# Identification of dynamic panel logit models with fixed effects

Jiaying Gu

University of Toronto

IASM-BIRS Workshop
December 11-15, 2023

Joint work with Christopher Dobronyi and Kyoo il Kim

# Introduction

- Dynamic panel logit models with fixed effects:

$$Y_{it} = 1\{\alpha_i + \beta Y_{it-1} + \gamma' X_{it} \geq \epsilon_{it}\}, \quad t = 1, 2, \ldots, T$$

  where $\epsilon_{it}$ are iid Logit error and $\alpha_i$ is a scalar random variable with unknown distribution $Q(\cdot|\boldsymbol{x}, y_0)$.
- This is a workhorse model in industrial organization in Economics, often used to analyze consumer purchase decisions, firm entry/exit decisions, or binary longitudinal data in general.
- Data: $y_{it}$ and $x_{it}$ for $t = 0, \ldots, T$.
- Parameter of interest: $\theta = \{\beta, \gamma\}$, distribution $Q$ or some functionals of $Q$.
- We may also introduce more lags (e.g. AR(2) model)

$$Y_{it} = 1\{\alpha_i + \beta_1 Y_{it-1} + \beta_2 Y_{it-2} + \gamma' X_{it} \geq \epsilon_{it}\}$$

# Introduction

Key challanges:

- Incidental parameter problem due to the presence of $\alpha_i$: if we include individual dummies, $\beta$ will be inconsistently estimated.
- Functional of $Q$: $Q$ is not point identified due to binary nature of $Y_{it}$.

# Introduction

- Chamberlain (1985): (without x) Point identification of $\beta$ and use conditional MLE with sufficient statistics $S(\boldsymbol{y_i}) = \{y_{i0}, \sum_{t=1}^{T-1} y_{it}, y_{iT}\}$,

$$P(\boldsymbol{y}_i|y_{i0},\beta) = \underbrace{P(\boldsymbol{y}_i|S(\boldsymbol{y}_i),\beta)}_{\text{free from } \alpha_i} \int P(S(\boldsymbol{y}_i)|\beta,\alpha_i)dQ(\alpha_i|y_{i0})$$

- Honoré and Kyriazidou (2000): extends sufficient statistics idea to allow covariates under some assumptions on $\boldsymbol{x}$ [i.e., $x_2 = x_3$ for $T = 3$]

- Sufficient Statistics method fails to identify $\beta_1$ and $\beta_2$ for AR(2) model.
- If the panel length is very short (i.e. $T = 2$), sufficient statistics also fails to identify $\beta$.

# Introduction

- In this paper, we conduct an identification analysis for both parameters and $Q$.
- We show that the identification problem has a connection to the *truncated moment problem* in mathematics [dates back to Chebyshev 1874].
- Truncated moment problem: Given the first $K$ raw moments of a random variable $X$, to characterize the set of probability measure that $X$ can have: existence and uniqueness.

# Introduction

Using this connection, we show two types of results:

- Identified set for structural parameters $\theta$ is characterized by a set of moment equality and inequality conditions.
- The number of moment equality conditions may be substantially more than those found by sufficient statistics approach.
    - e.g. moment equality conditions available for AR(2) model, more moment conditions available for AR(1) model with $x$.
- The inequality conditions can sharpen the identified set for $\theta$ when they are not point identified by moment equality conditions alone (i.e. AR(1) model with time trend)

- Identified set of the latent distribution $Q$ is characterized by a finite vector of *generalized moments*, and the number of moments grows linearly in $T$.
- Provide sufficient conditions on point identification of functionals of $Q$.

We then provide estimation and inference method with these new identification results.

