The Influence Function of Semiparametric Two-step Estimators with Estimated Control Variables

Jinyong Hahn[*]   Zhipeng Liao[†]   Geert Ridder[‡]   Ruoyao Shi[§]

UCLA   UCLA   USC   UC Riverside

May 10, 2021

Abstract

This paper studies semiparametric two-step estimators with a control variable estimated in a first-step parametric or nonparametric model. We provide the explicit influence function of the two-step estimator under an index restriction which is imposed directly on the unknown control variable. The index restriction is weaker than the commonly used identification conditions in the literature, which are imposed on all exogenous variables. An extra term shows up in the influence function of the semiparametric two-step estimator under the weaker identification condition. We illustrate our influence function formula in a mean regression example, a quantile regression example, and a sample selection example where the control variable approach is applied for identification and consistent estimation of structural parameters.

JEL Classification: C14, C31, C32

Keywords: Control Variable Approach; Generated Regressors; Influence Function; Semiparametric Two-step Estimation

1 Introduction

An attractive identification strategy if one or more regressors are endogenous in an econometric model is to use a moment restriction that conditions on (and averages over) control variables. These control variables typically need to be estimated in a first stage as the residuals in a parametric or nonparametric relation between the endogenous regressors and instruments. In a second step a conditional expectation of the dependent variable of the model on the endogenous regressors

---

[*]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.

[§]Department of Economics, UC Riverside, 3136 Sproul Hall, Riverside, CA, 92521. Email: ruoyao.shi@ucr.edu

and the control variables can be estimated nonparametrically as in Imbens and Newey (2009), and this conditional expectation can be averaged over the control variables to obtain the Average Structural Function (ASF). However in applied work, parametric or semiparametric specifications along the line of Rivers and Vuong (1988) or Blundell and Powell (2004) are likely to be adopted, and it is of interest to understand how statistical (asymptotic) inference about the estimated finite-dimensional parameters should be implemented. Li and Wooldridge (2002), Lee (2007), and Newey (2009) are some of the well-known papers that established the asymptotic distribution of such two-step estimators for some specific models. These papers all consider a second step specification which takes the form of a partially linear regression model, where the estimated control variable enters as an argument of a nonparametric function.

The purpose of our paper is to develop a unified framework to understand inferential issues arising from such two-step estimation. We are interested in estimating a finite dimensional vector of parameters $\beta_* \in \mathbb{R}^k$, which is identified together with an unknown function $\lambda_*(\cdot)$ as the unique solution of a minimization problem. That is,

$$(\beta_*, \lambda_*) \equiv \arg\min_{\beta, \lambda} \mathbb{E}\left[\psi(Z, \beta, \lambda(v(\pi_*)))\right], \tag{1}$$

where Z is a vector of all observable variables, $v(\pi) \equiv v(Z, \pi)$ is the control variable that is known up to π_* and π_* is a finite-dimensional parameter or a vector of unknown functions identified outside the model in a first-stage. The $v(\cdot)$ is determined by the procedure used to generate the control variable that enters as an argument of the nonparametric part $\lambda_*(\cdot)$. The criterion function $\psi(z, \beta, \lambda)$ is known.

We make three technical contributions. First, we consider criterion functions $\psi(z, \beta, \lambda)$ that are general enough to nest many specific models, such as the nonlinear regression and the quantile regression models, as special cases, providing a unified framework to understand the inferential problems. Second, we relax certain moment restrictions that were imposed in the previous literature. The previous literature assumed that the "error" in the second step is (mean or quantile) independent of the endogenous regressor given a set of instruments, whereas we relax the assumption and impose conditional independence given just the control variable. For example, let's consider a model $Y = X\beta_* + \lambda_*(v) + \varepsilon$, where $X = W^\top \pi_* + v$. A common assumption in the literature is $\mathbb{E}\left[\varepsilon | X, W\right] = 0$, which is stronger than $\mathbb{E}\left[\varepsilon | X, v\right] = 0$. We adopt variants of the latter assumption, and derive the asymptotic distribution under the weaker assumption. We show that the weaker assumption does make a difference in the asymptotic distribution. Although the stronger condition $\mathbb{E}\left[\varepsilon | X, W\right] = 0$ is commonly used in the early literature, recent applications of the control variable approach relax this condition and impose the weaker conditional independence restriction given the control variable to achieve identification (see, e.g., Johnsson and Moon (2015) and Auerbach (2019)). Therefore, our results can be applied to derive the asymptotic distribution of the two-step estimators in these recent works. Third, we follow Newey's (1994) path-derivative

calculations to characterize the influence function that takes account of the estimation noise of the control variable, and therefore, our result is invariant to the specific method of nonparametric estimation in the second step.

The rest of the paper is organized as follows. Section 2 gives several examples where the control variable approach can be applied to identify and estimate the parameters of interest. Section 3 derives the influence function of the semiparametric two-step estimator when π_* is estimated in a parametric first step. Section 4 extends the result in Section 3 and provides the influence function of the semiparametric two-step estimator when π_* is estimated by a nonparametric first-step estimation. Section 5 applies the influence function formula established in Sections 3 and 4 to the examples discussed in Section 2. Section 6 concludes. The appendix offers proofs.

Notation. For any real matrix A, we use $A^\top$ to denote the transpose of A. We use a_j to denote the jth component of a vector a. For any multivariate function $f(\cdot) : \mathbb{R}^{d_x} \longmapsto \mathbb{R}$, we use $\partial f(x)/\partial x$ to denote the $d_x \times 1$ vector $(\partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_{d_x})^\top$, $\partial f(x)/\partial x^\top$ to denote the transpose of $\partial f(x)/\partial x$ and $\partial^2 f(x)/\partial x \partial x^\top$ to denote the $d_x \times d_x$ matrix whose ith row and jth column component is $\partial^2 f(x)/\partial x_i \partial x_j$ for any $i, j = 1, \ldots, d_x$. For any multivariate vector-valued function $f(\cdot) : \mathbb{R}^{d_x} \longmapsto \mathbb{R}^{d_f}$, we use $\partial f(x)/\partial x^\top$ to denote the $d_f \times d_x$ matrix whose ith row and jth column component is $\partial f_i(x)/\partial x_j$ for any $i = 1, \ldots, d_f$ and any $j = 1, \ldots, d_x$. We use $A \equiv B$ to denote that A is defined as B.

2 Examples

In this section, we provide several examples where the control variable approach is applied to identify and to estimate the parameters of interest. The main theory established in the next two sections 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

$$Y = m(X, W_0, \beta_*) + u, \tag{2}$$

where Y is the dependent variable, X is an endogenous regressor and W_0 is a vector of exogenous regressors, u is the unobservable residual term, the function $m(x, w_0, \beta)$ is known up to β and β_* denotes the unknown parameter of interest. To achieve identification of β_*, we assume there exists a control variable $v \equiv v(X, W, \pi_*)$, where $W \equiv (W_0^\top, W_1^\top)^\top$ and $v(x, w, \pi)$ is known up to π, such that

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

The above condition is imposed on the control variable $v(X, W, \pi_*)$ which is an "index" function of X and W. Therefore, this condition is weaker than the condition (10) below which is imposed on X and W.

