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 11, 2022

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 different from 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 new 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, 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). In applied work, however, 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 $\beta_* \in \mathbb{R}^{d_\beta}$, which is identified together with an unknown function $\lambda_*(\cdot)$ as the unique solution of a minimization problem.[1] 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.[2] 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 two 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. 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. Second, we consider moment conditions which are different from those 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 impose conditional independence given just the control variable. For example, let's consider a model $Y = X\beta_* + \lambda_*(v) + \varepsilon$, where

[1] The unique identification of λ_* together with β_* is important since it enables us to characterize this nuisance parameter through the optimality condition, and then concentrate it out to derive the influence function of β_*. Similar assumption is imposed on the model listed in (3.11) of Newey (1994) which is a special case of (1) if the control variable v is observed.

[2] The generic function $\psi(z, \beta, \lambda(v(\pi_*)))$ may depend on different components of z in different ways. Some components of z may enter $\psi(z, \beta, \lambda)$ directly, while the rest may enter through the control variable $v(\pi_*)$. See the examples in Section 2 for specific illustrations.

$X = W^\top \pi_* + v$. A common assumption in the literature is $\mathbb{E}[\varepsilon | X, W] = 0$, which is different from $\mathbb{E}[\varepsilon | X, v] = 0$. We adopt variants of the latter assumption when deriving the influence function, and show that this type of assumptions does make a difference in the asymptotic distribution. Although the condition $\mathbb{E}[\varepsilon | X, W] = 0$ is commonly used in the early literature, recent applications of the control variable approach directly impose the conditional independence restriction given the control variable to achieve identification (see, e.g., Auerbach (2021) and Johnsson and Moon (2021)). Therefore, our results can be applied to derive the asymptotic distribution of the two-step estimators in these recent works.

The new moment condition, such as $\mathbb{E}[\varepsilon | X, v] = 0$ in the partially linear example above, is observed when we derive the influence function from the minimization problem (1) without specifying the control variable $v(\pi_*)$. Different economic motivations led to such a condition in Auerbach (2021) and Johnsson and Moon (2021), and we will treat it as a purely mathematical assumption in the current paper.[3] Therefore, we do not advocate that the new moment condition is in general weaker and/or better than the commonly used moment conditions in the control variable literature. Instead, what this paper tries to emphasize is that one should be aware of the different form of the asymptotic variance under the new moment condition.

This paper contributes to the line of research on semiparametric two-step estimation and inference with control variables. We next briefly discuss the mostly related works in the literature. First, Blundell and Powell (2004) apply the control variable method to address the endogeneity of regressors in the discrete choice models. Their identification condition leads to a matching based estimation procedure, which is different from (1). Escanciano, Jacho-Chávez, and Lewbel (2016) study a class of semiparametric model similar to Blundell and Powell (2004) but without excluded instruments. The finite dimensional parameter in Escanciano, Jacho-Chávez, and Lewbel (2016) is identified together with a nuisance function as the minimizer of a criterion which is fundamentally different from (1). Therefore, the influence function derived in this paper applies to a class of models which are different from Blundell and Powell (2004) and Escanciano, Jacho-Chávez, and Lewbel (2016). The other strand of research related to the control variable approach is on the inference of a finite dimensional parameter identified by some moment conditions which depend on some nuisance function. The nuisance parameter is separately identified, for example as the conditional expectation of a *known* dependent variable given some observed regressors and unobserved control variables (see, e.g., Hahn and Ridder (2013), Mammen, Rothe, and Schienle (2016) and Hahn and

[3] One possible intuition is based on the omitted variable argument. For example in the partially linear model above, one may view v as an omitted variable which causes the endogeneity of X and affects the response variable Y through a functional form $\lambda_*(v)$. When v is observed, one can estimate the parameter of interest β_* by the partially linear regression $Y = X\beta_* + \lambda_*(v) + \varepsilon$, where $\mathbb{E}[\varepsilon | X, v] = 0$ is one of the usual identification conditions in the literature. On the other hand, the control variable is often invoked when v is unobservable but it is obtained as the projection residual of X on W under the assumption $\mathbb{E}[\varepsilon | X, W] = 0$. Note that the condition $\mathbb{E}[\varepsilon | X, v] = 0$ is implied by $\mathbb{E}[\varepsilon | X, W] = 0$.

Ridder (2019)). Since the nuisance function and the finite dimensional parameter are jointly identified as the minimizer of the problem in (1), the models studied in these papers are also different from ours. Third, the parameter of interest in this paper is a root-n estimable finite dimensional parameter, which is different from the functional value of unknown functions studied in Newey, Powell, and Vella (1999), Das, Newey, and Vella (2003) and Mammen, Rothe, and Schienle (2012). Fourth, Hahn, Liao, and Ridder (2018) study general semiparametric two-step sieve estimation which nests (1). This paper and Hahn, Liao, and Ridder (2018) complement each other in the sense that: (i) Hahn, Liao, and Ridder (2018) provide a specific estimator of (β_*, λ_*) defined in (1) based on the sieve method and establish its root-n normality, but they do not provide the influence function of the sieve estimator because their asymptotic normality result is established using the sieve variance which is an approximation of the asymptotic variance; (ii) the influence function derived in this paper applies to the sieve estimator as well as other semi/nonparametric (such as kernel) estimators, and hence it is useful for finding the asymptotic variance and calculating the standard error for a general class of semiparametric methods. Finally as we shall discuss in the next section, the general model (1) nests and generalizes Li and Wooldridge (2002), Lee (2007), and Newey (2009) in specific examples.

