Supplemental Appendix to: "The Influence Function of Semiparametric Two-step Estimators with Estimated Control Variables"

Jinyong Hahn* UCLA | Zhipeng Liao† UCLA | Geert Ridder‡ USC | Ruoyao Shi§ UC Riverside

July 21, 2023

Abstract

This supplemental appendix provides the proofs of the main results presented in Hahn, Liao, Ridder, and Shi (2023), and three examples that serve to illustrate these results. Sections A and B contain the proof of their Theorems 1 and 2 respectively. Section C outlines the setups used for the three examples, while Section D presents the computation of the influence function of the semiparametric estimator within these examples.

A Proof of Theorem 1

The theorem is proved using the arguments in Sections 2 and 3 of Newey (1994). Under Assumptions 1(i) and 2(i) of Hahn et al. (2023), (3.10) in Newey (1994) shows that the influence function of $\hat{\beta}$ can be derived from (5) in Hahn et al. (2023), and it takes the following form

$$-\left(\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial \beta^\top}\right)^{-1}\left(J(Z,\beta_*,\pi_*)+\eta(Z)\right), \tag{A.1}$$

where $\eta(Z)$ satisfies $\mathbb{E}\left[\eta(Z)\right]=0$ and

$$\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_{*,\tau})\right]}{\partial \tau}=\mathbb{E}\left[\eta(Z)\frac{\partial \ln(f_{z,\tau}(Z))}{\partial \tau}\right], \tag{A.2}$$

*Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: hahn@econ.ucla.edu

†Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: zhipeng.liao@econ.ucla.edu

‡Department of Economics, University of Southern California, Los Angeles, CA 90089. Email: ridder@usc.edu.

§Corresponding author. Department of Economics, UC Riverside, 900 University Ave., Riverside, CA 92521, USA. Tel: +1(951)827-1494. Email: ruoyao.shi@ucr.edu.

where $f_{z,\tau}(\cdot)$ denotes any one-dimensional path of density of Z indexed by $\tau \in \mathbb{R}$ such that the path hits the true density at $\tau = 0$, and $\pi_{*,\tau}$ is the counterpart of π_* under the path density $f_{z,\tau}(\cdot)$. Let $\varphi_\pi(Z)$ denote the influence function of the first-step estimator, that is, $\mathbb{E}\left[\varphi_\pi(Z)\right] = 0$ and

$$\frac{\partial \pi_{*,\tau}}{\partial \tau} = \mathbb{E}\left[\varphi_\pi(Z)\frac{\partial \ln(f_{z,\tau}(Z))}{\partial \tau}\right]. \tag{A.3}$$

From (A.2) and (A.3), we get $\eta(Z) = \frac{\partial \mathbb{E}[J(Z,\beta_*,\pi_*)]}{\partial \pi^\top}\varphi_\pi(Z)$, and hence the influence function of $\hat{\beta}$ is

$$-\left(\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial \beta^\top}\right)^{-1}\left(J(Z,\beta_*,\pi_*) + \frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial \pi^\top}\varphi_\pi(Z)\right) \tag{A.4}$$

by Newey (1994). It remains to find the explicit forms of $J(Z,\beta_*,\pi_*)$, $\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]/\partial \pi^\top$ and $\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]/\partial \beta^\top$, which are calculated below in (A.11), (A.17) and (A.18) respectively.

The rest of the proof proceeds in three steps. Step 1 and Step 2 contain auxiliary results which are used in Step 3. The main result of the theorem is proved in Step 3.

Step 1. In this step, we show that

$$\mathbb{E}\left[\left.\psi_\lambda\left(Z,\beta_*,\lambda_*\left(v\right)\right)\right| v\right] = 0. \tag{A.5}$$

First note that $h\left(v\left(\pi\right);\beta,\pi\right)$ satisfies the first-order condition

$$\mathbb{E}\left[\psi_\lambda\left(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right)\right)\lambda\left(v\left(\pi\right)\right)\right] = 0 \tag{A.6}$$

for any function $\lambda\left(v\left(\pi\right)\right)$ of $v(\pi)$. Evaluating (A.6) at (β_*,π_*) and using $h\left(v\left(\pi_*\right);\beta_*,\pi_*\right) = \lambda_*\left(v\right)$, we obtain $\mathbb{E}\left[\psi_\lambda(Z,\beta_*,\lambda_*(v))\lambda(v)\right] = 0$ for any function $\lambda\left(v\right)$ of v, which immediately implies (A.5).

Step 2. In this step, we show that for any π,

$$\frac{\partial h\left(v\left(\pi\right);\beta_*,\pi\right)}{\partial \beta} = -\frac{\mathbb{E}\left[\left.\psi_{\lambda,\beta}\left(Z,\beta_*,h\left(v\left(\pi\right);\beta_*,\pi\right)\right)\right| v\left(\pi\right)\right]}{\mathbb{E}\left[\left.\psi_{\lambda,\lambda}\left(Z,\beta_*,h\left(v\left(\pi\right);\beta_*,\pi\right)\right)\right| v\left(\pi\right)\right]}. \tag{A.7}$$

Under Assumptions 1(i) and 2(i, v) of Hahn et al. (2023), we can differentiate (A.6) with respect to β and apply the chain rule to obtain

$$0 = \mathbb{E}\left[\left(\psi_{\lambda,\beta}(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right)) + \psi_{\lambda,\lambda}(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right))\frac{\partial h\left(v\left(\pi\right);\beta,\pi\right)}{\partial \beta}\right)\lambda\left(v\left(\pi\right)\right)\right] \tag{A.8}$$

for any function $\lambda\left(v\left(\pi\right)\right)$ of $v\left(\pi\right)$, which implies that

$$0 = \mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_*,h\left(v\left(\pi\right);\beta_*,\pi\right)\right) + \psi_{\lambda,\lambda}\left(Z,\beta_*,h\left(v\left(\pi\right);\beta_*,\pi\right)\right)\frac{\partial h\left(v\left(\pi\right);\beta_*,\pi\right)}{\partial \beta}\right| v\left(\pi\right)\right], \tag{A.9}$$

from which and the observation that $\partial h\left(v\left(\pi\right);\beta_*,\pi\right)/\partial \beta$ is a function of $v\left(\pi\right)$, we get

$$\begin{aligned} 0 = {} & \mathbb{E}\left[\left.\psi_{\lambda,\beta}\left(Z,\beta_*,h\left(v\left(\pi\right);\beta_*,\pi\right)\right)\right| v\left(\pi\right)\right] \\ & + \mathbb{E}\left[\left.\psi_{\lambda,\lambda}\left(Z,\beta_*,h\left(v\left(\pi\right);\beta_*,\pi\right)\right)\right| v\left(\pi\right)\right]\frac{\partial h\left(v\left(\pi\right);\beta_*,\pi\right)}{\partial \beta}. \end{aligned} \tag{A.10}$$

The claim in (A.7) follows from (A.10).