# Roadmap

- A simple example with $T = 2$.
- Identification of $\theta$ and $Q$ for AR(1) model with general $T$.
- Identification of a class of functionals of $Q$.
- Examples: time trend
- Empirical Illustration

# Identification Analysis

- Let $\mathcal{Y}$ be the set containing all choice histories $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^J$ with $J = 2^T$.

$$\begin{aligned}\mathcal{P}_j &= \mathbb{P}((Y_1 \ldots Y_T) = \boldsymbol{y}^j | Y_0 = y_0, X = \boldsymbol{x}, \alpha) \\ &= \mathcal{L}_j(\alpha, \theta, \boldsymbol{x}, y_0) = \prod_{t=1}^{T} \frac{\exp(\alpha + \beta y_{t-1} + \gamma x_t)^{y_t}}{1 + \exp(\alpha + \beta y_{t-1} + \gamma x_t)}\end{aligned}$$

- Denote the probability vector $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = (\mathcal{P}_1, \ldots, \mathcal{P}_J)$ and $\boldsymbol{\mathcal{L}}$ the vector that stacks $\mathcal{L}_j$.
- Let $A = \exp(\alpha)$ with distribution $Q(A|y_0, \boldsymbol{x})$ supported on $\mathcal{A} = [0, \infty)$.
- Define the set of probability measures with support $\mathcal{A}$:

$$\mathcal{Q}(\theta, y_0, \boldsymbol{x}) = \{Q : \boldsymbol{\mathcal{P}}_{\boldsymbol{x}} = \int_{\mathcal{A}} \boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0) dQ\}$$

Definition (Identified Set): The identified set of $\theta$ is

$$\Theta^* = \{\theta : \mathcal{Q}(\theta, y_0, \boldsymbol{x}) \neq \emptyset, \text{for all } \boldsymbol{x} \in \mathcal{X}\}$$

# Simplest example: $T = 2$ without x

- Consider $T = 2$, no covariates and fix $y_0 = 0$.
- We have $2^T = 4$ distinct elements in $\mathcal{Y}$ and

$$\mathcal{L}(A,\beta) = \begin{pmatrix} \mathbb{P}((0,0)|A,\beta) \\ \mathbb{P}((1,0)|A,\beta) \\ \mathbb{P}((0,1)|A,\beta) \\ \mathbb{P}((1,1)|A,\beta) \end{pmatrix} = \begin{pmatrix} \frac{1}{(1+A)^2} \\ \frac{A}{1+A}\frac{1}{1+AB} \\ \frac{A}{(1+A)^2} \\ \frac{A}{1+A}\frac{AB}{1+AB} \end{pmatrix} = \frac{1}{g(A,\beta)} \begin{pmatrix} (1+AB) \\ A(1+A) \\ A(1+AB) \\ A^2B(1+A) \end{pmatrix}$$

with $B = \exp(\beta)$ and $g(A,\beta) = (1+AB)(1+A)^2$.

- The right hand side consists polynomials of $A$ up to degree $2T - 1 = 3$.

# Simplest example: $T = 2$ without x

In particular, for any $(\beta, Q)$

$$\int_{\mathcal{A}} \boldsymbol{\mathcal{L}}(A,\beta) dQ(A) = \int_{\mathcal{A}} G(\beta) \begin{pmatrix} 1 \\ A \\ A^2 \\ A^3 \end{pmatrix} \frac{1}{g(A,\beta)} dQ(A) = G(\beta) \int_{\mathcal{A}} \begin{pmatrix} 1 \\ A \\ A^2 \\ A^3 \end{pmatrix} d\bar{Q}(A \mid \beta)$$

with

$$G(\beta) = \begin{pmatrix} 1 & B & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & B & 0 \\ 0 & 0 & B & B \end{pmatrix}$$

and $d\bar{Q}(A \mid \beta) = \frac{1}{g(A,\beta)} dQ(A)$.

- $Q$ is a probability measure on $\mathcal{A}$ and $1/g(A,\beta) \in (0,1]$ for all $A \in \mathcal{A}$,
- $\bar{Q}$ is a non-negative Borel measure on $\mathcal{A}$.

