Distributional Treatment Effect with Latent Rank Invariance

Myungkou Shin

University of Surrey

Midwest Econometrics Group annual conference

October 3, 2025

Distributional treatment effect

Potential outcome setup: with $D \in \{0,1\}$,

$$Y = D \cdot Y(1) + (1-D) \cdot Y(0).$$

We do not observe $Y(1)$ and $Y(0)$ simultaneously; focus on ATE, LATE, etc.

Some questions can only be answered with **distribution** of treatment effect $Y(1) - Y(0)$.
"How many people are better off under the treatment?"
"How heterogeneous is the treatment effect at the individual level?"

Various distributional treatment effect (DTE) parameters can be defined:
$\mathrm{Var}\left(Y(1) - Y(0)\right)$, $\Pr\{Y(1) - Y(0) \geq 0\}$, etc.

Distributional treatment effect

Existing approaches

- Partial identification: put a bound on $\Pr\{Y(1) - Y(0) \leq y\}$

  Heckman et al. (1997); Fan and Park (2010); Fan et al. (2014); Firpo and Ridder (2019)
  Frandsen and Lefgren (2021); Kaji and Cao (2023) and more

- Independence: assume $Y(1) \perp\!\!\!\perp Y(0)$ or $Y(0) \perp\!\!\!\perp \big(Y(1) - Y(0)\big)$

  Heckman et al. (1997); Noh (2023)

In this paper, I follow the latter, assuming a latent variable U such that

$$Y(1) \perp\!\!\!\perp Y(0) \,\big|\, U$$

and two proxy variables X and Z to identify the cond. dist. of $Y(d)$ given U.

Distributional treatment effect: setup

An econometrican observes $\{Y_i, D_i, X_i, Z_i\}_{i=1}^n$:

$$Y_i = D_i \cdot Y_i(1) + (1 - D_i) \cdot Y_i(0).$$

$Y_i, X_i, Z_i \in \mathbb{R}$, $D_i \in \{0, 1\}$ and $\big(Y_i(1), Y_i(0), D_i, X_i, Z_i, U_i\big) \sim iid$.
X_i and Z_i are proxy variables for U_i. $U_i \in \mathbb{R}$.

Assumption 1. $\big(Y_i(1), Y_i(0), X_i\big) \perp\!\!\!\perp \big(D_i, Z_i\big) \mid U_i$.
- One of the proxy Z_i and the latent variable U_i are confounders.
- In proximal inference terminology,
  X_i is outcome-aligned proxy and Z_i is treatment-aligned proxy.
  Hu and Schennach (2008); Miao et al. (2018); Deaner (2023); Nagasawa (2022) and more

Assumption 2. $Y_i(1), Y_i(0), X_i$ are mutually independent given U_i.

Distributional treatment effect: example 1 *(rank invariance)*

Assume **rank invariance** between $Y_i(1)$ and $Y_i(0)$:

$$\Pr\left\{F_{Y(1)}\big(Y_i(1)\big) = F_{Y(0)}\big(Y_i(0)\big)\right\} = 1.$$

When D_i is random, Assumptions 1-2 trivially hold with

$$U_i = X_i = Z_i = F_{Y(1)}\big(Y_i(1)\big) = F_{Y(0)}\big(Y_i(0)\big).$$

Rank invariance is a commonly used assumption in quantile treatment effect/IV literature:
Chernozhukov and Hansen (2005, 2006); Athey and Imbens (2006); Callaway and Li (2019) and more.

In this paper,

1. U_i is latent and not a deterministic function of $Y_i(1)$ or $Y_i(0)$;
2. *conditional expectation of $Y_i(d)$ given U_i* have the same rank.

Hence '*latent* rank invariance.'

Distributional treatment effect: example 2 *(hidden Markov model)*

Consider a panel where $T = 3$ and $D_i = 1$ means being treated for $t = 2, 3$.

$$Y_{it}(d) = g_d\big(U_{it}, \varepsilon_{it}(d)\big).$$

There are a common shock U_{it} and treatment-status-specific shocks $\big(\varepsilon_{it}(0), \varepsilon_{it}(1)\big)$.

Assumption 2 holds when 1) $\{U_{it}\}_{t=1}^{3}$ is first-order Markov given D_i and
2) $\left(\{U_{it}\}_{t=1}^{3}, D_i\right), \varepsilon_{i1}(0), \varepsilon_{i2}(1), \varepsilon_{i2}(0), \varepsilon_{i3}(1), \varepsilon_{i3}(0) \sim$ ind.

Hidden Markov model: Kasahara and Shimotsu (2009); Hu and Shum (2012) and more.

Panel data model with proxy: Deaner (2023)

In this paper,

1. treatment-status-specific shocks are introduced and assumed to be $\varepsilon_{it}(1) \perp\!\!\!\perp \varepsilon_{it}(0)$.

Identification

Assumption 3/4. full rank/completeness of $f_{X|Z}$ when U_i is discrete/continuous: A3 A4
"Both of the proxy variables are informative for the latent variable U_i."

Assumption 5. $\mathbf{E}[Y_i(1)|U_i = u]$ and $\mathbf{E}[Y_i(0)|U_i = u]$ are strictly increasing in u.
"Conditional expectation of $Y_i(1)$ given U_i and that of $Y_i(0)$ given U_i have the same rank."
U_i can be thought of as a 'latent' or 'interim' rank.

I apply Hu and Schennach (2008) to treated subpopulation and to untreated subpopulation. more

Theorem 1.
Assumptions 1-3 or Assumptions 1-2, 4-5 hold. Then, the distribution of $(Y_i(1), Y_i(0), D_i, X_i, Z_i)$ is identified.