Step 3. We prove the claim of the theorem in this step. First, by the definition of $J(Z,\beta_*,\pi_*)$ in (6) and the definition of $g_*(v)$ in (7) of Hahn et al. (2023), and the expression in (A.7), we get

$$\begin{aligned}
J(Z,\beta_*,\pi_*) &= \psi_\beta\left(Z,\beta_*,\lambda_*\left(v\right)\right)+\psi_\lambda\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\beta}\\
&= \psi_\beta\left(Z,\beta_*,\lambda_*\left(v\right)\right)-g_*(v)\psi_\lambda\left(Z,\beta_*,\lambda_*\left(v\right)\right)=\varphi_\beta(Z).
\end{aligned} \tag{A.11}$$

Next, from (5) of Hahn et al. (2023), we observe that

$$\begin{aligned}
\frac{\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial\pi^\top} &= \mathbb{E}\left[\psi_{\beta,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{d}{d\pi^\top}h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)\right]\\
&\quad+\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)g_*(v)\frac{d}{d\pi^\top}h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)\right]\\
&\quad-\mathbb{E}\left[\psi_\lambda\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{d}{d\pi^\top}g(v\left(\pi_*\right);\pi_*)\right],
\end{aligned} \tag{A.12}$$

where $g(v\left(\pi\right);\pi)\equiv-\partial h\left(v\left(\pi\right);\beta_*,\pi\right)/\partial\beta$. We recall that π enters $h\left(v\left(\pi\right);\beta,\pi\right)$ in two places, first as an argument of $v\left(\pi\right)$ and second as a way of changing the entire functional form of $h\left(v\left(\pi\right);\beta,\pi\right)$. We will use the following notation to distinguish the two roles played by π:

$$\frac{d}{d\pi^\top}h\left(v\left(\pi_1\right);\beta_*,\pi_2\right)\equiv\frac{\partial h\left(v\left(\pi_1\right);\beta_*,\pi_2\right)}{\partial\pi_1^\top}+\frac{\partial h\left(v\left(\pi_1\right);\beta_*,\pi_2\right)}{\partial\pi_2^\top}.$$

So we have

$$\frac{d}{d\pi^\top}h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)\equiv\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_1^\top}+\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}. \tag{A.13}$$

Moreover, because $h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)=\lambda_*\left(v\right)$, we can see that

$$\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_1^\top}=\frac{\partial\lambda_*\left(v(\pi_*)\right)}{\partial v}\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}. \tag{A.14}$$

We also note that $\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)/\partial\pi_2^\top$ is a function of $v\left(\pi_*\right)=v$, which together with (A.7), (A.9) and the definition of $g_*(v)$ implies that

$$\begin{aligned}
&\mathbb{E}\left[\psi_{\beta,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}\right]\\
&+\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\beta}\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_*,\lambda_*\left(v\right)\right)-g_*(v)\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\middle|\,v\right]\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}\right]=0,
\end{aligned} \tag{A.15}$$

where we also used Assumption 2(iv) in Hahn et al. (2023), i.e., $\psi_{\beta,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)=\psi_{\lambda,\beta}\left(Z,\beta_*,\lambda_*\left(v\right)\right)$ almost surely. Therefore, using (A.12), (A.13), (A.14) and (A.15), we get

$$\begin{aligned}
\frac{\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial\pi^\top} &= \mathbb{E}\left[\left[\psi_{\lambda,\beta}\left(Z,\beta_*,\lambda_*(v)\right)-g_*(v)\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*(v)\right)\right]\frac{\partial\lambda_*\left(v(\pi_*)\right)}{\partial v}\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right]\\
&\quad-\mathbb{E}\left[\psi_\lambda\left(Z,\beta_*,\lambda_*(v)\right)\frac{d}{d\pi^\top}g(v\left(\pi_*\right);\pi_*)\right].
\end{aligned} \tag{A.16}$$

Note that

$$\frac{d}{d\pi^\top} g(v(\pi);\pi) = \frac{\partial g(v(\pi_1);\pi_2)}{\partial \pi_1^\top} + \frac{\partial g(v(\pi_1);\pi_2)}{\partial \pi_2^\top}.$$

So we have

$$\frac{d}{d\pi^\top} g(v(\pi_*);\pi_*) = \frac{\partial g(v(\pi_*);\pi_*)}{\partial \pi_1^\top} + \frac{\partial g(v(\pi_*);\pi_*)}{\partial \pi_2^\top},$$

where $\partial g(v(\pi_*);\pi_*)/\partial\pi_2^\top$ is a function of $v(\pi_*) = v$. Therefore, by (A.5),

$$\begin{aligned}
&\mathbb{E}\left[\psi_\lambda(Z,\beta_*,\lambda_*(v)) \frac{d}{d\pi^\top} g(v(\pi_*);\pi_*)\right] = \mathbb{E}\left[\psi_\lambda(Z,\beta_*,\lambda_*(v)) \frac{\partial g(v(\pi_*);\pi_*)}{\partial \pi_1^\top}\right] \\
&= \mathbb{E}\left[\psi_\lambda(Z,\beta_*,\lambda_*(v)) \frac{\partial g(v(\pi_*);\pi_*)}{\partial v} \frac{\partial v(\pi_*)}{\partial \pi^\top}\right],
\end{aligned}$$

which together with (A.16), the definition of $g_*(v)$, and $\partial g(v(\pi_*),\pi_*)/\partial v = \partial g_*(v)/\partial v$ implies that

$$\begin{aligned}
\frac{\partial \mathbb{E}[J(Z,\beta_*,\pi_*)]}{\partial \pi^\top} &= \mathbb{E}\left[\left[\psi_{\lambda,\beta}(Z,\beta_*,\lambda_*(v)) - g_*(v)\psi_{\lambda,\lambda}(Z,\beta_*,\lambda_*(v))\right] \frac{\partial \lambda_*(v(\pi_*))}{\partial v} \frac{\partial v(\pi_*)}{\partial \pi^\top}\right] \\
&\quad - \mathbb{E}\left[\psi_\lambda(Z,\beta_*,\lambda_*(v)) \frac{\partial g_*(v)}{\partial v} \frac{\partial v(\pi_*)}{\partial \pi^\top}\right] = \Psi_{\beta,\pi}. \qquad \text{(A.17)}
\end{aligned}$$

Finally, we calculate $\partial \mathbb{E}[J(Z,\beta_*,\pi_*)]/\partial\beta^\top$. Specifically,

$$\begin{aligned}
\frac{\partial \mathbb{E}[J(Z,\beta_*,\pi_*)]}{\partial \beta^\top} &= \mathbb{E}\left[\psi_{\beta,\beta}(Z,\beta_*,\lambda_*(v)) + \psi_{\beta,\lambda}(Z,\beta_*,\lambda_*(v)) \frac{\partial h(v(\pi_*);\beta_*,\pi_*)}{\partial \beta^\top}\right] \\
&\quad + \mathbb{E}\left[\frac{\partial h(v(\pi_*);\beta_*,\pi_*)}{\partial \beta} \psi_{\lambda,\beta}(Z,\beta_*,\lambda_*(v))^\top\right] \\
&\quad + \mathbb{E}\left[\psi_{\lambda,\lambda}(Z,\beta_*,\lambda_*(v)) \frac{\partial h(v(\pi_*);\beta_*,\pi_*)}{\partial \beta} \frac{\partial h(v(\pi_*);\beta_*,\pi_*)}{\partial \beta^\top}\right] \\
&\quad + \mathbb{E}\left[\psi_\lambda(Z,\beta_*,\lambda_*(v)) \frac{\partial^2 h(v(\pi_*);\beta_*,\pi_*)}{\partial\beta\partial\beta^\top}\right],
\end{aligned}$$

which together with (A.5), (A.7) and the definition of $g_*(v)$ implies that

$$\begin{aligned}
\frac{\partial \mathbb{E}[J(Z,\beta_*,\pi_*)]}{\partial \beta^\top} &= \mathbb{E}\left[\psi_{\beta,\beta}(Z,\beta_*,\lambda_*(v)) - \psi_{\beta,\lambda}(Z,\beta_*,\lambda_*(v))\, g_*(v)^\top\right] \\
&\quad - \mathbb{E}\left[g_*(v)\psi_{\lambda,\beta}(Z,\beta_*,\lambda_*(v))^\top\right] + \mathbb{E}\left[\psi_{\lambda,\lambda}(Z,\beta_*,\lambda_*(v))\, g_*(v) g_*(v)^\top\right] \\
&= \mathbb{E}\left[\psi_{\beta,\beta}(Z,\beta_*,\lambda_*(v)) - \psi_{\lambda,\lambda}(Z,\beta_*,\lambda_*(v))\, g_*(v) g_*(v)^\top\right] = -\Psi_{\beta,\beta}. \qquad \text{(A.18)}
\end{aligned}$$