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. 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

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

$$X = \varphi(W, \pi_*) + v, \ \mathbb{E}[v|W] = 0, \tag{3}$$

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

$$\mathbb{E}[u|X, W_0, v] = \mathbb{E}[u|v]. \tag{4}$$

The above condition is imposed on the control variable $v = X - \varphi(W, \pi_*)$ which is an "index" function of X and W. Let $\lambda_*(v) \equiv \mathbb{E}[u|v]$ and $\varepsilon \equiv u - \lambda_*(v)$. Then we can write (2) as

$$Y = m(X, W_0, \beta_*) + \lambda_*(v) + \varepsilon. \tag{5}$$

By the definition of $\lambda_*(v)$ and the restriction in (4)

$$\mathbb{E}[\varepsilon|X, W_0, v] = 0, \tag{6}$$

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

$$\min_{\beta,\lambda} \mathbb{E}\left[2^{-1}|Y - m(X, W_0, \beta) - \lambda(v)|^2\right]. \tag{7}$$

To construct feasible estimators of unknown parameters β_* and λ_* based on (7), 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 (3), 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

$$X = \pi_*(W) + v, \text{ where } \mathbb{E}[v|W] = 0. \tag{8}$$

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

$$(\hat{\beta}^\top, \hat{\gamma}^\top)^\top \equiv \underset{\beta \in \mathbb{R}^{d_\beta}, \gamma \in \mathbb{R}^{k_n}}{\arg\min} \sum_{i=1}^n 2^{-1}\left|Y_i - m(X_i, W_{0,i}, \beta) - P_{k_n}(\hat{v}_i)^\top \gamma\right|^2. \tag{9}$$

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 (4) is different from 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 different identification condition in (4). □

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

$$Y = X\beta_{1,\alpha,*} + W_0^\top \beta_{2,\alpha,*} + u, \tag{12}$$

$$X = \varphi(W, \pi_{\tilde{\alpha},*}) + v,\ \ Q_{v|W}^{\tilde{\alpha}}(w) = 0, \tag{13}$$

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, $Q_{v|W}^{\tilde{\alpha}}(\cdot)$ 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

$$Q_{u|X,W_0,v}^{\alpha}(x, w_0, v) = Q_{u|v}^{\alpha}(v), \tag{14}$$

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

$$Y = X\beta_{1,\alpha,*} + W_0^\top \beta_{2,\alpha,*} + \lambda_{\alpha,*}(v) + \varepsilon. \tag{15}$$

By the definition of $\lambda_{\alpha,*}(v)$ and the restriction in (14),

$$Q_{\varepsilon|X,W_0,v}^{\alpha}(x, w_0, v) = Q_{u|X,W_0,v}^{\alpha}(x, w_0, v) - \lambda_{\alpha,*}(v) = 0, \tag{16}$$

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 (15) and (16), the finite dimensional parameter $\beta_{\alpha,*}$ is identified together with the unknown function $\lambda_{\alpha,*}(v)$ as the minimizer of the following problem

$$\min_{\beta_1,\beta_2,\lambda} \mathbb{E}\left[\rho_\alpha(Y - X\beta_1 - W_0^\top \beta_2 - \lambda(v))\right], \tag{17}$$

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 (17) 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

$$X = \pi_{\tilde{\alpha},*}(W) + v, \text{ where } Q^{\tilde{\alpha}}_{v|W}(w) = 0 \tag{18}$$

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

$$(\hat{\beta}_\alpha^\top, \hat{\gamma}_\alpha^\top)^\top \equiv \underset{\beta \in \mathbb{R}^{d_\beta}, \gamma \in \mathbb{R}^{k_n}}{\arg\min} \sum_{i=1}^n \rho_\alpha(Y_i - (X_i, W_{0,i}^\top)\beta - P_{k_n}(\hat{v}_i)^\top \gamma), \tag{19}$$

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 (14) is imposed directly on the control variable $v(X, W, \pi_{\tilde{\alpha},*})$ which is a function of X and W. Therefore, (14) is different from, but implied by the following condition

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

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 (14) is different from that under (20). □

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

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

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 parameter 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 $\mathbb{E}[\varepsilon|X, v, d = 1] = 0$ 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 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

$$(\hat{\beta}^\top, \hat{\gamma}^\top)^\top \equiv \arg \min_{\beta\in\mathbb{R}^{d_\beta}, \gamma\in\mathbb{R}^{k_n}} \sum_{i=1}^n 2^{-1}d_i \left|Y_i - m(X_i, \beta) - P_{k_n}(\hat{v}_i)^\top\gamma\right|^2 \tag{26}$$

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, Stock, and Stoker (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 (22) which is implied by 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).[4] □

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

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

for any β and any π. The properties of $h\left(v\left(\pi\right);\beta,\pi\right)$ are characterized by the optimality condition of the above minimization problem, which together with the optimality condition of β_* enables us to derive the influence function. The following assumption is needed to obtain the optimality conditions.

Assumption 1. *(i) For any β, any π and any square integrable functions $\lambda\left(\cdot\right)$ and $\lambda_1\left(\cdot\right)$ of $v\left(\pi\right)$, there exist functions $\psi_\lambda(\cdot)$, $\psi_\beta(\cdot)$, $\psi_{\beta,\lambda}(\cdot)$, $\psi_{\beta,\beta}(\cdot)$, $\psi_{\lambda,\beta}(\cdot)$ and $\psi_{\lambda,\lambda}(\cdot)$ of z, β and λ such that*

$$\begin{aligned}
\left.\frac{\partial \mathbb{E}\left[\psi(Z,\beta,\lambda\left(v\left(\pi\right)\right)+\tau\lambda_1\left(v\left(\pi\right)\right))\right]}{\partial\tau}\right|_{\tau=0} &= \mathbb{E}\left[\psi_\lambda(Z,\beta,\lambda\left(v\left(\pi\right)\right))\lambda_1\left(v\left(\pi\right)\right)\right], \\
\frac{\partial \mathbb{E}\left[\psi(Z,\beta,\lambda(v(\pi)))\right]}{\partial\beta} &= \mathbb{E}\left[\psi_\beta(Z,\beta,\lambda(v(\pi)))\right], \\
\left.\frac{\partial \mathbb{E}\left[\psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)+\tau\lambda_1\left(v\left(\pi\right)\right)\right)\right]}{\partial\tau}\right|_{\tau=0} &= \mathbb{E}\left[\psi_{\beta,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\lambda_1\left(v\left(\pi\right)\right)\right], \\
\frac{\partial \mathbb{E}\left[\psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\right]}{\partial\beta^\top} &= \mathbb{E}\left[\psi_{\beta,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\right], \\
\frac{\partial \mathbb{E}\left[\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\right]}{\partial\beta} &= \mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\right], \text{ and} \\
\left.\frac{\partial \mathbb{E}\left[\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)+\tau\lambda_1\left(v\left(\pi\right)\right)\right)\right]}{\partial\tau}\right|_{\tau=0} &= \mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\lambda_1\left(v\left(\pi\right)\right)\right];
\end{aligned}$$

(ii) $\mathrm{Var}\left(\psi_\beta\left(Z,\beta_*,\lambda_*(v)\right)\right)$ *and* $\mathrm{Var}\left(\psi_\lambda\left(Z,\beta_*,\lambda_*(v)\right)\right)$ *are non-singular.*

The functions in condition (i) are defined using expectation to make our influence function formula applicable to models with non-smooth criterion function $\psi(\cdot)$, such as the check function in the quantile regression. When $\psi(z,\beta,\lambda)$ is smooth in β and λ and one can take differentiation under expectation, these functions can be obtained directly from the derivatives of $\psi(z,\beta,\lambda)$.[5] Condition (ii) ensures that the scores of the unknown parameters are not degenerate.