Identification

From Theorem 1, the following two distributional treatment effect (DTE) parameters are identified:

$$F_{Y(0),Y(1)}(y,y') = \int_{\mathbb{R}} F_{Y(0),Y(1)|U}(y,y'|u) f_U(u)du = \int_{\mathbb{R}} F_{Y(0)|U}(y|u) \cdot F_{Y(1)|U}(y'|u) f_U(u)du,$$

$$F_{Y(1)-Y(0)}(\delta) = \int_{\mathbb{R}} F_{Y(1)-Y(0)|U}(\delta|u) f_U(u)du = \int_{\mathbb{R}} \int_{\mathbb{R}} F_{Y(1)|U}(y+\delta|u) \cdot f_{Y(0)|U}(y|u) f_U(u)dydu.$$

Other distributional parameters $\mathrm{Var}(Y_i(1) - Y_i(0))$ are identified and can be estimated as well.

To estimate DTEs, estimate $\{F_{Y(1)|U}(\cdot|u)\}_u$, $\{F_{Y(0)|U}(\cdot|u)\}_u$ and $f_U(\cdot)$.

Implementation: finite support assumption

To let U_i be continuous and apply the standard semiparametric estimation theory,

Shen (1997); Chen and Shen (1998); Ai and Chen (2003) and more

Need strong assumptions such as bounded support of Y_i and X_i.

Why? DTE parameters are inner products of densities. sieve

Instead, I assume $U_i \in \{u^1, \cdots, u^K\}$ with $K < \infty$. choice of K

Reasoning behind the finite support assumption:

1. Finite mixture: Henry et al. (2014) and more.
   Discretization as approximation: Bonhomme et al. (2022) and more.
2. DTE parameters are identified with quadratic moments;
   a limiting distribution is derived from U stat. theory and Neyman orthogonality.
3. Identification is not tied to the finite support;
   alternative asymptotic theory could be developed as well.

Implementation: premise

From Assumptions 1-2 and $U_i \in \{u^1, \cdots, u^K\}$, we have a finite mixture representation:

$$
\begin{aligned}
&\begin{pmatrix} F_{Y|D=d,Z}(y|\mathcal{Z}^1) & \cdots & F_{Y|D=d,Z}(y|\mathcal{Z}^K) \end{pmatrix} \\
&\quad = \begin{pmatrix} F_{Y(d)|U}(y|u^1) & \cdots & F_{Y(d)|U}(y|u^K) \end{pmatrix} \\
&\qquad \cdot \underbrace{\begin{pmatrix} \Pr\{U_i = u^1 | D_i = d, Z_i \in \mathcal{Z}^1\} & \cdots & \Pr\{U_i = u^1 | D_i = d, Z_i \in \mathcal{Z}^K\} \\ \vdots & \ddots & \vdots \\ \Pr\{U_i = u^K | D_i = d, Z_i \in \mathcal{Z}^1\} & \cdots & \Pr\{U_i = u^K | D_i = d, Z_i \in \mathcal{Z}^K\} \end{pmatrix}}_{=\Lambda_d}.
\end{aligned}
$$

With ${\Lambda_d}^{-1}$, the distribution of $Y_i(d)$ given U_i are linear in observed distributions.

Implementation: nonnegative matrix factorization

Fix K. Given some partitions $\{\mathcal{Y}^m\}_m, \{\mathcal{X}^{m'}\}_{m'}, \{\mathcal{Z}^l\}_l$, let

$$\mathbf{H}_d = \begin{pmatrix} \Pr\left\{Y_i \in \mathcal{Y}^1, X_i \in \mathcal{X}^1 \middle| D_i = d, Z_i \in \mathcal{Z}^1\right\} & \cdots & \Pr\left\{Y_i \in \mathcal{Y}^1, X_i \in \mathcal{X}^1 \middle| D_i = d, Z_i \in \mathcal{Z}^{M_Z}\right\} \\ \vdots & \ddots & \vdots \\ \Pr\left\{Y_i \in \mathcal{Y}^{M_Y}, X_i \in \mathcal{X}^{M_X} \middle| D_i = d, Z_i \in \mathcal{Z}^1\right\} & \cdots & \Pr\left\{Y_i \in \mathcal{Y}^{M_Y}, X_i \in \mathcal{X}^{M_X} \middle| D_i = d, Z_i \in \mathcal{Z}^{M_Z}\right\} \end{pmatrix}$$
$$= \Gamma_d \cdot \Lambda_d$$

where $\Gamma_d = \left(\Pr\left\{Y_i(d) \in \mathcal{Y}^m, X_i \in \mathcal{X}^{m'} \middle| U_i = u^k\right\}\right)_{(m,m'),k}$

$$\Lambda_d = \left(\Pr\left\{U_i = u^k \middle| D_i = d, Z_i \in \mathcal{Z}^l\right\}\right)_{k,l}.$$

$\mathbf{H}_d$ is a discretization of $f_{Y,X|D=d,Z}$.

The full rank condition implies $|\text{supp}_Z| \geq K$; if $|\text{supp}_Z| > K$, use partition $\{\mathcal{Z}^l\}_{l=1}^K$.

Implementation: nonnegative matrix factorization

Solve the following nonnegative matrix factorization problem:

$$\left(\widehat{\Gamma}_0, \widehat{\Gamma}_1, \widehat{\Lambda}_0, \widehat{\Lambda}_1\right) = \arg\min \|\mathbb{H}_0 - \Gamma_0 \cdot \Lambda_0\|_F + \|\mathbb{H}_1 - \Gamma_1 \cdot \Lambda_1\|_F \tag{1}$$

subject to 1) $\Gamma_0, \Gamma_1, \Lambda_0, \Lambda_1$ are nonnegative.

