Semi/Nonparametric IV estimation with first stage isotonic regression

Taisuke Otsu (LSE) Kazuhiko Shinoda (Nagoya) and Mengshan Xu (Mannheim)

ESEM 2024
Rotterdam
28-August-2024

Content

Content

Introduction: Background of SPIV and NPIV

- Terminologies
  - SPIV: Semiparametric instrumental variable estimator
  - NPIV: Nonparametric instrumental variable estimator
  - Y: dependent variable; X: endogenous variable; W: instrumental variable

---

- Recall: Linear IV model

Second stage $Y = \beta_1 + X\beta_2 + U, \quad \mathbb{E}[U|W] = 0,$

First stage $X = \gamma_1 + W\gamma_2 + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0.$

Estimated by 2SLS

- SPIV

$$
\begin{aligned}
Y &= \beta_1 + X\beta_2 + U, \quad \mathbb{E}[U|W] = 0, \\
X &= \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0.
\end{aligned}
$$

Estimate the first stage by nonparametric methods

Then plug in the fitted IV, $\hat{X} = \hat{\zeta}(\mathrm{X})$

- NPIV

$$
\begin{aligned}
Y &= g(X) + U, \quad \mathbb{E}[U|W] = 0, \\
X &= \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0,
\end{aligned}
$$

Estimate BOTH stages by nonparametric methods

Introduction: NPIV

X: endogenous variable; W: instrumental variable

- NPIV

Second stage $\quad Y = g(X) + U, \quad \mathbb{E}[U|W] = 0,$

First stage $\quad X = \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0,$

- X is endogenous $\Rightarrow \mathbb{E}(U|X) \neq 0$

$$\Rightarrow \mathbb{E}(Y - g(X)|X) \neq 0 \quad \Rightarrow \mathbb{E}(Y|X) \neq g(X)$$

$g(\cdot)$ cannot be estimated by regressing Y on X

- W is exogenous $\Rightarrow \mathbb{E}(U|W) = 0$

$$\Rightarrow \mathbb{E}(Y - g(X)|W) = 0 \quad \Rightarrow \mathbb{E}(Y|W) = \mathbb{E}(g(X)|W) = \sum_{i=1}^{\infty} \beta_i \, \mathbb{E}(p_i(X)|W)$$

$g(\cdot)$ can be identified, if $f_{X|W}(\cdot,\cdot)$ is complete

- In estimation, we focus on the "elements" of $g(x) = \sum_{i=1}^{\infty} \beta_i \, p_i(x)$
  - $\{p_i(\cdot)\}_{i=1}^{\infty}$ are known series functions

Introduction: NPIV

$\mathbb{E}(Y|W) = \mathbb{E}(g(X)|W)$

- The "elements" of a function: series expansion
- The historical result
  - Stone-Weierstrass Theorem (Weierstrass, 1885, Stone, 1948)
  - For a continuous $g(\cdot)$
  - With polynomial basis $p(x) = \left(1, x, \ldots, x^{K-1}\right)'$
  - For any (small) $\varepsilon > 0$, there is some sufficiently large K such that

$$\inf_{\beta} \sup_{x \in \mathcal{X}} |g(x) - p(x)'\beta| \leq \varepsilon$$

- With L_2 norm
  - For $g(\cdot) \in L_2(\mathcal{X})$ with $\sqrt{\int_{\mathcal{X}} h(x)^2 dx} < \infty$
  - With a complete basis in $L_2(\mathcal{X})$ $p(x) = (p_1(x), p_2(x), \ldots)'$
  - $g(x) = \sum_{i=1}^{\infty} \beta_i \, p_i(x)$
  - $p_i(x)$ is known basis function, and $\beta_i' s$ are unknown series coefficients

Introduction: NPIV

$\mathbb{E}(Y|W) = \mathbb{E}(g(X)|W)$

X: endogenous variable; W: instrumental variable

- $g(x) = \sum_{i=1}^{\infty} \beta_i \, p_i(x)$
- $p_i(x)$ is known basis function, and $\beta_i' s$ are unknown series coefficient

