# Binarization for panel models with fixed effects

Irene Botosaru and Chris Muris (SFU/Bristol)

2018 Annual Meeting, January 7, 2018

# Introduction

# Motivation

Panel models in microeconometrics typically include **fixed effects** to control for unobserved heterogeneity.

Fixed effects models are **pervasive** in applied econometrics.

In **nonlinear** panel models with **small** $T$:

1. Identification of **common parameters** is model-specific
2. If they are identified, **partial effects** may not be
3. If they are, they are restricted to be **time-invariant**

*We identify common parameters and time-varying partial effects in a large class of models.*

# Model

We study the fixed effects linear transformation (**FELT**) model:

$$
\begin{aligned}
Y_{it} &= h_t\left(\alpha_i + X_{it}\beta - U_{it}\right), && (1) \\
U_{it}|\alpha_i, X_i &\overset{d}{=} F(u|\alpha_i, X_i), && (2)
\end{aligned}
$$

where

- **transformation function** $h_t$ is
  - known, parametrized, or **unknown**
  - **time-varying**
  - **weakly** monotone
- dependent variable $Y_{it}$ can be **continuous or discrete**
- **fixed effects**: correlation $\alpha_i$ with $X_i$ is unrestricted
- error terms $U_{it}$ can be logistic or **nonparametric**

Because $h_t$ can have **flat parts and jumps**, this model nests many panel models that are important for empirical practice:

- linear $h_t(x) = x + \lambda_t$
- binary choice $h_t = 1\{x \geq \lambda_t\}$
- **ordered choice** $h_t = \sum_j 1\{x \geq \gamma_{jt}\}$
- censored regression
- duration models
- (Box-Cox) transformation models
- **nonlinear DiD**
- ...

and we consider their **extensions** to **time-varying** link functions

# Contributions

1. Identification of common parameters $(\beta, h_t)$
   - **general solution** to the incidental parameter problem
2. Identification of distribution of counterfactual outcomes
   - yields a menu of **partial effects**
   - distribution of counterfactual outcomes in **nonlinear DiD**
3. **Estimation** of $\beta$ and $h_t$
   - $\sqrt{n}-$consistent and AN
   - Exception: nonparametric $U$ **and** discrete $Y$

Our results require **two** time periods $T = 2$.

# Literature

## Incidental parameter problem

- model specific solutions (Chamberlain, 1980; Hausman et al., 1984; *long list*)
- parametric models (Lancaster, 2002; Bonhomme, 2012)
- continuous outcomes via Kotlarski's lemma (Evdokimov, 2011; Freyberger, 2017)
- random effects/large-$T$ (Arellano and Alvarez, 2003; Hahn and Kuersteiner, 2004; *long list*

**Our solution:** nonparametric $(h_t, F)$, $T = 2$, fixed effects.

## Partial effects in panel models

- Linear model with random coefficients:
  - Chamberlain, 1992; Graham, Hahn, Powell, 2009; Graham and Powell, 2012
- Nonlinear models, using time homogeneity
  - Chernozhukov et al., 2013: discrete outcomes
  - Hoderlein and White, 2012; Chernozhukov et al., 2015: continuous outcomes
- Nonlinear models, ID entire structure
  - Altonji and Matzkin, 2005: exchangeability, time invariance
  - Evdokimov, 2011; Freyberger, 2017: stronger conditions on the error term and unobserved heterogeneity, and $T \geq 3$.

**Our solution:** Nonlinear models with discrete/continuous outcomes, time-varying transformation function, using only $T = 2$. *We do not need the entire structure for partial effects.*

## Transformation models

- *Cross-sectional* transformation models (Horowitz, 1996; Chen, 2002; Chiappori, Komunjer, Kristensen, 2015)
- Closest papers on *panel* transformation model:

| | time var. | weak mono. | $\beta$ | $h_t$ |
|---|---|---|---|---|
| Abrevaya (1999) | $\surd$ | $\times$ | $\surd$ | $\times$ |
| Abrevaya (2000) | $\times$ | $\surd$ | $\surd$ | $\times$ |
| This paper | $\surd$ | $\surd$ | $\surd$ | $\surd$ |

- Extended to allow for censoring (Khan and Tamer, 2007; Chen, 2010).

**Our contribution:** identification of $h_t$ and partial effects.

# Identification of common parameters

# Model

Drop the $i$ subscripts, and set $T = 2$. Then FELT has:

$$Y_t^* = \alpha + X_t\beta - U_t, \tag{3}$$
$$Y_t = h_t(Y_t^*), \tag{4}$$
$$U_t|\alpha, X \sim F_t(u|\alpha, X), \tag{5}$$

where $Y_t^*$ is the **latent outcome** variable at time $t$.

Going forward, denote

- the supports of $Y_t$, $Y_t^*$, $X_t$ by $\mathcal{Y} \subseteq \mathbb{R}, \mathcal{Y}^* = \mathbb{R}$, and $\mathcal{X} \subseteq \mathbb{R}^K$,
- $Y = (Y_1', Y_2')'$,
- $X = (X_1', X_2')'$.

First result: **identification of the common parameters**

$$(\beta, h_1, h_2).$$

Key assumption:

**Assumption 1.** [Weak monotonicity] For each $t$, the transformation function $h_t : \mathcal{Y}^* \to \mathcal{Y}$ is unknown, non-decreasing, and right continuous.

# Proof sketch

**Step 1.** For an arbitrary pair $(y_1, y_2) \in \mathcal{Y}^2$, $\beta$ and

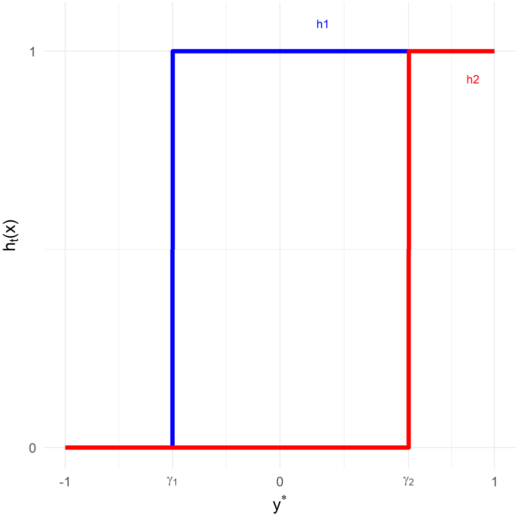
$$h_2^-(y_2) - h_1^-(y_1)$$

are identified.