Also, their columnwise sums are one. $\cdots$ *(linear constraints)*

2) Γ_0 and Γ_1 satisfy $Y_i(d) \perp\!\!\!\perp X_i \mid U_i$ $\cdots$ *(quadratic constraints)*

3) Γ_0 and Γ_1 imply the same marginal distribution of X_i $\cdots$ *(linear constraints)*

The objective becomes quadratic once we fix (Γ_0, Γ_1) or (Λ_0, Λ_1).

The quadratic constraint becomes linear once we fix Γ_X or $(\Gamma_{Y0}, \Gamma_{Y1})$.

(1) is solved iteratively. algorithm

Theorem 2. Assumptions 1-3 hold. Up to some permutation on $\{u^1, \cdots, u^K\}$,

$$\left\|\widehat{\Lambda}_0 - \Lambda_0\right\|_F = O_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \left\|\widehat{\Lambda}_1 - \Lambda_1\right\|_F = O_p\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$.

The convergence rate is $n^{-\frac{1}{2}}$.
A direct corollary is that $\left(\widehat{\Lambda}_d\right)^{-1}$ is consistent for $(\Lambda_d)^{-1}$ at the same rate.

Implementation: plug-in GMM

Then, quadratic moments identify DTE:

$$F_{Y(1),Y(0)}(y,y') = \sum_{k,l,m=1}^{K} w_{klm} \cdot \mathbf{E}\big[\mathbf{1}\{Y_i \le y, D_i = 1, Z_i \in \mathcal{Z}^m, Y_j \le y', D_j = 0, Z_j \in \mathcal{Z}^l\}\big]$$

$$F_{Y(1)-Y(0)}(\delta) = \sum_{k,l,m=1}^{K} w_{klm} \cdot \mathbf{E}\big[\mathbf{1}\{Y_i \le Y_j + \delta, D_i = 1, Z_i \in \mathcal{Z}^m, D_j = 0, Z_j \in \mathcal{Z}^l\}\big]$$

for all $(y,y') \in \mathbb{R}^2$ and $\delta \in \mathbb{R}$, with $(Y_i, D_i, Z_i) \perp\!\!\!\perp (Y_j, D_j, Z_j)$ and

- $w_{klm} = \dfrac{p_U(k)\tilde{\lambda}_{lk,0}\tilde{\lambda}_{mk,1}}{p_{D,Z}(0,l)p_{D,Z}(1,m)}$.
- $\tilde{\lambda}_{lk,d}$ is l-th row k-th column component of $(\Lambda_d)^{-1}$ $\quad \forall d = 0, 1$.
- $p_U(k) := \Pr\{U_i = u^k\} \quad \forall k = 1, \cdots, K$.
- $p_{D,Z}(d,l) := \Pr\{D_i = d, Z \in \mathcal{Z}^l\} \quad \forall d = 0, 1$ and $l = 1, \cdots, K$.

Implementation: plug-in GMM

The DTE estimators are plug-in GMM estimator from (orthogonalized) moment: Lemma 1.

$$\widehat{F}_{Y(1)-Y(0)}(\delta) = \sum_{k,l,m=1}^{K} \widehat{w}_{klm} \cdot \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{Y_i \leq Y_j + \delta, D_i = 1, Z_i \in \mathcal{Z}^m, D_j = 0, Z_j \in \mathcal{Z}^l\}$$

$+$ orthogonalization term

where $\widehat{w}_{klm} = \dfrac{\hat{p}_U(k)\widehat{\tilde{\lambda}}_{lk,0}\widehat{\tilde{\lambda}}_{mk,1}}{\hat{p}_{D,Z}(0,l)\hat{p}_{D,Z}(1,m)}$.

Theorem 3. Assumptions 1-3 hold. Then, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{F}_{Y(1),Y(0)}(y,y') - F_{Y(1),Y(0)}(y,y')\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma(y,y')^2\right)$$
$$\sqrt{n}\left(\widehat{F}_{Y(1)-Y(0)}(\delta) - F_{Y(1)-Y(0)}(\delta)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma(\delta)^2\right).$$

Empirical Illustration

I revisit Jones et al. (2019), which studies the effect of workplace wellness program.
The program *eligibility* was randomly assigned to employees at UIUC; intent-to-treat.
Using the University-provided health insurance data,
Jones et al. (2019) estimates its effect on medical spending.

The variables in the dataset are:

$$Y_i = \text{monthly medical spending over August 2016-July 2017}$$

$$D_i = \mathbf{1}\{\textit{eligible} \text{ for the wellness program starting in September 2016}\}$$

$$X_i = \text{monthly medical spending over July 2015-July 2016}$$

$$Z_i = \text{monthly medical spending over August 2017-January 2019}$$

“Underlying health status U_i depends on past health status, but not on medical spendings.”

Empirical Illustration: setup

The eigenvalue ratio estimator and the Kleibergen-Paap rank test suggest $K = 3$. choice of K

I use $K = 5$ and used 5-fold partitions for Y_i, X_i and Z_i:
e.g. $F_Y^{-1}(0), F_Y^{-1}(1/5), \cdots, F_Y^{-1}(1)$ and so on.

As a falsification test, I test the null hypothesis $f_{X|D=1,U}(\cdot|u) = f_{X|D=0,U}(\cdot|u)$ for all u:
with $W_n = \left(\hat{f}_{X|D=1,U}(\mathcal{X}^m|u) - \hat{f}_{X|D=0,U}(\mathcal{X}^m|u)\right)_{m,u} \in \mathbb{R}^{25}$,

$$nW_n^\intercal Avar(W)^{-1} W_n = 16.435$$

The p-value is 0.689.