Plugging the forms of $J(Z,\beta_*,\pi_*)$, $\partial\mathbb{E}[J(Z,\beta_*,\pi_*)]/\partial\pi^\top$ and $\partial\mathbb{E}[J(Z,\beta_*,\pi_*)]/\partial\beta^\top$ (which are obtained in (A.11), (A.17) and (A.18) respectively) in (A.4), and applying Assumptions 2(ii, iii) in Hahn et al. (2023), we obtain the influence function stated in the theorem.

B Proof of Theorem 2

Taking derivative with respect to τ in (16) and applying the chain rule and Assumption 4(i) in Hahn et al. (2023), we get

$$\frac{\partial \mathbb{E}_\tau\left[\mu_l(Z_l,\pi_{*,l})\pi_l(W_l)\right]}{\partial \tau} + \mathbb{E}\left[\mathbb{E}\left[\mu_{l,\pi}(Z_l,\pi_{*,l})|\, W_l\right]\pi_l(W_l)\frac{\partial \pi_{*,l,\tau}(W_l)}{\partial \tau}\right] = 0 \tag{B.1}$$

for any function $\pi_l(w_l)$, where derivatives with respect to τ are evaluated at $\tau = 0$ unless otherwise indicated. The finite dimensional parameter β_* still satisfies the first-order condition in (5) in Hahn et al. (2023). Using similar calculation as in (A.12), (A.15), (A.16) and (A.17) in the proof of Theorem 1, we obtain

$$\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_{*,\tau})\right]}{\partial \tau} = \frac{\partial \mathbb{E}\left[D(Z,\pi_{*,\tau})\right]}{\partial \tau}, \tag{B.2}$$

where $\pi_{*,\tau} \equiv (\pi_{*,1,\tau},\ldots,\pi_{*,L,\tau})^\top$ and

$$D(Z,\pi_\tau) \equiv \left[\left(\psi_{\lambda,\beta}(Z) - g_*(v)\,\psi_{\lambda,\lambda}(Z)\right)\frac{\partial \lambda_*(v)}{\partial v} - \psi_\lambda(Z)\frac{\partial g_*(v)}{\partial v}\right]\sum_{l=1}^{L}\frac{\partial v(\pi_*)}{\partial \pi_l}\pi_{l,\tau}(W_l),$$

which is linear in π_τ. Note that by (11), (12) and (17) of Hahn et al. (2023), $D(Z,\pi_\tau)$ thus defined satisfies

$$\mathbb{E}\left[D_\tau(Z,\pi)\right] = \mathbb{E}\left[\delta_\pi(W)^\top \pi_\tau(W)\right] \tag{B.3}$$

for any $\pi_\tau(W)$ by the law of iterated expectation. Combining (B.2) and (B.3), we deduce that

$$\begin{aligned}\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_{*,\tau})\right]}{\partial \tau} &= \frac{\partial \mathbb{E}\left[D(Z,\pi_{*,\tau})\right]}{\partial \tau} = \mathbb{E}\left[\delta_\pi(W)^\top\frac{\partial \pi_{*,\tau}(W)}{\partial \tau}\right] = \sum_{l=1}^{L}\mathbb{E}\left[\delta_{l,\pi}(W_l)\frac{\partial \pi_{*,l,\tau}(W_l)}{\partial \tau}\right]\\ &= \sum_{l=1}^{L}\frac{\partial}{\partial \tau}\mathbb{E}_\tau\left[-\frac{\mu_l(Z_l,\pi_{*,l})}{\mathbb{E}\left[\mu_{l,\pi}(Z_l,\pi_{*,l})|\,W_l\right]}\delta_{l,\pi}(W_l)\right] = \frac{\partial}{\partial \tau}\mathbb{E}_\tau\left[\delta_\pi(W)^\top\varphi_\pi(Z)\right], \end{aligned} \tag{B.4}$$

where the third equality follows from (B.1) by replacing $\pi_l(W_l)$ with $\delta_{l,\pi}(W_l)/\mathbb{E}\left[\mu_{l,\pi}(Z_l,\pi_{*,l})|\,W_l\right]$ for $l = 1,\ldots,L$. Therefore, (3.9) in Newey (1994) follows from (B.4) and Theorem 2 is directly implied by Theorem 2.1 of Newey (1994).

C Three Examples

In this section, we provide three examples where the control variable approach is applied to identify and to estimate the parameters of interest. The main theory established in Sections 2 and 3 of Hahn et al. (2023) can be used to derive the influence functions of the semiparametric two-step estimators in these examples.

Example 1 (Mean Regression). Consider the following nonlinear regression model

$$Y = m(X, W_0, \beta_*) + u, \tag{C.1}$$

$$X = \varphi(W, \pi_*) + v, \ \mathbb{E}\left[v | W\right] = 0, \tag{C.2}$$

where Y is the dependent variable, X is an endogenous regressor and W_0 is a vector of exogenous regressors, u and v are the unobservable residuals, $m(x, w_0, \beta)$ and $\varphi(w, \pi)$ are smooth functions and known up to β and π respectively, $W \equiv (W_0^\top, W_1^\top)^\top$ and W_1 is a vector of excluded variables.

To achieve identification of β_*, we assume that v is the control variable such that

$$\mathbb{E}\left[u | X, W_0, v\right] = \mathbb{E}\left[u | v\right]. \tag{C.3}$$

The above condition is imposed on the control variable $v = X - \varphi(W, \pi_*)$ which is an "index" function of X and W. Let $\lambda_*(v) \equiv \mathbb{E}\left[u | v\right]$ and $\varepsilon \equiv u - \lambda_*(v)$. Then we can write (C.1) as

$$Y = m(X, W_0, \beta_*) + \lambda_*(v) + \varepsilon. \tag{C.4}$$

By the definition of $\lambda_*(v)$ and the restriction in (C.3)

$$\mathbb{E}\left[\varepsilon | X, W_0, v\right] = 0, \tag{C.5}$$

which implies that the finite dimensional parameter β_* is identified together with the unknown function $\lambda_*(v)$ as the minimizer of the following problem

$$\min_{\beta, \lambda} \mathbb{E}\left[2^{-1} \left|Y - m(X, W_0, \beta) - \lambda(v)\right|^2\right]. \tag{C.6}$$