---

- NPIV estimation (NP 2SLS)
- Choice the tuning parameter for the second stage, the series length K_n
- First stage:
  - Regress $p_i(X)$ on W nonparametrically for each $i \in \{1\!:K_n\}$, obtaining $\widehat{\mathbb{E}}(p_i(X)|W)$
    - Nonparametric estimation requires additional tuning parameter(s)
  - $\mathbb{E}(Y|W) = \mathbb{E}(g(X)|W) \approx \sum_{i=1}^{K_n} \beta_i \, \widehat{\mathbb{E}}(p_i(X)|W)$
- Second stage:
  - Regress Y on $\widehat{\mathbb{E}}(p_i(X)|W)$ to obtain estimators of $\{\beta_i\}_i^{K_n}$, say $\{\hat{\beta}_i\}_i^{K_n}$
  - NPIV estimator: $\hat{g}(x) = \sum_{i=1}^{K_n} \hat{\beta}_i \, p_i(x)$

Introduction: NPIV

- First stage:
  - Regress $p_i(X)$ on W nonparametrically for each $i \in \{1: K_n\}$, obtaining $\widehat{\mathbb{E}}(p_i(X)|W)$
    - Nonparametric estimation implies additional tuning parameter(s)

- In the literature, the first stage is usually conducted by series estimation as well
  - Newey and Powell (2003), Blundell, Chen, and Christensen (2007), Horowitz (2011, 2012), among others.

- In this paper, we propose using Isotonic regression to estimate the first stage
  - for both SPIV and NPIV estimation
  - We call them the *monotone SPIV* and *monotone NPIV*

Content

Introduction: Isotonic regression

- Monotone increasing conditional mean function

$$Y = r(X) + \epsilon \qquad \text{where} \ \ x_1 < x_2 \Rightarrow r(x_1) \leq r(x_2)$$

- Isotonic estimation

  A nonparametric method to estimate this $r(\cdot)$

  under the assumption of monotonicity

Monotonicity assumption

- Natural conditions of many models applied widely in economics
  - Demand function, utility functions..
  - CDF function, and functions derived from CDF functions

Monotone instrumental variable

$$\begin{aligned} \text{Second stage} \quad Y &= g(X) + U, \quad \mathbb{E}[U|W] = 0, \\ \text{First stage} \quad X &= \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0, \end{aligned}$$

X: endogenous variable; W: instrumental variable

- Our approach requires that $\mathbb{E}(X|W = w)$ increases in w
- Many examples in economics
  - Härdle and Linton (1994), Blundell, Chen, and Kristensen (2007): $X =$ log family expenditure, $W =$ log gross earnings
  - Measurement error problem: $X =$ a measurement (with error) of X^*, $W =$ an independent repeated measurement of X^*
- Furthermore, any existing study involving a univariate endogenous variable and univariate instrumental variables, and utilizing the linear 2SLS methods, implicitly imposes the monotonicity of $\mathbb{E}(X|W = w)$

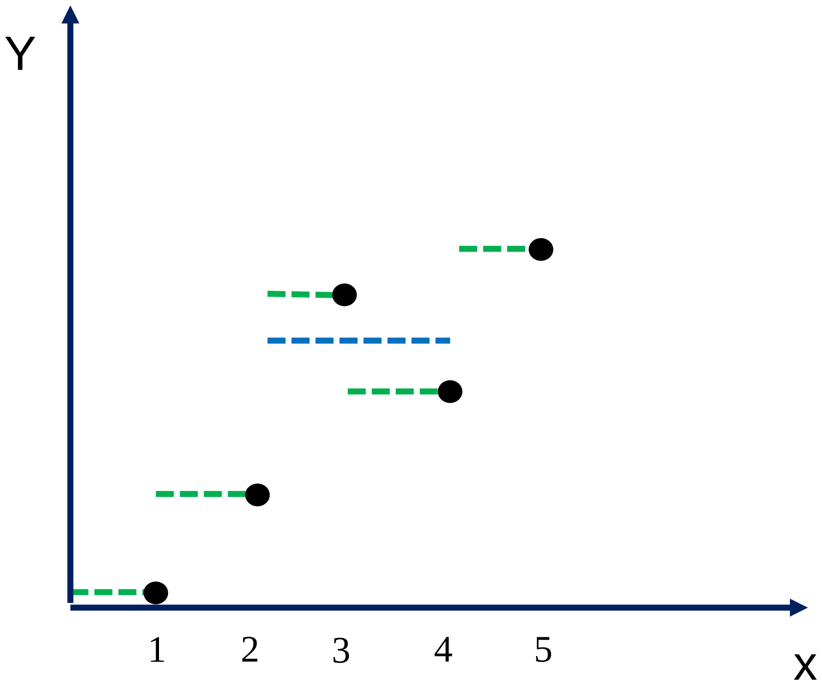
Implement isotonic estimation

$Y_i = r(x_i) + \epsilon_i,$ where $x_1 < x_2 < \cdots < x_n$, and $\epsilon_i s$ are independent