Empirical Illustration: joint distribution of potential outcomes

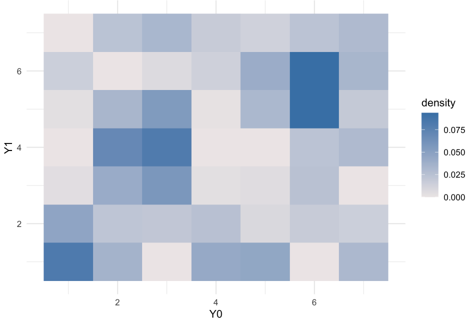

Figure 1: Joint density of $\big(Y(1), Y(0)\big)$.

y-axis is $Y(1)$ and x-axis $Y(0)$; each cell corresponds to $F_X^{-1}(0), F_X^{-1}(1/7), \cdots, F_X^{-1}(1)$.
No noticeable treatment effect; in Jones et al. (2019), p-values for ATE are 0.86-0.94.

Empirical Illustration: treatment effect distribution

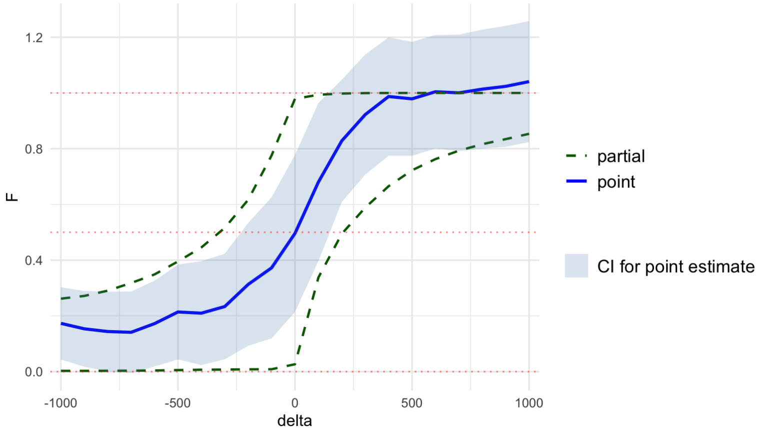

Figure 2: Marginal density of $Y(1) - Y(0)$.

Information gain from partial identification (Fan and Park, 2010).

Summary

- Assume a latent variable U such that

$$Y_i(1) \perp\!\!\!\perp Y_i(0) \mid U_i.$$

  This assumption could be thought of as a ‘latent rank invariance’ condition when $\mathbf{E}[Y_i(d)|U_i = u]$ is monotone increasing in u.

- Use two proxy variables X_i and Z_i to identify the distribution of $Y_i(d)|U_i$.

- Nonnegative matrix factorization estimates distribution of U_i given (D_i, Z_i).

- An asymptotic distribution is derived for the plug-in GMM estimator.

Identification à la Hu and Schennach (2008)

An essential building block in the identification argument: $f_{Y,X|D,Z}$.
Fix y and d and discretize X_i and Z_i:

$$\mathbf{H} = \Big(f_{Y=y,X|D=d,Z}(x|z)\Big)_{x,z} = \Big(f_{X|U}(x|u)\Big)_{x,u} \cdot \text{diag}\Big(f_{Y|U}(y|u)\Big)_{u} \cdot \Big(f_{U|D=d,Z}(u|z)\Big)_{u,z}.$$

H is a $|\text{supp}_X| \times |\text{supp}_Z|$ matrix whose rows correspond to X_i and columns to Z_i.
Likewise, define $\mathbf{H}_X = \Big(f_{X|D=d,Z}(x|z)\Big)_{x,z}$.

Under Assumptions 1-2 and **full rank/completeness** of $\mathbf{H}_X$, A3 A4

$$\mathbf{H} \cdot (\mathbf{H}_X)^{-1} = \Big(f_{X|U}(x|u)\Big)_{x,u} \cdot \text{diag}\Big(\{f_{Y(d)=y|U}(u)\}_u\Big) \cdot \left(\Big(f_{X|U}(x|u)\Big)_{x,u}\right)^{-1}$$

Spectral decomposition identifies $f_{X|U}$.

Spectral Theorem of Hu and Schennach (2008)

Several deviations from Hu and Schennach (2008):

1. Two decomposition results; treated population and untreated population.
   Need to connect $\{f_{Y(1)|U}(\cdot|u)\}_u$ to $\{f_{Y(0)|U}(\cdot|u)\}_u$.

2. Mapping from $\{f_{X|U}(\cdot|u)\}_u$ to u to have distribution of U_i.

1. is easily solved.

Firstly, split the sample into two subsamples $\{i : D_i = 1\}$ and $\{i : D_i = 0\}$ and we get $\left\{f_{Y(1)|U}(\cdot|u), f_{X=1|U}(\cdot|u)\right\}_u$ and $\left\{f_{Y(0)|U}(\cdot|u), f_{X=0|U}(\cdot|u)\right\}_u$.

Under Assumption 1, $f_{X|D=1,U}(\cdot|u)$ and $f_{X|D=0,U}(\cdot|u)$ should be the same.