In the fixed effects **binary choice** model:

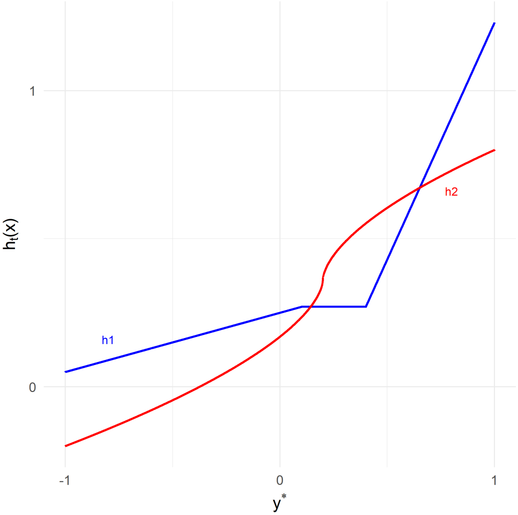
$$Y_{it} = 1\{\alpha_i + X_{it}\beta - U_{it} \geq \gamma_t\},$$

$(\beta, \gamma_2 - \gamma_1)$ are identified.



Compare the outcome equation for FELT:

$$Y_{it} = h_t\left(\alpha_i + X_{it}\beta - U_{it}\right)$$



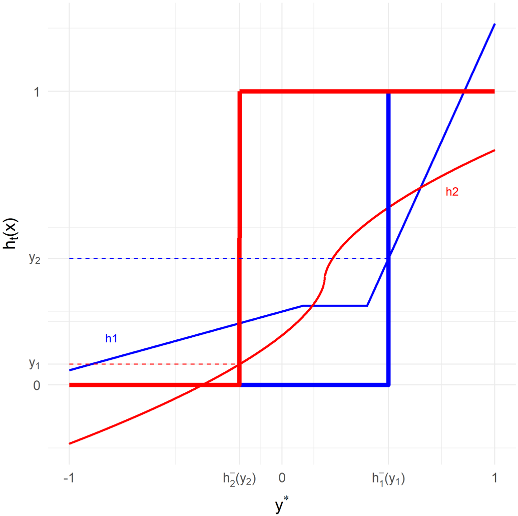
For each time period $t$, **pick** a point $y_t$ on the vertical axis, and set

$$D_t = 1\{Y_t \geq y_t\}.$$

This **transformed** outcome follows a FE binary choice model:

$$\begin{aligned} D_t &= 1\{Y_t \geq y_t\} \\ &= 1\{h_t(\alpha + X_t\beta - U_t) \geq y_t\} \\ &= 1\{\alpha + X_t\beta - U_t \geq h_t^-(y_t)\}, \end{aligned}$$

where $h_t^-$ denotes the generalized inverse of $h_t$.

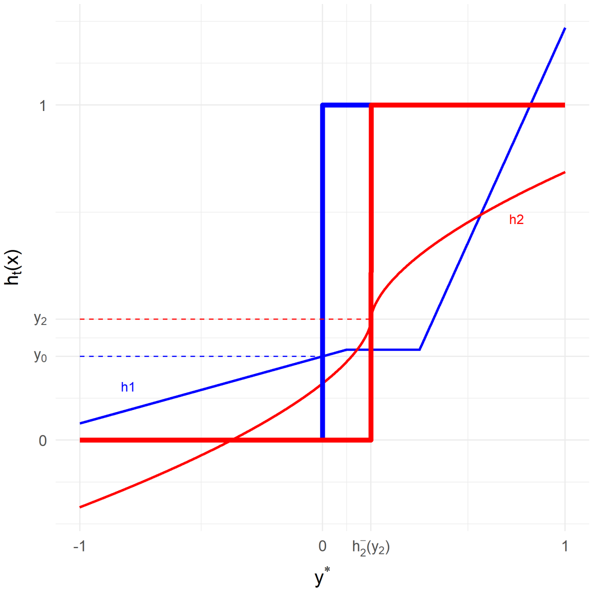



Identification of $(\beta, h_2^-(y_2) - h_1^-(y_1))$ follows by modifying existing results for binary choice. **End of step 1.**

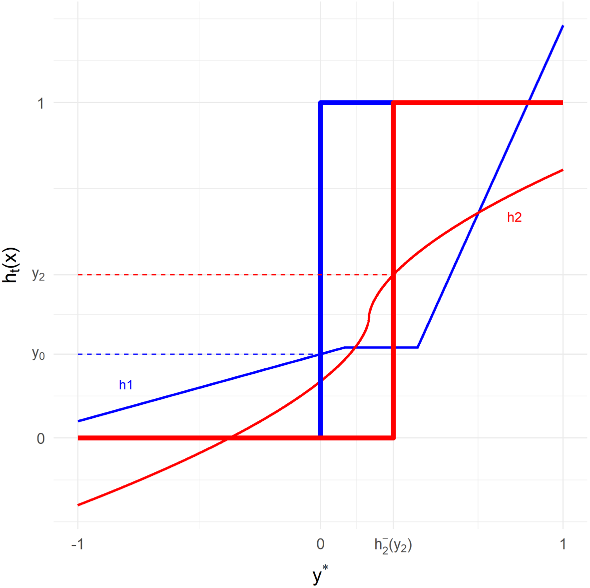
**Step 2.** Use 1 normalization to fix $h_1^-(y_0) = 0$. (Don't need one in other time periods.)

