SEMIPARAMETRIC AND NONPARAMETRIC INSTRUMENTAL VARIABLE ESTIMATION WITH FIRST-STAGE ISOTONIC REGRESSION

TAISUKE OTSU, KAZUHIKO SHINODA, AND MENGSHAN XU

Abstract. This paper proposes a semiparametric and a nonparametric instrumental variable (IV) estimators, under the assumption that the conditional mean of the endogenous variable, given the instrumental variable, is known to be monotone increasing. We employ isotonic estimation to obtain the fitted instruments in the first stage of a two-stage semiparametric or nonparametric estimation procedure. We show that the proposed semiparametric IV estimator is tuning-parameter-free and achieves the semiparametric efficiency bound. Moreover, we show that compared to the nonparametric two-stage least squares estimator (Blundell, Chen and Kristensen, 2007; Horowitz, 2011, 2012), our proposed nonparametric IV estimator requires notably fewer tuning parameters and achieves the same convergence rate. Additionally, it exhibits greater stability as evidenced by Monte Carlo simulations.

1. Introduction

In economic studies, the explanatory variables are often endogenous, i.e., they are correlated with unobservables, leading to inconsistent estimators. To address this endogeneity issue, linear instrumental variable estimation, also known as two-stage least squares (2SLS), is extensively applied; see, e.g., Wooldridge (2010, Chapter 7) for a review.

To mitigate the restrictive assumption of linearity and to enhance the robustness of estimation and inference processes, semiparametric and nonparametric IV estimations—henceforth referred to as SPIV and NPIV, respectively—have been developed. Significant contributions in this area include works by Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen and Kristensen (2007), Chen and Reiss (2011), and Horowitz (2011, 2012), among others. Except for Hall and Horowitz (2005), which employs a kernel-based method, the majority of the works mentioned above utilize series estimation in the first-stage regression of the endogenous variables on the instrumental variables (or in an implicit first stage of a one-step NPIV estimation procedure, which can be equivalently decomposed into two stages of separate series estimations). Once the fitted instrumental variables are obtained, they are then used in the second-stage estimation to recover a consistent estimator of conditional mean in the presence of endogenous explanatory variables. For the NPIV estimation, this two-step procedure is sometimes referred to as nonparametric 2SLS; see Horowitz (2011) for a detailed discussion of its properties.

In this paper, we propose employing isotonic estimation for the first-stage estimation within both SPIV and NPIV frameworks, and we refer to the proposed estimators as "monotone SPIV/NPIV". The isotonic estimator can be traced back to the mid-20th century, with foundational contributions from Ayer et al. (1955), Grenander (1956), Rao (1969, 1970), and Barlow and Brunk (1972), among others. The isotonic estimator for a regression function is defined through least squares estimation under the monotonicity constraint. Assuming the conditional expectation $\mathbb{E}[Y|X] = m(X)$ is monotone increasing, for an iid random sample $\{Y_i, X_i\}_{i=1}^n$, the isotonic estimator emerges as the solution that minimizes the sum of squared differences, $\min_{m\in\mathcal{M}} \sum_{i=1}^n \{Y_i - m(X_i)\}^2$, where $\mathcal{M}$ denotes the set of monotone increasing functions. The minimizer can be derived using the pool adjacent violators algorithm, as outlined by Barlow and Brunk (1972), or equivalently by identifying the greatest convex minorant in the cumulative sum diagram $\{(0,0), (i, \sum_{j=1}^i Y_j), i = 1, \ldots, n\}$, with $\{X_i\}_{i=1}^n$ arranged in sequence; see Groeneboom and Jongbloed (2014) for an exhaustive examination of isotonic regression's various facets.

The implementation of isotonic regression is predicated on the assumption that the conditional mean of the endogenous variable, given the instrumental variable, is monotone increasing. This assumption is common in the literature on IV estimation. For instance, in a classic economic problem that examines the effect of education level on future wages, it is reasonably argued that the conditional expectation of education level increases monotonically with the IV that reflects the result of IQ tests. Blundell, Chen and Kristensen (2007) uses log earnings as the instrument for log family expenditure; the conditional mean of the latter, given the former, is arguably monotone increasing (see Figures 4 and 5 in Blundell, Chen and Kristensen (2007) and Figure 2 in Härdle and Linton (1994) for nonparametric regression plots of family expenditure on earnings). Furthermore, in the errors-in-variables problem, the conditional mean of the mismeasured running variable is by construction monotone increasing in its IV, an independently repeated measurement of the same value. Broadly speaking, any economic study involving a univariate endogenous variable and univariate instrumental variables, and utilizing the linear 2SLS method, implicitly imposes the monotonicity assumption, as a univariate linear function is inherently monotone.[1]