[4] 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.

[5] One can take differentiation under expectation under some regularity conditions such that the dominated convergence theorem can be applied. See Section C in the Appendix for illustration in specific examples.

From the optimality of $h(v(\pi);\beta,\pi)$, we have

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

for any function $\lambda(v(\pi))$ of $v(\pi)$ 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 derivative $\partial h\left(v\left(\pi\right);\beta,\pi\right)/\partial\beta$ exists by Assumption 1 and Assumption 2 below.[6]

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 $\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]/\partial\beta^{\top}$.[7]

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

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

The following assumption is also needed.

Assumption 2. *(i) Differentiation under expectation of $\psi_{\lambda}(\cdot)$ and $\psi_{\beta}(\cdot)$ is allowed; (ii) the influence function of $\hat{\pi}$ is $\varphi_{\pi}(z)$; (iii) $\Psi_{\beta,\beta}$ is non-singular; (iv) $\psi_{\lambda,\beta}(Z)=\psi_{\beta,\lambda}(Z)$ almost surely; (v) $h(v(\pi);\beta,\pi)$ is continuously differentiable in β and π and $v(\pi)$ is continuously differentiable in π for any β and π.*

Assumption 2 imposes some mild regularity conditions. Differentiation under expectation allows one to switch the order of expectation and differentiation, which is explicitly assumed in Newey (1994). Condition (i) is needed since our derivation of the influence function closely follows

[6] See (90) in the Appendix for the expression of $\partial h\left(v\left(\pi\right);\beta,\pi\right)/\partial\beta$.

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

the arguments in Newey (1994). Condition (ii) is on the influence function of $\hat{\pi}$, which can be derived from the specific first-stage estimation problem. Condition (iii) is a local identification condition on β_*, and Condition (iv) holds when $\mathbb{E}[\psi(Z,\beta,\lambda)]$ is twice continuously differentiable. Condition (v) imposes smoothness conditions on the function h and the control variable v.

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

Theorem 1 (Main Result)**.** *Suppose that Assumptions 1 and 2 hold. 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 $\hat{\beta}$.

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

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

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 $v(\pi_*) \equiv v(X, W, \pi_*)$ 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 specific models in the literature. On the other hand, in view of (29) the influence function of $\hat{\beta}$ derived here only uses

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

and (31), which is different from (40). Although condition (40) is popular in the early literature, recent applications of the control variable approach such as Auerbach (2021) and Johnsson and Moon (2021) use variants of (41), which are imposed on the control variables directly. Under the weaker condition (41), Theorem 1 shows that the extra term

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

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

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. We first present the identification condition of $\pi_{*,l}$.

Assumption 3. *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 square integrable function $\pi_l(\cdot)$, where $\mathbb{E}_\tau[\cdot]$ denotes the conditional expectation taken under the path density indexed by τ.

Assumption 4. *(i) Suppose that there exists a function $\mu_{l,\pi}(z_l, \pi_l)$ such that*

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

for any square integrable function $\pi_l(\cdot)$ and $l = 1, \ldots, L$; (ii) $v(z, \pi)$ is differentiable in $\pi \equiv (\pi_1, \ldots, \pi_L)^\top$ and it depends on π only through its value $\pi(w)$; (iii) $|\mathbb{E}[\mu_{l,\pi}(Z_l, \pi_{*,l})|W_l]| > 0$ almost surely for $l = 1, \ldots, L$.

Assumption 4 is mainly used to derive the effect of the first step nonparametric estimator on the influence function of $\hat{\beta}$. Condition (ii) imposes smoothness on the control variable in terms of π. Condition (iii) is intuitively a local identification condition of the unknown parameters $\pi_{*,l}$ $(l = 1, \ldots, L)$.

Theorem 2 (Nonparametric First Step). *Suppose that Assumptions 1, 2(i, iii, iv, v), 3 and 4 hold. 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 $\delta_\pi(W) \equiv (\delta_{1,\pi}(W_l), \ldots, \delta_{L,\pi}(W_L))^\top$,

$$\delta_{l,\pi}(W_l) \equiv \mathbb{E}\left[[\delta_\beta(Z) - \delta_g(Z)] \frac{\partial v(\pi_*)}{\partial \pi_l} \middle| W_l \right], \text{ and} \tag{43}$$

$$\varphi_\pi(Z) \equiv - \left(\frac{\mu_1(Z_1, \pi_{*,1})}{\mathbb{E}[\mu_{1,\pi}(Z_1, \pi_{*,1})|W_1]}, \cdots, \frac{\mu_L(Z_L, \pi_{*,L})}{\mathbb{E}[\mu_{L,\pi}(Z_L, \pi_{*,L})|W_L]} \right)^\top, \tag{44}$$

$\Psi_{\beta,\beta}$, $\varphi_\beta(\cdot)$, $\delta_\beta(\cdot)$ and $\delta_g(\cdot)$ are defined in (33), (35), (37) and (38), 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.[8] 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 notation, we let $Z_0 \equiv (X, W_0^\top)^\top$. In this example, we have

$$\psi(Z, \beta, \lambda(v(\pi))) = 2^{-1} (Y - m(Z_0, \beta) - \lambda(v(\pi)))^2. \tag{45}$$

Using the above expression, it is easy to calculate that[9]

$$\psi_\lambda(Z, \beta, \lambda(v(\pi))) = -(Y - m(Z_0, \beta) - \lambda(v(\pi))), \tag{46}$$

$$\psi_\beta(Z, \beta, \lambda(v(\pi))) = -(Y - m(Z_0, \beta) - \lambda(v(\pi)))\, m_\beta(Z_0, \beta), \tag{47}$$

$$\psi_{\lambda,\beta}(Z, \beta, \lambda(v(\pi))) = m_\beta(Z_0, \beta) = \psi_{\beta,\lambda}(Z, \beta, \lambda(v(\pi))), \tag{48}$$

$$\psi_{\beta,\beta}(Z, \beta, \lambda(v(\pi))) = m_\beta(Z_0, \beta) m_\beta(Z_0, \beta)^\top - (Y - m(Z_0, \beta) - \lambda(v(\pi)))\, m_{\beta,\beta}(Z_0, \beta), \tag{49}$$

$$\psi_{\lambda,\lambda}(Z, \beta, \lambda(v(\pi))) = 1, \tag{50}$$

[8] Similar calculations apply to the third example in Section 2. The details can be found in Appendix D.
[9] See Section C in the Appendix for low-level sufficient conditions to justify Assumption 1.