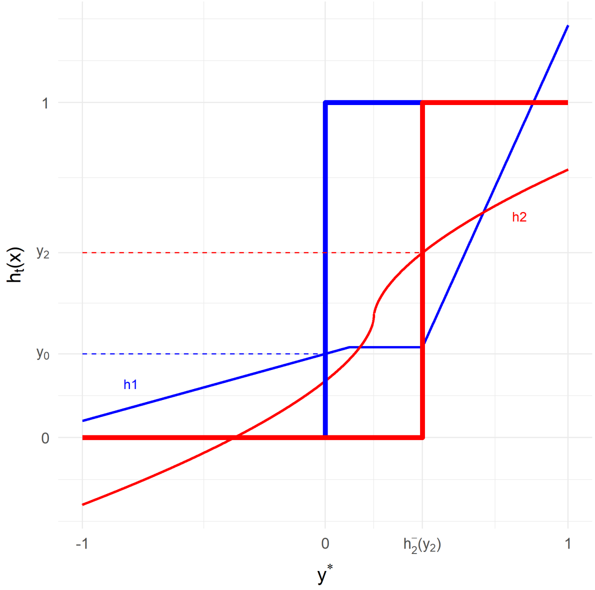
Consider all pairs $\{(y_0, y_2),\ y_2 \in \mathcal{Y}\}$ to trace out

$$h_2^-(y_2) = h_2^-(y_2) - 0 = h_2^-(y_2) - h_1^-(y_0).$$

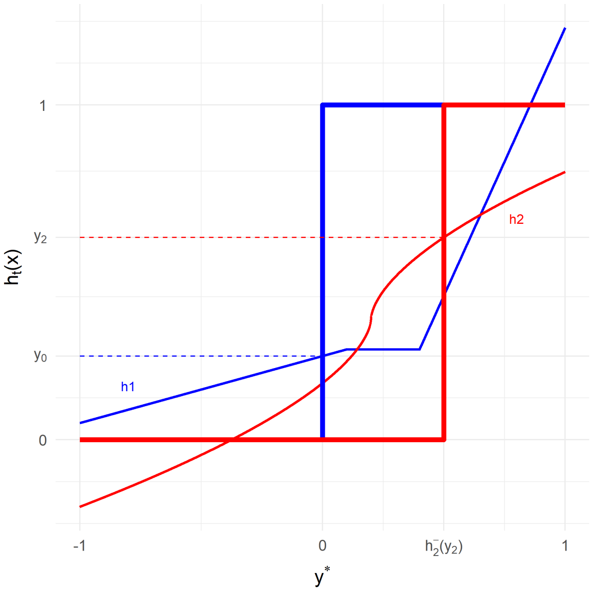


1
y2
y0
0
h1
h2
ht(x)
-1
0
h2⁻(y2)
1
y*

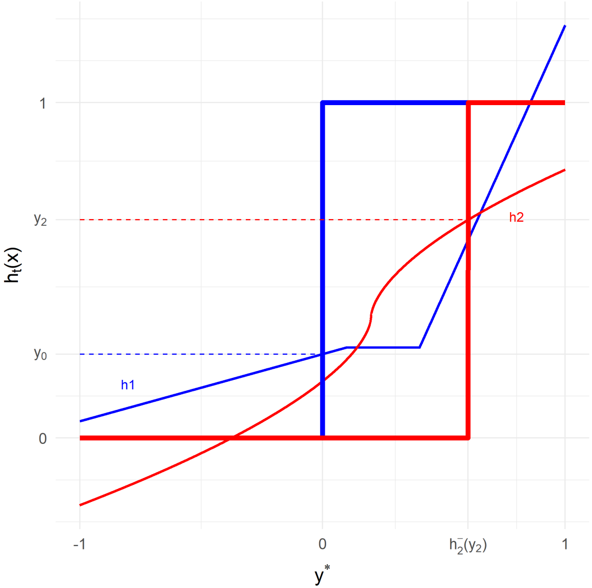

1
$y_2$
$y_0$
0
h1
h2
-1
0
$h_2^-(y_2)$
1
$h_t(x)$
$y^*$



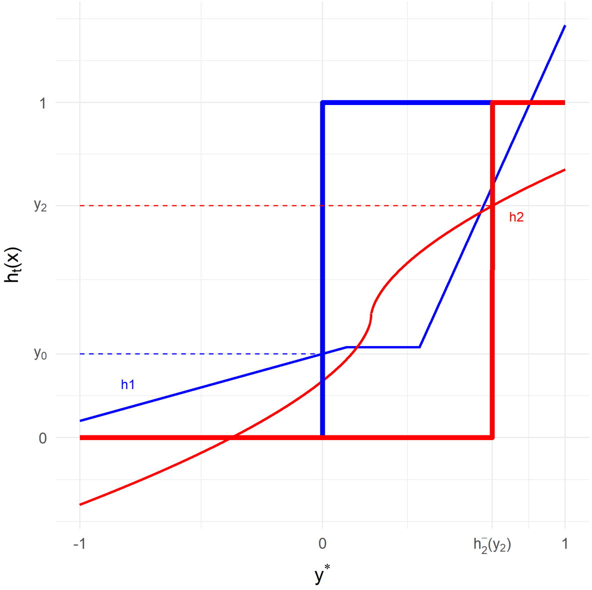
1
y2
y0
0
h1
h2
hₜ(x)
-1
0
h2⁻(y2)
1
y*



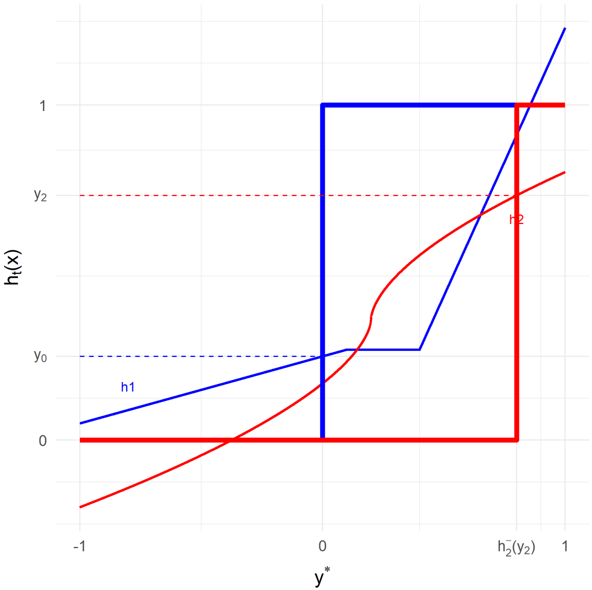
1
$y_2$
$y_0$
0
$h_t(x)$
h1
h2
-1
0
$h_2^-(y_2)$
1
$y^*$