Spectral Theorem of Hu and Schennach (2008)

A linear operator $L_{Y=y,X|D=d,X}$ maps a density of Z_i to a density of $(Y_i(d)=y, X_i)$:

$$\left(L_{Y=y,X|D=d,Z}\,g\right)(x) = \int_{\mathbb{R}} f_{Y(d),X|D,Z}(y,x|d,z)g(z)dz.$$

From the decomposition based on Assumption 2, we get

$$L_{Y=y,X|D=d,Z} = L_{X|U}\cdot\Delta_{Y=y|U}\cdot L_{U|D=d,Z}$$

with similarly defined operators $L_{X|U}$, $L_{U|D=d,Z}$ and a diagonal operator $\Delta_{Y=y|U}$. Thus,

$$\begin{aligned} L_{Y=y,X|D=d,Z}\left(L_{X|D=d,Z}\right)^{-1} &= L_{X|U}\cdot\Delta_{Y=y|U}\cdot L_{U|D=d,Z}\cdot\left(L_{X|U}\cdot L_{U|D=d,Z}\right)^{-1} \\ &= \underbrace{L_{X|U}\cdot\Delta_{Y=y|U}\cdot\left(L_{X|U}\right)^{-1}}_{\text{spectral decomposition}}. \end{aligned}$$

back

Assumption 3

Assumption 3.

a. *(finitely discrete U_i)* $U_i \in \{u^1, \cdots, u^K\}$.

b. *(full rank)* $\left(f_{U|D=1,Z}(u|z)\right)_{u,z}$, $\left(f_{U|D=0,Z}(u|z)\right)_{u,z}$ and $\left(f_{X|U}(x|u)\right)_{x,z}$ have rank K.

c. *(no repeated eigenvalue)* For any $k \neq k'$, there exist some $d \in \{0, 1\}$ and y such that

$$f_{Y(d)|U}(y|u^k) \neq f_{Y(d)|U}(y|u^{k'}).$$

"The latent heterogeneity U_i can be *at most* as rich/flexible as the proxy variables." back

Assumption 4

Assumption 4.

a. *(continuous U_i)* $U_i \in [0, 1]$.

b. *(bounded density)* All marginal and conditional densities of $\left(Y_i(1), Y_i(0), X_i, Z_i, U_i\right)$ are bounded.

c. *(completeness)* Let $f_{X|Z,d}$ denote the conditional density of X_i given $(D_i = d, Z_i)$.

$$\int_{\mathbb{R}} |g(x)|dx \quad \text{and} \quad \int_{\mathbb{R}} g(x) f_{X|Z,d}(x|z)d(x) = 0 \quad \forall d, z$$

implies $g(x) = 0$. Assume similarly for $f_{X|U}$.

d. *(no repeated eigenvalue)* $\forall u \neq u'$, there exists $d \in \{0, 1\}$ such that

$$\Pr\left\{f_{Y(d)|U}(Y_i(d)|u) \neq f_{Y(d)|U}(Y_i(d)|u')|D_i = d\right\} > 0.$$

back

Identification: implicit restriction

A crucial step in the identification argument is that there exists some w such that

$$\mathbf{E}[Y_i(1)|Y_i(0)=y] = \int_{\mathbb{R}} \frac{w(y,z)}{f_{Y(0)}(y)} \cdot \mathbf{E}[Y_i|D_i=1,Z_i=z]dz,$$

$$\mathbf{E}[Y_i(1)Y_i(0)] = \int_{\mathbb{R}} \int_{\mathbb{R}} w(y,z) \cdot y\mathbf{E}[Y_i|D_i=1,Z_i=z]dydz.$$

$\mathbf{E}[Y_i|D_i=1,Z_i]$ replaces $Y_i(1)$ and $w(y,z)$ replaces the joint density of $\big(Y_i(1),Y_i(0)\big)$.
"Proxy variable Z_i creates sufficient variation in the distribution of $Y_i(1)$."

The implicit restriction is that
"conditional distribution of $Y_i(1)$ given $Y_i(0)$ is a linear combination of $\{F_{Y|D=1,Z}(\cdot|z)\}_z$."

Identification: falsification test

The conditional independence of $Y_i(1)$ and $Y_i(0)$ is fundamentally untestable.

Instead, we can test the null

$$f_{X|D=1,U}(\cdot|u) = f_{X|D=0,U}(\cdot|u) \qquad \forall u.$$

with

$$\min_{g \text{ monotone}} \mathbf{E}\left[\int_{\mathbb{R}} \left(f_{X|D=1,U}(x|g(U_i)) - f_{X|D=0,U}(x|U_i)\right)^2 du \middle| D_i = 0\right] = 0.$$

"Can we construct a latent variable U_i that satisfies 1) conditional independence $X_i \perp\!\!\!\perp Y_i \mid U_i$ and 2) random treatment $X_i \perp\!\!\!\perp D_i \mid U_i$?"

In the short panel context,

- cannot test the conditional independence *across treatment regime*.
- can somewhat test the *intertemporal* conditional independence, given random treatment.

Nonnegative matrix factorization

The objective function in (1) is quadratic with linear constraints, once we fix two out of the three matrices $\Gamma_X, \Gamma_Y, \Lambda$.

Thus, find the (local) minima by iterating among three objects:

1. Given $\left(\Gamma_0^{(s)}, \Gamma_1^{(s)}\right)$, update (Λ_0, Λ_1).
2. Given $\left(\Lambda_0^{(s+1)}, \Lambda_1^{(s+1)}, {\Gamma_{Y0}}^{(s)}, {\Gamma_{Y1}}^{(s)}\right)$, update Γ_X.
3. Given $\left(\Lambda_0^{(s+1)}, \Lambda_1^{(s+1)}, {\Gamma_X}^{(s+1)}\right)$, update $(\Gamma_{Y0}, \Gamma_{Y1})$.
4. Iterate **1**-**3** until convergence.

