Supplement to "The influence function of semiparametric estimators"

(*Quantitative Economics*, Vol. 13, No. 1, January 2022, 29–61)

Hidehiko Ichimura
Department of Economics, University of Arizona and Department of Economics, University of Tokyo

Whitney K. Newey
Department of Economics, MIT

Appendix A: Validity of the influence function calculation

In this Appendix, we show validity of Steps I and II of the influence function calculation. Step I requires differentiability of $\theta(F_\tau)$ and the formula

$$\frac{d\theta(F_\tau)}{d\tau} = \int \psi(w)H(dw), \qquad E\big[\psi(W)\big] = 0, \qquad E\big[\psi(W)^2\big] < \infty. \tag{1}$$

Step II requires that evaluating the derivative at a point mass gives the influence function. We justify Step II as a limit as H approaches a point mass similar to Lebesgue differentiation from analysis. Lebesgue differentiation shows that the limit of an integral of a function over an interval divided by the length of the interval converges almost surely to the value of the function at a point as the interval collapses on that point. We give regularity conditions and classes of continuous, smooth probability distributions where the expectation of the influence function converges to its value at a point as the probability distribution collapses on the point.

The fundamental starting point for the influence function calculation is that the estimator is asymptotically linear with an influence function, that is, that it satisfies

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i) + o_p(1), \qquad E\big[\psi(W)\big] = 0, \qquad E\big[\psi(W)^T \psi(W)\big] < \infty.$$

We take a modern, high level approach to regularity conditions in assuming that the estimator is locally regular for a set of alternative distributions H that can approximate a point mass.

Definition A1. $\hat{\theta}$ is locally regular for F_τ if there is a fixed random variable Y such that for any $\tau_n = O(1/\sqrt{n})$ and $W_1, \dots, W_n$ i.i.d. with distribution F_{τ_n},

$$\sqrt{n}\big[\hat{\theta} - \theta(F_{\tau_n})\big] \xrightarrow{d} Y.$$

Hidehiko Ichimura: ichimura@arizona.edu
Whitney K. Newey: wnewey@mit.edu

© 2022 The Authors. Licensed under the Creative Commons Attribution-NonCommercial License 4.0.
Available at http://qeconomics.org. https://doi.org/10.3982/QE826

This local regularity condition is familiar from the efficient estimation literature. Local regularity of $\hat{\theta}$ is not a primitive condition but it is plausible when F_0 satisfies conditions for existence of $\theta(F)$ and H is well behaved relative to F_0. For example, F_0 could satisfy regularity conditions like some random variables being continuously distributed and expectations of certain functions existing and H could be a uniformly bounded, smooth deviation from F_0. In such settings, it is plausible many estimators $\hat{\theta}$ would be locally regular. We construct such H in this Appendix so that local regularity is plausibly satisfied for many semiparametric estimators $\hat{\theta}$.

We consider a sequence $(H_w^j)_{j=1}^{\infty}$ taking the form

$$H_w^j(\tilde{w}) = E\big[1(W \le \tilde{w})\delta_w^j(W)\big], \tag{2}$$

where for each j the random variable $\delta_w^j(W)$ is bounded with $E[\delta_w^j(W)] = 1$. In $H_w^j(\tilde{w})$, the variable $\tilde{w}$ represents a possible value of the random variable W. As we will discuss, this $H_w^j(\tilde{w})$ will have the needed properties when $\delta_w^j(W)$ is chosen appropriately. In particular, the support of $H_w^j(\tilde{w})$ will approach $\{w\}$ as the support of $\delta_w^j(w)$ does. Throughout, we will assume that w is a vector of real numbers of fixed dimension r. We impose the following properties.