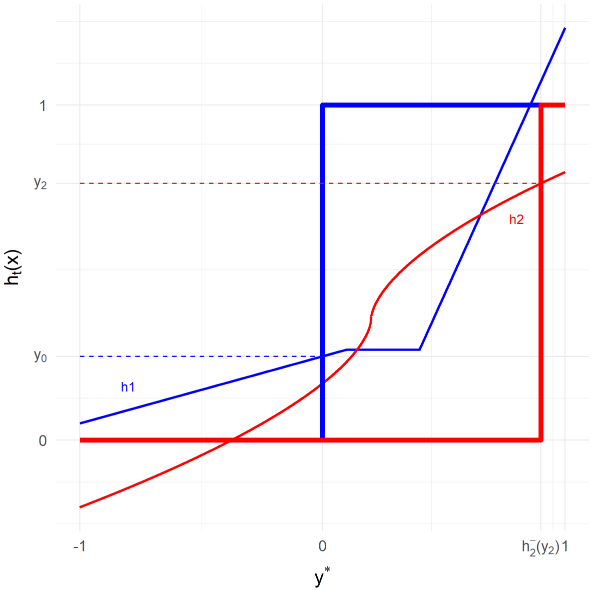
1
$y_2$
$y_0$
0
$h_t(x)$
h1
h2
-1
0
$h_2^-(y_2)$
1
$y^*$



1
y2
y0
0
h1
h2
ht(x)
-1
0
h2−(y2)
1
y*



1
$y_2$
$h_t(x)$
$y_0$
0
h1
h2
-1
0
$h_2^-(y_2)$
1
$y^*$



1
$y_2$
$y_0$
0
h2
h1
$h_t(x)$
-1
0
$h_2^-(y_2)$
1
$y^*$

**Step 3.** Recall the normalization $h_1^-(y_0) = 0$.

$$
\begin{aligned}
h_1^-(y_1) &= h_1^-(y_1) - h_1^-(y_0) \\
&= (h_1^-(y_1) - h_2^-(y_0')) + (h_2^-(y_0') - h_1^-(y_0))
\end{aligned}
$$

and both terms were identified in step 1. Trace out $h_1^-$ by considering the pairs $\{(y_1, y_0'),\ y_1 \in \mathcal{Y}\}$.

The functions $h_t$ are identified from $h_t^-$ because of monotonicity. **End of proof sketch.**

# Nonparametric errors

We provide identification and estimation results for two non-nested cases:

1. Non-parametric errors (a la Manski)
2. Logistic errors (a la Chamberlain)

The assumptions on the error terms and regressors in the first model are as in Manski (1987).

**Assumption 2.** (i) $F_1(u|\alpha, X) = F_2(u|\alpha, X) \equiv F(u|\alpha, X)$ for all $(\alpha, X)$; (ii) The support of $F(\, u|\, \alpha, X)$ is $\mathbb{R}$ for all $(\alpha, X)\,$.

Define $W = (\Delta X, -1)$.

**Assumption 3.** [Covariates] (i) The distribution of $\Delta X$ is such that at least one component of $\Delta X$ has positive Lebesgue density on $\mathbb{R}$ conditional on all the other components of $\Delta X$ with probability one. The corresponding component of $\beta$ is non-zero; (ii) The support of W is not contained in any proper linear subspace of $\mathbb{R}^{K+1}$.

We also require the following two normalizations.

**Assumption 4.** [Normalization-$\beta$] For any $(y_1, y_2) \in \underline{\mathcal{Y}}^2$, $\theta(y_1, y_2) \in \Theta = \mathcal{B} \times \mathbb{R}$, where $\mathcal{B} = \left\{\beta : \beta \in \mathbb{R}^K, \|\beta\| = 1\right\}$.

**Assumption 5.** [Normalization-$h_1$] For some $y_0 \in \underline{\mathcal{Y}}$, $h_1^-(y_0) = 0$.

Recall that $W = (\Delta X, -1)$. Denote its associated coefficient under transformation $(y_1, y_2)$ by $\theta(y_1, y_2) = (\beta, h_2^-(y_2) - h_1^-(y_1))$.

**Theorem 1.** Suppose that $(Y, X)$ follows the FELT model, and let the distribution of $(Y, X)$ be observed. Let Assumptions 1, 2, 3, and 4 hold. Then, for an arbitrary pair $(y_1, y_2) \in \underline{\mathcal{Y}}^2$, $\theta(y_1, y_2)$ is identified. With Assumption 5, the transformation functions $h_1(\cdot)$ and $h_2(\cdot)$ are identified.

## Proof step 1

For an arbitrary $y \in \underline{\mathcal{Y}}$, define the binary random variable

$$D_t\left(y\right) \equiv 1\left\{Y_t \geq y\right\}.$$

**Lemma 1.** Suppose that $(Y, X)$ follows the FELT model equations. Let Assumptions 1 and 2 hold. Then for all $(y_1, y_2) \in \underline{\mathcal{Y}}^2$,

$$\begin{aligned}
&\operatorname{med}\left(D_2\left(y_2\right) - D_1\left(y_1\right) \middle| X,\ D_1\left(y_1\right) + D_2\left(y_2\right) = 1\right) \\
&= \operatorname{sgn}\left(\Delta X \beta - \left(\left(h_2^-\left(y_2\right) - h_1^-\left(y_1\right)\right)\right)\right) \\
&\equiv \operatorname{sgn}\left(W\theta\left(y_1, y_2\right)\right),
\end{aligned}$$

where $\Delta X \equiv X_2 - X_1$.