Let $\lambda_*(v) \equiv \mathbb{E}\left[u|\,v\right]$ and $\varepsilon \equiv u - \lambda_*(v)$. Then we can write (2) as

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

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

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

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

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

To construct a feasible estimator of unknown parameters β_* and λ_* based on (6), we assume that there exists a first-step estimator $\hat{\pi}$ of π_* such that v_i is estimated by $\hat{v}_i \equiv v(X_i, W_i, \hat{\pi})$. For example, the first step could be a non-linear regression of the reduced form

$$X_i = \varphi(W_i, \pi_*) + v_i, \text{ where } \mathbb{E}\left[v_i|\,W_i\right] = 0 \tag{7}$$

where $\varphi(w, \pi)$ is known up to π. In this case, we have $v(X_i, W_i, \pi) \equiv X_i - \varphi(W_i, \pi)$, $\hat{\pi}$ is the non-linear regression estimator, and $\hat{v}_i$ is the fitted residual. The first step could also be a nonparametric regression of the reduced form

$$X_i = \pi_*(W_i) + v_i, \text{ where } \mathbb{E}\left[v_i|\,W_i\right] = 0. \tag{8}$$

In this case, we have $v(X_i, W_i, \pi) \equiv X_i - \pi(W_i)$, $\hat{\pi}$ is the nonparametric regression estimator of X_i on W_i, and $\hat{v}_i$ 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 the semiparametric two-step mean regression

$$(\hat{\beta}, \hat{\lambda}) \equiv \arg\min_{\beta, \lambda} \sum_{i=1}^n 2^{-1}\,|Y_i - m(X_i, W_{0,i}, \beta) - \lambda(\hat{v}_i)|^2. \tag{9}$$

The main result established in the next two sections can be applied to derive the influence function of $\hat{\beta}$.

This example nests the model studied in Li and Wooldridge (2002), where the control variable v_i is parametrically specified. More importantly, the identification condition (3) is weaker than the condition

$$\mathbb{E}\left[u|\,X, W\right] = \mathbb{E}\left[u|\,v\right], \tag{10}$$

which implies that

$$\mathbb{E}\left[\varepsilon|\,X, W\right] = 0. \tag{11}$$

Li and Wooldridge (2002) derive the root-n asymptotic normality of the two-step estimator under (11). As we shall see in Section 5, the influence function and the asymptotic variance of the two-step estimator $\hat{\beta}$ are different under the weaker identification condition in (3). $\qquad\square$

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

$$Y = X\theta_{\alpha,*} + W_0^\top \gamma_{\alpha,*} + u, \qquad (12)$$

where X is a continuously distributed endogenous variable, W_0 is a vector of exogenous variable, u is the unobservable error term, $\theta_{\alpha,*}$ and $\gamma_{\alpha,*}$ are the unknown parameters for some $\alpha \in (0,1)$. Due to the endogeneity of X, the quantile regression of Y on X and W_0 may produce inconsistent estimates of $\theta_{\alpha,*}$ and $\gamma_{\alpha,*}$.

To address the endogeneity issue, we assume that there exist variables W_1 excluded from (12) and a control variable $v \equiv v(X, W, \pi_*)$, where $v(x, w, \pi)$ is known up to π and $W \equiv (W_0^\top, W_1^\top)^\top$, such that

$$Q_{u|X,W_0,v}^\alpha(x, w_0, v) = Q_{u|v}^\alpha(v), \qquad (13)$$

where $Q_{u|X,W_0,v}^\alpha(x, w_0, v)$ and $Q_{u|v}^\alpha(v)$ denote the conditional α-quantile functions of u given $(X, W_0^\top, v)^\top$ and of u given v, respectively. Let $\lambda_{\alpha,*}(v) \equiv Q_{u|v}^\alpha(v)$ and $\varepsilon \equiv u - \lambda_{\alpha,*}(v)$. Then we can write (12) as

$$Y = X\theta_{\alpha,*} + W_0^\top \gamma_{\alpha,*} + \lambda_{\alpha,*}(v) + \varepsilon. \qquad (14)$$

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

$$Q_{\varepsilon|X,W_0,v}^\alpha(x, w_0, v) = Q_{u|X,W_0,v}^\alpha(x, w_0, v) - \lambda_{\alpha,*}(v) = 0, \qquad (15)$$

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

$$\min_{\theta,\gamma,\lambda} \mathbb{E}\left[\rho_\alpha(Y - X\theta - W_0^\top\gamma - \lambda(v)) \right], \qquad (16)$$

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

Estimation of $\beta_{\alpha,*}$ and $\lambda_{\alpha,*}(v)$ based on (16) is not feasible since $v \equiv v(x, w, \pi_*)$ depends on unknown π_*. We assume that there exists a preliminary estimator $\hat{\pi}$ of π_*. For example, Lee (2007) considers

$$X_i = W_i^\top \pi_* + v_i, \quad \text{where } Q_{v|W}^{\tilde{\alpha}}(w) = 0 \qquad (17)$$

where $Q_{v|W}^{\tilde{\alpha}}(w)$ denotes the conditional $\tilde{\alpha}$-quantile of v given W for some $\tilde{\alpha} \in (0,1)$. Under the above assumption, one can estimate π_* through the quantile regression of X_i on W_i and estimate

$v_i = X_i - W_i^\top \pi_*$ using the fitted residual. One may also consider a nonparametric specification

$$X_i = \pi_*(W_i) + v_i, \text{ where } Q_{v|W}^{\tilde{\alpha}}(w) = 0 \tag{18}$$

and estimate the conditional quantile function $\pi_*(W_i)$ 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 by the semiparametric second-step quantile regression

$$(\hat{\beta}_\alpha, \hat{\lambda}_\alpha) \equiv \arg\min_{\theta,\gamma,\lambda} n^{-1} \sum_{i=1}^n \rho_\alpha(Y_i - X_i\theta - W_{0,i}^\top \gamma - \lambda(\hat{v}_i)). \tag{19}$$

It is worth noting that the identification condition (13) is imposed directly on the control variable $v(X, W, \pi_*)$ which is a function of X and W. Therefore, (13) is weaker than the following condition

$$Q_{u|X,W}^\alpha(x, w) = Q_{u|v}^\alpha(v), \tag{20}$$

where $Q_{u|X,W}^\alpha(x, w)$ 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 $(\theta_{\alpha,*}, \gamma_{\alpha,*}^\top)^\top$ under (13) is different from that under (20). □

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

$$\begin{aligned} Y^* &= m(X, \beta_*) + u, \\ v(X, W, \pi_*) &\equiv \mathbb{E}\left[d | X, W \right], \end{aligned} \tag{21}$$

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 parameters of interest. To achieve identification, we assume that

$$\mathbb{E}[u|X, v, d = 1] = \mathbb{E}[u|v, d = 1] \tag{22}$$

where $v \equiv v(X, W, \pi_*)$. A basic implication of model (21) and condition (22) is that

$$\mathbb{E}[Y^*|X, v, d = 1] = m(X, \beta_*) + \lambda_*(v), \text{ where } \lambda_*(v) = \mathbb{E}[u|v, d = 1]. \tag{23}$$

In (23), $\lambda_*(v)$ 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 (21) can be written as

$$Y^* = m(X, \beta_*) + \lambda_*(v) + \varepsilon \tag{24}$$

where $\mathbb{E}[\varepsilon|v, d = 1] = 0$ by (22), which implies that the finite dimensional parameter β_* is identified together with the unknown function λ_* as the minimizer of the following problem

$$\min_{\beta, \lambda} \mathbb{E}\left[2^{-1}d\left|Y - m(X, \beta) - \lambda(v)\right|^2\right] \tag{25}$$

where $Y \equiv dY^*$.