## Simplest example: $T = 2$ without x

- Identifying condition: $\beta \in \Theta^* \Leftrightarrow \exists Q : \boldsymbol{\mathcal{P}} = G(\beta) \int_{\mathcal{A}} \begin{pmatrix} 1 \\ A \\ A^2 \\ A^3 \end{pmatrix} \frac{1}{g(A,\beta)} dQ(A).$
- Given $G(\beta)$ full rank, $\beta \in \Theta^* \Leftrightarrow \exists \bar{Q}$:

$$\boldsymbol{r}(\beta) := G(\beta)^{-1} \boldsymbol{\mathcal{P}} = \int_{\mathcal{A}} \begin{pmatrix} 1 & A & A^2 & A^3 \end{pmatrix}' d\bar{Q}(A \mid \beta)$$

- $\beta \in \Theta^* \Leftrightarrow$ the observed vector $\boldsymbol{r}(\beta)$ is a truncated moment sequence of some non-negative measure.

Definition (Moment Space): The moment space of any non-negative Borel measure $\mu$ on $\mathcal{A}$ is:

$$\mathcal{M}_K = \left\{ \boldsymbol{r} \in \mathbb{R}^{K+1} : \text{ there exists } \mu \text{ such that } r_k = \int_{\mathcal{A}} A^k d\mu(A), \text{ for all } k = 0, 1, \ldots, K \right\}$$

In this simple $T = 2$ example: $\Theta^* = \{\beta : \boldsymbol{r}(\beta) \in \mathcal{M}_3\}$.

# Moment Space

Moment space has unique geometric structure:

- $\mathcal{M}_K$ is a closed convex cone in $\mathbb{R}^{K+1}$.
- The moment space $\mathcal{M}_K$ does not take up the entire $\mathbb{R}^{K+1}$ since moments have dependency:
  - Cauchy-Schwartz inequality: $\mathbb{E}[A^2] \geq \mathbb{E}[A]^2$
  - More generally Hölder's inequlity has to hold.
- This implies: $\boldsymbol{r}(\beta) \in \mathcal{M}_3$ provides nontrivial constraints on $\beta$.

- Using the result of Karlin and Studden (1966), the restricitons boils down to nonnegativity of two matrices (Hankel matrices):

$$\begin{pmatrix} r_0(\beta) & r_1(\beta) \\ r_1(\beta) & r_2(\beta) \end{pmatrix}, \begin{pmatrix} r_1(\beta) & r_2(\beta) \\ r_2(\beta) & r_3(\beta) \end{pmatrix}$$