**Proof Lemma 1.** Abbreviate $D_t = D_t(y_t)$, and define $\overline{D} \equiv D_1 + D_2$ and $D = (D_1, D_2)$.

Note that

$$\begin{aligned} P(D_t = 1|X, \alpha) &= P(Y_t \geq y_t|X, \alpha) \\ &= P(\alpha + X_t\beta - U_t \geq h_t^-(y_t)|X, \alpha) \\ &= F(\alpha + X_t\beta - h_t^-(y_t)|X, \alpha). \end{aligned}$$

Then ...

$$
\begin{aligned}
&\text{med}\left(D_2 - D_1 | X, \overline{D} = 1\right) \\
&= \text{sgn}\left(P\left(D = (0,1)\,|X, \overline{D} = 1\right) - P\left(D = (1,0)\,|X, \overline{D} = 1\right)\right) \\
&= \text{sgn}\ \left(\frac{P\left(D = (0,1)\,, \overline{D} = 1|X\right)}{P(\overline{D} = 1|X)} - \frac{P\left(D = (1,0)\,, \overline{D} = 1|X\right)}{P(\overline{D} = 1|X)}\right) \\
&= \text{sgn}\ \left(P\left(D = (0,1)\,, \overline{D} = 1|X\right) - P\left(D = (1,0)\,, \overline{D} = 1|X\right)\right) \\
&= \text{sgn}\left(P\left(D = (0,1)\,|X\right) - P\left(D = (1,0)\,|X\right)\right) \\
&= \text{sgn}\left(P\left(D_2 = 1|X\right) - P\left(D_1 = 1|X\right)\right) \\
&= \text{sgn}\left(\Delta X\beta - \left(h_2^-(y_2) - h_1^-(y_1)\right)\right).
\end{aligned}
$$

Remainder of the proof of Step 1 is similar to Manski (1985). The proof of the other steps are as in the proof sketch.

# Logistic errors

Replace the previous assumptions on $(U_1, U_2, X)$ by

**Assumption 6.** [Logit] (i)

$$F_1(u|\alpha, X) = F_2(u|\alpha, X) = \Lambda(u) = \frac{\exp(u)}{1 + \exp(u)},$$

and $U_1$ and $U_2$ are independent; (ii) $E(W'W)$ is invertible.

This obtains a logit version of the previous result:

**Theorem 3.** Suppose that $(Y, X)$ follow the FELT model equations, and let the distribution of $(Y, X)$ be observed. Let Assumptions 1 and 6 hold. Then $\theta(y_1, y_2)$ is identified for any $(y_1, y_2) \in \underline{\mathcal{Y}}^2$. With Assumption 5, the transformation functions $h_1(\cdot)$ and $h_2(\cdot)$ are identified.

**Proof:**

First, the following Lemma establishes that $\bar{D} = D_1(y_1) + D_2(y_2)$ is sufficient for the fixed effect.

**Lemma 2.** Suppose that $(Y, X)$ follows the FELT model equations. Let Assumptions 1 and 6 hold. Then for all $(y_1, y_2) \in \underline{\mathcal{Y}}^2$,

$$\begin{aligned} P(D_2(y_2) = 1 | \overline{D} = 1, X, \alpha) &= \Lambda(\Delta X \beta - (h_2^-(y_2) - h_1^-(y_1)) \\ &\equiv \Lambda(W\theta(y_1, y_2)). \end{aligned}$$

Proof Lemma 2 modifies those for FE BC logit.

1. Denote

$$p(X, y_1, y_2) \equiv P(D_2(y_2) = 1 | \overline{D} = 1, X)$$

   and note that it is identified from the distribution of $(Y, X)$.
2. Identification of $\theta(y_1, y_2)$ follows from manipulating the expression in Lemma 2 and invertibility of $E(W'W)$:

$$\theta(y_1, y_2) = [E(W'W)]^{-1} E(W' \Lambda^{-1}(p(X, y_1, y_2))).$$

3. Identification of $(h_1, h_2)$ is as in the nonparametric case.

**End of proof.**

# So far. . .

We have set up a general class of panel models with

$$Y_t = h_t(\alpha + X_t\beta - U_t)$$

and obtained identification of $(\beta, h_1, h_2)$ under two distinct sets of assumptions on the errors.

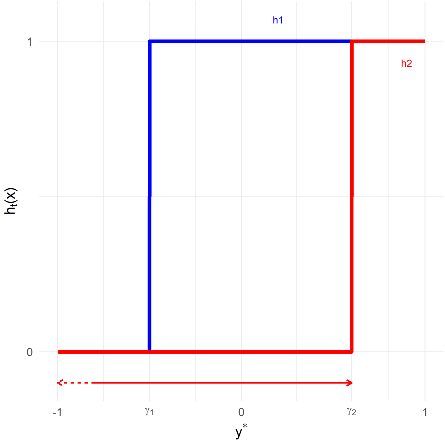
# Identification of time-varying partial effects

# Problem

Fixed effects and partial effects don't mix well.

**Example.** In the FE binary choice model,

$$Y_{it} = 0 \Rightarrow \alpha_i + X_{it}\beta - U_{it} < \lambda_t.$$



Identification of $(\beta, \lambda_t)$ does not pin down the magnitude of the effect of $X$, because $\alpha_i$ or its (conditional) distribution is not identified with $T = 2$.

# Solution

We show that identification of the common parameters $(\beta, h_t)$ is sufficient for (partial) identification of the **distribution of counterfactual outcomes**

$$P(Y_t(x) \leq y|X),$$

where

$$Y_t(x) = h_t(\alpha + x\beta - U_t).$$

**Intuition behind formal result.**

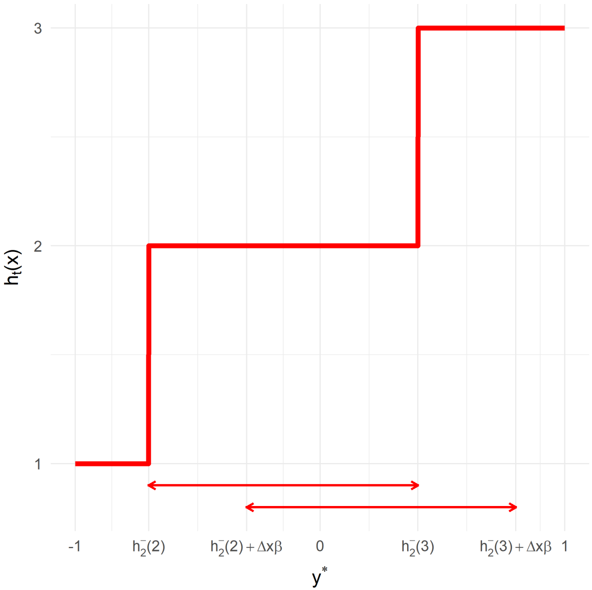
*First*, assume invertibility of $h_t$. The **observed** outcome can be turned into the **latent variable** which can be turned into a **counterfactual outcome:**

$$\begin{aligned} Y_{it}(x) &= h_t(\alpha + x\beta - U_t) \\ &= h_t(\alpha + X_{it}\beta - U_t + (x - X_{it})\beta) \\ &= h_t(h_t^{-1}(Y_{it}) + (x - X_{it})\beta) \end{aligned}$$

*Second*, if $h_t$ is not invertible (**discrete or censored outcomes**), we can still obtain bounds:

$$Y_{it}(x) \geq h_t(h_t^-(Y_{it}) + (x - X_{it})\beta),$$
$$Y_{it}(x) \leq h_t(h_t^+(Y_{it}) + (x - X_{it})\beta),$$

where $h^+$ denotes the right-inverse.