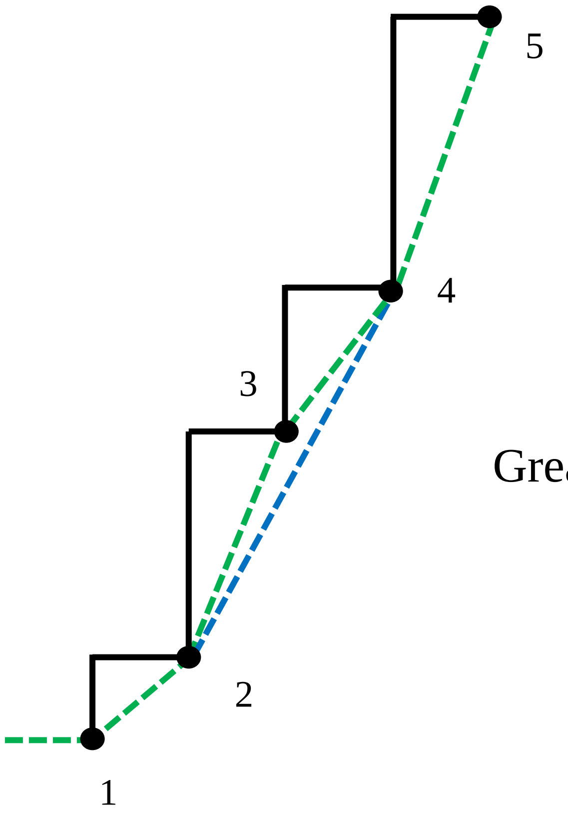
$$(\hat{r}(x_1), \ldots, \hat{r}(x_n)) = \operatorname{argmin} \sum_{i=1}^{n} (Y_i - r(x_i))^2$$

The algorithm: PAVA, the greatest convex minorant

Implement isotonic regression



The cumulative sum diagram



Greatest convex minorant

$$(\hat{r}(x_1), \ldots, \hat{r}(x_n)) = \underset{r(x_1) \leq r(x_2) \leq \cdots \leq r(x_n)}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{n} (Y_i - r(x_i))^2$$

No tuning parameters are involved

Content

Isotonic regression in a two-stage process

- As a nonparametric method, isotonic regression has
  - Strengths:
    - Tuning parameter free
    - Require minimum smoothness condition
  - Shortcomings:
    - a discrete estimator
    - slow convergence rate
- Using isotonic regression as the first stage in a two-stage procedure sometimes could allow us to
  - circumvent its shortcomings
  - while preserving its strengths

Motivation for monotone SPIV

- Compared to 2SLS, our proposed SPIV method is more efficient

$$\begin{aligned} \text{Second stage} \quad Y &= \beta_1 + X\beta_2 + U, \quad \mathbb{E}[U|W] = 0, \\ \text{First stage} \quad X &= \gamma_1 + W\gamma_2 + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0. \end{aligned}$$

  - 2SLS can give a consistent estimator of $\beta = (\beta_1, \beta_2)$,
  - even if $\mathbb{E}(X|W)$ is not linear in the first stage
  - but it will be less efficient
  - Under the monotonicity of $\mathbb{E}(X|W = w)$ in w, our monotone SPIV estimator is efficient
    - As efficient as ones in Ai and Chen (2003) and Blundell, Chen and Kristensen (2007)

- Compared to other SPIV estimators
  - Our method is tuning-parameter-free

Motivation for monotone NPIV

Second stage $Y = g(X) + U, \quad \mathbb{E}[U|W] = 0,$
First stage $X = \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0,$