Assumption A1. *F_0 is absolutely continuous with respect to a measure μ on $\mathbb{R}^r$ with pdf $f_0(w)$, $\delta_w^j(W)$ is not constant, bounded, and $E[\delta_w^j(W)] = 1$.*

By $\delta_w^j(W)$ bounded $F_\tau^j = (1-\tau)F_0 + \tau H_w^j$ will be a CDF for small enough τ with pdf with respect to μ given by

$$f_\tau(\tilde{w}) = f_0(\tilde{w})\big[1-\tau+\tau\delta_w^j(\tilde{w})\big] = f_0(\tilde{w})\big[1+\tau S(\tilde{w})\big], S(\tilde{w}) = \delta_w^j(\tilde{w}) - 1, \tag{3}$$

where we suppress the j superscript and w subscript on $f_\tau(\tilde{w})$ and $S(\tilde{w})$ for notational convenience. Note that by $S(\tilde{w})$ bounded there is C such that for small enough τ,

$$(1-\tau)f_0/C \le f_\tau \le Cf_0, \tag{4}$$

so that f_τ and f_0 will be absolutely continuous with respect to each other. Thus, variables that are continuously distributed under F_0 will also be continuously distributed under F_τ^j. Also objects that have expectation close to zero for F_0 will also have expectation close to zero under F_τ^j and vice versa. If $\theta(F)$ being well-defined depends on existence of derivatives of the pdf for F then that restriction can be imposed by choosing $\delta_w^j(\tilde{w})$ so its derivatives exist. In these ways, we can choose $\delta_w^j(w)$ so that $f_\tau(\tilde{w})$ satisfies the restrictions needed for $\theta(F_\tau^j)$ to be well- defined.

We assume that the sequence $(\delta_w^j)_{j=1}^{\infty}$ satisfies a condition leading to

$$\lim_{j\longrightarrow\infty}\int \psi(\tilde{w})H_w^j(d\tilde{w}) \longrightarrow \psi(w), \tag{5}$$

thus justifying Step II of the influence function calculation. Define a function $a(\tilde{w})$ to be almost surely continuous at w in μ if for any $\varepsilon > 0$ there is a neighborhood N of w and a subset N_μ of N such that $\mu(N_\mu) = \mu(N)$ and $|a(\tilde{w}) - a(w)| < \varepsilon$ for all $\tilde{w} \in N_\mu$.

Assumption A2. *If $a(\tilde{w})$ is μ almost surely continuous at w and $E[a(W)^2] < \infty$, then $\delta_w^j(W)$ satisfies* $\lim_{j\longrightarrow\infty} E[a(W)\delta_w^j(W)] = a(w)$.

This assumption will be sufficient for equation (5). There are a variety of ways that $\delta_w^j(W)$ can be chosen so that Assumption A2 will be satisfied. The basic idea is to consider w where $f_0(\tilde{w})$ is bounded away from zero on a neighborhood of w in the support of W and choose $\delta_w^j(\tilde{w}) = g_w^j(\tilde{w})/f_0(\tilde{w})$ where $g_w^j(\tilde{w})$ is a bounded pdf and the support of $g_w^j(\tilde{w})$ to converge to $\{w\}$. A choice of $g_w^j(\tilde{w})$ that will lead to equation (5) in many cases can be based on a nonnegative kernel $K(u)$ with bounded support S, as in the following result.

Lemma A1. (*i*) *If*$K(u) \geq 0$, $\int K(u)\,du = 1$,*and* $K(u)$ *has bounded support* S ; (*ii*) *there is a neighborhood* N *of* w *and* $C > 0$ *such that* $f_0(\tilde{w}) \geq C$ *almost surely* μ *for* $\tilde{w} \in N$; (*iii*) $\mu(w + \sigma S) > 0$ *for all* $\sigma > 0$; *then for any* $(\sigma(j))_{j=1}^{\infty}$ *with* $\sigma(j) > 0$, $\sigma(j) \longrightarrow 0$, *and* $w + \sigma(j)S \subseteq N$ *for all* $\sigma(j)$, *Assumptions* A1 *and* A2 *are satisfied for*

$$\delta_w^j(W) = f_0(W)^{-1}\left[\int 1\big(\tilde{w} \in w + \sigma(j)S\big)\sigma(j)^{-r}K\left(\frac{\tilde{w} - w}{\sigma(j)}\right)\mu(d\tilde{w})\right]^{-1}\sigma(j)^{-r}K\left(\frac{W - w}{\sigma(j)}\right).$$

Note that if W has the Lebesgue density f_0, then the expression for δ_w^j simplifies to

$$\delta_w^j(\tilde{w}) = f_0(\tilde{w})^{-1}\sigma(j)^{-r}K\left(\frac{\tilde{w} - w}{\sigma(j)}\right).$$

Proof. Note that

$$\int 1\big(\tilde{w} \in W + \sigma(j)S\big)\sigma(j)^{-r}K\left(\frac{\tilde{w} - W}{\sigma(j)}\right)\mu(d\tilde{w}) > 0$$

by (i) and (iii). Also, $K((W - w)/\sigma(j))$ is nonzero only on a subset of N so that $\delta_w^j(W)$ is bounded by (i) and (ii). In addition, $E[\delta_w^j(W)] = 1$ by construction.