To construct feasible estimators of unknown parameters β_* and λ_* based on (C.6), we assume that there exists a first-step estimator $\hat{\pi}$ of π_* such that v is estimated by $\hat{v} \equiv X - \varphi(W, \hat{\pi})$. For example, the first step could be a non-linear regression of the reduced form (C.2), if π is a finite dimensional parameter. In this case, $\hat{\pi}$ is the non-linear regression estimator, and $\hat{v}$ is the fitted residual. The first step could also be a nonparametric regression of the reduced form

$$X = \pi_*(W) + v, \text{ where } \mathbb{E}\left[v | W\right] = 0. \tag{C.7}$$

In this case, we have $v(X, W, \pi) \equiv X - \pi(W)$, $\hat{\pi}$ is the nonparametric regression estimator of X on W, and $\hat{v}$ is the fitted residual from the nonparametric estimation. Given a random sample $\left\{(Y_i, X_i, W_i^\top)^\top\right\}_{i=1}^n$ and the estimate $\hat{v}_i$ from the first step, β_* and $\lambda_*(v)$ can be estimated by many popular semi/nonparametric methods. For example, one may approximate $\lambda_*(\cdot)$ by k_n approximating functions $P_{k_n}(\cdot) \equiv (p_1(\cdot), \ldots, p_{k_n}(\cdot))^\top$ and estimate β_* along with λ_* through the semiparametric series regression

$$(\hat{\beta}^\top, \hat{\gamma}^\top)^\top \equiv \operatorname*{arg\,min}_{\beta \in \mathbb{R}^{d_\beta}, \gamma \in \mathbb{R}^{k_n}} \sum_{i=1}^n 2^{-1} \left|Y_i - m(X_i, W_{0,i}, \beta) - P_{k_n}(\hat{v}_i)^\top \gamma\right|^2. \tag{C.8}$$

The main results established in the paper can be applied to derive the influence function of $\hat{\beta}$ based on the series method, as well as other nonparametric (e.g., kernel) methods.

This example nests the model studied in Li and Wooldridge (2002), where the control variable v_i is parametrically specified. Moreover, the identification condition (C.3) is different from the condition

$$\mathbb{E}\left[u|\, X, W\right] = \mathbb{E}\left[u|\, v\right], \tag{C.9}$$

which implies that

$$\mathbb{E}\left[\varepsilon|\, X, W\right] = 0. \tag{C.10}$$

Li and Wooldridge (2002) derive the root-n asymptotic normality of the two-step estimator under (C.10). As we shall see in Section D, the influence function and the asymptotic variance of the two-step estimator $\hat{\beta}$ are different under the different identification condition in (C.3). $\square$

Example 2 (Quantile Regression). Suppose that we are interested in estimating the quantile structural effect of a set of explanatory variables on a dependent variable Y through the following model

$$Y = X\beta_{1,\alpha,*} + W_0^\top \beta_{2,\alpha,*} + u, \tag{C.11}$$

$$X = \varphi(W, \pi_{\tilde{\alpha},*}) + v, \;\; Q^{\tilde{\alpha}}_{v|W}(w) = 0, \tag{C.12}$$

where X is a continuously distributed endogenous variable, W_0 is a vector of exogenous variables, u and v are the unobservable error terms, $\beta_{\alpha,*} \equiv (\beta_{1,\alpha,*}, \beta_{2,\alpha,*}^\top)^\top$ are the unknown parameters for some $\alpha \in (0,1)$, $\varphi(w,\pi)$ is known up to π, $\pi_{\tilde{\alpha},*}$ is an unknown parameter for some $\tilde{\alpha} \in (0,1)$, $W \equiv (W_0^\top, W_1^\top)^\top$ and W_1 is a vector of excluded variables, $Q^{\tilde{\alpha}}_{v|W}(\cdot)$ denotes the conditional $\tilde{\alpha}$-quantile function of v given W. Due to the endogeneity of X, the quantile regression of Y on X and W_0 may inconsistently estimate $\beta_{\alpha,*}$.

To address the endogeneity issue, we assume that v is the control variable such that

$$Q^{\alpha}_{u|X,W_0,v}(x, w_0, v) = Q^{\alpha}_{u|v}(v), \tag{C.13}$$

where $Q^{\alpha}_{u|X,W_0,v}(\cdot)$ and $Q^{\alpha}_{u|v}(\cdot)$ denote the conditional α-quantile functions of u given $(X, W_0^\top, v)^\top$ and of u given v, respectively. Let $\lambda_{\alpha,*}(v) \equiv Q^{\alpha}_{u|v}(v)$ and $\varepsilon \equiv u - \lambda_{\alpha,*}(v)$. Then we can write (C.11) as

$$Y = X\beta_{1,\alpha,*} + W_0^\top \beta_{2,\alpha,*} + \lambda_{\alpha,*}(v) + \varepsilon. \tag{C.14}$$

By the definition of $\lambda_{\alpha,*}(v)$ and the restriction in (C.13),

$$Q^{\alpha}_{\varepsilon|X,W_0,v}(x, w_0, v) = Q^{\alpha}_{u|X,W_0,v}(x, w_0, v) - \lambda_{\alpha,*}(v) = 0, \tag{C.15}$$

where $Q^{\alpha}_{\varepsilon|X,W_0,v}(x,w_0,v)$ denotes the conditional α-quantile function of ε given $(X,W_0^{\top},v)^{\top}$. In view of (C.14) and (C.15), the finite dimensional parameter $\beta_{\alpha,*}$ is identified together with the unknown function $\lambda_{\alpha,*}(v)$ as the minimizer of the following problem

$$\min_{\beta_1,\beta_2,\lambda} \mathbb{E}\left[\rho_\alpha(Y - X\beta_1 - W_0^{\top}\beta_2 - \lambda(v))\right], \tag{C.16}$$

where $\rho_\alpha(\varepsilon) \equiv (\alpha - 1\{\varepsilon \leq 0\})\varepsilon$ for any $\varepsilon \in \mathbb{R}$ denotes the check function.

Estimation of $\beta_{\alpha,*}$ and $\lambda_{\alpha,*}(v)$ based on (C.16) is not feasible since $v = X - \varphi(W, \pi_{\tilde{\alpha},*})$ depends on the unknown $\pi_{\tilde{\alpha},*}$. We assume that there exists a preliminary estimator $\hat{\pi}$ of $\pi_{\tilde{\alpha},*}$. For example, Lee (2007) considers $\varphi(W,\pi_{\tilde{\alpha},*}) = W^{\top}\pi_{\tilde{\alpha},*}$ where $\pi_{\tilde{\alpha},*}$ is a finite dimensional parameter. Under this parametric specification, one can estimate $\pi_{\tilde{\alpha},*}$ through the quantile regression of X on W and estimate $v = X - W^{\top}\pi_{\tilde{\alpha},*}$ using the fitted residual. One may also consider a nonparametric specification

$$X = \pi_{\tilde{\alpha},*}(W) + v, \text{ where } Q^{\tilde{\alpha}}_{v|W}(w) = 0 \tag{C.17}$$

and estimate the conditional quantile function $\pi_{\tilde{\alpha},*}(W)$ nonparametrically. Given a random sample $\left\{(Y_i, X_i, W_i^{\top})^{\top}\right\}_{i=1}^{n}$ and the estimate $\hat{v}_i$ from the first step, $\beta_{\alpha,*}$ and $\lambda_{\alpha,*}(v)$ can be estimated, for example, by the semiparametric series quantile regression

$$(\hat{\beta}_\alpha^{\top}, \hat{\gamma}_\alpha^{\top})^{\top} \equiv \underset{\beta\in\mathbb{R}^{d_\beta},\gamma\in\mathbb{R}^{k_n}}{\arg\min} \sum_{i=1}^{n} \rho_\alpha(Y_i - (X_i, W_{0,i}^{\top})\beta - P_{k_n}(\hat{v}_i)^{\top}\gamma), \tag{C.18}$$

where $P_{k_n}(\cdot) \equiv (p_1(\cdot),\ldots,p_{k_n}(\cdot))^{\top}$ denotes the vector of k_n approximating functions. The main results established in the next two sections can be applied to derive the influence function of $\hat{\beta}_\alpha$ based on the above series method as well as other nonparametric (e.g., kernel) method.