In practice, use may initial values to find the global minimum. back

Sieve MLE

To allow for a continuous U_i, we can directly construct a likelihood using sieves:

$$f_{Y,X|D=d,Z,n}(y,x|z;\theta) = \int_{\mathbb{R}} f_{Y(d)|U,n}(y|u;\theta) \cdot f_{X|U,n}(x|u;\theta) \cdot f_{U|D=d,Z,n}(u|z;\theta)du.$$

The latent rank interpretation is simple to impose with Bernstein polynomials:
a Bernstein polynomial of degree m is

$$g_m(u) = \sum_{k=0}^{m} \theta_k u^k (1-u)^{m-k}.$$

Then, monotonicity of $\int_0^1 u g_m(u)du$ is a set of linear constraints on $\{\theta_k\}_{k=0}^m$.

back

Sieve MLE

Theorem 4. Let Assumptions 1-2,4-6 hold. Then,

$$\left\| \hat{f}_{Y(1),Y(0)} - f_{Y(1),Y(0)} \right\|_\infty \xrightarrow{p} 0$$

as $n \to \infty$ and for any $(y, y') \in \mathbb{R}^2$ and $\delta \in \mathbb{R}$,

$$\sqrt{n}\left(\hat{f}_{Y(1),Y(0)}(y, y') - f_{Y(1),Y(0)}(y, y')\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma(y, y')^2\right)$$

$$\sqrt{n}\left(\mathrm{Pr}\left\{Y_i(1\widehat{) - Y_i}(0) \leq \delta\right\} - \mathrm{Pr}\left\{Y_i(1) - Y_i(0) \leq \delta\right\}\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma(\delta)^2\right)$$

as $n \to \infty$.

Assumption 6 I

Assumption 6

a. Functions in $\{\Theta_n\}_{n=1}^{\infty} \cup \Theta$ is uniformly bounded. Θ is convex.

b. $f_{Y(1)|U}, f_{Y(0)|U}, f_{X|U}, f_{U|D=1,Z}, f_{U|D=0,Z}$ are in the interior of $\Lambda_c^{\gamma_1}([0,1]^2)$ with $\gamma_1 > 1$. Also, for any $\theta \in \Theta_n$ for some n,

$$f_{Y(1)|U,n}(\cdot;\theta), f_{Y(0)|U,n}(\cdot;\theta), f_{X|U,n}(\cdot;\theta), f_{U|D=1,Z,n}(\cdot;\theta), f_{U|D=0,Z,n}(\cdot;\theta) \in \Lambda_c^{\gamma_1}([0,1]^2)$$

and $\log f_{Y,X|D,Z}(\cdot;\theta) \in \Lambda_c^{\gamma}([0,1]^4)$ with $\gamma > 2$.

c. $\mathbf{E}\left[\left(\log f_{Y,X|D,Z}(Y_i, X_i|D_i, Z_i)\right)^2\right] < \infty$. There exists measurable functions h_1, h_2 such that

$$\begin{aligned} & h_1(y,d,x,z) \\ & \leq \frac{1}{f_{Y,X|D,Z}(y,x|d,z;\theta)} \left(\int_0^1 \frac{f_{Y(d)|U}(y|u;\theta) f_{X|U}(x|u;\theta) f_{U|D=d,Z}(u|z;\theta)}{f_{Y(d)|U}(y|u;\theta) + f_{X|U}(x|u;\theta) + f_{U|D=d,Z}(u|z;\theta)} du \right) \\ & \leq h_2(y,d,x,z) \end{aligned}$$

Assumption 6 II

for all $\theta \in \Theta$ and $\mathbf{E}\left[(h_1(Y_i, D_i, X_i, Z_i,))^2\right], \mathbf{E}\left[(h_2(Y_i, D_i, X_i, Z_i))^2\right] < \infty$. ALso, There exist a measurable function h_3 such that

$$
\begin{aligned}
&\frac{1}{2f_{Y,X|D,Z}(y,x|d,z;\theta)^2}\left(\int_0^1 \frac{f_{Y(d)|U}(y|u;\theta)f_{X|U}(x|u;\theta)f_{U|D=d,Z}(u|z;\theta)}{f_{Y(d)|U}(y|u;\theta)+f_{X|U}(x|u;\theta)+f_{U|D=d,Z}(u|z;\theta)}du\right)^2 \\
&\quad+\frac{1}{f_{Y,X|D,Z}(y,x|d,z;\theta)}\int_0^1\left(f_{Y(d)|U}(y|u;\theta)+f_{X|U}(x|u;\theta)+f_{U|D=d,Z}(u|z;\theta)\right)du \\
&\leq h_3(y,d,x,z)
\end{aligned}
$$

for all $\theta \in \Theta$ and $\mathbf{E}\left[(h_3(Y_i, D_i, X_i, Z_i,))^2\right] < \infty$.

d. $\|\Pi_n\theta^0 - \theta^0\|_\infty = o\left(n^{-\frac{1}{4}}\right)$ as $n \to \infty$ where

$$\Pi_n\theta^0 = \arg\max_{\theta\in\Theta_n} \mathbf{E}\left[\log f_{Y,X|D,Z}(Y_i, X_i|D_i, Z_i;\theta)\right]$$

Also, $p_n \to \infty$, $\frac{p_n \log n}{\sqrt{n}} \to 0$ as $n \to \infty$.

Assumption 6 III

e. With some $c_1, c_2 > 0$,

$$c_1 \mathbf{E}\left[\log \frac{f_{Y,X|D,Z}(Y_i, X_i|D_i, Z_i; \theta^0)}{f_{Y,X|D,Z}(Y_i, X_i|D_i, Z_i; \theta)}\right] \leq \left\|\theta - \theta^0\right\|^2 \leq c_2 \mathbf{E}\left[\log \frac{f_{Y,X|D,Z}(Y_i, X_i|D_i, Z_i; \theta^0)}{f_{Y,X|D,Z}(Y_i, X_i|D_i, Z_i; \theta)}\right]$$

holds for any $\theta \in \Theta_n$ such that $\|\theta - \theta^0\|_\infty = o(1)$.

f. Let p_1 be the degree of a tensor product Bernstein polynomial used in approximating $f_{Y(1)|U}$ to Θ_n and similarly define p_0, p_X, p_{1Z} and p_{0Z}; for example, $p_1 = (p^y + 1) \cdot (p^u + 1)$. With some abuse of notation, let $\{\theta_{j,1}\}_{j=1}^{p_1}$ denote the basis functions used in approximating $f_{Y(1)|U}$ and similarly define $\{p_{j,0}\}_{j=1}^{p_0}, \cdots, \{p_{j,0Z}\}_{j=1}^{p_{0Z}}$.