The advantages of substituting the first-stage series estimation with isotonic regression convince us of its worth. First, it is widely acknowledged that the choice of tuning parameters significantly influences the performance of semiparametric and nonparametric estimations. The isotonic estimator, as a nonparametric method without involving tuning parameters, helps to alleviate this issue of choice for both SPIV and NPIV. In the context of the SPIV, we obtain a tuning-parameter-free SPIV estimator that attains the semiparametric efficiency bound, within the framework of Ai and Chen (2003). Regarding the NPIV, our approach simplifies the selection of tuning parameters, reducing the count from $K_n + 1$ – where K_n represents the series length in the second stage – to merely one.[2] Second, imposing the monotonicity assumption in the first-stage estimation can help to stabilize the resulting NPIV estimator. In Section 4 of simulations, we illustrate that compared to the nonparametric 2SLS, our proposed monotone

[1] We will discuss the multivariate SPIV and NPIV in Section 3, where we employ the monotone partially linear models and the monotone single index models to address the issue of the coexistence of monotonicity and multivariate regression.

[2] It is worth noting that although the NPIV methods by Blundell, Chen and Kristensen (2007) and Horowitz (2011, 2012) require the selection of only one tuning parameter, this simplicity is due to the enforcement of a universal series length across both stages of series estimations. The implicit number of tuning parameters of their approaches remains $K_n + 1$.

NPIV method exhibits a significantly more stable performance across various K_n. Similar stabilizing effects of the monotonicity constraint have also been observed in another NPIV problem within a different context, as discussed by Chetverikov and Wilhelm (2017).

Our proposed monotone NPIV method contributes to the literature on plugging nonparametric estimators into another nonparametric estimation procedure. Examples include Rilstone (1996), Song (2018), Mammen, Rothe and Schienle (2012), and Hahn and Ridder (2013), among others. Additionally, the studies on nonparametric 2SLS methods mentioned previously also fall into this category. Contrary to these works, our proposed method incorporates a non-smooth first-stage isotonic estimator into a smooth second-stage series estimator, making theoretical developments substantially different from the existing approaches.

This paper is organized as follows. In Section 2, we consider univariate monotone SPIV and NPIV estimation. The multivariate case will be discussed in Section 3. Section 4 illustrates the proposed method by a simulation study.

2. Univariate case

In this section, we introduce the setups and study our estimator in the univariate case. The multivariate case will be discussed in Section 3. Considering the following instrumental variable regression model:

$$\begin{aligned} Y &= g(X) + U, \quad \mathbb{E}[U|W] = 0, \\ X &= \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0, \end{aligned} \tag{2.1}$$

where Y is a scalar dependent variable, X is a scalar endogenous regressor, W is a scalar instrumental variable, and U and ϵ are unobservable error terms. Function g can be either parametric or nonparametric, and these two cases will be discussed in the following two subsections, respectively. Throughout the paper, function ζ is unknown, and we are particularly interested in estimating function g by using a random sample $\{Y_i, X_i, W_i\}_{i=1}^n$ of $(Y, X, W) \in \mathbb{R} \times \mathcal{X} \times \mathcal{W}$, under the shape constraint that ζ is a monotone increasing function. The case of decreasing ζ can be investigated in the same manner.

2.1. **Semiparametric IV estimation.**

In this section, we discuss the case where the second stage is a linear model, i.e.,

$$g(X) = \beta_1 + X\beta_2,$$

and we aim to estimate the unknown linear coefficients $\beta := (\beta_1, \beta_2)'$ under the potential endogeneity, $\mathbb{E}\,(U|X) \neq 0$. In this setup, model (2.1) becomes a semiparametric instrument variable model (SPIV).

Let $\mathbb{E}_n[\cdot] = n^{-1}\sum_{i=1}^n \cdot$. By utilizing monotonicity of ζ, we propose to implement the first stage regression by the isotonic regression

$$\min_{\zeta \in \mathcal{M}} \mathbb{E}_n[\{X - \zeta(W)\}^2], \tag{2.2}$$

where $\mathcal{M}$ is the class of monotone increasing functions. The minimizer of (2.2) can be calculated with the pool adjacent violators algorithm (Barlow and Brunk, 1972), or equivalently by solving the greatest convex minorant of the cumulative sum diagram. Define $Z = \zeta(W)$, $\hat{Z} = \hat{\zeta}(W)$ to be the estimator solving (2.2), and $v(x) = (1, x)'$. Then our semiparametric instrumental variable estimator of β is defined as

$$\hat{\beta} = \mathbb{E}_n[v(\hat{Z})v(X)']^{-1}\mathbb{E}_n[v(\hat{Z})Y]. \tag{2.3}$$

To study asymptotic properties of the proposed estimator, we impose the following assumptions.

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) $Var(\mathbb{E}[X|W]) \neq 0$; (ii) $\mathbb{E}\left[U^2|W = w\right] = \sigma_u^2$ for any $w \in \mathcal{W}$.*

Remark 2.1. Assumption 2.1.1 establishes the basic setup for data generating process. Assumption 2.1.2 imposes the monotonicity, the key property we aim to integrate into our SPIV and NPIV frameworks. Assumption 2.1.3(i) ensures the identification of β, and it serves a role analogous to $Cov(X, W) \neq 0$ in a linear IV model. Assumption 2.1.3(ii) enforces homoscedasticity conditional on instruments, ensuring the efficiency of the proposed estimator.