It is worth noting that the identification condition (C.13) is imposed directly on the control variable $v(X,W,\pi_{\tilde{\alpha},*})$ which is a function of X and W. Therefore, (C.13) is different from, but implied by the following condition

$$Q^{\alpha}_{u|X,W}(x,w) = Q^{\alpha}_{u|v}(v), \tag{C.19}$$

where $Q^{\alpha}_{u|X,W}(x,w)$ denotes the conditional α-quantile function of u given $(X,W^{\top})^{\top}$, which is commonly maintained in the literature (see, e.g., Lee (2007)). As we shall see in the next section, the influence function of the estimator of $\beta_{\alpha,*}$ under (C.13) is different from that under (C.19). □

Example 3 (Sample Selection Model). Consider the sample selection model

$$\begin{aligned} Y^* &= m(X,\beta_*) + u, \\ v(X,W,\pi_*) &\equiv \mathbb{E}\left[d|\, X,W\right], \end{aligned} \tag{C.20}$$

where $d \in \{0,1\}$ is the indicator of selection, Y^* is the dependent variable which is observed only when $d = 1$, X is a vector of regressors, u is the unobservable residual term, W is a vector of explanatory variables and $v(X, W, \pi)$ denotes the conditional selection probability function known up to π. The function $m(x, \beta)$ is known up to β and β_* denotes the unknown parameter of interest. To achieve identification, we assume that

$$\mathbb{E}[u|X, v, d = 1] = \mathbb{E}[u|v, d = 1], \tag{C.21}$$

where $v \equiv v(X, W, \pi_*)$. A basic implication of model (C.20) and condition (C.21) is that

$$\mathbb{E}[Y^*|X, v, d = 1] = m(X, \beta_*) + \lambda_*(v), \text{ where } \lambda_*(v) = \mathbb{E}[u|v, d = 1]. \tag{C.22}$$

In (C.22), $\lambda_*(v)$ stands for the sample selection bias which takes different forms under different modeling assumptions. For example, Heckman (1976) assumes that the error terms in the outcome equation and the selection equation are jointly normally distributed. In this case, $\lambda_*(v)$ is the inverse of Mill's ratio. Newey (2009) relaxes the parametric assumption on the joint distribution of the error terms and models $\lambda_*(v)$ nonparametrically.

Let $\varepsilon \equiv u - \lambda_*(v)$. Then the structural equation in (C.20) can be written as

$$Y^* = m(X, \beta_*) + \lambda_*(v) + \varepsilon \tag{C.23}$$

where $\mathbb{E}[\varepsilon|X, v, d = 1] = 0$ by (C.21), which implies that the finite dimensional parameter β_* is identified together with the unknown function λ_* as the minimizer of the following problem

$$\min_{\beta,\lambda} \mathbb{E}\left[2^{-1}d\left|Y - m(X, \beta) - \lambda(v)\right|^2\right] \tag{C.24}$$

where $Y \equiv dY^*$.

To construct feasible estimators of β_* and $\lambda_*(v)$ based on (C.24), we assume that there exists a first-step estimator $\hat{\pi}$ of π_* such that v is estimated by $\hat{v} \equiv v(X, W, \hat{\pi})$. Given a random sample $\{(Y_i, d_i, X_i^\top, W_i^\top)^\top\}_{i=1}^n$ and the estimate $\hat{v}_i$ from the first step, β_* and $\lambda_*(v)$ can be estimated, for example by the semiparametric series regression

$$(\hat{\beta}^\top, \hat{\gamma}^\top)^\top \equiv \arg\min_{\beta\in\mathbb{R}^{d_\beta},\gamma\in\mathbb{R}^{k_n}} \sum_{i=1}^n 2^{-1}d_i\left|Y_i - m(X_i, \beta) - P_{k_n}(\hat{v}_i)^\top\gamma\right|^2 \tag{C.25}$$

where $P_{k_n}(\cdot) \equiv (p_1(\cdot), \ldots, p_{k_n}(\cdot))^\top$ denotes the vector of k_n approximating functions. The main results established in the paper can be applied to derive the influence function of $\hat{\beta}$ based on the above series method, or other nonparametric (e.g., kernel) methods.

In the literature, the function $m(x, \beta_*)$ is usually assumed to be linear, i.e., $m(x, \beta_*) \equiv x^\top\beta_*$ (see, e.g., Heckman (1976) and Newey (2009)), and π_* is a finite dimensional parameter which is estimated by parametric methods such as Probit (see, e.g., Heckman (1976)) or semiparametric methods (see, e.g., Powell et al. (1989), Ichimura (1993), and Cavanagh and Sherman (1998)). The theory established in this paper allows for parametric, semiparametric and nonparametric first-step estimation of π_*, and it can be applied to derive the influence function of $\hat{\beta}$ under the index restriction (C.21) which is implied by the condition

$$\mathbb{E}[u|X, W, d = 1] = \mathbb{E}[u|v, d = 1], \tag{C.26}$$

employed in the literature such as Newey (2009).[1]

D The Influence Function in the Three Examples

In this section, we provide the influence functions of the two-step estimators discussed in the three examples of Section C. In view of Theorem 1 and Theorem 2 in Hahn et al. (2023), it is sufficient to calculate the quantities $\varphi_\beta(Z)$, $\varphi_\pi(Z)$, $\delta_\pi(W)$, $\Psi_{\beta,\pi}$ and $\Psi_{\beta,\beta}$ in these examples.

Example 1 (Mean Regression Continued). For ease of notation, we let $Z_0 \equiv (X, W_0^\top)^\top$. In this example, we have

$$\psi\left(Z, \beta, \lambda(v(\pi))\right) = 2^{-1}\left(Y - m(Z_0, \beta) - \lambda(v(\pi))\right)^2. \tag{D.1}$$

Using the above expression, it is easy to calculate that

$$\psi_\lambda\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right) = -(Y - m(Z_0, \beta) - \lambda\left(v\left(\pi\right)\right)), \tag{D.2}$$

$$\psi_\beta\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right) = -\left(Y - m(Z_0, \beta) - \lambda\left(v\left(\pi\right)\right)\right) m_\beta(Z_0, \beta), \tag{D.3}$$

$$\psi_{\lambda,\beta}\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right) = m_\beta(Z_0, \beta) = \psi_{\beta,\lambda}\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right), \tag{D.4}$$

$$\psi_{\beta,\beta}\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right) = m_\beta(Z_0, \beta) m_\beta(Z_0, \beta)^\top - \left(Y - m(Z_0, \beta) - \lambda\left(v\left(\pi\right)\right)\right) m_{\beta,\beta}(Z_0, \beta), \tag{D.5}$$

$$\psi_{\lambda,\lambda}\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right) = 1, \tag{D.6}$$

for any function $\lambda\left(v\left(\pi\right)\right)$ of $v\left(\pi\right)$, where

$$m_\beta(Z_0, \beta) \equiv \frac{\partial m(Z_0, \beta)}{\partial \beta} \text{ and } m_{\beta,\beta}(Z_0, \beta) \equiv \frac{\partial^2 m(Z_0, \beta)}{\partial \beta \partial \beta^\top}. \tag{D.7}$$

By (D.2), the first-order condition of the profiled nonparametric function $h\left(v\left(\pi\right); \beta, \pi\right)$ can be written as

$$\mathbb{E}\left[\left(Y - m(Z_0, \beta) - h\left(v\left(\pi\right); \beta, \pi\right)\right) \lambda\left(v\left(\pi\right)\right)\right] = 0$$

for any function $\lambda\left(v\left(\pi\right)\right)$ of $v\left(\pi\right)$ and any β, which immediately implies that in this example

$$h\left(v\left(\pi\right); \beta, \pi\right) = \mathbb{E}\left[Y - m(Z_0, \beta)\middle| v\left(\pi\right)\right], \tag{D.8}$$