To construct feasible estimators of β_* and $\lambda_*(v)$ based on (25), we assume that there exists a first-step estimator $\hat{\pi}$ of π_* such that v_i is estimated by $\hat{v}_i \equiv v(X_i, W_i, \hat{\pi})$. 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 by the semiparametric mean regression

$$(\hat{\beta}, \hat{\lambda}) \equiv \arg\min_{\beta, \lambda} \sum_{i=1}^n 2^{-1}d_i \left|Y_i - m(X_i, \beta) - \lambda(\hat{v}_i)\right|^2. \tag{26}$$

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., Cavanagh and Sherman (1998), Ichimura (1993), and Powell, Stock, and Stoker (1989)). 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 (22) which is weaker than the condition

$$\mathbb{E}[u|X, W, d = 1] = \mathbb{E}[u|v, d = 1], \tag{27}$$

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

3   Two-step Estimation with a Parametric First Step

In this section, we derive the influence function of the semiparametric two-step estimator $\hat{\beta}$ when π_* is parametrically specified. Since the focus is on β_*, we profile out the nonparametric component λ by solving

$$h\left(v\left(\pi\right); \beta, \pi\right) \equiv \arg\min_{\lambda} \mathbb{E}\left[\psi(Z, \beta, \lambda(v(\pi)))\right] \tag{28}$$

---

[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.

for any β and any π. To introduce the optimality condition of $h\left(v\left(\pi\right);\beta,\pi\right)$, we assume that there exists $\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)$ with

$$\mathrm{Var}\left(\psi_\lambda\left(Z,\beta_*,\lambda_*(v)\right)\right) > 0$$

such that

$$\left.\frac{\partial\mathbb{E}\left[\psi(Z,\beta,\lambda\left(v\left(\pi\right)\right)+\tau\lambda_1\left(v\left(\pi\right)\right))\right]}{\partial\tau}\right|_{\tau=0} = \mathbb{E}\left[\psi_\lambda(Z,\beta,\lambda\left(v\left(\pi\right)\right))\lambda_1\left(v\left(\pi\right)\right)\right],$$

where $\lambda\left(v\left(\pi\right)\right)$ and $\lambda_1\left(v\left(\pi\right)\right)$ are any functions of $v\left(\pi\right)$. Then $h\left(v\left(\pi\right);\beta,\pi\right)$ satisfies

$$\mathbb{E}\left[\psi_\lambda\left(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right)\right)\lambda\left(v\left(\pi\right)\right)\right] = 0 \tag{29}$$

for any function $\lambda\left(v\left(\pi\right)\right)$ of $v\left(\pi\right)$ and any β.

The profiled version of the minimization problem (1) becomes

$$\min_\beta\mathbb{E}\left[\psi(Z,\beta,h\left(v\left(\pi_*\right);\beta,\pi_*\right))\right]. \tag{30}$$

Therefore, β_* satisfies the following first-order condition

$$\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right] = 0, \tag{31}$$

where

$$J(Z,\beta,\pi) \equiv \psi_\beta\left(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right)\right) + \psi_\lambda\left(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right)\right)\frac{\partial h\left(v\left(\pi\right);\beta,\pi\right)}{\partial\beta}, \tag{32}$$

and the function $\psi_\beta(Z,\beta,\lambda(v(\pi)))$ is assumed to satisfy $\mathrm{Var}\left(\psi_\beta\left(Z,\beta_*,\lambda_*(v)\right)\right) > 0$ and

$$\mathbb{E}\left[\psi_\beta(Z,\beta,\lambda(v(\pi)))\right] = \frac{\partial\mathbb{E}\left[\psi(Z,\beta,\lambda(v(\pi)))\right]}{\partial\beta}$$

for any function $\lambda(v(\pi))$ of $v(\pi)$ and any β. The influence function of $\hat{\beta}$ is calculated using the arguments in Newey (1994), which shows that the function $J(Z,\beta,\pi)$ is the key for the calculation, because: (i) $J(Z,\beta_*,\pi_*)$ is the score of $\hat{\beta}$ when π_* is known; (ii) the impact of estimating π_* on the score function of $\hat{\beta}$ is the derivative $\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]/\partial\pi^\top$ times the influence function of $\hat{\pi}$; (iii) the Hessian matrix of $\hat{\beta}$ is given by $\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]/\partial\beta^\top$.[2]

Some notations are needed to derive the derivatives of $J(Z,\beta,\pi)$ with respect to β or π. For any functions $\lambda\left(v\left(\pi\right)\right)$ and $\lambda_1\left(v\left(\pi\right)\right)$ of $v\left(\pi\right)$, we assume that there exist $\psi_{\beta,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)$, $\psi_{\beta,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)$, $\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)$ and $\psi_{\lambda,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)$ such that

$$\left.\frac{\partial\mathbb{E}\left[\psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)+\tau\lambda_1\left(v\left(\pi\right)\right)\right)\right]}{\partial\tau}\right|_{\tau=0} = \mathbb{E}\left[\psi_{\beta,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\lambda_1\left(v\left(\pi\right)\right)\right],$$

---

[2]A more formal discussion on the relevance of $J(Z,\beta,\pi)$ can be found at the beginning of Section A in the Appendix.

$$\frac{\partial \mathbb{E}\left[\psi_\beta\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right)\right]}{\partial \beta^\top} = \mathbb{E}\left[\psi_{\beta,\beta}\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right)\right],$$

$$\frac{\partial \mathbb{E}\left[\psi_\lambda\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right)\right]}{\partial \beta} = \mathbb{E}\left[\psi_{\lambda,\beta}\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right)\right],$$

and

$$\left.\frac{\partial \mathbb{E}\left[\psi_\lambda\left(Z, \beta, \lambda\left(v\left(\pi\right)\right) + \tau\lambda_1\left(v\left(\pi\right)\right)\right)\right]}{\partial \tau}\right|_{\tau=0} = \mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z, \beta, \lambda\left(v\left(\pi\right)\right)\right)\lambda_1\left(v\left(\pi\right)\right)\right].$$

These functions are defined using expectation to make our influence function formula applicable to models with non-smooth criterion function $\psi(z, \beta, \lambda(v(\pi)))$, such as the check function in the quantile regression. When $\psi(z, \beta, \lambda(v(\pi)))$ is smooth, such as the square loss function in the mean regression, these functions can be defined directly using the derivatives of $\psi(z, \beta, \lambda(v(\pi)))$.

For ease of notations, we suppress the dependence of the derivatives of $\psi(z, \beta, \lambda(v(\pi)))$ on the parameters when they are evaluated at the true parameter values. Therefore $v \equiv v(\pi_*)$, $\psi_\beta\left(Z\right) \equiv \psi_\beta\left(Z, \beta_*, \lambda_*(v)\right)$, $\psi_\lambda\left(Z\right) \equiv \psi_\lambda\left(Z, \beta_*, \lambda_*(v)\right)$ and the other notations are understood similarly. Define

$$g_*(v) \equiv \frac{\mathbb{E}\left[\psi_{\beta,\lambda}(Z)\mid v\right]}{\mathbb{E}\left[\psi_{\lambda,\lambda}(Z)\mid v\right]} \text{ and } \Psi_{\beta,\beta} \equiv -\mathbb{E}\left[\psi_{\beta,\beta}(Z) - \psi_{\lambda,\lambda}\left(Z\right)g_*(v)g_*(v)^\top\right]. \tag{33}$$

The influence function of $\hat\beta$ with a parametric first-step estimation is provided in the following theorem.