Assumption 2.1.3(ii) can be relaxed. In this case, to achieve the efficiency, we shall plug-in an estimated conditional variance matrix obtained by regressing residual $\hat{U}^2$, which can be derived from a consistent estimator (2.3), on instrument W. If $\mathbb{E}\left[U^2|W = w\right]$ is assumed to be monotone increasing in w, we can also employ isotonic regression, as in Arai, Otsu and Xu (2022). The resulting estimator remains tuning-parameter-free and efficient.

Theorem 2.1. *Under Assumptions 2.1.1 to 2.1.3, it holds*

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

where $\Omega = \sigma_U^2 \mathbb{E}\left[v(Z)v(Z)'\right]^{-1}$ attains the efficiency bound of estimating β (see, e.g., formula (22) on p. 1814 of Ai and Chen, 2003). Furthermore, the estimation of β does not involve any tuning parameter. In addition, a tuning-parameter-free estimator of the variance-covariance matrix is given by $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 \left[v(\hat{Z})v(\hat{Z})'\right]^{-1}$.

2.2. **Nonparametric IV estimation.** In this section, we no longer impose a parametric structure on function $g(X)$ in (2.1). The estimation of $g(\cdot)$ becomes a NPIV problem, where the unobserved U might be correlated to the regressor X.

Let $p(x) = (p_1(x), \ldots, p_{K_n}(x))'$ be a vector of complete basis functions for $L_2(\mathcal{X}) := \{h : \mathcal{X} \to \mathbb{R}, \sqrt{\int_{\mathcal{X}} h(x)^2 dx} < \infty\}$, and $\{K_n\}$ be a positive sequence such that $K_n \to \infty$ as $n \to \infty$. To apply our monotone NPIV estimator, we choose those types of bases that for any $k \in \mathbb{N}$, $\mathbb{E}(p_k(X)|W = w)$ is a monotone increasing function of w. This condition can be satisfied by polynomials under certain assumptions (see the remark following Assumption 2.2.4 below), and it can be extended to series bases that are transformed from polynomials, including various types of splines.

Now we can define

$$q_k(w) = \mathbb{E}(p_k(X)|W = w),$$
$$q(w) = (q_1(w), \ldots, q_{K_n}(w))'. \tag{2.4}$$

Note that $q_k(\cdot)$ is a function defined on $\mathcal{W}$ while $p_k(\cdot)$ is a function defined on $\mathcal{X}$.

Assuming the monotonicity of $q_k(\cdot)$, we propose to employ isotonic regression to estimate the basis functions $q(w)$ directly: For each $k \in \{1 : K_n\}$,

$$\hat{q}_k(\cdot) = \arg\min_{\zeta \in \mathcal{M}} \mathbb{E}_n[\{p_k(X) - \zeta(W)\}^2],$$
$$\hat{q}(w) = (\hat{q}_1(w), \ldots, \hat{q}_{K_n}(w))'. \tag{2.5}$$

Recall $\mathcal{M}$ is the class of monotone increasing functions. Note that $p(\cdot)$, $q(\cdot)$, $\hat{p}(\cdot)$, and $\hat{q}(\cdot)$ have the same dimension of K_n, which should increase with the sample size n. This information about their dimension is suppressed in their notations provided there is no confusion in the context.

Then our proposed NPIV estimator of g is defined as

$$\hat{g}(\cdot) = p(\cdot)' \mathbb{E}_n\left[\hat{q}(W)\, p(X)'\right]^{-1} \mathbb{E}_n\left[\hat{q}(W)\, Y\right]. \tag{2.6}$$

To study the asymptotic properties of $\hat{g}$, let $f_{X|W}$ be the Lebesgue density functions of X conditional on W and f_W be the marginal density function of W, respectively. Then, we define the reduced-form conditional mean to be

$$m(w) = \mathbb{E}[Y|W = w],$$

and define $T : L_2(\mathcal{X}) \to L_2(\mathcal{W})$ to be an operator such that for any $w \in \mathcal{W}$,

$$(T\nu)(w) = \int_{\mathcal{X}} \nu(x) f_{X|W=w}(x, w) dx.$$

Moreover, for $\mathcal{A} \in \{\mathcal{X}, \mathcal{W}\}$, let $\|\cdot\|$ denote $L_2(\mathcal{A})$ norm, $\|h\| := \sqrt{\int_{\mathcal{A}} h(a)^2 da}$. For a non-negative integers r and c, let $\Lambda_c^r(\mathcal{A})$ denote the Hölder space that is defined, for example, on p. 1623 of Blundell, Chen and Kristensen (2007). Then for some positive constant K, we define

$$\mathcal{H}_n = \left\{ h : \mathcal{X} \to [0, K], h \in \Lambda_c^r(\mathcal{X}), h(X) = \sum_{k=1}^{K_n} b_k p_k(X) \right\}, \tag{2.7}$$

where $b = (b_1, \ldots, b_{K_n})$ is a vector of unknown sieve coefficients.