- Our method requires choosing only one tunning parameter (in the case of one endogenous variable)
- "Series+series" (NP2SLS) methods (implicitly) require more
- Recall the first stage of NPIV estimation
  - After the tuning parameter for the second stage, the series length K_n, has been chosen
  - Regress $p_i(X)$ on W nonparametrically for **each** $i \in \{1{:}\, K_n\}$, obtaining $\widehat{\mathbb{E}}(p_i(X)|W)$
- Let $l(w) = (l_1(w), l_2(w), \dots)'$ be a complete basis in $L_2(\mathcal{W})$
- The first-stage series estimation is

$$p_1(X) = \sum_{i=1}^{J_n^{(1)}} \gamma_i^{(1)} l_i(W) + e_1$$

$$p_2(X) = \sum_{i=1}^{J_n^{(2)}} \gamma_i^{(2)} l_i(W) + e_2$$

$$\dots$$

$$p_{K_n}(X) = \sum_{i=1}^{J_n^{(K_n)}} \gamma_i^{(K_n)} l_i(W) + e_{K_n}$$

There are potentially $K_n + 1$ tuning parameters

Motivation for monotone NPIV

Second stage	Y	$=$	$g(X) + U,$	$\mathbb{E}[U\|W] = 0,$
First stage	X	$=$	$\zeta(W) + \epsilon,$	$\mathbb{E}[\epsilon\|W] = 0,$

- Let $l(w) = (l_1(w), l_2(w), \dots)'$ be a complete basis in $L_2(\mathcal{W})$
- The first stage-series estimation is

$$p_1(X) = \sum_{i=1}^{J_n^{(1)}} \gamma_i^{(1)} l_i(W) + e_1$$

$$p_2(X) = \sum_{i=1}^{J_n^{(2)}} \gamma_i^{(2)} l_i(W) + e_2$$

$$\cdots$$

$$p_{K_n}(X) = \sum_{i=1}^{J_n^{(K_n)}} \gamma_i^{(K_n)} l_i(W) + e_{K_n}$$

They are potentially $K_n + 1$ tuning parameters

- In the literature of NPIV, people usually enforce $K_n = J_n^{(1)} = J_n^{(2)} = \cdots = J_n^{(K_n)}$
  - But this practice is rarely justified theoretically
  - It is an implicit choice of $K_n + 1$ tuning parameters
- In comparison, our proposed method only needs to choose one, K_n
  - Our first stage is adaptively handled by isotonic regression
  - Our first stage is truly tuning-parameter-free

Content

Theory for monotone SPIV

$$\begin{aligned} \text{Second stage} \quad Y &= \beta_1 + X\beta_2 + U, \quad \mathbb{E}[U|W] = 0, \\ \text{First stage} \quad X &= \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0. \end{aligned}$$

Assumption 2.1.1. *[Sampling] (i)* $\{Y_i, W_i, X_i\}_{i=1}^n$ *is an iid sample of* $(Y, X, W) \in \mathbb{R} \times \mathcal{X} \times \mathcal{W}$*, where* $\mathcal{X} \subseteq \mathbb{R}$*, and* $\mathcal{W}$ *is a compact subset of* $\mathbb{R}$*; (ii)* X *and* W *are jointly continuously distributed.*

Assumption 2.1.2. *[Monotonicity]* $\zeta(w) = \mathbb{E}[X|W = w]$ *is a monotone increasing function of* $w \in \mathcal{W}$*.*

Assumption 2.1.3. *[Instrument relevance and homoscedasticity] (i)* $\mathrm{Var}(\mathbb{E}[X|W]) \neq 0$*; (ii)* $\mathbb{E}[U^2|W = w] = \sigma_U^2$ *for each* $w \in \mathcal{W}$*.*

Assumption 2.1.4. *[Moment conditions] there exist positive constants* c_0 *and* M_0 *such that* $\mathbb{E}[|U|^m|W = w] \leq m!\, M_0^{m-2} c_0$ *and* $\mathbb{E}[|X|^m|W = w] \leq m!\, M_0^{m-2} c_0$ *hold for all integers* $m \geq 2$ *and every* $w \in \mathcal{W}$*.*

Notations:

$\beta := (\beta_1, \beta_2)'$ $\quad Z = \zeta(W)$ $\quad \hat{Z} = \hat{\zeta}(W)$ is the isotonic estimator of $\zeta(W)$