Theorem 1 (Main Result). *Suppose that: (i) the influence function of $\hat\pi$ is $\varphi_\pi(z)$; (ii) $\Psi_{\beta,\beta}$ is non-singular; and (iii) $\psi_{\lambda,\beta}\left(Z\right) = \psi_{\beta,\lambda}\left(Z\right)$ almost surely. Then the influence function of $\hat\beta$ is*

$$\Psi_{\beta,\beta}^{-1}\left(\varphi_\beta(Z) + \Psi_{\beta,\pi}\varphi_\pi(Z)\right), \tag{34}$$

where

$$\varphi_\beta(Z) \equiv \psi_\beta\left(Z\right) - g_*(v)\psi_\lambda\left(Z\right), \tag{35}$$

$$\Psi_{\beta,\pi} \equiv \mathbb{E}\left[\left(\delta_\beta(Z) - \delta_g(Z)\right)\frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}\right], \tag{36}$$

$$\delta_\beta(Z) \equiv \left[\psi_{\lambda,\beta}\left(Z\right) - g_*(v)\psi_{\lambda,\lambda}\left(Z\right)\right]\frac{\partial \lambda_*\left(v\right)}{\partial v}, \tag{37}$$

$$\delta_g(Z) \equiv \psi_\lambda\left(Z\right)\frac{\partial g_*\left(v\right)}{\partial v}. \tag{38}$$

Remark 1 (Asymptotic Variance of $\hat\beta$). By Theorem 1, the asymptotic variance of $\hat\beta$ takes the sandwich form

$$\mathrm{AsyVar}(\hat\beta) = \Psi_{\beta,\beta}^{-1}\Omega_*\Psi_{\beta,\beta}^{-1},$$

where

$$\Omega_* \equiv \lim_{n\to\infty} \mathrm{Var}\left(n^{-1/2}\sum_{i=1}^n\left(\varphi_\beta\left(Z_i\right) + \Psi_{\beta,\pi}\varphi_\pi\left(Z_i\right)\right)\right)$$

denotes the asymptotic variance of the score function of $\hat{\beta}$. $\qquad\square$

Remark 2 (Index Restriction). The adjustment in the score function of $\hat{\beta}$ can be simplified under an extra assumption

$$\mathbb{E}\left[\psi_\lambda\left(Z\right)\bigg|v\left(\pi_*\right),\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right]=0, \tag{39}$$

because in this case,

$$\Psi_{\beta,\pi}=\mathbb{E}\left[\delta_\beta(Z)\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right].$$

Condition (39) is further implied by

$$\mathbb{E}\left[\psi_\lambda\left(Z\right)|X,W\right]=0 \tag{40}$$

since $v\left(\pi_*\right)\equiv v(X,W,\pi_*)$ is a function of X and W. As we shall discuss in Section 5, condition (40) becomes the commonly used identification condition when the control variable approach is applied to specified models in the literature. On the other hand, in view of (29) the influence function of $\hat{\beta}$ derived here only uses

$$\mathbb{E}\left[\psi_\lambda\left(Z\right)|v\left(\pi_*\right)\right]=0 \tag{41}$$

and (31), which is weaker than (40). Although condition (40) is popular in the early literature, recent applications of the control variable approach such as Johnsson and Moon (2015) and Auerbach (2019), use variants of (41) which are imposed on the control variables directly. Under the weaker condition (41), Theorem 1 shows that the extra term

$$\mathbb{E}\left[\delta_g(Z)\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right]$$

in the influence function of $\hat{\beta}$ may not be negligible, when the strong assumption (39) does not hold. $\qquad\square$

4 Two-step Estimation with a Nonparametric First Step

In this section, we extend the influence function formula of $\hat{\beta}$ obtained in the previous section to the case where π_* is nonparametrically specified in the first step. Suppose that there are L functions $\pi_{*,l}(w_l)$ $(l=1,\ldots,L)$ estimated separately in the first step. For each $l=1,\ldots,L$, $\pi_{*,l}$ is identified by the following conditional moment condition

$$\mathbb{E}\left[\mu_l(Z_l,\pi_{*,l})|W_l\right]=0,$$

where $\mu_l(z_l,\pi_l)$ is a first-step residual function, Z_l is a sub-vector of Z and W_l is a sub-vector (of exogenous variables) of Z_l.

To derive the influence function of $\hat{\beta}$ in this case, we follow Newey (1994) and consider any one-dimensional path of densities of Z indexed by $\tau \in \mathbb{R}$ such that the path hits the true density at $\tau = 0$. Let $\pi_{*,l,\tau}$ denote the counterpart of $\pi_{*,l}$ under the path τ, i.e., $\pi_{*,l,\tau}$ satisfies

$$\mathbb{E}_\tau \left[\mu_l(Z_l, \pi_{*,l,\tau}) \pi_l(W_l) \right] = 0 \tag{42}$$

for any $\pi_l(w_l)$, where $\mathbb{E}_\tau [\cdot]$ denotes the conditional expectation taken under the path density indexed by τ.

Suppose that there exists $\mu_{l,\pi}(Z_l, \pi)$ such that

$$\frac{\partial \mathbb{E} \left[\mu_l(Z_l, \pi_{*,l,\tau}) \pi_l(W_l) \right]}{\partial \tau} = \mathbb{E} \left[\mu_{l,\pi}(Z_l, \pi_{*,l}) \pi_l(W_l) \frac{\partial \pi_{*,l,\tau}(W_l)}{\partial \tau} \right]$$

for any $\pi_l(w_l)$, where derivatives with respect to τ are evaluated at $\tau = 0$ unless otherwise indicated. Taking derivative with respect to τ in (42) and applying the chain rule, we get

$$\frac{\partial \mathbb{E}_\tau \left[\mu_l(Z_l, \pi_{*,l}) \pi_l(W_l) \right]}{\partial \tau} + \mathbb{E} \left[\mathbb{E} \left[\mu_{l,\pi}(Z_l, \pi_{*,l}) | W_l \right] \pi_l(W_l) \frac{\partial \pi_{*,l,\tau}(W_l)}{\partial \tau} \right] = 0 \tag{43}$$

for any function $\pi_l(w_l)$.

The finite dimensional parameter β_* still satisfies the first-order condition in (31). We assume that $v(Z, \pi)$ is smooth and it depends on $\pi \equiv (\pi_1, \ldots, \pi_L)^\top$ only through its value $\pi(W)$. Using similar calculation in (99), (102), (103) and (104) in the Appendix, we obtain

$$\frac{\partial \mathbb{E} \left[J(Z, \beta_*, \pi_\tau) \right]}{\partial \tau} = \frac{\partial \mathbb{E} \left[D(Z, \pi_\tau) \right]}{\partial \tau}, \tag{44}$$

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. Moreover, for any $\pi(W)$

$$\mathbb{E} \left[D(Z, \pi) \right] = \mathbb{E} \left[\delta_\pi(W)^\top \pi(W) \right] \tag{45}$$

by the law of iterated expectation, where $\delta_\pi(W) \equiv (\delta_{1,\pi}(W_1), \ldots, \delta_{L,\pi}(W_L))^\top$ and

$$\delta_{l,\pi}(W_l) \equiv \mathbb{E} \left[\left[\delta_\beta(Z) - \delta_g(Z) \right] \frac{\partial v(\pi_*)}{\partial \pi_l} \Bigg| W_l \right] \tag{46}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (37) and (38) respectively. Combining (44) and (45), we

deduce that

$$
\begin{aligned}
\frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_\tau)\right]}{\partial \tau} = \frac{\partial \mathbb{E}\left[D(Z, \pi_\tau)\right]}{\partial \tau} &= \mathbb{E}\left[\delta_\pi(W)^\top \frac{\partial \pi_\tau(W)}{\partial \tau}\right] \\
&= \sum_{l=1}^{L} \mathbb{E}\left[\delta_{l,\pi}(W_l) \frac{\partial \pi_{l,\tau}(W_l)}{\partial \tau}\right] \\
&= \sum_{l=1}^{L} \frac{\partial}{\partial \tau} \mathbb{E}_\tau\left[-\frac{\mu_l(Z_l, \pi_{*,l})}{\mathbb{E}\left[\mu_{l,\pi}(Z_l, \pi_{*,l})\middle| W_l\right]} \delta_{l,\pi}(W_l)\right] \\
&= \frac{\partial}{\partial \tau} \mathbb{E}_\tau\left[\delta_\pi(W)^\top \varphi_\pi(Z)\right],
\end{aligned}
\tag{47}
$$