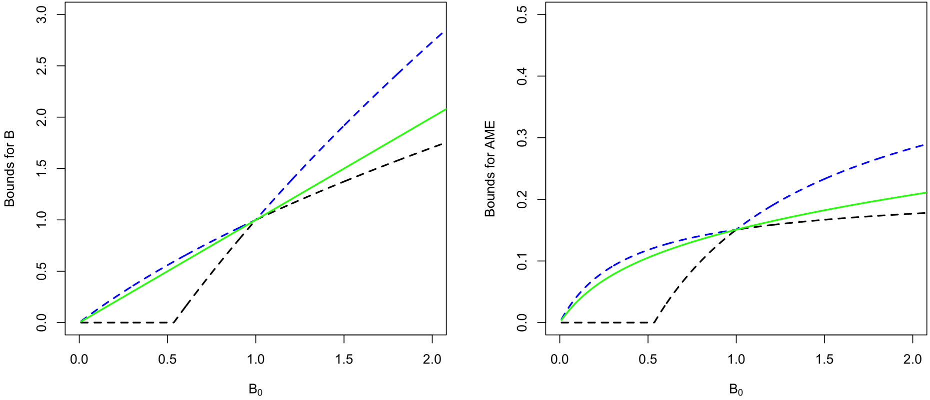
# Simplest example: $T = 2$ without $x$

- Non-negativity of square matrices $\Leftrightarrow$ non-negativity of all principal minors: this gives us *moment inequalities* for $\beta$.
- $\beta \in \Theta^* \Leftrightarrow \boldsymbol{r}(\beta) \geq 0$, $r_0(\beta)r_2(\beta) - r_1(\beta)^2 \geq 0$ and $r_1(\beta)r_3(\beta) - r_2(\beta)^2 \geq 0$.
- For $T = 2$ and no $x$: $\mathcal{P} := (p_0, p_1, p_2, p_3) = \mathbb{P}(Y = \boldsymbol{y} | Y_0 = 0)$:

$$\boldsymbol{r}(\beta) := \begin{pmatrix} r_0(\beta) \\ r_1(\beta) \\ r_2(\beta) \\ r_3(\beta) \end{pmatrix} = \begin{pmatrix} p_0 - \frac{B^2}{B-1}p_1 + \frac{B}{B-1}p_2 \\ \frac{Bp_1 - p_2}{B-1} \\ \frac{p_2 - p_1}{B-1} \\ \frac{p_1 - p_2}{B-1} + \frac{p_3}{B} \end{pmatrix}$$

# Illustration: T = 2, no x, $y_0 = 0$

- DGP: $\alpha_i \sim \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_1$, we vary $\exp(\beta_0)$ from 0.01 to 2.
- Left: identified set for $\exp(\beta)$



What does the analysis say about identification of distribution $Q$ and its functionals?

# Identified set for $Q$

- Recall

$$\boldsymbol{r}(\beta) = G(\beta)^{-1}\boldsymbol{\mathcal{P}} = \int_{\mathcal{A}} \begin{pmatrix}1 & A & A^2 & A^3\end{pmatrix}' \frac{1}{g(A,\beta)} dQ$$

- The identified set of the distribution $Q$ is characterized by the *generalized moments*:

$$\mathcal{Q}(\beta, y_0) = \left\{ Q : \boldsymbol{r}(\beta) = \int_{\mathcal{A}} \begin{pmatrix}1 & A & A^2 & A^3\end{pmatrix}' \frac{1}{g(A,\beta)} dQ \right\}$$

  for each $\beta \in \Theta^*$.

- $\boldsymbol{r}(\beta)$ is the dimension reduction from the data information $\boldsymbol{\mathcal{P}}$ to the information on $Q$.

## General $T$

With the results from the simple example, we now generalize:

- The simple example reveals a polynomial structure of the dynamic panel logit model with fixed effects, which generalizes to any finite $T$ with or without $x$.

# General Results for AR(1) model

- For general $T$, for each given $\boldsymbol{x} \in \mathcal{X}$ and $y_0$, we can construct $g(A, \theta, \boldsymbol{x}, y_0)$ as a polynomial of $A$ of degree $2T - 1$ such that:

$$\boldsymbol{\mathcal{L}}(A, \theta, \boldsymbol{x}, y_0) = G(\theta, \boldsymbol{x}) \begin{pmatrix} 1 \\ A \\ \vdots \\ A^{2T-1} \end{pmatrix} \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)}$$

  where $G(\theta, \boldsymbol{x})$ is of dimension $2^T \times 2T$.

- When $T > 2$, we obtain moment equalities in addition to moment inequalities. [Because $G$ is of dimension $2^T \times 2T$].
  - The number of moment equalities available is determined by the dimension of the left null space of $G(\theta)$.
  - The form of the moment equalities can be constructed analytically with the basis of the left null space of $G(\theta)$.

- Define the set (left null space of $G(\theta, \boldsymbol{x})$):

$$\boldsymbol{M}_{\boldsymbol{x}}(\theta) = \{\boldsymbol{v}_{\boldsymbol{x}}(\theta) \in \mathbb{R}^{2^T} : \boldsymbol{v}_{\boldsymbol{x}}(\theta)' G(\theta, \boldsymbol{x}) = 0\}$$

- Moment equality conditions: $\mathbb{E}[\boldsymbol{v}_{\boldsymbol{x}}(\theta)_j 1\{Y = y^j, X = x\}] = 0, \forall j$.