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

$$m_\beta(Z_0,\beta) \equiv \frac{\partial m(Z_0,\beta)}{\partial \beta} \text{ and } m_{\beta,\beta}(Z_0,\beta) \equiv \frac{\partial^2 m(Z_0,\beta)}{\partial \beta \partial \beta^\top}. \tag{51}$$

By (46), the first-order condition of the profiled nonparametric function $h(v(\pi);\beta,\pi)$ can be written as

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

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

$$h(v(\pi);\beta,\pi) = \mathbb{E}\left[Y - m(Z_0,\beta)|\, v(\pi)\right], \tag{52}$$

and therefore

$$h(v(\pi_*);\beta_*,\pi_*) = \mathbb{E}\left[u|\, v\right] = \lambda_*(v). \tag{53}$$

Let $m_\beta(Z_0) \equiv m_\beta(Z_0,\beta_*)$. Using the expressions in (46)-(50) and (53), we get

$$\varphi_\beta(Z) = -\varepsilon(m_\beta(Z_0) - \mathbb{E}\left[m_\beta(Z_0)|v\right]), \tag{54}$$

$$\delta_\beta(Z) = (m_\beta(Z_0) - \mathbb{E}\left[m_\beta(Z_0)|\, v\right])\frac{\partial \mathbb{E}\left[u|v\right]}{\partial v}, \tag{55}$$

$$\delta_g(Z) = -\varepsilon\frac{\partial \mathbb{E}\left[m_\beta(Z_0)|\, v\right]}{\partial v} \text{ and} \tag{56}$$

$$\Psi_{\beta,\pi} = \mathbb{E}\left[\left((m_\beta(Z_0) - \mathbb{E}\left[m_\beta(Z_0)|\, v\right])\frac{\partial \mathbb{E}\left[u|v\right]}{\partial v} + \varepsilon\frac{\partial \mathbb{E}\left[m_\beta(Z_0)|\, v\right]}{\partial v}\right)\frac{\partial v(\pi_*)}{\partial \pi^\top}\right] \tag{57}$$

when π_* is a finite-dimensional parameter vector. From (6), (46)-(50), the Hessian matrix takes the following form

$$\Psi_{\beta,\beta} = -\mathbb{E}\left[m_\beta(Z_0)m_\beta(Z_0)^\top - \mathbb{E}\left[m_\beta(Z_0)|\, v\right]\mathbb{E}\left[m_\beta(Z_0)|\, v\right]^\top\right]. \tag{58}$$

From the components in (54), (57) and (58), 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.

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

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

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

$$\varphi_\pi(Z) = X - \pi_*(W) \tag{59}$$

and

$$\delta_\pi(W) = -\mathbb{E}\left[\delta_\beta(Z) - \delta_g(Z)|\, W\right], \tag{60}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (55) and (56) respectively. Using the components in (54), (58), (59) and (60), 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 condition (10) holds, we obtain

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

in the case with a parametric first step, and

$$\delta_\pi(W)=-\mathbb{E}\left[(m_\beta(Z_0)-\mathbb{E}\left[m_\beta(Z_0)|\,v\right])\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v}\middle|\,W\right] \tag{62}$$

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 $v\left(\pi_*\right)=X-W^\top\pi_*$ to derive the main results. Under these extra conditions,

$$\varphi_\beta(Z)=-\varepsilon\left(W_0-\mathbb{E}\left[W_0|\,v\right]\right), \tag{63}$$

$$\Psi_{\beta,\pi}=-\mathbb{E}\left[\left(W_0-\mathbb{E}\left[W_0|\,v\right]\right)\frac{\partial\mathbb{E}\left[u|v\right]}{\partial v}W^\top\right], \tag{64}$$

$$\Psi_{\beta,\beta}=-\mathbb{E}\left[W_0W_0^\top-\mathbb{E}\left[W_0|\,v\right]\mathbb{E}\left[W_0|\,v\right]^\top\right]. \tag{65}$$

The influence function of the two-step estimator $\hat{\beta}$ can be calculated using Theorem 1, the items in (63)-(65) 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

$$\psi\left(Z,\beta,\lambda(v(\pi))\right)=\rho_\alpha(Y-Z_0^\top\beta-\lambda(v)). \tag{66}$$

Using the above expression, it is easy to calculate that[10]

$$\left.\frac{\partial\mathbb{E}\left[\rho_\alpha(Y-Z_0^\top\beta-\lambda\left(v\left(\pi\right)\right)-\tau\lambda_1\left(v\left(\pi\right)\right))\right]}{\partial\tau}\right|_{\tau=0}$$
$$=\mathbb{E}\left[\left(1\left\{Y\leq Z_0^\top\beta+\lambda\left(v\left(\pi\right)\right)\right\}-\alpha\right)\lambda_1\left(v\left(\pi\right)\right)\right] \tag{67}$$

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

$$\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=1\left\{Y\leq Z_0^\top\beta+\lambda\left(v\left(\pi\right)\right)\right\}-\alpha. \tag{68}$$

[10] See Section C in the Appendix for low-level sufficient conditions to justify Assumption 1.

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[11]

$$h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)=Q^{\alpha}_{u|v}\left(v\right)=\lambda_*(v). \tag{69}$$

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 (66), it is easy to calculate that

$$\psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=\left(1\left\{Y\leq Z_0^\top\beta+\lambda\left(v\left(\pi\right)\right)\right\}-\alpha\right)Z_0, \tag{70}$$

$$\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=f_\varepsilon\left(Z_0^\top\left(\beta-\beta_*\right)+\lambda\left(v\left(\pi\right)\right)-\lambda_*(v)\middle|\,Z_0,v\left(\pi\right)\right)Z_0, \tag{71}$$

$$\psi_{\beta,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=f_\varepsilon\left(Z_0^\top\left(\beta-\beta_*\right)+\lambda\left(v\left(\pi\right)\right)-\lambda_*(v)\middle|\,Z_0,v\left(\pi\right)\right)Z_0Z_0^\top, \tag{72}$$

$$\psi_{\lambda,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)=f_\varepsilon\left(Z_0^\top\left(\beta-\beta_*\right)+\lambda\left(v\left(\pi\right)\right)-\lambda_*(v)\middle|\,Z_0,v\left(\pi\right)\right), \tag{73}$$