Following Blundell, Chen and Kristensen (2007), we define the sieve measure of ill-posedness:

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

In addition, we define

$$\begin{aligned}\hat{\beta} &= \mathbb{E}_n\left[\hat{q}\left(W\right)p\left(X\right)'\right]^{-1}\mathbb{E}_n\left[\hat{q}\left(W\right)Y\right],\\ \beta_{x,n} &= \mathbb{E}\left[p(X)p(X)'\right]^{-1}\mathbb{E}\left[p(X)g(X)\right],\\ \beta_{w,n} &= \mathbb{E}\left[q(W)q(W)'\right]^{-1}\mathbb{E}\left[q(W)Y\right],\\ g_{x,n}(\cdot) &= p(\cdot)'\beta_{x,n},\ g_{w,n}(\cdot) = p(\cdot)'\beta_{w,n}.\end{aligned} \tag{2.8}$$

Note that by definition, $\hat{g}(\cdot) = p(\cdot)'\hat{\beta}$. Although both $g_{x,n}$ and $g_{w,n}$ are functions defined on $\mathcal{X}$, the latter is constructed with the projection coefficients of projecting Y on $q(W)$. This $g_{w,n}$ is required to handle the possible case where $\mathbb{E}\left[q(W)\left(Y - g_{x,n}(X)\right)\right] \neq 0$, despite $\mathbb{E}\left[q(W)U\right] = 0$. (See the remark following Assumption 2.2.4 below). Note that we do *not* defined $\beta_{x,n}$ as $\mathbb{E}\left[p(X)p(X)'\right]^{-1}\mathbb{E}\left[p(X)Y\right]$ since X is an endogenous variable.

To study asymptotic properties of the proposed estimator, we impose the following assumptions.

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 $\mathbb{E}\left[p(X)p(X)'\right]$ and $\mathbb{E}\left[q(W)q(W)'\right]$ 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] $\rho_n \cdot \|T\left(g_{w,n} - g_{x,n}\right)\| \leq const \cdot \|g - g_{x,n}\|$.*

Remark 2.2. Assumption 2.2.1(i) establishes the basic setup for data generating process. Assumptions 2.2.1 (ii) and (iii) are adapted from Assumption 1 of Blundell, Chen and Kristensen (2007), restricting the true function of interest, g, to a function space subject to certain conditions of smoothness and boundedness.

Assumption 2.2.2 (i) requires that the conditional mean of each basis function, $\mathbb{E}[p_k(X)|W = w]$, is monotone increasing in w. Combined with the following conditions, this assumption can be satisfied by the polynomial basis.

A1: For all $w \in \mathcal{W}$, $\zeta(w)$ is non-negative and monotone increasing in w.

A2: For all $w \in \mathcal{W}$, the conditional density of $\epsilon|W = w$ is symmetrically distributed around 0.

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

A1 imposes the non-negativity, which can be satisfied transforming the original data. A2 can be satisfied by many well-known distributions. A3 is satisfied if ϵ is mean-independent of W. Assumptions similar to A2 and A3 are not uncommon in the literature addressing different but related problems, for example, Newey and Steigerwald (1997).

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

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

Under assumption A1-A3, we have $I = 0$, and $II = \sum_{0\leq i\leq k, i \text{ is even}}\begin{pmatrix} k \\ i \end{pmatrix}\zeta(w)^{k-i}\sigma^i(w)$ is a sum of monotone increasing functions of w. As a result, $\mathbb{E}[p_k(X)|W = w]$ is monotone increasing in w.

Assumption 2.2.3(i) is the nonparametric version of Assumption 2.1.3(i). Assumption 2.2.3(ii) is a standard condition for series estimation. Assumption 2.2.4 is adapted from Assumption 6 of Blundell, Chen and Kristensen (2007). It is required to control the size of $\mathbb{E}\left[q(W)\left(g(X) - g_{x,n}(X)\right)\right]$. The projection error, $g(X) - g_{x,n}(X)$, is uncorrelated to the basis function $p_k(X)$, but not necessarily uncorrelated to $q_k(W) = \mathbb{E}\left(p_k(X)|W\right)$. For more details, see (A.9), (A.10), and the relevant discussions in Appendix A.2.

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

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

3. Multivariate case

(Forthcoming: first-stage partially linear model; first-stage monotone single index model)

4. Simulation

Here we conduct a simulation study to assess the performance of our proposed NPIV estimator. Considering the following DGP

$$\begin{aligned} Y &= X^2 + \epsilon, \qquad (4.1) \\ X &= \exp(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0, \\ W &\sim \mathrm{U}[-1.2, 1.3], \\ \epsilon &\sim N(0,1). \end{aligned}$$

In the following Table 1, we compare the proposed method (labeled as "isotonic+series") with that proposed in Blundell, Chen and Kristensen (2007) and Horowitz (2011,2012) labeled as "series+series"). For the series estimation in both methods, we test the polynomial order $K_n = 2, 3, 4,$ and 5. Among these choices, $K_n = 3$ is an appropriate choice given the data generating process of the second-stage equation (4.1). The number of Monte-Carlo simulations is 500 for each sample size. We evaluate perfomances of two estimators by the sample mean and median of intergrated square error (ISE).