Suppose $a(W)$ has finite second moment and is continuous at w a.s. μ. Then for any $\varepsilon > 0$, there is j_ε large enough such that for $j \geq j_\varepsilon$,

$$a(w) - \varepsilon \leq a(W) \leq a(w) + \varepsilon$$

a.s. μ for $W \in w + \sigma(j)S$. Since $\delta_w^j(W)$ is nonnegative and nonzero only on $W \in w + \sigma(j)S$ we have

$$a(w) - \varepsilon = E\big[\big\{a(w) - \varepsilon\big\}\delta_w^j(W)\big] \leq E\big[a(W)\delta_w^j(W)\big] \leq E\big[\big\{a(w) + \varepsilon\big\}\delta_w^j(W)\big] = a(w) + \varepsilon,$$

for all $j \geq j_\varepsilon$. The conclusion follows by ε being any positive number. □

The choice of $\delta_w^j(W)$ in Lemma A1 is simply a device to help the limit of the Gateaux derivative exist under as general conditions as possible. The limit, and hence the influence function, does not depend on the kernel. Also, we could replace the continuity of $a(\tilde{w})$ at w in Assumption A2 with other conditions that are sufficient for equation (5) on a set of w with probability one under F_0. Equation (5) is analogous to the Lebesgue differentiation theorem that is known to hold under quite general conditions on $a(\tilde{w})$. For example, for the $\delta^j_w(w)$ of Lemma A1 equation (5) can be shown to hold for any measurable $a(\tilde{w})$ if μ is the sum of Lebesgue measure and a measure with a finite number of atoms. We use the continuity condition of Assumption A2 because it is relatively simple to state and because many influence functions will be μ almost sure continuous on a set of w that has probability one.

The next result shows that the influence function formula (5) is valid for H^j_w as specified in equation (2).

THEOREM A2. *If Assumptions A1 and A2 are satisfied, $\hat{\theta}$ is asymptotically linear with influence function $\psi(\tilde{w})$, $\hat{\theta}$ is locally regular for $F^j_\tau(\tilde{w}) = (1-\tau)F_0(\tilde{w}) + \tau H^j_w(\tilde{w})$ for each integer j and $H^j_w(\tilde{w}) = E[1(W \le \tilde{w})\delta^j(W)]$, and $\psi(\tilde{w})$ is μ almost surely continuous at w, then $d\theta(F^j_\tau)/d\tau$ exists, $d\theta(F^j_\tau)/d\tau = \int \psi(\tilde{w})H^j_w(d\tilde{w})$, and equation (5) is satisfied.*

PROOF. By $S(\tilde{w}) = \delta^j_w(\tilde{w}) - 1$ bounded, there is an open set T containing zero such that for all $\tau \in T$, $1 + \tau S(\tilde{w})$ is positive, bounded away from zero, and $f_\tau(\tilde{w})^{1/2} = f_0(\tilde{w})^{1/2}[1 + \tau S(\tilde{w})]^{1/2}$ is continuously differentiable in τ with

$$s_\tau(\tilde{w}) = \frac{d}{d\tau} f_0(\tilde{w})^{1/2}\big[1+\tau S(\tilde{w})\big]^{1/2} = \frac{1}{2}\frac{f_0(\tilde{w})^{1/2}S(\tilde{w})}{\big[1+\tau S(\tilde{w})\big]^{1/2}} \le Cf_0(\tilde{w})^{1/2}S(\tilde{w}).$$