Assumption 6 IV

Let

$$\frac{d}{d\theta_1}\log f_{Y,X|D,Z}(Y_i,X_i|D_i,Z_i;\theta^0)\left[\{\theta_{j,1}\}_{j=1}^{p_1}\right]=\begin{pmatrix}\frac{d}{d\theta_1}\log f_{Y,X|D,Z}(Y_i,X_i|D_i,Z_i;\theta^0)\left[\theta_{1,1}\right]\\ \vdots\\ \frac{d}{d\theta_1}\log f_{Y,X|D,Z}(Y_i,X_i|D_i,Z_i;\theta^0)\left[\theta_{p_1,1}\right]\end{pmatrix}$$

$$W_n(Y_i,D_i,X_i,Z_i)=\begin{pmatrix}\frac{d}{d\theta_1}\log f_{Y,X|D,Z}(Y_i,X_i|D_i,Z_i;\theta^0)\left[\{\theta_{j,1}\}_{j=1}^{p_1}\right]\\ \vdots\\ \frac{d}{d\theta_{0Z}}\log f_{Y,X|D,Z}(Y_i,X_i|D_i,Z_i;\theta^0)\left[\{\theta_{j,0Z}\}_{j=1}^{p_{0Z}}\right]\end{pmatrix}$$

and

$$\Omega_n=\mathbf{E}\left[W_n(Y_i,D_i,X_i,Z_i)\left(W_n(Y_i,D_i,X_i,Z_i)\right)^{\intercal}\right].$$

Then, the smallest eigenvalue of Ω_n is bounded away from zero uniformly across n.

Choice of K

Under Assumption 3, the rank of the following $M_X \times 2M_Z$ matrix is K:

$$\mathbf{H}_X = \begin{pmatrix} \Pr\left\{X_i \in \mathcal{X}^1 \middle| D_i = 0, Z_i \in \mathcal{Z}^1\right\} & \cdots & \Pr\left\{X_i \in \mathcal{X}^1 \middle| D_i = 1, Z_i \in \mathcal{Z}^{M_Z}\right\} \\ \vdots & \ddots & \vdots \\ \Pr\left\{X_i \in \mathcal{X}^{M_X} \middle| D_i = 0, Z_i \in \mathcal{Z}^1\right\} & \cdots & \Pr\left\{X_i \in \mathcal{X}^{M_X} \middle| D_i = 1, Z_i \in \mathcal{Z}^{M_Z}\right\} \end{pmatrix}$$

We can apply the Kleibergen-Paap rank test or the eigenvalue ratio test. back

Kleibergen and Paap (2006); Ahn and Horenstein (2013)

Simulation

Monte Carlo simulations with a simple DGP with $K = 3$ and $Y_i, X_i, Z_i \in \{1, 2, 3\}$.
Nonnegative matrix factorization is applied to two 9×3 matrices.

Informativeness of the two proxy variables:

$$\Gamma_X = \left(\Pr\{X_i = x | U_i = u^k\} \right)_{x,k} = \begin{pmatrix} 0.800 & 0.100 & 0.067 \\ 0.133 & 0.800 & 0.133 \\ 0.067 & 0.100 & 0.800 \end{pmatrix}$$

$$\Lambda = \left(\Pr\{U_i = u^k | Z_i = z\} \right)_{z,k} = \begin{pmatrix} 0.840 & 0.091 & 0.040 \\ 0.077 & 0.772 & 0.055 \\ 0.083 & 0.137 & 0.905 \end{pmatrix}.$$

Their smallest singular values are 0.665 and 0.701.

Simulation

As we shift Λ, estimation worsens:

	$\widehat{F}_{Y(1)-Y(0)}$					
	$\sigma_{\min}(\Lambda) = 0.701$		$\sigma_{\min}(\Lambda) = 0.501$		$\sigma_{\min}(\Lambda) = 0.310$	
δ	bias	rMSE	bias	rMSE	bias	rMSE
-2	0.000	0.006	0.001	0.010	0.001	0.025
-1	-0.000	0.017	0.000	0.025	-0.002	0.052
0	-0.007	0.028	-0.012	0.040	-0.014	0.076
1	-0.009	0.025	-0.014	0.040	-0.015	0.084

Table 1: Bias and rMSE of DTE estimator, $B = 200$.

First step NMF worsens as Z_i gets less informative.

Simulation

	$\widehat{F}_{Y(1)-Y(0)}$		
	$\sigma_{\min}(\Lambda) = 0.701$	$\sigma_{\min}(\Lambda) = 0.501$	$\sigma_{\min}(\Lambda) = 0.310$
$\Pr\left\{F_{Y(1)-Y(0)}(-2) \in \widehat{CI}\right\}$	0.968	0.970	0.990
$\Pr\left\{F_{Y(1)-Y(0)}(-1) \in \widehat{CI}\right\}$	0.978	0.960	0.970
$\Pr\left\{F_{Y(1)-Y(0)}(0) \in \widehat{CI}\right\}$	0.960	0.975	0.990
$\Pr\left\{F_{Y(1)-Y(0)}(1) \in \widehat{CI}\right\}$	0.970	0.970	0.980
$\Pr\left\{\text{reject } F_{X\mid D=1,U} = F_{X\mid D=0,U}\right\}$	0.070	0.063	0.049

Table 2: Coverage of CI and type I error of falsification test, $B = 200$.