where

$$
\varphi_\pi(Z) \equiv -\left(\frac{\mu_1(Z_1, \pi_{*,1})}{\mathbb{E}\left[\mu_{1,\pi}(Z_1, \pi_{*,1})\middle| W_1\right]}, \cdots, \frac{\mu_L(Z_L, \pi_{*,L})}{\mathbb{E}\left[\mu_{L,\pi}(Z_L, \pi_{*,L})\middle| W_L\right]}\right)^\top
\tag{48}
$$

and the third equality in (47) follows from (43) by replacing $\pi_l(W_l)$ with $\delta_{l,\pi}(W_l)/\mathbb{E}\left[\mu_{l,\pi}(Z_l, \pi_{*,l})\middle| W_l\right]$ for $l = 1, \ldots, L$. Therefore, (3.9) in Newey (1994) follows from (47). The following theorem is directly implied by Theorem 2.1 of Newey (1994).

Theorem 2 (Nonparametric First Step). *Suppose that: (i)* $\left|\mathbb{E}\left[\mu_{l,\pi}(Z_l, \pi_{*,l})\middle| W_l\right]\right| > 0$ *almost surely for* $l = 1, \ldots, L$; *(ii)* $\Psi_{\beta,\beta}$ *is non-singular; and (iii)* $\psi_{\lambda,\beta}(Z) = \psi_{\beta,\lambda}(Z)$ *almost surely. Then the influence function of* $\hat{\beta}$ *is*

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

where $\Psi_{\beta,\beta}$ *and* $\varphi_\beta(Z)$ *are defined in (33) and (35), respectively.*

5 Applications

In this section, we provide the influence functions of the two-step estimators discussed in the first two examples of Section 2.[3] In view of Theorem 1 and Theorem 2, 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 notations, we let $Z_0 \equiv (X, W_0^\top)^\top$. In this example, we have

$$
\psi\left(Z, \beta, \lambda(v(\pi))\right) = 2^{-1}\left(Y - m(Z_0, \beta) - \lambda(v(\pi))\right)^2.
\tag{49}
$$

---

[3]Similar calculations apply to the third example in Section 2. The details can be found in Appendix B.

Using the above expression, it is easy to calculate that

$$\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = -(Y - m(Z_0,\beta) - \lambda\left(v\left(\pi\right)\right)), \tag{50}$$

$$\psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = -\left(Y - m(Z_0,\beta) - \lambda\left(v\left(\pi\right)\right)\right)m_\beta(Z_0,\beta), \tag{51}$$

$$\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = m_\beta(Z_0,\beta) = \psi_{\beta,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right), \tag{52}$$

$$\psi_{\beta,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = m_\beta(Z_0,\beta)m_\beta(Z_0,\beta)^\top - \left(Y - m(Z_0,\beta) - \lambda\left(v\left(\pi\right)\right)\right)m_{\beta,\beta}(Z_0,\beta), \tag{53}$$

$$\psi_{\lambda,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = 1, \tag{54}$$

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

$$m_\beta(Z_0,\beta) \equiv \frac{\partial m(Z_0,\beta)}{\partial\beta} \text{ and } m_{\beta,\beta}(Z_0,\beta) \equiv \frac{\partial^2 m(Z_0,\beta)}{\partial\beta\partial\beta^\top}. \tag{55}$$

By (50), the first-order condition of the profiled nonparametric function $h\left(v\left(\pi\right);\beta,\pi\right)$ can be written as

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

which immediately implies that in this example

$$h\left(v\left(\pi\right);\beta,\pi\right) = \mathbb{E}\left[Y - m(Z_0,\beta)\middle| v\left(\pi\right)\right] \tag{56}$$

and therefore

$$h\left(v\left(\pi_*\right);\beta_*,\pi_*\right) = \mathbb{E}\left[u\middle| v\right] = \lambda_*(v). \tag{57}$$

Let $m_\beta(Z_0) \equiv m_\beta(Z_0,\beta_*)$. Using the expressions in (50)-(54) and (57), we get

$$\varphi_\beta(Z) = -\varepsilon\left[m_\beta(Z_0) - \mathbb{E}\left[m_\beta(Z_0)|v\right]\right], \tag{58}$$

$$\delta_\beta(Z) = \left(m_\beta(Z_0) - \mathbb{E}\left[m_\beta(Z_0)\middle| v\right]\right)\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v}, \tag{59}$$

$$\delta_g(Z) = -\varepsilon\frac{\partial\mathbb{E}\left[m_\beta(Z_0)\middle| v\right]}{\partial v} \text{ and} \tag{60}$$

$$\Psi_{\beta,\pi} = \mathbb{E}\left[\left(\left(m_\beta(Z_0) - \mathbb{E}\left[m_\beta(Z_0)\middle| v\right]\right)\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v} + \varepsilon\frac{\partial\mathbb{E}\left[m_\beta(Z_0)\middle| v\right]}{\partial v}\right)\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right] \tag{61}$$

when π_* is a finite-dimensional parameter vector. From (5), (50)-(54), the Hessian matrix takes the following form

$$\Psi_{\beta,\beta} = -\mathbb{E}\left[m_\beta(Z_0)m_\beta(Z_0)^\top - \mathbb{E}\left[m_\beta(Z_0)|v\right]\mathbb{E}\left[m_\beta(Z_0)|v\right]^\top\right]. \tag{62}$$

From the components in (58), (61) and (62), the influence function of $\hat{\beta}$, when the influence function of the estimator $\hat{\pi}$ of π_* is $\varphi_\pi(Z_i)$, can be readily computed using Theorem 1.

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

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

13

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

$$\varphi_\pi(Z) = X - \pi_*(W) \tag{63}$$

and

$$\delta_\pi(W) = -\mathbb{E}\left[\delta_\beta(Z) - \delta_g(Z)|W\right] \tag{64}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (59) and (60) respectively. Using the components in (58), (62), (63) and (64), Theorem 2 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 the strong condition (10) holds, we obtain

$$\Psi_{\beta,\pi} = \mathbb{E}\left[[m_\beta(Z_0) - \mathbb{E}[m_\beta(Z_0)|v]]\frac{\partial\mathbb{E}[u|v]}{\partial v}\frac{\partial v(\pi_*)}{\partial\pi^\top}\right] \tag{65}$$

in the case with a parametric first step, and

$$\delta_\pi(W) = -\mathbb{E}\left[m_\beta(Z_0) - \mathbb{E}[m_\beta(Z_0)|v]\frac{\partial\mathbb{E}[u|v]}{\partial v}\bigg| W\right] \tag{66}$$

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 (10) 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,

$$\varphi_\beta(Z) = -\varepsilon\left(W_0 - \mathbb{E}[W_0|v]\right), \tag{67}$$

$$\Psi_{\beta,\pi} = -\mathbb{E}\left[(W_0 - \mathbb{E}[W_0|v])\frac{\partial\mathbb{E}[u|v]}{\partial v}W^\top\right], \tag{68}$$

$$\Psi_{\beta,\beta} = -\mathbb{E}\left[W_0 W_0^\top - \mathbb{E}[W_0|v]\mathbb{E}[W_0|v]^\top\right]. \tag{69}$$

