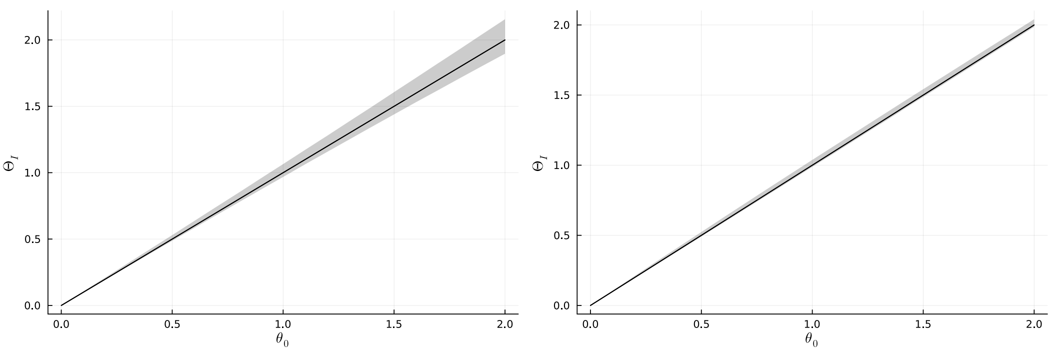
(a) $T(\theta)$ for the probit model.



(b) Θ_I for various error distributions.

Figure 8: 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 9: Identified sets for regression coefficient in static binary choice probit.

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

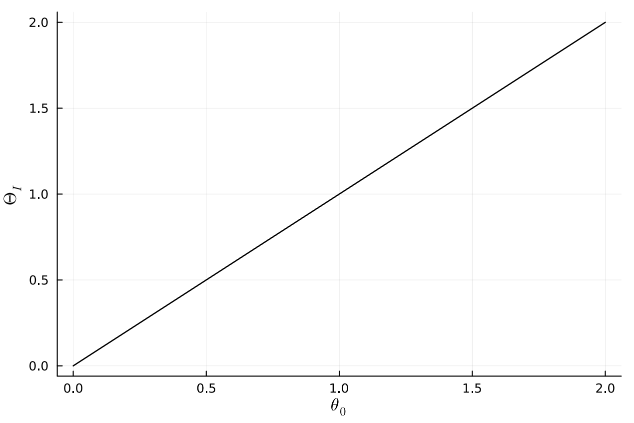
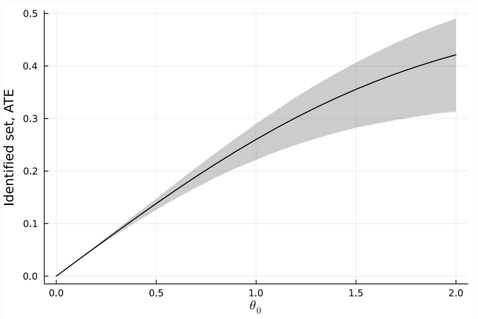


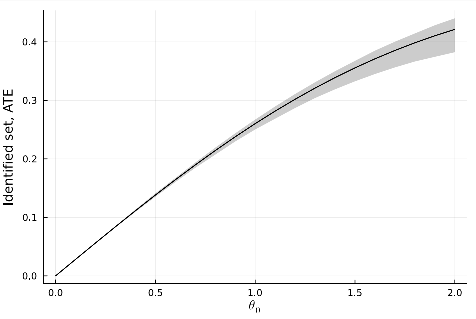
Figure 10: Identified sets for regression coefficient in static binary choice probit, $T = 4$

- average treatment effect of moving a randomly selected individual's x_t from 0 to 1, i.e.

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



(a) $T = 2$



(b) $T = 3$

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

Binary choice, strict exogeneity

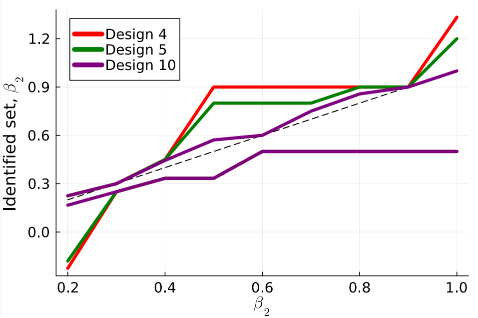
SPBC with fixed effects:

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

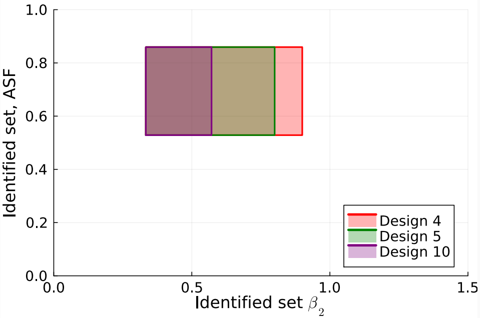
- outcomes $Y_t \in \{0, 1\}$ and regressors $X_t \in \mathcal{X}_t$
- fixed effect $\alpha \in \mathbb{R}$ and error terms $V_t \in \mathbb{R}$
- assume strict stationarity

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

- 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:

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

- $\Gamma_\theta(\mathcal{W})$ **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)
  - ???

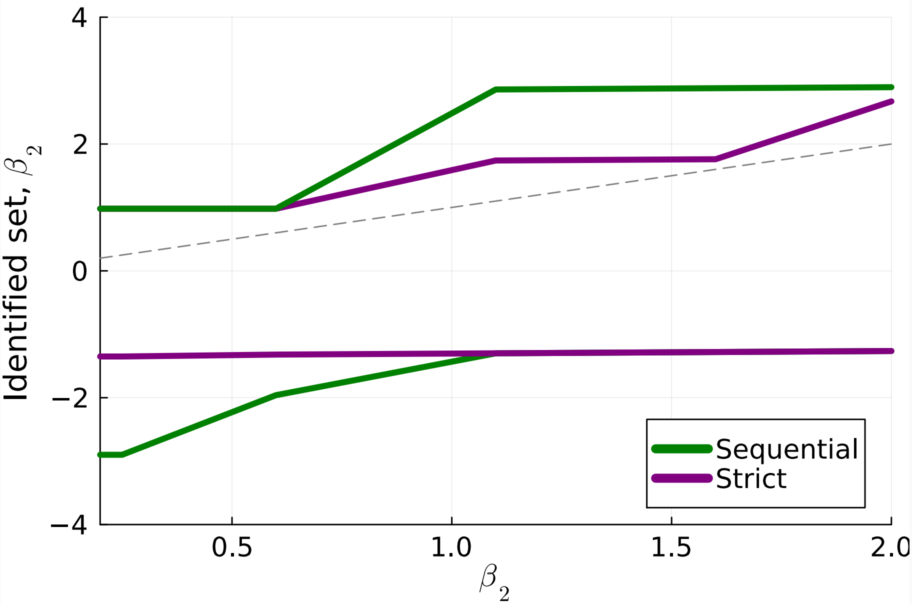


Figure 13: DGP2: Identified set for β_2.

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”