$v(x) = (1, x)', v(z) = (1, z)'$

Theory for monotone SPIV

Second stage $Y = \beta_1 + X\beta_2 + U, \quad \mathbb{E}[U|W] = 0,$
First stage $X = \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0.$

Notations:

$\beta := (\beta_1, \beta_2)' \quad Z = \zeta(W) \quad \hat{Z} = \hat{\zeta}(W)$ is the isotonic estimator of $\zeta(W)$

$v(x) = (1, x)', v(z) = (1, z)'$

The monotone SPIV: $\hat{\beta} = \mathbb{E}_n[v(\hat{Z})v(X)']^{-1}\mathbb{E}_n[v(\hat{Z})Y]$ $\quad \mathbb{E}_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n}(\cdot)$

Theorem 2.1. *Under Assumptions 2.1.1 to 2.1.4, it holds*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Omega),$$

where $\Omega = \sigma_U^2 \mathbb{E}[v(Z)v(Z)']^{-1}$ *attains the semiparametric efficiency bound of estimating* β (Ai and Chen, 2003)

Content

Theory for monotone NPIV

Second stage $Y = g(X) + U, \quad \mathbb{E}[U|W] = 0,$

First stage $X = \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0,$

- Notations and definitions:

$p(x) = (p_1(x), \ldots, p_{K_n}(x))'$ a vector of complete basis functions for $L_2(\mathcal{X})$

$q_k(w) = \mathbb{E}(p_k(X)|W = w),$ $\quad \hat{q}_k(\cdot) = \arg\min_{\zeta \in \mathcal{M}} \mathbb{E}_n[\{p_k(X) - \zeta(W)\}^2],$

$q(w) = (q_1(w), \ldots, q_{K_n}(w))'.$ $\quad \hat{q}(w) = (\hat{q}_1(w), \ldots, \hat{q}_{K_n}(w))'.$

$\Lambda_c^r(\mathcal{X})$ denotes the Hölder space defined in Blundell, Chen and Kristensen (2007)

The sieve space: for a vector of unknown sieve coefficient $b = (b_1, \ldots, b_{K_n})$

$$\mathcal{H}_n = \left\{ h : \mathcal{X} \to [0, K], h \in \Lambda_c^r(\mathcal{X}), h(X) = \sum_{k=1}^{K_n} b_k p_k(X) \right\}$$

T is the conditional mean operator: $(T\nu)(w) = \int_{\mathcal{X}} \nu(x) f_{X|W=w}(x, w) dx.$

Recall $\mathbb{E}(Y|W) = \mathbb{E}(g(X)|W)$

The sieve measure of ill-posedness: $\rho_n = \sup_{h \in \mathcal{H}_n} \frac{\|h\|}{\|Th\|}$

Theory for monotone NPIV

Assumption 2.2.1. *[Data generating] (i)* $\{Y_i, W_i, X_i\}_{i=1}^n$ *is an iid sample of* $(Y, X, W) \in \mathbb{R} \times \mathcal{X} \times \mathcal{W}$, *where* $\mathcal{X} \subseteq \mathbb{R}$, *and* $\mathcal{W}$ *is a compact subset of* $\mathbb{R}$; *(ii)* T *is nonsingular, and the equation* $Tv = m$ *has a unique solution* $v = g$ *almost surely; (iii) for the true function of interest* g, *it holds that* $g \in \mathcal{H} \equiv \Lambda_c^r(\mathcal{X})$ *for some* $r \geq 2$ *and* $\sup_{x \in \mathcal{X}} g(x) \leq K$, *where* K *is the same positive constant in* (2.7).

Assumption 2.2.2. *[Monotonicity and continuity] For each* $k \in \{1 : K_n\}$: *(i)* $q_k(w) = \mathbb{E}[p_k(X)| W = w]$ *is a monotone increasing function of* w; *(ii)* (X, W) *has a Lebesgue density function* f_{XW}, *and the marginal density of* W, $f_W(\cdot)$, *satisfies that for some positive constants* $\overline{f}$ *and* $\underline{f}$, *it holds* $\underline{f} < f_W(\cdot) < \overline{f}$ *all* $w \in \mathcal{W}$; *(iii) there exist* $b > 0$ *and* $M > 0$ *such that* $\mathbb{E}[|X|^m|W = w] \leq m!\, M^{m-2} b$ *for all integers* $m \geq 2$ *and almost every* w.