3
2
1
$h_t(x)$
-1
$h_2^-(2)$
$h_2^-(2)+\Delta x\beta$
0
$h_2^-(3)$
$h_2^-(3)+\Delta x\beta$
1
$y^*$

*Third*, observations from **other time periods** are informative:

$$\begin{aligned} Y_{it}(x) &= h_t(\alpha + x\beta - U_{it}) \\ &\overset{d}{=} h_t(\alpha + x\beta - U_{is}) \\ &= h_t(\alpha + X_{is}\beta - U_{is} + (x - X_{is})\beta) \\ &= h_t(h_s^{-1}(Y_{is}) + (x - X_{is})\beta), \end{aligned}$$

where $\overset{d}{=}$ denotes equality in distribution conditional on $X_i$.

This is particularly useful when outcomes are discrete, since $Y_{it} \in \{\min \mathcal{Y}, \max \mathcal{Y}\}$ leads to uninformative bounds.

# Result

**Corollary 1.** Let the conditions of Theorem 1 or 3 hold. Then, for $s, t \in \{1, 2\}$,

$$\begin{aligned} & \max_s L_s(x, y; \beta, h_s, h_t) \\ & \leq P(Y_t(x) \leq y | X = x) \\ & \leq \min_s U_s(x, y; \beta, h_s, h_t), \end{aligned}$$

where

$$L_s(x, y; \beta, h_s, h_t) \equiv P\left( Y_s \leq h_s\left(h_t^-(y) + (X_s - x)\beta\right) \middle| X = x\right),$$
$$U_s(x, y; \beta, h_s, h_t) \equiv P\left( Y_s \leq h_s\left(h_t^+(y) + (X_s - x)\beta\right) \middle| X = x\right).$$

**Proof.**

Formalize the intuition above, at the population level. For $s, t \in \{1, 2\}$,

$$\begin{aligned}
P\left(Y_t(x) \leq y \middle| \alpha, X\right) &= P\left(h_t\left(\alpha + x\beta - U_t\right) \leq y \middle| \alpha, X\right) \\
&\overset{d}{=} P\left(h_t\left(\alpha + x\beta - U_s\right) \leq y \middle| \alpha, X\right) \\
&\geq P\left(\alpha + x\beta - U_s \leq h_t^-(y) \middle| \alpha, X\right) \\
&= P\left(\alpha + X_s\beta - U_s \leq h_t^-(y) + (X_s - x)\beta \middle| \alpha, X\right) \\
&= P\left(Y_s \leq h_s\left(h_t^-(y) + (X_s - x)\beta\right) \middle| \alpha, X\right).
\end{aligned}$$

Complete the proof by

1. obtaining the upper bound using the right inverse,
2. integrating out wrt $\alpha | X$
3. taking the minimum/maximum across $s$, conditional on $X$.

**Remarks:**

1. Bounds are more informative for larger $|\mathcal{Y}|$
2. Bounds are more informative for larger $T$
3. Counterfactual distributions lead to results for

$$P\left( Y_t\left(x\right) \leq y \middle| X \in \overline{\mathcal{X}}\right),$$

   marginal effects, or ...
4. Useful in a difference-in-differences setting

# Estimation

# Overview

| Errors | Outcome | Estimator | Rate |
|---|---|---|---|
| Logistic | | composite CMLE | $\sqrt{n}$ |
| Nonparametric | continuous | two-step rank | $\sqrt{n}$ |
| Nonparametric | discrete | maximum score | $n^{1/3}$ |

1. Results are uniform over compact subsets.
2. Results: see paper.
3. We recommend **composite CMLE** for applied practice, and use it in the simulations below.

The CMLE is

$$\widehat{\theta}_n\left(y_1, y_2\right) = \text{argmax}_{\theta \in \mathbb{R}^{K+1}} \frac{1}{n} \sum_{i=1}^{n} l_i\left(\theta, y_1, y_2\right),$$

based on the conditional log-likelihood contribution $l_i(\theta, y_1, y_2)$:

$$\overline{D}_i\left(y_1, y_2\right)\left[D_{i2}(y_2) \ln \Lambda\left(W_i \theta\right) + \left(1 - D_{i2}(y_2)\right) \ln\left(1 - \Lambda\left(W_i \theta\right)\right)\right],$$

with information matrix $J(y_1, y_2)$.

**Theorem 7.** Under the identification conditions for logit FELT and a random sampling assumption

$$\sqrt{n}\left(\widehat{\theta}\left(y_1, y_2\right) - \theta_0\left(y_1, y_2\right)\right) \xrightarrow{d} \mathcal{N}\left(0, J^{-1}\left(y_1, y_2\right)\right)$$

as $n \to \infty$.

For discrete outcomes, the CMLEs can be combined into estimators for $(\beta, h_1, h_2)$. For continuous outcomes, we need a **functional CLT.**

**Assumption 9.** (i) $E\left\|\Delta X_i\right\|^{2+\epsilon}<\infty$ for some $\epsilon>0$; (ii) the conditional density $f_{Y_t}\left(y|\Delta X_i=x\right),\ t=1,2$, exists, and it is bounded and uniformly continuous in $y$, uniformly in $x$ over the support of $\Delta X_i$; (iii) $h_t$ is continuous for each $t=1,2$.

**Theorem 8.** Assume that the conditions for Theorem 7 hold, and let Assumption 8 hold. Then

$$\sqrt{n}\left(\widehat{\theta}\left(\cdot\right)-\theta\left(\cdot\right)\right)\Rightarrow z\left(\cdot\right)\ \text{ in } \ell^{\infty}\left(\left[\underline{y},\overline{y}\right]^2\right)$$

as $n\rightarrow\infty$ where $z\left(\cdot\right)$ is a Gaussian process with covariance function $\Sigma\left(y_1,y_2,y_1^{'},y_2^{'}\right)$.