Table 1. series+series vs. isotonic+series

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

As evident from Table 1, for $K_n = 3, 4, 5$, the proposed isotonic IV+series outperforms the series+series method across all the sample size. For $K_n = 2$, the series+series performs marginally better. However, in this case, the model is misspecified as a linear model while in the other three cases, the model is correctly specified but has some redundant regressors. Another remarkable feature of our proposed estimator is that it gives much more stable results than the series+series method: In all the setup in Table 1, it never gives extreme results as the series+series method does in the case of $K_n = 4{,}5$ and $n = 1000$.

Overall, the Monte-Carlo simulation results support the proposed monotone NPIV method.

Appendix A. Mathematical appendix

A.1. **Proof of Theorem 2.1.** We have the semiparametric model

$$\begin{aligned} Y &= \beta_1 + X\beta_2 + U, \quad \mathbb{E}[U|W] = 0, \\ X &= \zeta(W) + \epsilon, \quad \mathbb{E}[\epsilon|W] = 0. \end{aligned} \tag{A.1}$$

Note that for $v(x) = (1, x)'$, we have $\beta_1 + X\beta_2 = v(X)'\beta$. And from (2.3), we have

$$\mathbb{E}_n[v(\hat{Z})Y] - \mathbb{E}_n[v(\hat{Z})v(X)']\hat{\beta} = 0.$$

Then

$$\begin{aligned} &\mathbb{E}\left[v(Z)v(X)'\right]\left(\hat{\beta} - \beta\right) + 0 \\ &= \mathbb{E}\left[v(Z)v(X)'\right]\left(\hat{\beta} - \beta\right) + \mathbb{E}_n[v(\hat{Z})Y] - \mathbb{E}_n[v(\hat{Z})v(X)']\hat{\beta} \\ &= \mathbb{E}\left[v(Z)v(X)'\right]\left(\hat{\beta} - \beta\right) + \mathbb{E}_n[v(\hat{Z})\left(v(X)'\beta + U\right)] - \mathbb{E}_n[v(\hat{Z})v(X)']\hat{\beta} \\ &= \mathbb{E}_n\left[v(\hat{Z})U\right] - \left\{\mathbb{E}_n[v(\hat{Z})v(X)'] - \mathbb{E}\left[v(Z)v(X)'\right]\right\}\left(\hat{\beta} - \beta\right) \\ &= \mathbb{E}_n\left[v(Z)U\right] - \left\{\mathbb{E}_n\left[v(Z)v(X)'\right] - \mathbb{E}\left[v(Z)v(X)'\right]\right\}\left(\hat{\beta} - \beta\right) \\ &+ \mathbb{E}_n\left[\left(v(\hat{Z}) - v(Z)\right)U\right] - \mathbb{E}_n\left[\left(v(\hat{Z}) - v(Z)\right)v(X)'\right]\left(\hat{\beta} - \beta\right) \\ &= \mathbb{E}_n\left[v(Z)U\right] + \mathbb{E}_n\left[\left(v(\hat{Z}) - v(Z)\right)U\right] + o_p(\hat{\beta} - \beta). \end{aligned} \tag{A.2}$$

The last equality follows from

$$\begin{aligned} \mathbb{E}_n\left[v(Z)v(X)'\right] - \mathbb{E}\left[v(Z)v(X)'\right] &= o_p(1), \\ \mathbb{E}_n\left[\left(v(\hat{Z}) - v(Z)\right)v(X)'\right] &= o_p(1). \end{aligned} \tag{A.3}$$

Note that $v(\hat{Z}) - v(Z) = \left(0, \hat{\zeta}(W) - \zeta(W)\right)$, and both $\hat{\zeta}(\cdot)$ and $\zeta(\cdot)$ belongs to $\mathcal{M}$, which is a Donsker class. Thus,

$$\mathbb{E}_n\left[\left(v(\hat{Z}) - v(Z)\right)U\right] = o_p(n^{-1/2}). \tag{A.4}$$

Furthermore, we note that $v(Z) = \mathbb{E}\left[v(X)|W\right]$ due to $Z = \zeta(W) = \mathbb{E}\left[X|W\right]$. By Law of iterated expectation,

$$\mathbb{E}\left[v(Z)v(X)'\right] = \mathbb{E}\left[v(Z)v(Z)'\right].$$

Note that

$$\mathbb{E}\left[v(Z)v(Z)'\right] = \mathbb{E}\begin{pmatrix} 1 & \mathbb{E}[X|W] \\ \mathbb{E}[X|W] & \mathbb{E}[X|W]^2 \end{pmatrix}.$$

Therefore, $\mathbb{E}\left[v(Z)v(Z)'\right]$ is invertible if

$$\begin{aligned} \det\left(\mathbb{E}\left[v(Z)v(Z)'\right]\right) &= \mathbb{E}\left(\mathbb{E}[X|W]^2\right) - \left\{\mathbb{E}\left(\mathbb{E}[X|W]\right)\right\}^2 \\ &= Var\left(\mathbb{E}[X|W]\right) \neq 0. \end{aligned}$$

This is satisfied by Assumption 2.1.3(i).