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 (68)-(73), we get

$$\varphi_\beta(Z)=\left(1\left\{\varepsilon\leq 0\right\}-\alpha\right)\left(Z_0-g_*(v)\right), \tag{74}$$

$$\delta_\beta(Z)=f_\varepsilon\left(0|Z_0,v\right)\left(Z_0-g_*(v)\right)\frac{\partial Q^{\alpha}_{u|v}\left(v\right)}{\partial v}, \tag{75}$$

$$\delta_g(Z)=\left(1\left\{\varepsilon\leq 0\right\}-\alpha\right)\frac{\partial g_*\left(v\right)}{\partial v},\text{ and} \tag{76}$$

$$\Psi_{\beta,\pi}=\mathbb{E}\left[\left(f_\varepsilon\left(0|Z_0,v\right)\left(Z_0-g_*(v)\right)\frac{\partial Q^{\alpha}_{u|v}\left(v\right)}{\partial v}-\left(1\left\{\varepsilon\leq 0\right\}-\alpha\right)\frac{\partial g_*\left(v\right)}{\partial v}\right)\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right] \tag{77}$$

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 (69)-(73), the Hessian matrix takes the following form

$$\Psi_{\beta,\beta}=-\mathbb{E}\left[f_\varepsilon\left(0|Z_0,v\right)\left(Z_0Z_0^\top-g_*(v)g_*(v)^\top\right)\right]. \tag{78}$$

Using the components in (74), (77) and (78), 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.

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

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

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

Therefore in this case,

$$\varphi_\pi(Z) = -\frac{1\{X \leq \pi_*(W)\} - \tilde{\alpha}}{f_{X|W}(\pi_*(W))}, \tag{79}$$

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

$$\delta_\pi(W) = -\mathbb{E}\left[\delta_\beta(Z) - \delta_g(Z)|\, W\right], \tag{80}$$

where $\delta_\beta(Z)$ and $\delta_g(Z)$ are defined in (75) and (76) respectively. Using the components in (74), (78), (79) and (80), Theorem 2 implies that the influence function of the two-step estimator in this case is

$$\Psi_{\beta,\beta}^{-1}\left(\varphi_\beta(Z) - \delta_\pi(W)\frac{1\{X_i \leq \pi_*(W)\} - \tilde{\alpha}}{f_{X|W}(\pi_*(W))}\right). \tag{81}$$

If condition (20) holds, we get

$$\Psi_{\beta,\pi} = \mathbb{E}\left[f_\varepsilon\left(0|Z_0, v\right)\left(Z_0 - g_*(v)\right)\frac{\partial Q^\alpha_{u|v}\left(v\right)}{\partial v}W^\top\right] \tag{82}$$

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

$$\delta_\pi(W) = -\mathbb{E}\left[f_\varepsilon\left(0|Z_0, v\right)\left(Z_0 - g_*(v)\right)\frac{\partial Q^\alpha_{u|v}\left(v\right)}{\partial v}\middle| W\right] \tag{83}$$

in the case with a nonparametric first step. Therefore the influence function of $\hat{\beta}$ is slightly simplified in both cases. Moreover, under (13) and (20), the asymptotic variance of $\hat{\beta}$ implied by its influence function (which can be calculated using (74), (78) and (82)) 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 an 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 different from 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 may not be negligible. 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 taken for identification and consistent estimation of structural parameters with endogenous explanatory variables.

In the three examples discussed in the paper, the index restriction used for deriving the influence function turns out to be implied by the commonly adopted conditional moment condition given all exogenous variables, which indicates that the latter conditional moment condition may provide over-identification restriction to the unknown parameters. Therefore, the commonly used estimation procedure based on (1) may be inefficient since it is invariant under the index restriction and the conditional moment condition. To the best of our knowledge, semiparametric efficiency and efficient estimation haven't been explored in the control variable literature. Some fruitful results may be obtained through generalizing the existing methodology on the semiparametric efficiency bound calculation (see, e.g., Newey (1990) and Chamberlain (1992), and also see Severini and Tripathi (2013) for a recent survey), to allow for generated control variables in the conditional moment condition. The semiparametric efficiency, as well as the influence function of the efficient estimator, is left for future investigations.

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. (2021): "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.

Chamberlain, G. (1992): "Efficiency bounds for semiparametric regression," *Econometrica: Journal of the Econometric Society*, pp. 567–596.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric estimation of sample selection models," *The Review of Economic Studies*, 70(1), 33–58.

Escanciano, J. C., D. Jacho-Chávez, and A. Lewbel (2016): "Identification and estimation of semiparametric two-step models," *Quantitative Economics*, 7(2), 561–589.

Hahn, J., Z. Liao, and G. Ridder (2018): "Nonparametric two-step sieve M estimation and inference," *Econometric Theory*, 34(6), 1281–1324.

Hahn, J., and G. Ridder (2013): "Asymptotic variance of semiparametric estimators with generated regressors," *Econometrica*, 81(1), 315–340.

——— (2019): "Three-stage semi-parametric inference: Control variables and differentiability," *Journal of econometrics*, 211(1), 262–293.

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 (2021): "Estimation of peer effects in endogenous social networks: Control function approach," *Review of Economics and Statistics*, 103(2), 328–345.

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.

Mammen, E., C. Rothe, and M. Schienle (2012): "Nonparametric regression with nonparametrically generated covariates," *The Annals of Statistics*, 40(2), 1132–1170.

——— (2016): "Semiparametric estimation with generated covariates," *Econometric Theory*, 32(5), 1140–1177.

Newey, W. K. (1990): "Semiparametric efficiency bounds," *Journal of applied econometrics*, 5(2), 99–135.

——— (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.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67(3), 565–603.

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.

Severini, T. A., and G. Tripathi (2013): "Semiparametric efficiency bounds for microeconometric models: A survey," *Foundations and Trends® in Econometrics*, 6(3–4), 163–397.