# General $T$

Theorem 2: If $G(\theta, \boldsymbol{x})$ is full rank, then $\theta \in \Theta^*$ if and only if the following conditions hold:

(a) For all $\boldsymbol{x} \in \mathcal{X}$, we have $\boldsymbol{v_x}(\theta)'\boldsymbol{\mathcal{P}_x} = 0$ for all $\boldsymbol{v_x}(\theta) \in \boldsymbol{M_x}(\theta)$.

(b) For all $\boldsymbol{x} \in \mathcal{X}$, we have $\boldsymbol{r}(\theta, \boldsymbol{x}) \in \mathcal{M}_{2T-1}$, where $\boldsymbol{r}(\theta, \boldsymbol{x}) = H(\theta, \boldsymbol{x})\boldsymbol{\mathcal{P}_x}$ and $H(\theta, \boldsymbol{x})$ is a matrix of dimension $2T \times 2^T$ such that $H(\theta, \boldsymbol{x})G(\theta, \boldsymbol{x}) = I_{2T}$.

- Condition (a) provides moment equalities and condition (b) provides moment inequalities.
- The number of non-redundant moment equalities available: $2^T - rank(G) = 2^T - 2T$.

▸ Details on Inequalities

# Identification of $Q$

For general $T$,

$$\mathcal{Q}(\theta, y_0, \boldsymbol{x}) = \left\{ Q : \boldsymbol{r}(\theta, \boldsymbol{x}) = \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta, \boldsymbol{x}, y_0)} dQ \right\}$$

Theorem 4: For each $\boldsymbol{x} \in \mathcal{X}$ and each value of $\theta \in \Theta^*$, the sharp identified set $\mathcal{Q}(\theta, y_0, \boldsymbol{x})$ of the latent distribution are those $Q$ that has its generalized moments, $\mathbb{E}_Q[A^j / g(A, \theta, \boldsymbol{x}, y_0)] = \boldsymbol{r}_j(\theta, \boldsymbol{x})$ for $j = 0, 1, 2, \ldots, 2T - 1$.

- $Q$ is in general not point identified from the $2T - 1$ generalized moments.
- But some functional of $Q$ may be point identified.

- If $\theta$ is point identified (i.e. $\Theta^* = \{\theta_0\}$), then the generalized moments $\boldsymbol{r}(\theta_0, \boldsymbol{x}) = \int_{\mathcal{A}} \begin{pmatrix} 1 & A & \cdots A^{2T-1} \end{pmatrix}' \frac{1}{g(A, \theta_0, \boldsymbol{x}, y_0)} dQ_0(A \mid y_0, \boldsymbol{x})$ is also point identified.

# Point Identification of Functionals $\int \psi(A,\theta,x)dQ$

Theorem (Point Identification of Functionals of $Q$) If $\theta$ is point identified and the product $\psi(A,\theta_0,\boldsymbol{x})g(A,\theta_0,\boldsymbol{x},y_0)$ is a polynomial of $A$ with a degree that is no larger than $2T-1$ such that:

$$\psi(A,\theta_0,\boldsymbol{x})g(A,\theta_0,\boldsymbol{x},y_0) = \sum_{j=0}^{2T-1} \eta_j(\theta_0,\boldsymbol{x})A^j,$$

for some vector $\boldsymbol{\eta}(\theta_0,\boldsymbol{x}) = (\eta_0(\theta_0,\boldsymbol{x}),\eta_1(\theta_0,\boldsymbol{x}),\ldots,\eta_{2T-1}(\theta_0,\boldsymbol{x}))$, then $\mathbb{E}_{Q_0(A|y_0,\boldsymbol{x})}[\psi(A,\theta_0,\boldsymbol{x})]$ is point identified and equal to $\boldsymbol{\eta}(\theta_0,\boldsymbol{x})'\boldsymbol{r}(\theta_0,\boldsymbol{x})$.