[1] A similar condition is employed in Ahn and Powell (1993) (see, their condition (2.3)). The model studied here does not strictly nest that in Ahn and Powell (1993), since they also allow X to be endogenous. On the other hand, the influence function derived in this paper applies to Ahn and Powell (1993) when X is exogenous.

and therefore

$$h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)=\mathbb{E}\left[u|\,v\right]=\lambda_*(v). \tag{D.9}$$

Let $m_\beta(Z_0)\equiv m_\beta(Z_0,\beta_*)$. Using the expressions in (D.2)-(D.6) and (D.9), we get

$$\varphi_\beta(Z)=-\varepsilon(m_\beta(Z_0)-\mathbb{E}\left[m_\beta(Z_0)|v\right]), \tag{D.10}$$

$$\delta_\beta(Z)=(m_\beta(Z_0)-\mathbb{E}\left[m_\beta(Z_0)|\,v\right])\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v}, \tag{D.11}$$

$$\delta_g(Z)=-\varepsilon\frac{\partial\mathbb{E}\left[m_\beta(Z_0)|\,v\right]}{\partial v}\text{ and} \tag{D.12}$$

$$\Psi_{\beta,\pi}=\mathbb{E}\left[\left((m_\beta(Z_0)-\mathbb{E}\left[m_\beta(Z_0)|\,v\right])\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v}+\varepsilon\frac{\partial\mathbb{E}\left[m_\beta(Z_0)|\,v\right]}{\partial v}\right)\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right] \tag{D.13}$$

when π_* is a finite-dimensional parameter vector. From (C.5), (D.2)-(D.6), the Hessian matrix takes the following form

$$\Psi_{\beta,\beta}=-\mathbb{E}\left[m_\beta(Z_0)m_\beta(Z_0)^\top-\mathbb{E}\left[m_\beta(Z_0)|\,v\right]\mathbb{E}\left[m_\beta(Z_0)|\,v\right]^\top\right]. \tag{D.14}$$

From the components in (D.10), (D.13) and (D.14), the influence function of $\hat{\beta}$, when the influence function of the estimator $\hat{\pi}$ of π_* is $\varphi_\pi(Z)$, can be readily computed using Theorem 1 in Hahn et al. (2023).

When the control variable $v\left(\pi_*\right)$ is nonparametrically specified as the residual in the reduced form, i.e.,

$$v\left(\pi_*\right)=X-\pi_*(W),$$

where $\pi_*(W)\equiv\mathbb{E}\left[X|\,W\right]$, the general residual function is $\mu(Z,\pi_*)=X-\pi_*(W)$. In this case, it is easy to calculate that

$$\varphi_\pi(Z)=X-\pi_*(W) \tag{D.15}$$

and

$$\delta_\pi(W)=-\mathbb{E}\left[\delta_\beta(Z)-\delta_g(Z)|\,W\right], \tag{D.16}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (D.11) and (D.12) respectively. Using the components in (D.10), (D.14), (D.15) and (D.16), Theorem 2 of Hahn et al. (2023) implies that the influence function of the two-step estimator $\hat{\beta}$ is

$$\Psi_{\beta,\beta}^{-1}\left(\varphi_\beta(Z)+\delta_\pi(W)(X-\pi_*(W))\right).$$

If condition (C.9) holds, we obtain

$$\Psi_{\beta,\pi}=\mathbb{E}\left[(m_\beta(Z_0)-\mathbb{E}\left[m_\beta(Z_0)|\,v\right])\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v}\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right] \tag{D.17}$$

in the case with a parametric first step, and

$$\delta_\pi(W)=-\mathbb{E}\left[(m_\beta(Z_0)-\mathbb{E}\left[m_\beta(Z_0)|\,v\right])\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v}\middle|\,W\right] \tag{D.18}$$

in the case with a nonparametric first step. Therefore, the influence function of $\hat{\beta}$ is slightly simplified in both cases. Li and Wooldridge (2002) impose (C.9) and assume that $m(Z_0, \beta) = W_0^\top \beta$ and $v(\pi_*) = X - W^\top \pi_*$ to derive the main results. Under these extra conditions,

$$\varphi_\beta(Z) = -\varepsilon \left(W_0 - \mathbb{E}\left[W_0 | v\right]\right), \tag{D.19}$$

$$\Psi_{\beta,\pi} = -\mathbb{E}\left[\left(W_0 - \mathbb{E}\left[W_0 | v\right]\right) \frac{\partial \mathbb{E}\left[u|v\right]}{\partial v} W^\top\right], \tag{D.20}$$

$$\Psi_{\beta,\beta} = -\mathbb{E}\left[W_0 W_0^\top - \mathbb{E}\left[W_0 | v\right] \mathbb{E}\left[W_0 | v\right]^\top\right]. \tag{D.21}$$

The influence function of the two-step estimator $\hat{\beta}$ can be calculated using Theorem 1 of Hahn et al. (2023), the items in (D.19)-(D.21) and the influence function $\varphi_\pi(Z)$ from the first-step estimation of π_*. In this case, the influence function implies the same asymptotic variance-covariance matrix of the trimming-based estimator proposed in Li and Wooldridge (2002) as indicated in their Conjecture 2.1. □

Example 2 (Quantile Regression Continued). For ease of notation, we suppress the dependence of β and λ on α, and of π on $\tilde{\alpha}$. Let $Z_0 \equiv (X, W_0^\top)^\top$ and $\beta \equiv (\beta_1, \beta_2^\top)^\top$. In this example, we have

$$\psi\left(Z, \beta, \lambda(v(\pi))\right) = \rho_\alpha(Y - Z_0^\top \beta - \lambda(v)). \tag{D.22}$$

Using the above expression, it is easy to calculate that

$$\left.\frac{\partial \mathbb{E}\left[\rho_\alpha(Y - Z_0^\top \beta - \lambda\left(v\left(\pi\right)\right) - \tau \lambda_1\left(v\left(\pi\right)\right))\right]}{\partial \tau}\right|_{\tau=0}$$
$$= \mathbb{E}\left[\left(1\left\{Y \leq Z_0^\top \beta + \lambda\left(v\left(\pi\right)\right)\right\} - \alpha\right) \lambda_1\left(v\left(\pi\right)\right)\right] \tag{D.23}$$

for any functions $\lambda\left(v\left(\pi\right)\right)$ and $\lambda_1\left(v\left(\pi\right)\right)$ of $v\left(\pi\right)$, which implies that

$$\psi_\lambda\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right) = 1\left\{Y \leq Z_0^\top \beta + \lambda\left(v\left(\pi\right)\right)\right\} - \alpha. \tag{D.24}$$

Applying the above expression to the first-order condition (3) in Hahn et al. (2023), we see that $h\left(v\left(\pi\right); \beta, \pi\right)$ is the conditional α-quantile function of $Y - Z_0^\top \beta$ given $v\left(\pi\right)$, and therefore[2]

$$h\left(v\left(\pi_*\right); \beta_*, \pi_*\right) = Q_{u|v}^\alpha\left(v\right) = \lambda_*(v). \tag{D.25}$$

Let $f_\varepsilon(\cdot|Z_0, v\left(\pi\right))$ denote the conditional density function of ε given $(Z_0^\top, v\left(\pi\right))^\top$. By (D.22), it is