The influence function of the two-step estimator $\hat\beta$ can be calculated using Theorem 1, the items in (67)-(69) 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. $\qquad\square$

Example 2 (Quantile Regression Continued). For ease of notations, let $Z_0 \equiv (X, W_0^\top)^\top$ and $\beta \equiv (\theta, \gamma^\top)^\top$. In this example, we have

$$\psi(Z, \beta, \lambda(v(\pi))) = \rho_\alpha(Y - Z_0^\top\beta - \lambda(v)). \tag{70}$$

Using the above expression, it is easy to calculate that

$$\frac{\partial\mathbb{E}\left[\rho_\alpha(Y - Z_0^\top\beta - \lambda(v(\pi)) - \tau\lambda_1(v(\pi)))\right]}{\partial\tau}\bigg|_{\tau=0}$$
$$= \mathbb{E}\left[\left(1\left\{Y \le Z_0^\top\beta + \lambda(v(\pi))\right\} - \alpha\right)\lambda_1(v(\pi))\right] \tag{71}$$

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

$$\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = 1\left\{Y \leq Z_0^\top\beta + \lambda\left(v\left(\pi\right)\right)\right\} - \alpha. \tag{72}$$

Applying the above expression to the first-order condition (29), 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[4]

$$h\left(v\left(\pi_*\right);\beta_*,\pi_*\right) = Q_{u|v}^\alpha\left(v\right) = \lambda_*(v). \tag{73}$$

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 (70), it is easy to calculate that

$$\psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = \left(1\left\{Y \leq Z_0^\top\beta + \lambda\left(v\left(\pi\right)\right)\right\} - \alpha\right)Z_0, \tag{74}$$

$$\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = f_\varepsilon\left(\left.Z_0^\top\left(\beta - \beta_*\right) + \lambda\left(v\left(\pi\right)\right) - \lambda_*(v)\right|Z_0,v\left(\pi\right)\right)Z_0, \tag{75}$$

$$\psi_{\beta,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = f_\varepsilon\left(\left.Z_0^\top\left(\beta - \beta_*\right) + \lambda\left(v\left(\pi\right)\right) - \lambda_*(v)\right|Z_0,v\left(\pi\right)\right)Z_0Z_0^\top, \tag{76}$$

$$\psi_{\lambda,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) = f_\varepsilon\left(\left.Z_0^\top\left(\beta - \beta_*\right) + \lambda\left(v\left(\pi\right)\right) - \lambda_*(v)\right|Z_0,v\left(\pi\right)\right), \tag{77}$$

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 (73)-(77), we get

$$\varphi_\beta(Z) = \left(1\left\{\varepsilon \leq 0\right\} - \alpha\right)\left(Z_0 - g_*(v)\right), \tag{78}$$

$$\delta_\beta(Z) = f_\varepsilon\left(0|Z_0,v\right)\left(Z_0 - g_*(v)\right)\frac{\partial Q_{u|v}^\alpha\left(v\right)}{\partial v}, \tag{79}$$

$$\delta_g(Z) = \left(1\left\{\varepsilon \leq 0\right\} - \alpha\right)\frac{\partial g_*\left(v\right)}{\partial v}, \text{ and} \tag{80}$$

$$\Psi_{\beta,\pi} = \mathbb{E}\left[\left(f_\varepsilon\left(0|Z_0,v\right)\left(Z_0 - g_*(v)\right)\frac{\partial Q_{u|v}^\alpha\left(v\right)}{\partial v} - \left(1\left\{\varepsilon \leq 0\right\} - \alpha\right)\frac{\partial g_*\left(v\right)}{\partial v}\right)\frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}\right] \tag{81}$$

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

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

From (73)-(77), the Hessian matrix takes the following form

$$\Psi_{\beta,\beta} = -\mathbb{E}\left[f_\varepsilon\left(0|Z_0,v\right)\left(Z_0Z_0^\top - g_*(v)g_*(v)^\top\right)\right]. \tag{82}$$

Using the components in (78), (81) and (82), the influence function of $\hat{\beta}$, when the influence function of the estimator $\hat{\pi}$ of π_* is $\varphi_\pi(Z_i)$, can be readily computed using Theorem 1.

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

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

---

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

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

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

Therefore in this case,

$$\varphi_\pi(Z) = -\frac{1\{X \leq \pi_*(W)\} - \tilde{\alpha}}{f_{X|W}(\pi_*(W))} \tag{83}$$

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

$$\delta_\pi(W) = -\mathbb{E}\left[\delta_\beta(Z) - \delta_g(Z) | W\right] \tag{84}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (79) and (80) respectively. Using the components in (78), (82), (83) and (84), Theorem 2 implies that the influence function of the two-step estimator in this case is

$$\Psi^{-1}_{\beta,\beta}\left(\varphi_\beta(Z) - \delta_\pi(W)\frac{1\{X_i \leq \pi_*(W)\} - \tilde{\alpha}}{f_{X|W}(\pi_*(W))}\right). \tag{85}$$

If the strong condition (20) holds, we get

$$\Psi_{\beta,\pi} = \mathbb{E}\left[f_\varepsilon(0|Z_0, v)(Z_0 - g_*(v))\frac{\partial Q^\alpha_{u|v}(v)}{\partial v}W^\top\right] \tag{86}$$

in the case with a parametric first step, and

$$\delta_\pi(W) = -\mathbb{E}\left[f_\varepsilon(0|Z_0, v)(Z_0 - g_*(v))\frac{\partial Q^\alpha_{u|v}(v)}{\partial v}\bigg| W\right] \tag{87}$$

in the case with a nonparametric first step. Therefore the influence function of $\hat{\beta}$ is slightly simplified in both cases. Moreover, under (17) and (20), the asymptotic variance of $\hat{\beta}$ implied by its influence function (which can be calculated using (78), (82) and (86)) is similar to the one stated in Theorem 3.1 in Lee (2007). □

6 Conclusion

In this paper, we derive the influence function of semiparametric two-step estimators where a unknown function/control variable is estimated in a first-step estimation which can be parametric, semiparametric or fully nonparametric. The influence function is derived under an index restriction that is imposed directly on the control variable and hence is weaker than the commonly adopted identification condition in the literature, which is imposed on all exogenous variables. As a result, the influence function formula derived in this paper contains an additional term which is not negligible in general. The general influence function formula is illustrated in a mean regression example,

a quantile regression example and a sample selection example where the control variable approach is applied for identification and consistent estimation of structural parameters with endogenous explanatory variables.

References

AHN, H., AND J. L. POWELL (1993): "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," Journal of Econometrics, 58(1-2), 3–29.

AUERBACH, E. (2019): "Identification and estimation of a partially linear regression model using network data," arXiv preprint arXiv:1903.09679.

BLUNDELL, R. W., AND J. L. POWELL (2004): "Endogeneity in semiparametric binary response models," The Review of Economic Studies, 71(3), 655–679.

CAVANAGH, C., AND R. P. SHERMAN (1998): "Rank estimators for monotonic index models," Journal of Econometrics, 84(2), 351–382.

HECKMAN, J. 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, volume 5, number 4, pp. 475–492. NBER.

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

IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and estimation of triangular simultaneous equations models without additivity," Econometrica, 77(5), 1481–1512.

JOHNSSON, I., AND H. R. MOON (2015): "Estimation of peer effects in endogenous social networks: control function approach," Review of Economics and Statistics, pp. 1–51.

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

LI, Q., AND J. M. WOOLDRIDGE (2002): "Semiparametric estimation of partially linear models for dependent data with generated regressors," Econometric Theory, 18(3), 625–645.