Choice of K

In the empirical application, both rank test and eigenvalue ratio estimator suggest $K = 3$.

K	1	2	3	4	5	6	7	8
eigenvalue ratio	3.505	3.991	4.029	2.721	1.653	1.863	1.418	3.309
growth ratio	0.964	1.135	1.472	1.353	0.893	0.956	0.580	1.035

Table 3: Eigenvalue ratios and growth ratios

K	1	2	3	4	5	6
test statistic	884.82	116.23	35.75	20.08	13.80	7.94
p-value	0.000	0.001	0.984	0.998	0.995	0.992

Table 4: Kleibergen-Paap rank test statistics for H_0 : rank $= K$ and their p-values

back

Additional moments

The quadratic moment is

$$\sum_{l=1}^{K} \frac{\tilde{\lambda}_{lk,d}}{p_{D,Z}(d,l)} \cdot \mathbf{E}\left[\frac{1}{2}\mathbf{1}\{Y_i \in \mathcal{Y}, D_i = d, X_i \in \mathcal{X}, Z_i \in \mathcal{Z}^l\}\right]$$
$$+ \sum_{m=1}^{K} \frac{\tilde{\lambda}_{mk,d}}{p_{D,Z}(d,m)} \cdot \mathbf{E}\left[\frac{1}{2}\mathbf{1}\{Y_j \in \mathcal{Y}, D_j = d, X_j \in \mathcal{X}, Z_i \in \mathcal{Z}^m\}\right]$$
$$- \sum_{l=1}^{K} \sum_{m=1}^{K} \frac{\tilde{\lambda}_{lk,d}\tilde{\lambda}_{mk,d}}{p_{D,Z}(d,l) \cdot p_{D,Z}(d,m)} \mathbf{E}[\mathbf{1}\{Y_i \in \mathcal{Y}, D_i = d, Z_i \in \mathcal{Z}^l, X_j \in \mathcal{X}, D_j = d, Z_j \in \mathcal{Z}^m\}] = 0$$

with $(Y_i, D_i, Z_i) \perp\!\!\!\perp (Y_j, D_j, Z_j)$. back

References I

Ahn, Seung C and Alex R Horenstein, "Eigenvalue ratio test for the number of factors," *Econometrica*, 2013, *81* (3), 1203–1227.

Ai, Chunrong and Xiaohong Chen, "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 2003, *71* (6), 1795–1843.

Athey, Susan and Guido W Imbens, "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, 2006, *74* (2), 431–497.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "Discretizing unobserved heterogeneity," *Econometrica*, 2022, *90* (2), 625–643.

Callaway, Brantly and Tong Li, "Quantile treatment effects in difference in differences models with panel data," *Quantitative Economics*, 2019, *10* (4), 1579–1618.

Chen, Xiaohong and Xiaotong Shen, "Sieve extremum estimates for weakly dependent data," *Econometrica*, 1998, pp. 289–314.

Chernozhukov, Victor and Christian Hansen, "An IV model of quantile treatment effects," *Econometrica*, 2005, *73* (1), 245–261.

Chernozhukov, Victor and Christian Hansen, "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 2006, *132* (2), 491–525.

Deaner, Ben, "Proxy controls and panel data," 2023.

Fan, Yanqin and Sang Soo Park, "Sharp bounds on the distribution of treatment effects and their statistical inference," *Econometric Theory*, 2010, *26* (3), 931–951.

References II

Fan, Yanqin, Robert Sherman, and Matthew Shum, "Identifying treatment effects under data combination," *Econometrica*, 2014, *82* (2), 811–822.

Firpo, Sergio and Geert Ridder, "Partial identification of the treatment effect distribution and its functionals," *Journal of Econometrics*, 2019, *213* (1), 210–234.

Frandsen, Brigham R and Lars J Lefgren, "Partial identification of the distribution of treatment effects with an application to the Knowledge is Power Program (KIPP)," *Quantitative Economics*, 2021, *12* (1), 143–171.

Heckman, James J, Jeffrey Smith, and Nancy Clements, "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *The Review of Economic Studies*, 1997, *64* (4), 487–535.

Henry, Marc, Yuichi Kitamura, and Bernard Salanié, "Partial identification of finite mixtures in econometric models," *Quantitative Economics*, 2014, *5* (1), 123–144.

Hu, Yingyao and Matthew Shum, "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 2012, *171* (1), 32–44.

Hu, Yingyao and Susanne M Schennach, "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 2008, *76* (1), 195–216.

Jones, Damon, David Molitor, and Julian Reif, "What do workplace wellness programs do? Evidence from the Illinois workplace wellness study," *The Quarterly Journal of Economics*, 2019, *134* (4), 1747–1791.

Kaji, Tetsuya and Jianfei Cao, "Assessing Heterogeneity of Treatment Effects," 2023.

References III

Kasahara, Hiroyuki and Katsumi Shimotsu, "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 2009, *77* (1), 135–175.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of econometrics*, 2006, *133* (1), 97–126.

Miao, Wang, Zhi Geng, and Eric J Tchetgen Tchetgen, "Identifying causal effects with proxy variables of an unmeasured confounder," *Biometrika*, 2018, *105* (4), 987–993.

Nagasawa, Kenichi, "Treatment effect estimation with noisy conditioning variables," *arXiv preprint arXiv:1811.00667*, 2022.

Noh, Sungho, "Nonparametric identification and estimation of heterogeneous causal effects under conditional independence," *Econometric Reviews*, 2023, *42* (3), 307–341.

Shen, Xiaotong, "On methods of sieves and penalization," *The Annals of Statistics*, 1997, *25* (6), 2555–2591.