[2]To make the notation consistent to Thereom 1, we suppress the dependence of β_* and $\lambda_*(v)$ on α.

easy to calculate that

$$\psi_{\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=\left(1\left\{Y\leq Z_0^{\top}\beta+\lambda\left(v\left(\pi\right)\right)\right\}-\alpha\right)Z_0, \tag{D.26}$$

$$\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=f_{\varepsilon}\left(Z_0^{\top}(\beta-\beta_*)+\lambda\left(v\left(\pi\right)\right)-\lambda_*(v)\middle|\,Z_0,v\left(\pi\right)\right)Z_0, \tag{D.27}$$

$$\psi_{\beta,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=f_{\varepsilon}\left(Z_0^{\top}(\beta-\beta_*)+\lambda\left(v\left(\pi\right)\right)-\lambda_*(v)\middle|\,Z_0,v\left(\pi\right)\right)Z_0Z_0^{\top}, \tag{D.28}$$

$$\psi_{\lambda,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=f_{\varepsilon}\left(Z_0^{\top}(\beta-\beta_*)+\lambda\left(v\left(\pi\right)\right)-\lambda_*(v)\middle|\,Z_0,v\left(\pi\right)\right), \tag{D.29}$$

and $\psi_{\beta,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)$ for any function $\lambda\left(v\left(\pi\right)\right)$ of $v\left(\pi\right)$.

Using (D.24)-(D.29), we get

$$\varphi_{\beta}(Z)=\left(1\left\{\varepsilon\leq 0\right\}-\alpha\right)\left(Z_0-g_*(v)\right), \tag{D.30}$$

$$\delta_{\beta}(Z)=f_{\varepsilon}\left(0|Z_0,v\right)\left(Z_0-g_*(v)\right)\frac{\partial Q^{\alpha}_{u|v}\left(v\right)}{\partial v}, \tag{D.31}$$

$$\delta_{g}(Z)=\left(1\left\{\varepsilon\leq 0\right\}-\alpha\right)\frac{\partial g_*\left(v\right)}{\partial v},\text{ and} \tag{D.32}$$

$$\Psi_{\beta,\pi}=\mathbb{E}\left[\left(f_{\varepsilon}\left(0|Z_0,v\right)\left(Z_0-g_*(v)\right)\frac{\partial Q^{\alpha}_{u|v}\left(v\right)}{\partial v}-\left(1\left\{\varepsilon\leq 0\right\}-\alpha\right)\frac{\partial g_*\left(v\right)}{\partial v}\right)\frac{\partial v\left(\pi_*\right)}{\partial\pi^{\top}}\right] \tag{D.33}$$

when π_* is a finite-dimensional parameter vector, where

$$g_*\left(v\right)=\frac{\mathbb{E}\left[f_{\varepsilon}\left(0|Z_0,v\right)Z_0|\,v\right]}{\mathbb{E}\left[f_{\varepsilon}\left(0|Z_0,v\right)|\,v\right]}.$$

From (D.25)-(D.29), the Hessian matrix takes the following form

$$\Psi_{\beta,\beta}=-\mathbb{E}\left[f_{\varepsilon}\left(0|Z_0,v\right)\left(Z_0Z_0^{\top}-g_*(v)g_*(v)^{\top}\right)\right]. \tag{D.34}$$

Using the components in (D.30), (D.33) and (D.34), the influence function of $\hat{\beta}$, when the influence function of the estimator $\hat{\pi}$ of π_* is $\varphi_{\pi}(Z)$, can be readily computed using Theorem 1 in Hahn et al. (2023).

When the control variable $v\left(\pi_*\right)$ is nonparametrically specified as the residual from the reduced form, i.e.,

$$v\left(\pi_*\right)=X-\pi_*(W),$$

where $\pi_*(w)=Q^{\tilde{\alpha}}_{X|W}\left(w\right)$ denotes the conditional $\tilde{\alpha}$-quantile function of X given W for some $\tilde{\alpha}\in(0,1)$, the first-stage residual function becomes

$$\mu(Z,\pi_*)=1\left\{X\leq\pi_*(W)\right\}-\tilde{\alpha}.$$

Therefore in this case,

$$\varphi_{\pi}(Z)=-\frac{1\left\{X\leq\pi_*(W)\right\}-\tilde{\alpha}}{f_{X|W}(\pi_*(W))}, \tag{D.35}$$

where $f_{X|W}(\cdot)$ denotes the conditional density of X given W, and

$$\delta_\pi(W) = -\mathbb{E}\left[\,\delta_\beta(Z) - \delta_g(Z)|\,W\right], \tag{D.36}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (D.31) and (D.32) respectively. Using the components in (D.30), (D.34), (D.35) and (D.36), Theorem 2 of Hahn et al. (2023) implies that the influence function of the two-step estimator in this case is

$$\Psi^{-1}_{\beta,\beta}\left(\varphi_\beta(Z) - \delta_\pi(W)\frac{1\{X_i \leq \pi_*(W)\} - \tilde{\alpha}}{f_{X|W}(\pi_*(W))}\right). \tag{D.37}$$

If condition (C.19) holds, we get

$$\Psi_{\beta,\pi} = \mathbb{E}\left[f_\varepsilon\left(0|Z_0, v\right)\left(Z_0 - g_*(v)\right)\frac{\partial Q^\alpha_{u|v}\left(v\right)}{\partial v}W^\top\right] \tag{D.38}$$

in the case with the parametric first step $v(\pi_*) = X - W^\top\pi_*$, and

$$\delta_\pi(W) = -\mathbb{E}\left[f_\varepsilon\left(0|Z_0, v\right)\left(Z_0 - g_*(v)\right)\frac{\partial Q^\alpha_{u|v}\left(v\right)}{\partial v}\right| W\right] \tag{D.39}$$

in the case with a nonparametric first step. Therefore the influence function of $\hat{\beta}$ is slightly simplified in both cases. Moreover, under (C.12) and (C.19), the asymptotic variance of $\hat{\beta}$ implied by its influence function (which can be calculated using (D.30), (D.34) and (D.38)) is similar to the one stated in Theorem 3.1 in Lee (2007). □

Example 3 (Sample Selection Model Continued). In this example, we have

$$\psi\left(Z, \beta, \lambda(v(\pi))\right) = 2^{-1}d\left(Y - m(X, \beta) - \lambda(v(\pi))\right)^2. \tag{D.40}$$

It is easy to calculate that

$$\psi_\lambda\left(Z, \beta, \lambda(v(\pi))\right) = -d\left(Y - m(X, \beta) - \lambda(v(\pi))\right), \tag{D.41}$$

$$\psi_\beta\left(Z, \beta, \lambda(v(\pi))\right) = -d\left(Y - m(X, \beta) - \lambda(v(\pi))\right)m_\beta(X, \beta), \tag{D.42}$$

$$\psi_{\beta,\beta}\left(Z, \beta, \lambda(v(\pi))\right) = dm_\beta(X, \beta)m_\beta(X, \beta)^\top - d\left(Y - m(X, \beta) - \lambda(v(\pi))\right)m_{\beta,\beta}(X, \beta), \tag{D.43}$$

$$\psi_{\lambda,\beta}\left(Z, \beta, \lambda(v(\pi))\right) = dm_\beta(X, \beta)\ = \psi_{\beta,\lambda}\left(Z, \beta, \lambda(v(\pi))\right), \text{ and} \tag{D.44}$$

$$\psi_{\lambda,\lambda}\left(Z, \beta, \lambda(v(\pi))\right) = d, \tag{D.45}$$

for any function $\lambda(v(\pi))$ of $v(\pi)$, where

$$m_\beta(X, \beta) \equiv \frac{\partial m(X, \beta)}{\partial \beta} \text{ and } m_{\beta,\beta}(X, \beta) \equiv \frac{\partial^2 m(X, \beta)}{\partial \beta \partial \beta^\top}.$$