By $S(\tilde{w})$ bounded, $\int [Cf_0(\tilde{w})^{1/2}S(\tilde{w})]^2\, d\mu < \infty$. Then by the dominated convergence theorem $f_0(\tilde{w})^{1/2}[1+\tau S(\tilde{w})]^{1/2}$ is mean-square differentiable and $I(\tau) = \int s_\tau(\tilde{w})^2\, d\mu$ is continuous in τ on a neighborhood of zero. By Assumption A1, $S(W)$ is not zero so that $I(\tau) > 0$. Then by Theorem A2 and Example 6.5 of Van der Vaart (1998), it follows that for any $\tau_n = O(1/\sqrt{n})$ a vector of n observations $(W_1, \dots, W_n)$ that is i.i.d. with pdf $f_{\tau_n}(\tilde{w})$ is contiguous to $(W_1, \dots, W_n)$ that is i.i.d. with pdf $f_0(\tilde{w})$. Therefore,

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(W_i) + o_p(1)$$

holds when $(W_1, \dots, W_n)$ are i.i.d. with pdf $f_{\tau_n}(\tilde{w})$.

Next define $\mu^j_w = E[\psi(W)S(W)] = E[\psi(W)\delta^j_w(W)]$. Then by $E[\psi(W)] = 0$,

$$E_\tau\big[\psi(W)\big] = \tau\mu^j_w.$$

Suppose $(W_1, \dots, W_n)$ are i.i.d. with pdf $f_{\tau_n}(\tilde{w})$. Let $\theta(\tau) = \theta((1-\tau)F_0 + \tau G^j_w)$, $\theta_n = \theta(\tau_n)$, and $\breve{\psi}_n(W) = \psi(W) - \tau_n\mu^j_w$. Adding and subtracting terms,

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_n) &= \sqrt{n}(\hat{\theta} - \theta_0) - \sqrt{n}(\theta_n - \theta_0)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(W_i) + o_p(1) - \sqrt{n}(\theta_n - \theta_0)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\breve{\psi}_n(W_i) + o_p(1) + \sqrt{n}\tau_n\mu^j_w - \sqrt{n}(\theta_n - \theta_0).\end{aligned}$$

Note that $E_{\tau_n}[\breve{\psi}_n(W)] = 0$. Also, by τ_n bounded,

$$\begin{aligned} E_\tau\big[1\big(\|\breve{\psi}_n(W)\| \geq M\big)\|\breve{\psi}_n(W)\|^2\big] &\leq CE\big[1\big(\|\breve{\psi}_n(W)\| \geq M\big)\|\breve{\psi}_n(W)\|^2\big] \\ &\leq CE\big[1\big(\|\breve{\psi}_n(W)\| \geq M\big)\big(\|\psi(W)\|^2 + C\big)\big] \\ &\leq CE\big[1\big(\|\psi(W)\| \geq M - C\big)\big(\|\psi(W)\|^2 + C\big)\big] \longrightarrow 0, \end{aligned}$$

as $M \longrightarrow \infty$, so the Lindbergh–Feller condition for a central limit theorem is satisfied. Furthermore, it follows by similar calculations that $E_{\tau_n}[\breve{\psi}_n(W)\breve{\psi}_n(W)^T] \longrightarrow V$. Therefore, by the Lindbergh–Feller central limit theorem, $\sum_{i=1}^n \breve{\psi}_n(W_i)/\sqrt{n} \stackrel{d}{\longrightarrow} N(0, V)$. By local regularity, $\sqrt{n}(\hat{\theta} - \theta_n) \stackrel{d}{\longrightarrow} N(0, V)$ implying that

$$\sqrt{n}\tau_n\mu_w^j - \sqrt{n}(\theta_n - \theta_0) \longrightarrow 0. \tag{6}$$

Next, we follow the proof of Theorem 2.1 of Van der Vaart (1991). The above argument shows that local regularity implies that equation (6) holds for all $\tau_n = O(1/\sqrt{n})$. Consider any sequence $r_m \longrightarrow 0$. Let n_m be the subsequence such that

$$(1 + n_m)^{-1/2} < r_m \leq n_m^{-1/2}.$$

Let $\tau_n = r_m$ for $n = n_m$ and $\tau_n = n^{-1/2}$ for $n \notin \{n_1, n_2, \ldots\}$. By construction, $\tau_n = O(1/\sqrt{n})$, so that eq (6) holds. Therefore, it also holds along the subsequence n_m, so that

$$\sqrt{n_m}r_m\left\{\mu_z^j - \frac{\theta(r_m) - \theta_0}{r_m}\right\} = \sqrt{n_m}r_m\mu_z^j - \sqrt{n_m}\big[\theta(r_m) - \theta_0\big] \longrightarrow 0.$$