Combining (A.2), (A.4), and the invertibility of $\mathbb{E}\left[v(Z)v(Z)'\right]$, we have

$$\sqrt{n}\left(\hat{\beta}-\beta\right)=\mathbb{E}\left[v(Z)v(Z)'\right]^{-1}\sqrt{n}\mathbb{E}_n\left[v(Z)U\right]+o_p\left[\sqrt{n}\left(\hat{\beta}-\beta\right)\right],$$

and under Assumption 2.1.3(ii), we have

$$\sqrt{n}\left(\hat{\beta}-\beta\right)\sim_d N(0,\sigma_u^2\mathbb{E}\left[v(Z)v(Z)'\right]^{-1}),$$
$$\Omega=\sigma_u^2\mathbb{E}\left[v(Z)v(Z)'\right]^{-1}.$$

Note that under Assumption 2.1.3(ii), Ω is the efficient variance matrix for estimating β. (Ai and Chen, 2003).

A.2. **Proof of Theorem 2.2.** By the triangle inequality

$$\begin{aligned}\|\hat{g}-g\| &\le \|\hat{g}-g_{x,n}\|+\|g-g_{x,n}\| \\ &\le \|\hat{g}-g_{w,n}\|+\|g_{w,n}-g_{x,n}\|+\|g-g_{x,n}\|. \end{aligned} \tag{A.5}$$

Note that both $g_{w,n}$ and $g_{x,n}$ belong to $\mathcal{H}_n$. By the definition of ρ_n and Assumption 2.2.4,

$$\begin{aligned}\|g_{w,n}-g_{x,n}\| &\le \rho_n\cdot\|T\left(g_{w,n}-g_{x,n}\right)\| \\ &\le \text{const}\cdot\|g-g_{x,n}\|.\end{aligned}$$

Furthermore, by Assumptions (2.2.1)(iii) and (2.2.2) (iii), and the properties of the sieve space $\mathcal{H}_n$ given by (2.7), we have

$$\|g-g_{x,n}\|=O_p\left(K_n^{-r}\right). \tag{A.6}$$

In addition, by the definitions of $g_{w,n}$, $\hat{g}$, operator T, and $q(\cdot)$,

$$\begin{aligned}T\left(g_{w,n}\right)&=T\left(p(\cdot)'\beta_{w,n}\right)=q(\cdot)'\beta_{w,n}, \\ T\left(\hat{g}\right)&=T\left(p(\cdot)'\hat{\beta}\right)=q(\cdot)'\hat{\beta}.\end{aligned} \tag{A.7}$$

Now we focus on $\|\hat{g}-g_{w,n}\|$. Define $\mathbf{Q}_n^X=\mathbb{E}\left[p(X)p(X)'\right]$ and $\mathbf{Q}_n^W=\mathbb{E}\left[q(W)q(W)'\right]$:

$$\begin{aligned}\|\hat{g}-g_{w,n}\|^2 &= \frac{\|\hat{g}-g_{w,n}\|^2}{\|T\left(\hat{g}-g_{w,n}\right)\|^2}\|T\left(\hat{g}-g_{w,n}\right)\|^2 \\ &= \frac{\|\hat{g}-g_{w,n}\|^2}{\|T\left(\hat{g}-g_{w,n}\right)\|^2}\times\left(\hat{\beta}-\beta_{w,n}\right)'\mathbf{Q}_n^W\left(\hat{\beta}-\beta_{w,n}\right) \\ &\le \sup_{h\in\mathcal{H}_n}\frac{\|h\|^2}{\|Th\|^2}\times\left(\hat{\beta}-\beta_{w,n}\right)'\mathbf{Q}_n^W\left(\hat{\beta}-\beta_{w,n}\right) \\ &= \rho_n^2\times\left(\hat{\beta}-\beta_{w,n}\right)'\mathbf{Q}_n^W\left(\hat{\beta}-\beta_{w,n}\right),\end{aligned} \tag{A.8}$$

where the second equality follows from (A.7); the first inequality follows from the fact that both $\hat{g}$ and $g_{w,n}$ belong to $\mathcal{H}_n$.

Define

$$e=Y-g_{w,n}(X)=Y-p(X)'\beta_{w,n}. \tag{A.9}$$

By the definition of $q(\cdot)$ and Law of iterated expectation, we have

$$\begin{aligned}\mathbb{E}\left(q(W)e\right) &= \mathbb{E}\left[q(W)\mathbb{E}\left(Y - p(X)'\beta_{w,n}|W\right)\right] \\ &= \mathbb{E}\left[q(W)\mathbb{E}\left(Y - q(W)'\beta_{w,n}|W\right)\right] \\ &= \mathbb{E}\left[q(W)\left(Y - q(W)'\beta_{w,n}\right)\right] \\ &= 0, \end{aligned} \tag{A.10}$$

where the last equality follows by (2.8) and the property of linear projection.