With that result in hand, we analyze the behavior of the **composite CMLE**, which maximizes:

$$\widetilde{l}_i\left(\beta, h_2^-(\cdot), h_1^-(\cdot)\right) = \int_{[\underline{y},\overline{y}]}\int_{[\underline{y},\overline{y}]} w(y_1,y_2)\, l_i(\theta,y_1,y_2)\, dy_1 dy_2,$$

which imposes the equality constrainst.

- See paper for details.
- $w = 1$ works well!

# Nonlinear DiD

# Literature

Few papers on nonlinear difference-in-differences:

- **Discrete and continuous outcomes:** Athey and Imbens (2006) - CiC
- **Continuous outcomes:** Bonhomme and Sauder (2011) and D'Haultfoeuille et al. (2015).
- **Quantile difference-in-differences:** Callaway and Li (2017).

# Our contribution

**Identification:**

- distribution of counterfactual outcomes of treated
- accommodates both continuous and discrete outcomes
- extends CiC to continuous outcomes with censoring, and to discrete outcomes with fixed effects
- applies to panel data only

**Estimation:**

- easy to implement
- $\sqrt{n}-$consistent and asymptotically normal
- trivial to include regressors

# Model

**Standard setup**

- Before ($t = 1$) and after ($t = 2$)
- Treated ($S_1 = 0, S_2 = 1$) and control ($S_1 = S_2 = 0$)
- Potential outcomes
  - in absence of treatment: $Y_t(0)$
  - under treatment: $Y_t(1)$
- Observed outcome: $Y_t = S_t Y_t(1) + (1 - S_t) Y_t(0)$

**Control outcomes follow FELT.**

$$Y_t(0) = h_t(\alpha + X_t\beta - U_t(0))$$
$$U_t(0)|\alpha, X \overset{d}{=} F$$

**Parameter of interest:**

Distribution of the counterfactual outcome for the treated,

$$\tau(y; X) = P(Y_2(0) \leq y|X, S_1 = 0, S_2 = 1).$$

Can be turned into ATT.

**Corollary 2.** The bounds on the distribution of counterfactual outcomes are given by:

$$
\begin{aligned}
& P\left(\widetilde{Y}_2^l(0) \leq y | X, S_1 = 0, S_2 = 1\right) \\
\leq & \tau(y; X) \\
\leq & P\left(\widetilde{Y}_2^u(0) \leq y | X, S_1 = 0, S_2 = 1\right)
\end{aligned}
$$

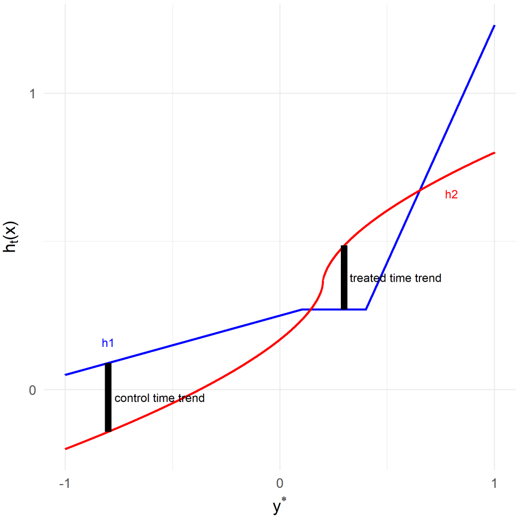
where

$$
\begin{aligned}
\widetilde{Y}_2^l(0) &\equiv h_2\left(h_1^-(Y_1) + (X_2 - X_1)\beta\right) \\
\widetilde{Y}_2^u(0) &\equiv h_2\left(h_1^+(Y_1) + (X_2 - X_1)\beta\right)
\end{aligned}
$$

*Subtract the time period 1 time trend, adjust the covariates, add the period 2 time trend.*

# Linear DiD predicts

$$E(Y_2(0)|\text{treated}) = E(Y_1(0)|\text{treated}) + \text{control time trend}.$$



# Simulations

**Control group**. Potential outcomes $Y_{it}(0)$ follow FELT with

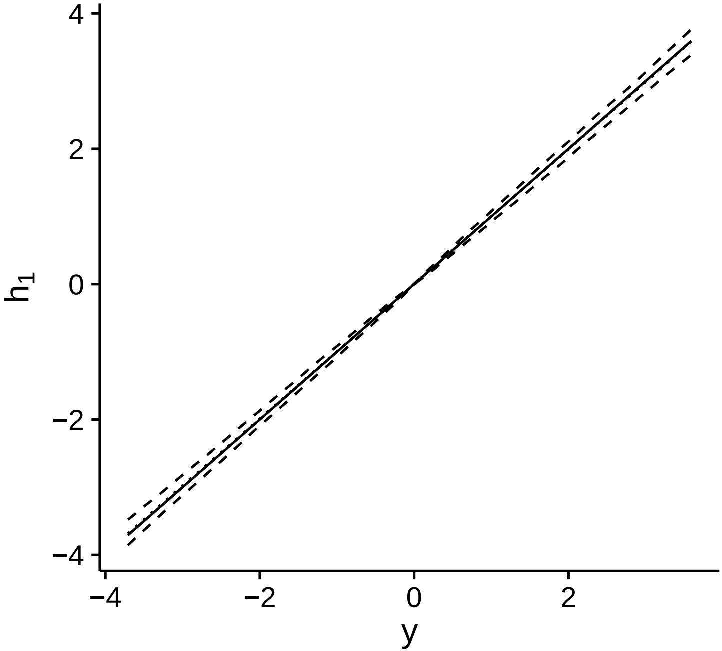
$$h_1(y^*) = y^*$$
$$h_2(y^*) = \Phi\left(\frac{y^* - 1}{0.5}\right)$$