By construction $\sqrt{n_m}r_m$ is bounded away from zero, so that $\mu_z^h - [\theta(r_m) - \theta_0]/r_m \longrightarrow 0$. Since r_m is any sequence converging to zero, it follows that $\theta(\tau)$ is differentiable at $\tau = 0$ with derivative μ_z^j. The conclusion then follows by Assumption A2. □

Let H_w^∞ be the CDF with $\Pr(W = w) = 1$. Theorem A2 gives sufficient conditions for equation (5), which is

$$\psi(w) = \int \psi(\tilde{w})H_w^\infty(d\tilde{w}) = \lim_{H_w^j \longrightarrow H_w^\infty} \int \psi(\tilde{w})H_w^j(d\tilde{w}),$$

where the first equality holds by definition of H_w^∞. The second equality states that $\psi(w)$ is the Lebesgue derivative of $\int \psi(\tilde{w})H(d\tilde{w})$ based on the regularity conditions of Assumptions A1 and A2 and the sequences of functions detailed there. This Lebesgue differentiation conclusion justifies Step II of the Gateaux derivative calculation as simply evaluating the Lebesgue derivative at a point. This evaluation will be valid with probability one under Assumptions A1 and A2.

We emphasize that the purpose of Theorem A2 is quite different than the results of Bickel, Klaasen, Ritov, and Wellner (1993), Van der Vaart (1991), and other important contributions to the semiparametric efficiency literature. Here, $\theta(F)$ is not a parameter of some semiparametric model. Instead, $\theta(F)$ is associated with an *estimator* $\hat{\theta}$, being the probability limit of that estimator when F is a distribution that is unrestricted except for regularity conditions, as in Newey (1994). Our goal is to use $\theta(F)$ to calculate the influence function of $\hat{\theta}$ under the assumption that $\hat{\theta}$ is asymptotically linear. The purpose of Theorem A2 is to justify Steps I and II as a way to do that calculation. In contrast, the goal of the semiparametric efficiency literature is to find the efficient influence function for a parameter of interest when F belongs to a family of distributions.

To highlight this contrast, note that the Gateaux derivative limit calculation can be applied to obtain the influence function under misspecification while efficient influence function calculations generally impose correct specification. Indeed, the definition of $\theta(F)$ *requires* that misspecification be allowed for, because $\theta(F)$ is limit of the estimator θ under all distributions F that are unrestricted except for regularity condition. Of course, correct specification may lead to simplifications in the form of the influence function. Such simplifications will be incorporated automatically when the Gateaux derivative limit is taken at an F_0 that satisfies model restrictions.

Theorem A2 is like Van der Vaart (1991, Theorem 2.1) in having differentiability of $\theta(F_\tau)$ as a conclusion. It differs in restricting the paths to have the form $(1-\tau)F_0 + \tau H_w^j$. Such a restriction on the paths actually weakens the local regularity hypothesis because θ only has to be locally regular for a particular kind of path rather than the general class of paths in Van der Vaart (1991). We note that this result allows for the distribution of W to have discrete components because the dominating measure μ may have atoms.

The weak nature of the local regularity condition highlights the strength of the asymptotic linearity hypothesis. Primitive conditions for asymptotic linearity can be quite strong and complicated. For example, it is known that asymptotic linearity of estimators with a nonparametric first step often requires some degree of smoothness in the functions being estimated; see Ritov and Bickel (1990). Our purpose here is to bypass those conditions in order to justify the Gateaux derivative formula for the influence function. The formula for the influence function can then be used in all the important ways outlined in Section 2.

It is also common to bypass regularity conditions when calculating the influence function or asymptotic variance of parametric estimators. There are well-known formulae that allow us to do this, such as Hansen (1982) for GMM estimators. The Gateaux derivative limit provides such a formula for semiparametric estimators. It provides an influence function formula that will be valid "under sufficient regularity conditions" analogously to the GMM formula for parametric estimators.