By (D.41), the first-order condition of the profiled nonparametric function $h(v(\pi);\beta,\pi)$ can be written as

$$\mathbb{E}\left[d\left(Y-m(X,\beta)-h(v(\pi);\beta,\pi)\right)\lambda(v(\pi))\right]=0,$$

which implies that in this example

$$h(v(\pi);\beta,\pi)=\frac{\mathbb{E}\left[d\left(Y-m(X,\beta)\right)\middle|\,v(\pi)\right]}{\mathbb{E}\left[d\middle|\,v(\pi)\right]}=\mathbb{E}\left[Y-m(X,\beta)\middle|\,v(\pi),d=1\right],$$

where the second equality is by

$$\begin{aligned}\mathbb{E}\left[d\left(Y-m(X,\beta)\right)\middle|\,v(\pi)\right]&=\mathbb{E}\left[d\mathbb{E}\left[Y-m(X,\beta)\middle|\,v(\pi),d\right]\middle|\,v(\pi)\right]\\&=\mathbb{E}\left[Y-m(X,\beta)\middle|\,v(\pi),d=1\right]\mathbb{E}\left[d\middle|\,v(\pi)\right].\end{aligned}$$

Recall that $v\equiv v(\pi_*)$, therefore

$$h(v(\pi_*);\beta_*,\pi_*)=\mathbb{E}\left[u\middle|\,v,d=1\right]=\lambda_*(v)$$

by the definition of $\lambda_*(v)$.

Let $m_\beta(X)\equiv m_\beta(X,\beta_*)$. By (7) in Hahn et al. (2023), (D.44) and (D.45), we get

$$g_*(v)=\frac{\mathbb{E}\left[dm_\beta(X)|v\right]}{\mathbb{E}\left[d|v\right]}=\mathbb{E}\left[m_\beta(X)|v,d=1\right], \tag{D.46}$$

where the second equality is by

$$\mathbb{E}\left[dm_\beta(X)|v\right]=\mathbb{E}\left[d\mathbb{E}\left[m_\beta(X)|v,d\right]|v\right]=\mathbb{E}[m_\beta(X)|v,d=1]\mathbb{E}\left[d|v\right].$$

By (7), (9) - (12) in Hahn et al. (2023), and (D.40)-(D.46), we have

$$\varphi_\beta(Z)=-d\varepsilon\left(m_\beta(X)-\mathbb{E}\left[m_\beta(X)|v,d=1\right]\right), \tag{D.47}$$

$$\delta_\beta(Z)=d\left[m_\beta(X)-\mathbb{E}\left[m_\beta(X)|v,d=1\right]\right]\frac{\partial\mathbb{E}\left[u|\,v,d=1\right]}{\partial v}, \tag{D.48}$$

$$\delta_g(Z)=-d\varepsilon\frac{\partial\mathbb{E}\left[m_\beta(X)|v,d=1\right]}{\partial v}, \tag{D.49}$$

$$\Psi_{\beta,\pi}=\mathbb{E}\left[\left(\delta_\beta(Z)-\delta_g(Z)\right)\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right],\text{ and} \tag{D.50}$$

$$\Psi_{\beta,\beta}=-\mathbb{E}\left[d\left(m_\beta(X)m_\beta(X)^\top-\mathbb{E}\left[m_\beta(X)|v,d=1\right]\mathbb{E}\left[m_\beta(X)|v,d=1\right]^\top\right)\right] \tag{D.51}$$

when π_* is parametrically specified. Using the components in (D.47), (D.50) and (D.51), the influence function of $\hat{\beta}$ in this example can be readily computed using Theorem 1 of Hahn et al. (2023).

When π_* is nonparametrically specified, $\pi_*(X,W)=\mathbb{E}\left[d|\,X,W\right]$. In this case $v(X,W,\pi_*)=\pi_*(X,W)$ and the general residual function in the first step is

$$\mu(Z,\pi_*)=d-\pi_*(X,W).$$

Therefore in this case

$$\varphi_\pi(Z) = d - \pi_*(X, W) \tag{D.52}$$

and

$$\delta_\pi(W) \equiv \mathbb{E}\left[\delta_\beta(Z) - \delta_g(Z) \middle| X, W\right], \tag{D.53}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (D.48) and (D.49) respectively. Using the components in (D.47), (D.51), (D.52) and (D.53), Theorem 2 of Hahn et al. (2023) implies that the influence function of the two-step estimator in this case is

$$\Psi_{\beta,\beta}^{-1}\left(\varphi_\beta(Z) + \delta_\pi(W)(d - \pi_*(X, W))\right).$$

When the identification condition (C.26) holds,

$$\mathbb{E}[\varepsilon|X, W, d = 1] = \mathbb{E}[u|X, W, d = 1] - \lambda_*(v) = \mathbb{E}[u|v, d = 1] - \lambda_*(v) = 0,$$

which immediately implies that

$$\mathbb{E}\left[d\varepsilon \frac{\partial \mathbb{E}\left[m_\beta(X, \beta_*)|v, d = 1\right]}{\partial v} \frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}\right] = 0 \tag{D.54}$$

in the parametric case since $\partial v\left(\pi_*\right)/\partial \pi^\top$ is a function of $(X^\top, W^\top)^\top$, and

$$\mathbb{E}\left[d\varepsilon \frac{\partial \mathbb{E}\left[m_\beta(X)|v, d = 1\right]}{\partial v} \middle| X, W\right] = 0$$

in the nonparametric case. Therefore the influence function of $\hat{\beta}$ is slightly simplified in both cases. Moreover, in the parametric case, if one further assumes that $m(x, \beta_*) = x^\top \beta_*$ and the influence function from the first-step estimation is $\varphi_\pi(Z)$, the influence function computed using Theorem 1, and the items in (D.47), (D.50), (D.51) and (D.54) becomes identical to that in Newey (2009). □

References

Ahn, Hyungtaik and James L Powell (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," Journal of Econometrics, 58 (1-2), 3–29.

Cavanagh, Christopher and Robert P Sherman (1998): "Rank estimators for monotonic index models," Journal of Econometrics, 84 (2), 351–382.

Hahn, Jinyong, Zhipeng Liao, Geert Ridder, and Ruoyao Shi (2023): "The influence function of semiparametric two-step estimators with estimated control variables," forthcoming at Economics Letters.

Heckman, James J (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," in Annals of Economic and Social Measurement, NBER, vol. 5, 475–492.

Ichimura, Hidehiko (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," Journal of Econometrics, 58 (1-2), 71–120.

Lee, Sokbae (2007): "Endogeneity in quantile regression models: A control function approach," Journal of Econometrics, 141 (2), 1131–1158.

Li, Qi and Jeffrey M Wooldridge (2002): "Semiparametric estimation of partially linear models for dependent data with generated regressors," Econometric Theory, 18 (3), 625–645.

Newey, Whitney K (1994): "The asymptotic variance of semiparametric estimators," Econometrica, 62 (6), 1349–1382.

——— (2009): "Two-step series estimation of sample selection models," The Econometrics Journal, 12, S217–S229.

Powell, James L, James H Stock, and Thomas M Stoker (1989): "Semiparametric estimation of index coefficients," Econometrica, 1403–1430.