Appendix

A Proof of Theorem 1

The theorem is proved using the arguments in Sections 2 and 3 of Newey (1994). Specifically under Assumptions 1(i) and 2(i), (3.10) in Newey (1994) shows that the influence function of $\hat{\beta}$ can be derived from (31), and it takes the following form

$$-\left(\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial \beta^\top}\right)^{-1}\left(J(Z,\beta_*,\pi_*)+\eta(Z)\right), \tag{84}$$

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

$$\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_{*,\tau})\right]}{\partial \tau}=\mathbb{E}\left[\eta(Z)\frac{\partial \ln(f_{z,\tau}(Z))}{\partial \tau}\right], \tag{85}$$

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

$$\frac{\partial \pi_{*,\tau}}{\partial \tau}=\mathbb{E}\left[\varphi_\pi(Z)\frac{\partial \ln(f_{z,\tau}(Z))}{\partial \tau}\right]. \tag{86}$$

From (85) and (86), 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

$$-\left(\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial \beta^\top}\right)^{-1}\left(J(Z,\beta_*,\pi_*)+\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial \pi^\top}\varphi_\pi(Z)\right) \tag{87}$$

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 (94), (100) and (101) 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

$$\mathbb{E}\left[\left.\psi_\lambda\left(Z,\beta_*,\lambda_*\left(v\right)\right)\right|v\right]=0. \tag{88}$$

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

$$\mathbb{E}\left[\psi_\lambda\left(Z,\beta,h\left(v\left(\pi\right);\beta,\pi\right)\right)\lambda\left(v\left(\pi\right)\right)\right]=0 \tag{89}$$

for any function $\lambda\left(v\left(\pi\right)\right)$ of $v(\pi)$. Evaluating (89) 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 $\lambda(v)$ of v, which immediately implies (88).

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

$$\frac{\partial h(v(\pi);\beta_*,\pi)}{\partial\beta}=-\frac{\mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_*,h(v(\pi);\beta_*,\pi)\right)\middle|\,v(\pi)\right]}{\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_*,h(v(\pi);\beta_*,\pi)\right)\middle|\,v(\pi)\right]}. \tag{90}$$

Under Assumptions 1(i) and 2(i, v), we can differentiate (89) with respect to β and apply the chain rule to obtain

$$0=\mathbb{E}\left[\left(\psi_{\lambda,\beta}(Z,\beta,h(v(\pi);\beta,\pi))+\psi_{\lambda,\lambda}(Z,\beta,h(v(\pi);\beta,\pi))\frac{\partial h(v(\pi);\beta,\pi)}{\partial\beta}\right)\lambda(v(\pi))\right] \tag{91}$$

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

$$0=\mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_*,h(v(\pi);\beta_*,\pi)\right)+\psi_{\lambda,\lambda}\left(Z,\beta_*,h(v(\pi);\beta_*,\pi)\right)\frac{\partial h(v(\pi);\beta_*,\pi)}{\partial\beta}\middle|\,v(\pi)\right], \tag{92}$$

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

$$\begin{aligned}0&=\mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_*,h(v(\pi);\beta_*,\pi)\right)\middle|\,v(\pi)\right]\\&\quad+\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_*,h(v(\pi);\beta_*,\pi)\right)\middle|\,v(\pi)\right]\frac{\partial h(v(\pi);\beta_*,\pi)}{\partial\beta}.\end{aligned} \tag{93}$$

The claim in (90) follows from (93).

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 $g_*(v)$ in (33), and the expression in (90), we get

$$\begin{aligned}J(Z,\beta_*,\pi_*)&=\psi_\beta\left(Z,\beta_*,\lambda_*(v)\right)+\psi_\lambda\left(Z,\beta_*,\lambda_*(v)\right)\frac{\partial h(v(\pi_*);\beta_*,\pi_*)}{\partial\beta}\\&=\psi_\beta\left(Z,\beta_*,\lambda_*(v)\right)-g_*(v)\psi_\lambda\left(Z,\beta_*,\lambda_*(v)\right)=\varphi_\beta(Z).\end{aligned} \tag{94}$$

Next from (31), we observe that

$$\begin{aligned}\frac{\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial\pi^\top}&=\mathbb{E}\left[\psi_{\beta,\lambda}\left(Z,\beta_*,\lambda_*(v)\right)\frac{d}{d\pi^\top}h(v(\pi_*);\beta_*,\pi_*)\right]\\&\quad+\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*(v)\right)g_*(v)\frac{d}{d\pi^\top}h(v(\pi_*);\beta_*,\pi_*)\right]\\&\quad-\mathbb{E}\left[\psi_\lambda\left(Z,\beta_*,\lambda_*(v)\right)\frac{d}{d\pi^\top}g(v(\pi_*);\pi_*)\right],\end{aligned} \tag{95}$$

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

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

So we have

$$\frac{d}{d\pi^\top} h\left(v\left(\pi_*\right);\beta_*,\pi_*\right) \equiv \frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_1^\top} + \frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}. \tag{96}$$

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

$$\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_1^\top} = \frac{\partial\lambda_*\left(v(\pi_*)\right)}{\partial v}\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}. \tag{97}$$

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 (90), (92) and the definition of $g_*(v)$ implies that

$$\begin{aligned}
&\mathbb{E}\left[\psi_{\beta,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}\right] \\
&+\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\beta}\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}\right] \\
&=\mathbb{E}\left[\mathbb{E}\left[\psi_{\lambda,\beta}\left(Z,\beta_*,\lambda_*\left(v\right)\right) - g_*(v)\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\middle| v\right]\frac{\partial h\left(v\left(\pi_*\right);\beta_*,\pi_*\right)}{\partial\pi_2^\top}\right] = 0,
\end{aligned} \tag{98}$$

where we also used Assumption 2(iv), 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 (95), (96), (97) and (98), we get

$$\begin{aligned}
\frac{\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial\pi^\top} &= \mathbb{E}\left[\left[\psi_{\lambda,\beta}\left(Z,\beta_*,\lambda_*(v)\right) - g_*(v)\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*(v)\right)\right]\frac{\partial\lambda_*\left(v(\pi_*)\right)}{\partial v}\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right] \\
&\quad - \mathbb{E}\left[\psi_\lambda\left(Z,\beta_*,\lambda_*(v)\right)\frac{d}{d\pi^\top}g(v\left(\pi_*\right);\pi_*)\right].
\end{aligned} \tag{99}$$

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 (88),

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

which together with (99), the definition of $g_*(v)$, and $\partial g\left(v\left(\pi_*\right),\pi_*\right)/\partial v = \partial g_*(v)/\partial v$ implies that

$$\begin{aligned}
\frac{\partial\mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial\pi^\top} &= \mathbb{E}\left[\left[\psi_{\lambda,\beta}\left(Z,\beta_*,\lambda_*\left(v\right)\right) - g_*(v)\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)\right]\frac{\partial\lambda_*\left(v(\pi_*)\right)}{\partial v}\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right] \\
&\quad - \mathbb{E}\left[\psi_\lambda\left(Z,\beta_*,\lambda_*\left(v\right)\right)\frac{\partial g_*\left(v\right)}{\partial v}\frac{\partial v\left(\pi_*\right)}{\partial\pi^\top}\right] = \Psi_{\beta,\pi}.
\end{aligned} \tag{100}$$