NEWEY, W. K. (1994): "The asymptotic variance of semiparametric estimators," Econometrica: Journal of the Econometric Society, pp. 1349–1382.

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

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric estimation of index coefficients," Econometrica: Journal of the Econometric Society, pp. 1403–1430.

RIVERS, D., AND Q. H. VUONG (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," Journal of econometrics, 39(3), 347–366.

Appendix

A   Proof of Theorem 1

The theorem is proved using the arguments in Sections 2 and 3 of Newey (1994). Specifically, (3.10) in Newey (1994) shows that the influence function of $\hat{\beta}$ can be derived from (31), and it takes the following form

$$- \left(\frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_*)\right]}{\partial \beta^\top} \right)^{-1} (J(Z, \beta_*, \pi_*) + \eta(Z)) \tag{88}$$

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

$$\frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_{*,\tau})\right]}{\partial \tau} = \mathbb{E}\left[\eta(Z) \frac{\partial \ln(f_{z,\tau}(Z))}{\partial \tau}\right] \tag{89}$$

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

$$\frac{\partial \pi_{*,\tau}}{\partial \tau} = \mathbb{E}\left[\varphi_\pi(Z) \frac{\partial \ln(f_{z,\tau}(Z))}{\partial \tau}\right]. \tag{90}$$

From (89) and (90), we get

$$\eta(Z) = \frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_*)\right]}{\partial \pi^\top} \varphi_\pi(Z)$$

and hence the influence function of $\hat{\beta}$ is

$$- \left(\frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_*)\right]}{\partial \beta^\top} \right)^{-1} \left(J(Z, \beta_*, \pi_*) + \frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_*)\right]}{\partial \pi^\top} \varphi_\pi(Z) \right) \tag{91}$$

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 (98), (104) and (105) 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

$$\mathbb{E}\left[\psi_\lambda\left(Z, \beta_*, \lambda_*(v)\right)\middle| v\right] = 0. \tag{92}$$

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

$$\mathbb{E}\left[\psi_\lambda\left(Z, \beta, h\left(v\left(\pi\right); \beta, \pi\right)\right) \lambda\left(v\left(\pi\right)\right)\right] = 0 \tag{93}$$

for any function $\lambda\left(v\left(\pi\right)\right)$ of $v(\pi)$. Evaluating (93) at (β_*, π_*) and using $h\left(v\left(\pi_*\right); \beta_*, \pi_*\right) = \lambda_*(v)$, we obtain

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

18

for any function $\lambda(v)$ of v, which immediately implies (92).

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

$$\frac{\partial h\left(v\left(\pi\right);\beta_{*},\pi\right)}{\partial\beta} = -\frac{\mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_{*},h\left(v\left(\pi\right);\beta_{*},\pi\right)\right)|\,v\left(\pi\right)\right]}{\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_{*},h\left(v\left(\pi\right);\beta_{*},\pi\right)\right)|\,v\left(\pi\right)\right]}. \tag{94}$$

Differentiating (93) with respect to β and applying the chain rule, we obtain

$$0 = \mathbb{E}\left[\left(\psi_{\lambda,\beta}(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right)) + \psi_{\lambda,\lambda}(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right))\frac{\partial h\left(v\left(\pi\right);\beta,\pi\right)}{\partial\beta}\right)\lambda\left(v\left(\pi\right)\right)\right] \tag{95}$$

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

$$0 = \mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_{*},h\left(v\left(\pi\right);\beta_{*},\pi\right)\right) + \psi_{\lambda,\lambda}\left(Z,\beta_{*},h\left(v\left(\pi\right);\beta_{*},\pi\right)\right)\frac{\partial h\left(v\left(\pi\right);\beta_{*},\pi\right)}{\partial\beta}\bigg|\,v\left(\pi\right)\right], \tag{96}$$

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

$$0 = \mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_{*},h\left(v\left(\pi\right);\beta_{*},\pi\right)\right)|\,v\left(\pi\right)\right]$$

$$+ \mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_{*},h\left(v\left(\pi\right);\beta_{*},\pi\right)\right)|\,v\left(\pi\right)\right]\frac{\partial h\left(v\left(\pi\right);\beta_{*},\pi\right)}{\partial\beta}. \tag{97}$$

The claim in (94) follows from (97).

Step 3. We prove the claim of the theorem in this step. First, by the definition of $J(Z,\beta_{*},\pi_{*})$ in (32) and the definition of $g_{*}(v_{i})$ in (33), and the expression in (94), we get

$$J(Z,\beta_{*},\pi_{*}) = \psi_{\beta}\left(Z,\beta_{*},\lambda_{*}\left(v\right)\right) + \psi_{\lambda}\left(Z,\beta_{*},\lambda_{*}\left(v\right)\right)\frac{\partial h\left(v\left(\pi_{*}\right);\beta_{*},\pi_{*}\right)}{\partial\beta}$$

$$= \psi_{\beta}\left(Z,\beta_{*},\lambda_{*}\left(v\right)\right) - g_{*}(v)\psi_{\lambda}\left(Z,\beta_{*},\lambda_{*}\left(v\right)\right) = \varphi_{\beta}(Z_{i}). \tag{98}$$

Next from (31), we observe that

$$\frac{\partial\mathbb{E}\left[J(Z,\beta_{*},\pi_{*})\right]}{\partial\pi^{\top}} = \mathbb{E}\left[\psi_{\beta,\lambda}\left(Z,\beta_{*},\lambda_{*}\left(v\right)\right)\frac{d}{d\pi^{\top}}h\left(v\left(\pi_{*}\right);\beta_{*},\pi_{*}\right)\right]$$

$$+ \mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_{*},\lambda_{*}\left(v\right)\right)\frac{\partial h\left(v\left(\pi_{*}\right);\beta_{*},\pi_{*}\right)}{\partial\beta}\frac{d}{d\pi^{\top}}h\left(v\left(\pi_{*}\right);\beta_{*},\pi_{*}\right)\right]$$

$$- \mathbb{E}\left[\psi_{\lambda}\left(Z,\beta_{*},\lambda_{*}\left(v\right)\right)\frac{d}{d\pi^{\top}}g(v\left(\pi_{*}\right);\pi_{*})\right], \tag{99}$$

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

$$\frac{d}{d\pi^{\top}}h\left(v\left(\pi_{*}\right);\beta_{*},\pi_{*}\right) \equiv \frac{\partial h\left(v\left(\pi_{*}\right);\beta_{*},\pi_{*}\right)}{\partial\pi_{1}^{\top}} + \frac{\partial h\left(v\left(\pi_{*}\right);\beta_{*},\pi_{*}\right)}{\partial\pi_{2}^{\top}}. \tag{100}$$

19

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

$$\frac{\partial h\left(v\left(\pi_*\right); \beta_*, \pi_*\right)}{\partial \pi_1^\top} = \frac{\partial \lambda_*\left(v(\pi_*)\right)}{\partial v} \frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}. \tag{101}$$