Assumption 2.2.3. *[Instrument relevance and series order] (i) For each* K_n, *the largest eigenvalues of both* $\mathbb{E}[p(X)p(X)']$ *and* $\mathbb{E}[q(W)q(W)']$ *are bounded, and their smallest eigenvalues are bounded away from zero; (ii)* $K_n \to \infty$ *and* $K_n^3/n \to 0$.

Assumption 2.2.4. *[Bounds of projection errors]* $\rho_n \cdot \|T(g_{w,n} - g_{x,n})\| \leq \text{const} \cdot \|g - g_{x,n}\|$.

Theory for monotone NPIV

Second stage $Y = g(X) + U, \quad \mathbb{E}[U|W] = 0,$
First stage $X = \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0,$

K_n is the series order of the second stage

$p(x) = (p_1(x), \ldots, p_{K_n}(x))'$ a vector of complete basis functions for $L_2(\mathcal{X})$

$q_k(w) = \mathbb{E}(p_k(X)|W = w)$, $\quad \hat{q}_k(\cdot) = \arg\min_{\zeta \in \mathcal{M}} \mathbb{E}_n[\{p_k(X) - \zeta(W)\}^2]$,

$q(w) = (q_1(w), \ldots, q_{K_n}(w))'$. $\quad \hat{q}(w) = (\hat{q}_1(w), \ldots, \hat{q}_{K_n}(w))'$.

The sieve measure of ill-posedness:

$$\rho_n = \sup_{h \in \mathcal{H}_n} \frac{\|h\|}{\|Th\|}$$

The monotone NPIV: $\hat{g}(\cdot) = p(\cdot)' \mathbb{E}_n\left[\hat{q}(W) p(X)'\right]^{-1} \mathbb{E}_n\left[\hat{q}(W) Y\right]$ $\qquad \mathbb{E}_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n}(\cdot)$

Theorem 2.2. *Suppose Assumptions 2.2.1 to 2.2.4 hold, then*

$$\|\hat{g} - g\|_2 = O_p\left(K_n^{-r} + \rho_n \sqrt{\frac{K_n}{n}}\right).$$

The same rate as in Blundell, Chen, Christensen (2007) and Horowitz (2011, 2012)

Content

Simulation results

X: endogenous variable; W: instrumental variable

An NPIV model:

Second stage $\quad Y = X^2 + \epsilon,$

First stage $\quad X = \exp(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0, \quad W \sim \mathrm{U}[-1.2, 1.3],$

$$\epsilon \sim N(0,1).$$

We employ the polynomial basis:

$$p(x) = (1, x, \ldots, x^{K_n - 1})'$$

We try $K_n = 2, 3, 4,$ and 5, while the optimal order (for the 2nd stage) should be $K_n = 3$

(No misspecification, no (unnecessarily) redundant regressors)

The number of Monte-Carlo samples in each setting is 500.

n	Methods	K_n	ISE mean	ISE median	K_n	ISE mean	ISE median
1000	series+series	2	6.3849	6.3423	3	0.0186	0.0099
	isotonic+series	2	6.7102	6.6843	3	0.0129	0.0093
5000	series+series	2	6.3703	6.3516	3	0.0037	0.0022
	isotonic+series	2	6.6582	6.6374	3	0.0029	0.0020
10000	series+series	2	6.3851	6.3834	3	0.0019	0.0011
	isotonic+series	2	6.6671	6.6566	3	0.0015	0.0010

n	Methods	K_n	ISE mean	ISE median	K_n	ISE mean	ISE median
1000	series+series	4	76.9304	0.0840	5	6366.3067	0.6196
	isotonic+series	4	0.0544	0.0442	5	0.1286	0.0894
5000	series+series	4	0.0432	0.0179	5	402.7918	0.6034
	isotonic+series	4	0.0202	0.0140	5	0.3070	0.2305
10000	series+series	4	0.0229	0.0136	5	8.0989	1.1050
	isotonic+series	4	0.0151	0.0121	5	1.0894	0.6934