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

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

which together with (88), (90) and the definition of $g_*(v)$ implies that

$$\begin{aligned}\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_*)\right]}{\partial\beta^\top} &= \mathbb{E}\left[\psi_{\beta,\beta}\left(Z,\beta_*,\lambda_*\left(v\right)\right)-\psi_{\beta,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)g_*(v)^\top\right]\\ &-\mathbb{E}\left[g_*(v)\psi_{\lambda,\beta}\left(Z,\beta_*,\lambda_*\left(v\right)\right)^\top\right]+\mathbb{E}\left[\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)g_*(v)g_*(v)^\top\right]\\ &= \mathbb{E}\left[\psi_{\beta,\beta}\left(Z,\beta_*,\lambda_*\left(v\right)\right)-\psi_{\lambda,\lambda}\left(Z,\beta_*,\lambda_*\left(v\right)\right)g_*(v)g_*(v)^\top\right]=-\Psi_{\beta,\beta}. \qquad (101)\end{aligned}$$

Plugging the 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 obtained in (94), (100) and (101) respectively) in (87), and applying Assumptions 2(ii, iii), we obtain the influence function stated in the theorem.

B Proof of Theorem 2

Taking derivative with respect to τ in (42) and applying the chain rule and Assumption 4(i), we get

$$\frac{\partial \mathbb{E}_\tau\left[\mu_l(Z_l,\pi_{*,l})\pi_l(W_l)\right]}{\partial\tau}+\mathbb{E}\left[\mathbb{E}\left[\mu_{l,\pi}(Z_l,\pi_{*,l})\middle| W_l\right]\pi_l(W_l)\frac{\partial\pi_{*,l,\tau}(W_l)}{\partial\tau}\right]=0 \qquad (102)$$

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 (31). Using similar calculation as in (95), (98), (99) and (100) in the proof of Theorem 1, we obtain

$$\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_{*,\tau})\right]}{\partial\tau}=\frac{\partial \mathbb{E}\left[D(Z,\pi_{*,\tau})\right]}{\partial\tau}, \qquad (103)$$

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

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

which is linear in π_τ. Note that by (37), (38) and (43), $D(Z,\pi_\tau)$ thus defined satisfies

$$\mathbb{E}\left[D_\tau(Z,\pi)\right]=\mathbb{E}\left[\delta_\pi(W)^\top\pi_\tau(W)\right] \qquad (104)$$

for any $\pi_\tau(W)$ by the law of iterated expectation. Combining (103) and (104), we deduce that

$$\begin{aligned}\frac{\partial \mathbb{E}\left[J(Z,\beta_*,\pi_{*,\tau})\right]}{\partial \tau} = \frac{\partial \mathbb{E}\left[D(Z,\pi_{*,\tau})\right]}{\partial \tau} &= \mathbb{E}\left[\delta_\pi(W)^\top \frac{\partial \pi_{*,\tau}(W)}{\partial \tau}\right] \\ &= \sum_{l=1}^{L} \mathbb{E}\left[\delta_{l,\pi}(W_l)\frac{\partial \pi_{*,l,\tau}(W_l)}{\partial \tau}\right] \\ &= \sum_{l=1}^{L} \frac{\partial}{\partial \tau}\mathbb{E}_\tau\left[-\frac{\mu_l(Z_l,\pi_{*,l})}{\mathbb{E}\left[\mu_{l,\pi}(Z_l,\pi_{*,l})\middle| W_l\right]}\delta_{l,\pi}(W_l)\right] \\ &= \frac{\partial}{\partial \tau}\mathbb{E}_\tau\left[\delta_\pi(W)^\top \varphi_\pi(Z)\right], \end{aligned} \tag{105}$$

where the third equality follows from (102) 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 $l = 1,\dots,L$. Therefore, (3.9) in Newey (1994) follows from (105) and Theorem 2 is directly implied by Theorem 2.1 of Newey (1994).

C Low-level Conditions of Assumption 1 in Examples 1 and 2

Example 1 (Mean Regression Revisited). To verify Assumption 1(i) in this example, we assume that $\lambda(v(\pi))$ and $\mathbb{E}\left[Y|Z_0 = z_0\right]$ are uniformly bounded, $m(z_0,\beta)$ and its first and second order derivatives in β are also uniformly bounded.

From the definition of $\psi(\cdot)$, we can write

$$\mathbb{E}\left[\psi(Z,\beta,\lambda\left(v\left(\pi\right)\right))\right] = 2^{-1}(\mathbb{E}[Y^2] + \mathbb{E}[(m(Z_0,\beta)+\lambda(v(\pi)))^2]) - \mathbb{E}\left[Y(m(Z_0,\beta)+\lambda(v(\pi)))\right].$$

Since $\lambda(v(\pi))$, $\mathbb{E}\left[Y|Z_0 = z_0\right]$ and the first order derivative of $m(z_0,\beta)$ in β are uniformly bounded, we can use the dominated convergence theorem to find $\psi_\lambda(\cdot)$ and $\psi_\beta(\cdot)$ directly from the derivatives of $\psi\left(Z,\beta,\lambda(v(\pi))\right)$. That is,

$$\begin{aligned}\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) &= -(Y - m(Z_0,\beta) - \lambda\left(v\left(\pi\right)\right)), \\ \psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) &= -\left(Y - m(Z_0,\beta) - \lambda\left(v\left(\pi\right)\right)\right) m_\beta(Z_0,\beta).\end{aligned}$$

The same arguments obtain $\psi_{\lambda,\beta}(\cdot)$ and $\psi_{\lambda,\lambda}(\cdot)$ from the derivatives of $\psi_\lambda\left(\cdot\right)$:

$$\begin{aligned}\psi_{\lambda,\beta}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) &= m_\beta(Z_0,\beta), \\ \psi_{\lambda,\lambda}\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right) &= 1.\end{aligned}$$

Since $\psi_\beta\left(\cdot\right)$ is linear in $\lambda\left(v\left(\pi\right)\right)$, we immediately have