In particular

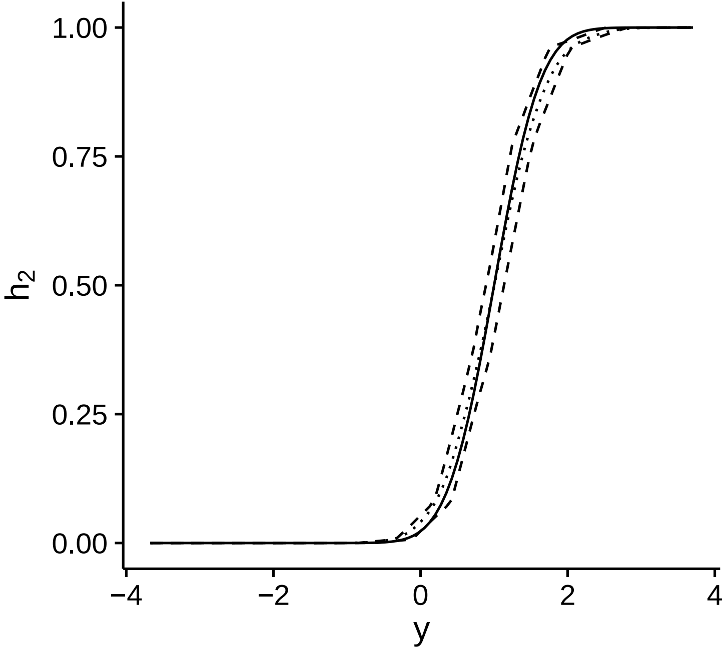
$$Y_{i1} = Y_{i1}(0) = \alpha_i + X_{i1}\beta - U_{i1}(0)$$
$$Y_{i2} = Y_{i2}(0) = \Phi\left(\frac{\alpha_i + X_{i2}\beta - U_{i2}(0) - 1}{0.5}\right)$$

- $F(u|\alpha_i, X_i)$ logistic
- $X_{it} \sim N(0,1)$
- $\alpha_i \sim N(0,1) + \frac{1}{2}(X_{i1} + X_{i2})$
- $\beta = 1$
- $n = 500, S = 1000.$
- Discretize $\mathcal{Y}$: 12 points at quantiles of $Y_t$ (for $h_t^{-1}$)

**Result** for $\widehat{\beta}$: $bias\left(\widehat{\beta}\right) = 0.01$, $RMSE\left(\widehat{\beta}\right) = 0.1$



4
2
0
−2
−4
h1
−4
−2
0
2
y



1.00
0.75
0.50
0.25
0.00
h2
−4
−2
0
2
4
y

## Treatment group

Heterogeneous treatment effects through $\gamma_i$:

$$
\begin{aligned}
Y_{i1} &= Y_{i1}(0) = \alpha_i + X_{i1}\beta - U_{i1}(0) = h_1(y^*) \\
Y_{i2}(0) &= \Phi\left(\frac{\alpha_i + X_{i2}\beta - U_{i2}(0) - 1}{\sigma}\right) \\
Y_{i2}(1) &= \Phi\left(\frac{\alpha_i + X_{i2}\beta - U_{i2}(1) + \gamma_i - 1}{\sigma}\right) \\
\alpha_i &\sim \mathcal{N}(\mu, 1) + \frac{1}{2}(X_{i0} + X_{i1}) \\
\mu &= 1 > 0, \\
\sigma &= 0.5 \\
\gamma_i &\sim \mathcal{N}(1, 1)
\end{aligned}
$$

## Comparison with linear DiD

- Panel regression for DiD

$$Y_{it} = \alpha_i + \delta_t + X_{it}\beta + \tau S_{it} + \varepsilon_{it}$$

- Design is difficult for linear DiD:
  - nonlinearity in $h_2$
  - location shift in $Y_{it}^*$ ($\alpha_i$ has mean $\mu = 1 > 0$)
- Run $S = 1000$, $n^{control} = n^{treat} = 500$.

**Results**:

- True ATT $= 0.1403$.
- DID $\widehat{\tau} = -0.7126$
- FELT $\widehat{ATT} = 0.1412$

Design (0): benchmark design described above.

Design (1): as (0) but with 6 points of discretization.

- FELT estimate of $h_2$ worse, increased ATT bias.

Design (2): $\sigma = 0.25$ (steeper $h_2$)

- relative performance unchanged.

Design (3): $\mu = 0$ (same cdf of $y^*$ for treated and control)

- DiD gets the trend despite the (not-so-severe) nonlinearity

Design (4): $h_2(y^*) = y^*$ (standard DiD framework)

- DiD consistent, and outperforms FELT

| Design | | $\beta$ | | | | ATT | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | DiD | | FELT | |
| | $S$ | 100b | rmse | true | 100b | rmse | 100b | rmse |
| (0) | 1000 | 1.00 | 0.10 | 0.14 | -85.00 | 0.15 | 0.08 | 0.03 |
| (1) | 1000 | 1.14 | 0.10 | | -85.00 | 0.15 | 5.83 | 0.04 |
| (2) | 100 | 1.75 | 0.10 | | -87.00 | 0.15 | 0.50 | 0.03 |
| (3) | 100 | 1.56 | 0.10 | 0.15 | -2.39 | 0.15 | -0.19 | 0.03 |
| (4) | 100 | 1.57 | 0.10 | 1.00 | -1.49 | 0.13 | -3.90 | 0.18 |

# Conclusion

# Conclusion

We consider the class of FELT models with **fixed-T** and:

- provide a **general solution** to the incidental parameter problem.
  - existing solutions are model-specific or likelihood-based.
- show identification of **distribution of counterfactual outcomes** at time $t$
  - current fixed-T results rely on time-homogeneity.
- extend our results to **FELT with RC**; apply our results to **nonlinear DiD**
- provide **estimators, parametric rate and AN**
  - except for nonparametric discrete

# Extensions: Random coefficients

Consider the extension to random coefficients.

$$Y_{it} = h_t(\alpha_i + X_{it}\beta + Z_{it}\gamma_i - U_{it}).$$

Assume that

- $h_t$ is invertible
- $U_{it}|\alpha_i, X_{it}, Z_{it} \sim LOG(0,1)$

Then

$$P(D_{it}(y_t) = 1|\bar{D}_i = 1, X_i, Z_i, \Delta Z_i = 0) = \Lambda(\Delta X_i \beta - (h_2^{-1}(y_2) - h_1^{-1}(y_1)))$$

and we can use the tools in this paper to identify $h_t, \beta$. Then

$$h_t^{-1}(Y_{it}) - X_{it}\beta = \alpha_i + Z_{it}\gamma_i - U_{it}$$

and we can use the tools in Graham and Powell (2012) to obtain partial effects.