- The Theorem provides *sufficient conditions* on the function $\psi$ under which $\mathbb{E}_{Q_0}[\psi(A,\theta_0,\boldsymbol{x})]$ is point identified.
- Examples:
  - Average marginal effect of lagged choice when $T \geq 3$.
  - Posterior expectation of $A$ when $T \geq 3$: $\mathbb{E}_{Q_0}[A|\boldsymbol{y}]$ for $\boldsymbol{y} \in \mathcal{Y}$.
  - Counterfactual choice probability with no dynamics:
    i.e. AR(1) model without $x$, we can compare the counterfactual
    $\mathbb{P}(Y=(1,1,1)|\beta=0)$ with $\mathbb{P}_0(1,1,1)$ in the data.

# Example: Average Marginal Effect

- For models without covariates, $T = 3$, and $y_0 = 0$ (we know $\beta$ is point identified)

$$AME = \int \frac{AB_0}{1 + AB_0} dQ_0(A) - \int \frac{A}{1 + A} dQ_0(A)$$

- Let $\psi(A, \beta_0) = \frac{AB_0}{1+AB_0} - \frac{A}{1+A}$, and $g(A, \beta_0) = (1 + A)^3(1 + AB_0)^2$.

$$\psi(A, \beta_0)g(A, \beta_0) = (B_0 - 1)A(1 + AB_0)(1 + A)^2 = \boldsymbol{\eta}(\beta_0)'(1, A, \ldots, A^5)'$$

with $\boldsymbol{\eta}(\beta_0) = (0, B_0 - 1, (2 + B_0)(B_0 - 1), (1 + 2B_0)(B_0 - 1), B_0(B_0 - 1), 0)'$.

- Point identification of AME:

$$AME = \int \psi(A, B_0) dQ_0(A) = \boldsymbol{\eta}(\beta_0)' H(\beta_0) \boldsymbol{\mathcal{P}}$$

$$= (B_0 - 1)\left\{\frac{1}{2}(\mathbb{P}_0(1, 0, 0) + \mathbb{P}_0(0, 1, 0)) + \frac{1}{B_0 + 1}(\mathbb{P}_0(1, 0, 1) + \mathbb{P}_0(0, 1, 1))\right\}$$

# Example: Bounding AME for $T = 2$

- For models without x, $T = 2$, and $y_0 = 0$ (we know $\beta$ is not point identified)

$$AME = \int \frac{AB_0}{1 + AB_0} dQ_0(A) - \int \frac{A}{1 + A} dQ_0(A)$$

with $\psi(A, \beta) = \frac{AB}{1+AB} - \frac{A}{1+A}$ and $g(A, \beta) = (1 + A)^2(1 + AB)$.

- It is easy to verify that $\psi(A, \beta)g(A, \beta) = \boldsymbol{\eta}' \begin{pmatrix} 1 & A & A^2 & A^3 \end{pmatrix}'$ with $\boldsymbol{\eta}' = \begin{pmatrix} 0 & B - 1 & B - 1 & 0 \end{pmatrix}'$.