From (A.9), we can derive

$$\begin{aligned}\hat{\beta} - \beta_{w,n} &= \mathbb{E}_n\left[\hat{q}(W)p(X)'\right]^{-1}\mathbb{E}_n\left[\hat{q}(W)Y\right] - \beta_{w,n} \\ &= \mathbb{E}_n\left[\hat{q}(W)p(X)'\right]^{-1}\mathbb{E}_n\left[\hat{q}(W)e\right].\end{aligned}$$

Then

$$\begin{aligned}&\left(\hat{\beta} - \beta_{w,n}\right)'\mathbf{Q}_n^W\left(\hat{\beta} - \beta_{w,n}\right) \\ &= \mathbb{E}_n\left[\hat{q}(W)e\right]'\mathbb{E}_n\left[p(X)\hat{q}(W)'\right]^{-1}\mathbf{Q}_n^W\mathbb{E}_n\left[\hat{q}(W)p(X)'\right]^{-1}\mathbb{E}_n\left[\hat{q}(W)e\right] \\ &= \mathbb{E}_n\left[\hat{q}(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\left(\mathbf{Q}_n^W\right)^{\frac{1}{2}}\mathbb{E}_n\left[p(X)\hat{q}(W)'\right]^{-1}\left(\mathbf{Q}_n^W\right)^{\frac{1}{2}} \\ &\times\left(\mathbf{Q}_n^W\right)^{\frac{1}{2}}\mathbb{E}_n\left[\hat{q}(W)p(X)'\right]^{-1}\left(\mathbf{Q}_n^W\right)^{\frac{1}{2}}\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\mathbb{E}_n\left[\hat{q}(W)e\right] \\ &= \mathbb{E}_n\left[\hat{q}(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\left(\mathbf{A}_n'\right)^{-1}\mathbf{A}_n^{-1}\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\mathbb{E}_n\left[\hat{q}(W)e\right],\end{aligned} \tag{A.11}$$

where

$$\begin{aligned}\mathbf{A}_n &:= \left\{\left(\mathbf{Q}_n^W\right)^{\frac{1}{2}}\mathbb{E}_n\left[\hat{q}(W)p(X)'\right]^{-1}\left(\mathbf{Q}_n^W\right)^{\frac{1}{2}}\right\}^{-1} \\ &= \left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\mathbb{E}_n\left[\hat{q}(W)p(X)'\right]\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}.\end{aligned}$$

is a $K_n \times K_n$ square matrix.

With some abuse of notations, in the following, we also use $\|\cdot\|$ to denote the matrix norm provided there is no confusion in the context.

Lemma A.1. *Under Assumptions 2.2.1 to 2.2.3, it holds: (i)* $\|\mathbf{A}_n - \boldsymbol{I}_{K_n}\|; \xrightarrow{p} 0$; *(ii)* $\left\|\mathbf{A}_n^{-1} - \boldsymbol{I}_{K_n}\right\| \xrightarrow{p} 0$; *(iii)* $\lambda_{max}\left[\left(\mathbf{A}_n'\right)^{-1}\mathbf{A}_n^{-1}\right] = 1/\lambda_{min}\left[\mathbf{A}_n\mathbf{A}_n'\right] \xrightarrow{p} 1$.

Consequently,

$$\begin{aligned}&\left(\hat{\beta} - \beta_{w,n}\right)'\mathbf{Q}_n^W\left(\hat{\beta} - \beta_{w,n}\right) \\ &= \mathbb{E}_n\left[\hat{q}(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\left(\mathbf{A}_n'\right)^{-1}\mathbf{A}_n^{-1}\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\mathbb{E}_n\left[\hat{q}(W)e\right] \\ &\le \left\|\left(\mathbf{A}_n'\right)^{-1}\mathbf{A}_n^{-1}\right\|\mathbb{E}_n\left[\hat{q}(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\left(\mathbf{Q}_n^W\right)^{-\frac{1}{2}}\mathbb{E}_n\left[\hat{q}(W)e\right] \\ &\le \lambda_{max}\left[\left(\mathbf{A}_n'\right)^{-1}\mathbf{A}_n^{-1}\right]\mathbb{E}_n\left[\hat{q}(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-1}\mathbb{E}_n\left[\hat{q}(W)e\right] \\ &= O_P(1)\mathbb{E}_n\left[\hat{q}(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-1}\mathbb{E}_n\left[\hat{q}(W)e\right],\end{aligned} \tag{A.12}$$

where the first equality is the last row of (A.11); the first inequality follows from Quadratic inequality; the second inequality and the last equality follow from Schwarz matrix inequality, the definition of the matrix norm, and Lemma A.1.