Content

Literature

- NPIV model
  - Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Chen and Reiss (2011), Horowitz (2011, 2012), Freyberger and Horowitz (2015), Freyberger (2017), Chetverikov and Wilhelm (2017), Chen and Christensen (2018), among others
- Inverse problem in statistics
  - Textbook: Kress (2014)
- Plug in nonparametric into nonparametric
  - Rilstone (1996), Song (2008), Sperlich (2009), Mammen, Rothe, and Schienle (2012), Hahn and Ridder (2013), among others
- Isotonic regression
  - Ayer et al. (1955), Grenander (1956), Rao (1969,1970), Barlow and Brunk (1972), Wright (1981), among others;
  - Textbook: Groeneboom and Jongbloed (2014)
  - Durot, Kulikov and Lopuhaä (2013), Groeneboom and Hendrickx (2018), Balabdaoui, Groeneboom and Hendrickx (2019), Balabdaoui, Durot and Jankowski (2019), Fang, Guntuboyina, and Sen (2021), Deng, Han, and Zhang (2021), among others

Concluding Remarks

- We develop a SPIV and an NPIV estimators, in which the first-stage conditional means are assumed to be monotone increasing and are estimated using isotonic regression
- The proposed SPIV estimator is tuning-parameter-free, and it is efficient in terms of Ai and Chen (2003)
- The proposed NPIV estimator requires choosing only one tuning parameter for the second stage. It achieves the same rate as Blundell, Chen, Christensen (2007)
- For multiple IV, we model the first stage by
  - the monotone partially linear model $X = W_1'\delta + \zeta(W_2) + \epsilon$, or
  - the monotone single index model $X = \zeta(\delta' W) + \epsilon$.

Thank you!

Additional results

- Justification of assumption on the monotone basis function

Theory for monotone NPIV

Second stage $Y = g(X) + U, \quad \mathbb{E}[U|W] = 0,$

First stage $X = \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0,$

- Justification of assumption on the monotone basis function

Assumption 2.2.2. *[Monotonicity and continuity] For each* $k \in \{1 : K_n\}$*: (i)* $q_k(w) = \mathbb{E}[p_k(X)|\, W = w]$ *is a monotone increasing function of* w*;*

This assumption can be satisfied by polynomial basis if we assume additionally

A1: For all $w \in \mathcal{W}$, the conditional density of $\epsilon|W = w$ is symmetrically distributed around 0.

A2: For all $w \in \mathcal{W}$, $\zeta(w)$ is non-negative and monotone increasing in w.

A3: For all $i \in \{1 : K_n\}$, $\sigma^i(w) := \mathbb{E}(\epsilon^i|W = w)$ is finite and non-decreasing in w.

Theory for monotone NPIV

Second stage $Y = g(X) + U, \quad \mathbb{E}[U|W] = 0,$

First stage $X = \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0,$

A1: For all $w \in \mathcal{W}$, the conditional density of $\epsilon|W = w$ is symmetrically distributed around 0.

A2: For all $w \in \mathcal{W}$, $\zeta(w)$ is non-negative and monotone increasing in w.

A3: For all $i \in \{1 : K_n\}$, $\sigma^i(w) := \mathbb{E}(\epsilon^i|W = w)$ is finite and non-decreasing in w.

For the polynomial bases, $p_k(X) = X^k$, we have

$$
\begin{aligned}
\mathbb{E}[p_k(X)|W = w] &= \mathbb{E}[X^k|W = w] \\
&= \mathbb{E}[(\zeta(W) + \epsilon)^k\,|W = w] = \mathbb{E}\left(\sum_{i=0}^{k} \binom{k}{i} \zeta(W)^{k-i}\epsilon^i|W = w\right) \\
&= \mathbb{E}\left(\sum_{0 \le i \le k, i \text{ is odd}} \binom{k}{i} \zeta(W)^{k-i}\epsilon^i|W = w\right) \quad = 0 \\
&\quad + \mathbb{E}\left(\sum_{0 \le i \le k, i \text{ is even}} \binom{k}{i} \zeta(W)^{k-i}\epsilon^i|W = w\right)
\end{aligned}
$$

Is monotone increasing in w

The same arguments can be extended to other series bases derived from polynomial, such as different kinds of splines