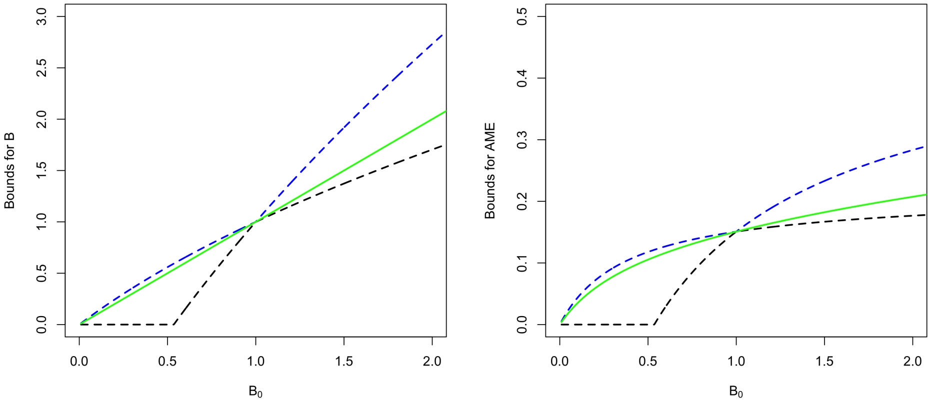
- Sharp bound of AME is:

$$\left[ \inf_{\beta \in \Theta^*} \boldsymbol{\eta}(\beta)' \boldsymbol{r}(\beta), \sup_{\beta \in \Theta^*} \boldsymbol{\eta}(\beta)' \boldsymbol{r}(\beta) \right]$$

- Since $\boldsymbol{\eta}(\beta)' \boldsymbol{r}(\beta) = (B - 1)\mathbb{P}_0(1, 0)$.
- As soon as we have the identified set $\Theta^*$, sharp bounds for AME is mapped directly from that.

# Illustration: T = 2, no x, $y_0 = 0$

- DGP: $Q_0 = \frac{1}{2}\delta_{\exp(-2)} + \frac{1}{2}\delta_{\exp(1)}$, we vary $\exp(\beta_0)$ from 0.01 to 2.
- Left: identified set for $\exp(\beta)$; Right: identified set for AME

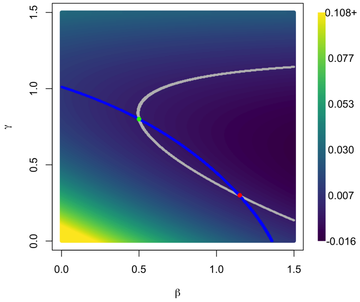


# Time trend model with $T = 3$

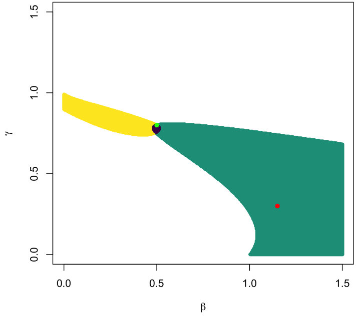
- $Y_{it} = 1\{\alpha_i + \beta Y_{it-1} + \gamma t \geq \epsilon_{it}\}$ and $T = 3$.
- There are two moment conditions for $(\beta, \gamma)$, which always give two solutions.
- Using the moment inequality, we demonstrate inequalities can be used to rule out the false solution.

# Time trend model with $T = 3$

- $\alpha \sim G = 0.5\delta_{-2} + 0.5\delta_1$, and $\beta_0 = 0.5$ and $\gamma_0 = 0.8$.
- The false root, roughly at $\tilde{\theta} = (1.15, 0.3)$.
- $r_1(\tilde{\theta}) \approx -0.24 < 0$.




(a) moment equality




(b) moment inequality

# Estimation

- We need an estimation and inference framework that allows for both moment equality and inequalities.
- We can use minimum distance framework to combine moment equality and inequalities: (Bajari, Benkard, Levin (2007), Shi and Shum (2015))

$$Q(\theta, \mathcal{P}) = h^e(\theta, \mathcal{P})' W \ h^e(\theta, \mathcal{P}) + \sum_{j=1}^{K} \Big( \min\{h^{ie}(\theta, \mathcal{P}), 0\} \Big)^2$$

- If $\Theta^*$ is singleton: $Q(\theta_0, \mathcal{P}_0) = 0 = \min_{\theta \in \Theta} Q(\theta, \mathcal{P}_0)$, and $Q(\theta, \mathcal{P}_0) > 0$ for $\theta \neq \theta_0$: CAN estimator under suitable conditions:

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \ Q_n(\theta, \hat{\mathcal{P}}_n)$$

  with $Q_n(\theta, \hat{\mathcal{P}}_n) = h^e(\theta, \hat{\mathcal{P}}_n)' \hat{W}_n \ h^e(\theta, \hat{\mathcal{P}}_n) + \sum_{j=1}^{K} \Big( \min\{h^{ie}(\theta, \hat{\mathcal{P}}_n), 0\} \Big)^2$ and $\hat{\mathcal{P}}_n$ a CAN estimator of $\mathcal{P}_0$.
- If $\Theta^*$ is a set: consistent estimation via Manski and Tamer (2002):

$$\hat{\Theta}_n = \Big\{ \theta : Q_n(\theta, \hat{\mathcal{P}}_n) \leq \min_{\theta \in \Theta} Q_n(\theta, \hat{\mathcal{P}}_n) + \kappa_n \Big\}$$

  with $\kappa_n > 0$ and $\kappa_n \to 0$.

## Empirical Application

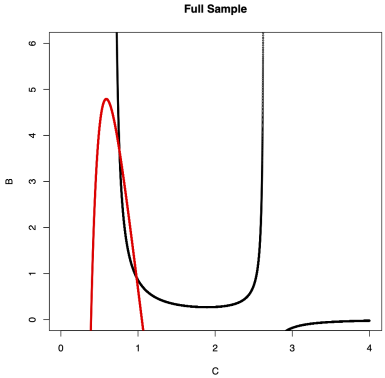
- We revisit Fitzmaurice and Laird (1993) on modeling children's respiratory conditions with data over a short period of time.
- Data: Observe wheezing conditions (binary) of 537 children from Steubenville, Ohio between the ages of 7 and 10.
- Model: time trend model with $T = 3$

$$y_{it} = 1\{\alpha_i + \beta y_{it-1} + \gamma t \geq \epsilon_{it}\}, \quad t = 1, 2, 3$$

- Focus on children with $y_0 = 0$ (85% of the sample).

- Time trend is crucial to distinguish age effect and persistence.
- Fixed effects are crucial to distinguish unobserved heterogeneity from true dynamics.

# Empirical Application




- Using only moment equality conditions, we have two solutions: $(\hat{\beta}, \hat{\gamma}) = (1.301, -0.276)$ and $(\tilde{\beta}, \tilde{\gamma}) = (-0.088, -0.019)$, indistinguishable for the GMM criteria with a diagonal weighting matrix.
- For the second root: $r_1(\tilde{\beta}, \tilde{\gamma}) = -4.82 < 0$.

# Empirical Application

| | *Logit Full* | *Logit Full* | *Logit* | *Logit* | *Logit FE ML* | *Logit FE ML* |
|---|---|---|---|---|---|---|
| | All Sample (n = 450) | | | | | |
| lagged y | 1.301** | 0.693 | 2.08*** | 1.772*** | −2.918*** | −2.736*** |
| | (0.671) | (0.707) | (0.258) | (0.238) | (0.690) | (0.503) |
| time trend | -0.276 | - | 1.05*** | - | 1.666*** | - |
| | (0.321) | - | (0.162) | - | (0.260) | - |

- *Logit Full*: our proposed method combining equality and inequalities.
- *Logit*: models without fixed effects.
- *Logit FE ML*: models with fixed effects estimated through full MLE (incidental parameter problem).

# Discussions (Finite sample issues)

When parameters are point identified by moment equalities, can inequalities improve finite sample efficiency?

- When inequalities are binding, they will act like moment equalities, hence incorporating them should improve efficiency.
- We need a way to detect binding inequalities.
- It may also be that incorporating all inequalities is not suited in practice, since some of them will be very noisy in finite sample.