Now we focus on the last row of (A.12):

$$\begin{aligned}
&\mathbb{E}_n\left[\hat{q}(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-1}\mathbb{E}_n\left[\hat{q}(W)e\right]\\
=&\mathbb{E}_n\left[\left(\hat{q}(W)+q(W)-q(W)\right)e\right]'\left(\mathbf{Q}_n^W\right)^{-1}\mathbb{E}_n\left[\left(\hat{q}(W)+q(W)-q(W)\right)e\right]\\
=&\mathbb{E}_n\left[q(W)e\right]'\left(\mathbf{Q}_n^W\right)^{-1}\mathbb{E}_n\left[q(W)e\right]\\
+&2\mathbb{E}_n\left[\left(\hat{q}(W)-q(W)\right)e\right]'\left(\mathbf{Q}_n^W\right)^{-1}\mathbb{E}_n\left[q(W)e\right]\\
+&\mathbb{E}_n\left[\left(\hat{q}(W)-q(W)\right)e\right]'\left(\mathbf{Q}_n^W\right)^{-1}\mathbb{E}_n\left[\left(\hat{q}(W)-q(W)\right)e\right]\\
=&A+2B+C.
\end{aligned}$$

Lemma A.2. *Under Assumptions 2.2.1 to 2.2.3, it holds: (i)* $A=O_p\left(\frac{K_n}{n}\right)$ *;(ii)* $B=O_p\left(\frac{K_n}{n}\right)$*; (iii)* $C=O_p\left(\frac{K_n}{n}\right)$.

As a result,

$$\left(\hat{\beta}-\beta_{w,n}\right)'\mathbf{Q}_n^W\left(\hat{\beta}-\beta_{w,n}\right)=O_p\left(\frac{K_n}{n}\right). \tag{A.13}$$

Combining (A.6), (A.8), and (A.13), we conclude that

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

References

[1] Ai, C. and Chen, X. (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6), 1795-1843.

[2] Arai, Y., Otsu, T. and Xu, M. (2022) GLS under monotone heteroskedasticity. arXiv preprint arXiv:2210.13843.

[3] Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and E. Silverman (1955) An empirical distribution function for sampling with incomplete information, *Annals of Mathematical Statistics*, 26, 641-647.

[4] Barlow, R. and H. Brunk (1972) The isotonic regression problem and its dual, *Journal of the American Statistical Association*, 67, 140-147.

[5] Chen, X. and Reiss, M. (2011) On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27(3), 497-521.

[6] Chetverikov, D. and Wilhelm, D. (2017) Nonparametric instrumental variable estimation under monotonicity. *Econometrica,* 85(4), 1303-1320.

[7] Blundell, R., Chen, X. and Kristensen, D. (2007) Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, 75(6), 1613-1669.

[8] Grenander, U. (1956) On the theory of mortality measurement. II., *Skand. Aktuarietidskr*, 39, 125-153.

[9] Hall, P. and Horowitz, J. L. (2005) Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics,* Ann. Statist. 33(6), 2904-2929.

[10] Hansen, B. (2022) *Econometrics*. Princeton University Press.

[11] Hahn, J. and Ridder, G. (2013) Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica*, 81(1), 315-340.

[12] Härdle, W. and Linton, O. (1994) Applied nonparametric methods. *Handbook of econometrics*, 4, 2295-2339.

[13] Horowitz, J. L. (2011) Applied nonparametric instrumental variables estimation. *Econometrica*, 79(2), 347-394.

[14] Horowitz, J. L. (2012) Specification testing in nonparametric instrumental variable estimation. *Journal of Econometrics*, 167(2), 383-396.

[15] Mammen, E., Rothe, C. and Schienle, M. (2012) Nonparametric regression with nonparametrically generated covariates. *The Annals of Statistics,* Ann. Statist. 40(2): 1132-1170.

[16] Newey, W. K. (1993) Efficient estimation of models with conditional moment restrictions. *Handbook of Statistics*, Vol. 11, 419-454.

[17] Newey, W. K. and Powell, J. L. (2003) Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5), 1565-1578.

[18] Newey, W. K., and Steigerwald, D. G. (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica: Journal of the Econometric Society*, 587-599.

[19] Rao, B. P. (1969) Estimation of a unimodal density, *Sankhyā*, A 31, 23-36.

[20] Rao, B. P. (1970) Estimation for distributions with monotone failure rate, *Annals of Mathematical Statistics*, 41, 507-519.

[21] Rilstone, P. (1996) Nonparametric estimation of models with generated regressors. *International Economic Review*, 299-313.

[22] Song, K. (2008) Uniform convergence of series estimators over function spaces. *Econometric Theory*, 24(6), 1463-1499.

[23] Wooldridge, J. M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK, and Keio Economic Observatory (KEO), 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

Email address: t.otsu@lse.ac.uk

Faculty of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

Email address: k-shinoda@keio.jp

Department of Economics, University of Mannheim, 68161 Mannheim, Germany.

Email address: mengshan.xu@uni-mannheim.de