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

Moreover since the second derivative of $m(z_0,\beta)$ in β is also uniformly bounded, we can again use the dominated convergence theorem to find $\psi_{\beta,\beta}(\cdot)$ from the derivatives of $\psi_\beta\left(\cdot\right)$:

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

Since $\psi_\lambda(Z, \beta_*, \lambda_*(v)) = -\varepsilon$ and $\psi_\beta(Z, \beta_*, \lambda_*(v)) = -m_\beta(Z_0, \beta_*)\varepsilon$ in this example, Assumption 1(ii) holds if $\mathbb{E}[\varepsilon^2] > 0$ and $\mathbb{E}[\varepsilon^2 m_\beta(Z_0, \beta_*) m_\beta(Z_0, \beta_*)^\top]$ is positive definite. $\square$

Example 2 (Quantile Regression Revisited). To verify Assumption 1(i) in this example, we assume that β is in some bounded subset of $\mathbb{R}^{d_\beta}$, Z_0 has compact support, $\lambda(v(\pi))$ is uniformly bounded, the conditional probability density function of Y given Z_0 and $v(\pi)$, denoted as $f_Y(y|Z_0, v(\pi))$ and its first order derivative in y are uniformly bounded.

From the definition of $\psi(\cdot)$ in this example, we can write

$$\begin{aligned}\mathbb{E}\left[\rho_\alpha(Y - Z_0^\top\beta - \lambda(v(\pi)))\right] &= \mathbb{E}\left[(\alpha - 1\{Y \le Z_0^\top\beta + \lambda(v(\pi))\})(Y - Z_0^\top\beta - \lambda(v(\pi)))\right] \\ &= \alpha\mathbb{E}[Y] - \mathbb{E}\left[Y1\{Y \le Z_0^\top\beta + \lambda(v(\pi))\}\right] \\ &\quad - \mathbb{E}\left[(\alpha - F_Y(Z_0^\top\beta + \lambda(v(\pi))|Z_0, v(\pi)))(Z_0^\top\beta + \lambda(v(\pi)))\right],\end{aligned}$$

where $F_Y(\cdot|Z_0, v(\pi))$ denotes the conditional cumulative distribution of Y given Z_0 and $v(\pi)$. Under the maintained low-level conditions, we can use the dominated convergence theorem to get

$$\begin{aligned}&\left.\frac{\partial\mathbb{E}\left[Y1\{Y \le Z_0^\top\beta + \lambda(v(\pi)) + \tau\lambda_1(v(\pi))\}\right]}{\partial\tau}\right|_{\tau=0} \\ &= \mathbb{E}\left[f_Y(Z_0^\top\beta + \lambda(v(\pi))|Z_0, v(\pi))(Z_0^\top\beta + \lambda(v(\pi)))\lambda_1(v(\pi))\right]\end{aligned}$$

and

$$\begin{aligned}&\left.\frac{\partial\mathbb{E}\left[(\alpha - F_Y(Z_0^\top\beta + \lambda(v(\pi)) + \tau\lambda_1(v(\pi))|Z_0, v(\pi)))(Z_0^\top\beta + \lambda(v(\pi)) + \tau\lambda_1(v(\pi)))\right]}{\partial\tau}\right|_{\tau=0} \\ &= \mathbb{E}\left[(\alpha - F_Y(Z_0^\top\beta + \lambda(v(\pi))|Z_0, v(\pi)))\lambda_1(v(\pi))\right] \\ &- \mathbb{E}\left[f_Y(Z_0^\top\beta + \lambda(v(\pi))|Z_0, v(\pi))(Z_0^\top\beta + \lambda(v(\pi)))\lambda_1(v(\pi))\right].\end{aligned}$$

Therefore

$$\begin{aligned}&\left.\frac{\partial\mathbb{E}\left[\rho_\alpha(Y - Z_0^\top\beta - \lambda(v(\pi)) - \tau\lambda_1(v(\pi)))\right]}{\partial\tau}\right|_{\tau=0} \\ &= \mathbb{E}\left[(F_Y(Z_0^\top\beta + \lambda(v(\pi))|Z_0, v(\pi)) - \alpha)\lambda_1(v(\pi))\right] \\ &= \mathbb{E}\left[(1\{Y \le Z_0^\top\beta + \lambda(v(\pi))\} - \alpha)\lambda_1(v(\pi))\right],\end{aligned}$$

which shows (67). Similarly

$$\begin{aligned}\frac{\partial\mathbb{E}\left[\rho_\alpha(Y - Z_0^\top\beta - \lambda(v(\pi)))\right]}{\partial\beta} &= \mathbb{E}\left[(F_Y(Z_0^\top\beta + \lambda(v(\pi))|Z_0, v(\pi)) - \alpha)Z_0\right] \\ &= \mathbb{E}\left[(1\{Y \le Z_0^\top\beta + \lambda(v(\pi))\} - \alpha)Z_0\right],\end{aligned}$$

which justifies (70). Using

$$\mathbb{E}\left[\psi_\lambda\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\right]=\mathbb{E}\left[F_Y(Z_0^\top\beta+\lambda\left(v\left(\pi\right)\right)|Z_0,v\left(\pi\right))-\alpha\right]$$

and

$$\mathbb{E}\left[\psi_\beta\left(Z,\beta,\lambda\left(v\left(\pi\right)\right)\right)\right]=\mathbb{E}\left[(F_Y(Z_0^\top\beta+\lambda\left(v\left(\pi\right)\right)|Z_0,v\left(\pi\right))-\alpha)Z_0\right],$$

and invoking the dominated convergence theorem under the maintained low level conditions, one can directly find the expressions of $\psi_{\lambda,\beta}(\cdot)$, $\psi_{\beta,\beta}(\cdot)$ and $\psi_{\lambda,\lambda}(\cdot)$ in (71), (72) and (73), respectively.

Since $\psi_\lambda\left(Z,\beta_*,\lambda_*\left(v\right)\right)=1\left\{\varepsilon\leq 0\right\}-\alpha$ and $\psi_\beta\left(Z,\beta_*,\lambda_*\left(v\right)\right)=\left(1\left\{\varepsilon\leq 0\right\}-\alpha\right)Z_0$ in this example, Assumption 1(ii) holds if $\mathbb{E}[(1\left\{\varepsilon\leq 0\right\}-\alpha)^2]>0$ and $\mathbb{E}[(1\left\{\varepsilon\leq 0\right\}-\alpha)^2Z_0Z_0^\top]$ is positive definite. □

D 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

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

which implies that in this example

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

where the second equality is by

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

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 (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) - (38), 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

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

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 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). □