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 (96) and the definition of $g_*(v)$ implies that

$$\mathbb{E}\left[\psi_{\beta,\lambda}\left(Z, \beta_*, \lambda_*\left(v\right)\right) \frac{\partial h\left(v\left(\pi_*\right); \beta_*, \pi_*\right)}{\partial \pi_2^\top}\right]$$
$$+ \mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z, \beta_*, \lambda_*\left(v\right)\right) \frac{\partial h\left(v\left(\pi_*\right); \beta_*, \pi_*\right)}{\partial \beta} \frac{\partial h\left(v\left(\pi_*\right); \beta_*, \pi_*\right)}{\partial \pi_2^\top}\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\psi_{\lambda,\beta}\left(Z, \beta_*, \lambda_*\left(v\right)\right) - g_*(v)\psi_{\lambda,\lambda}\left(Z, \beta_*, \lambda_*\left(v\right)\right)\vert\, v\right] \frac{\partial h\left(v\left(\pi_*\right); \beta_*, \pi_*\right)}{\partial \pi_2^\top}\right] = 0, \tag{102}$$

where we also used $\psi_{\beta,\lambda}\left(Z, \beta_*, \lambda_*\left(v\right)\right) = \psi_{\lambda,\beta}\left(Z, \beta_*, \lambda_*\left(v\right)\right)$ almost surely. Therefore using (99), (100), (101) and (102), we get

$$\frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_*)\right]}{\partial \pi^\top} = \mathbb{E}\left[\left[\psi_{\lambda,\beta}\left(Z, \beta_*, \lambda_*(v)\right) - g_*(v)\psi_{\lambda,\lambda}\left(Z, \beta_*, \lambda_*(v)\right)\right] \frac{\partial \lambda_*\left(v(\pi_*)\right)}{\partial v} \frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}\right]$$
$$- \mathbb{E}\left[\psi_\lambda\left(Z, \beta_*, \lambda_*(v)\right) \frac{d}{d\pi^\top} g(v\left(\pi_*\right); \pi_*)\right]. \tag{103}$$

Note that

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

So we have

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

where $\partial g(v\left(\pi_*\right); \pi_*)/\partial \pi_2^\top$ is a function of $v\left(\pi_*\right) = v$. Therefore by (92),

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

which together with (103), the definition of $g_*(v)$, and $\partial g\left(v\left(\pi_*\right), \pi_*\right)/\partial v = \partial g_*(v)/\partial v$ implies that

$$\frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_*)\right]}{\partial \pi^\top} = \mathbb{E}\left[\left[\psi_{\lambda,\beta}\left(Z, \beta_*, \lambda_*\left(v\right)\right) - g_*(v)\psi_{\lambda,\lambda}\left(Z, \beta_*, \lambda_*\left(v\right)\right)\right] \frac{\partial \lambda_*\left(v(\pi_*)\right)}{\partial v} \frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}\right]$$
$$- \mathbb{E}\left[\psi_\lambda\left(Z, \beta_*, \lambda_*\left(v\right)\right) \frac{\partial g_*\left(v\right)}{\partial v} \frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}\right] = \Psi_{\beta,\pi}. \tag{104}$$

20

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

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

which together with (92), (94) and the definition of $g_*(v)$ implies that

$$
\begin{aligned}
\frac{\partial \mathbb{E}\left[J(Z, \beta_*, \pi_*)\right]}{\partial \beta^\top} = & \ \mathbb{E}\left[\psi_{\beta,\beta}\left(Z, \beta_*, \lambda_*(v)\right) - \psi_{\beta,\lambda}\left(Z, \beta_*, \lambda_*(v)\right) g_*(v)^\top\right] \\
& - \mathbb{E}\left[g_*(v)\psi_{\lambda,\beta}\left(Z, \beta_*, \lambda_*(v)\right)^\top\right] + \mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z, \beta_*, \lambda_*(v)\right) g_*(v)g_*(v)^\top\right] \\
= & \ \mathbb{E}\left[\psi_{\beta,\beta}\left(Z, \beta_*, \lambda_*(v)\right) - \psi_{\lambda,\lambda}\left(Z, \beta_*, \lambda_*(v)\right) g_*(v)g_*(v)^\top\right] = -\Psi_{\beta,\beta}, \quad (105)
\end{aligned}
$$

which finishes the proof.

B Sample Selection Model Continued

This section provides the details on calculating the influence function of the two-step estimator in the sample selection model discussed in Section 2. In this example, we have

$$
\psi\left(Z, \beta, \lambda(v(\pi))\right) = 2^{-1} d\left(Y - m(X, \beta) - \lambda(v(\pi))\right)^2. \tag{106}
$$

It is easy to calculate that

$$
\psi_\lambda\left(Z, \beta, \lambda(v(\pi))\right) = -d\left(Y - m(X, \beta) - \lambda(v(\pi))\right), \tag{107}
$$

$$
\psi_\beta\left(Z, \beta, \lambda(v(\pi))\right) = -d\left(Y - m(X, \beta) - \lambda(v(\pi))\right) m_\beta(X, \beta), \tag{108}
$$

$$
\psi_{\beta,\beta}\left(Z, \beta, \lambda(v(\pi))\right) = dm_\beta(X, \beta)m_\beta(X, \beta)^\top - d\left(Y - m(X, \beta) - \lambda(v(\pi))\right) m_{\beta,\beta}(X, \beta), \tag{109}
$$

$$
\psi_{\lambda,\beta}\left(Z, \beta, \lambda(v(\pi))\right) = dm_\beta(X, \beta) \ = \psi_{\beta,\lambda}\left(Z, \beta, \lambda(v(\pi))\right), \text{ and} \tag{110}
$$

$$
\psi_{\lambda,\lambda}\left(Z, \beta, \lambda(v(\pi))\right) = d, \tag{111}
$$

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

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

21

which implies that in this example

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

where the second equality is by

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

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

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

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

Let $m_\beta(X) \equiv m_\beta(X, \beta_*)$. By (33), (110) and (111), we get

$$g_*(v) = \frac{\mathbb{E}\left[dm_\beta(X)v\right]}{\mathbb{E}\left[d|v\right]} = \mathbb{E}\left[m_\beta(X)|v, d = 1\right] \tag{112}$$

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 (33), (35), and (106)-(112), we have

$$\varphi_\beta(Z) = -d\varepsilon\left(m_\beta(X) - \mathbb{E}\left[m_\beta(X)|v, d = 1\right]\right), \tag{113}$$

$$\delta_\beta(Z) = d\left[m_\beta(X) - \mathbb{E}\left[m_\beta(X)|v, d = 1\right]\right]\frac{\partial\mathbb{E}\left[u| v, d = 1\right]}{\partial v}, \tag{114}$$

$$\delta_g(Z) = -d\varepsilon\frac{\partial\mathbb{E}\left[m_\beta(X)|v, d = 1\right]}{\partial v}, \tag{115}$$

$$\Psi_{\beta,\pi} = \mathbb{E}\left[\left(\delta_\beta(Z) - \delta_g(Z)\right)\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right], \text{ and} \tag{116}$$

$$\Psi_{\beta,\beta} = -\mathbb{E}\left[d\left(m_\beta(X)m_\beta(X)^\top - \mathbb{E}\left[m_\beta(X)|v, d = 1\right]\mathbb{E}\left[m_\beta(X)|v, d = 1\right]^\top\right)\right] \tag{117}$$

when π_* is parametrically specified. Using the components in (113), (116) and (117), the influence function of $\hat{\beta}$ in this example can be readily computed using Theorem 1.

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

$$\varphi_\pi(Z) = d - \pi_*(X, W) \tag{118}$$

and

$$\delta_\pi(W) \equiv -\mathbb{E}\left[\delta_\beta(Z) - \delta_g(Z)\middle|\, X, W\right] \tag{119}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (114) and (115) respectively. Using the components in (113), (117), (118) and (119), Theorem 2 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 strong identification condition (27) 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

$$\mathbb{E}\left[d\varepsilon \frac{\partial\mathbb{E}\left[m_\beta(X, \beta_*)|v, d = 1\right]}{\partial v} \frac{\partial v\left(\pi_*\right)}{\partial \pi^\top}\right] = 0 \tag{120}$$

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 (113), (116), (117) and (120) becomes identical to that in Newey (2009). $\quad\square$