Appendix B: The influence function of semiparametric M estimators

In this Appendix, we give the general structure of the influence function for a semiparametric M-estimator and show that the first step influence function (FSIF) is zero for any first step that maximizes the same objective function as does the parameter of interest. A maximization (M) estimator satisfies

$$\hat{\theta} = \arg\max_{\theta \in B} \hat{Q}(\theta),$$

for a function $\hat{Q}(\theta)$ that depends on the data and parameters. M estimators have long been studied. A more general type that is useful when $\hat{Q}(\theta)$ is not continuous has $\hat{Q}(\theta) \geq \sup_{\theta \in B} \hat{Q}(\theta) - \hat{R}$, where the remainder $\hat{R}$ is small in large samples. The plim $\theta(F)$ of $\hat{\theta}$ will be the maximizer of the probability limit of $\hat{Q}(\theta)$ under standard regularity conditions. Thus, the influence function will depend only on the limit of the objective function and so is not affected by whether $\hat{\theta}$ is an approximate or exact maximizer of $\hat{Q}(\theta)$. The way we give of calculating the influence function will work for many estimators of this form, including those maximizing U-processes as considered by Sherman (1993).

We can use the Gateaux derivative to characterize the influence function for semiparametric M-estimators. Let $Q_\tau(\theta)$ denote the plim of the objective function $\hat{Q}(\theta)$ when the CDF of W_i is F_τ. Then under standard regularity conditions the plim of $\hat{\theta}$ is

$$\theta_\tau = \arg\max_{\theta \in \Theta} Q_\tau(\theta).$$

Suppose that $Q_\tau(\theta)$ is twice continuously differentiable in θ and θ_τ is in the interior of the parameter set. Then θ_τ satisfies the first-order conditions $dQ_\tau(\theta_\tau)/d\theta = 0$. By the implicit function theorem, for $\Lambda = \partial^2 Q(\theta_0)/\partial\theta\partial\theta'$ we have

$$\frac{d\theta_\tau}{d\tau} = -\Lambda^{-1} \frac{\partial^2 Q_\tau(\theta_0)}{\partial\tau\partial\theta}\bigg|_{\tau=0} = -\Lambda^{-1} \frac{\partial}{\partial\tau}\left\{\frac{\partial Q_\tau(\theta_0)}{\partial\theta}\right\}.$$

Comparing this equation with equation (1), we see that the influence function $\psi(w)$ of a semiparametric M estimator can be calculated by evaluating the derivative with respect to τ of $dQ_\tau(\theta_0)/d\theta$ at the distribution H_w^∞ with $W = w$ and premultiplying by $-\Lambda^{-1}$. For $\xi(W)$, such that $dQ_\tau(\theta_0)/d\theta = \int \xi(w)H(dw)$ the influence function of $\hat{\theta}$ will be

$$\psi(W) = -\Lambda^{-1}\xi(W).$$

This formula generalizes that of Newey (1994) for semiparametric GMM to M-estimation.

For M-estimators, certain nonparametric components of $\hat{Q}(\theta)$ can be ignored in deriving the influence function. The ignorable components are those that have been "concentrated out," meaning they have a plim that maximizes the plim of $\hat{Q}(\theta)$. In such cases, the dependence of these functions on θ captures the whole asymptotic effect of their estimation. To show this result, suppose that there is a function γ that depends on θ and possibly other functions and a function $\tilde{Q}_\tau(\theta, \gamma)$ such that $Q_\tau(\theta) = \tilde{Q}_\tau(\theta, \gamma_\tau)$ where

$$\gamma_\tau = \arg\max_{\gamma} \tilde{Q}_\tau(\theta, \gamma).$$

Here, $\tilde{Q}_\tau(\theta, \gamma_\tau)$ is the plim of $\hat{Q}(\theta)$ and γ_τ the plim of a nonparametric estimator on which $\hat{Q}(\theta)$ depends, when W has CDF F_τ. Since γ_τ maximizes over all γ it must maximize over $\tilde{\tau}$ as the function $\gamma_{\tilde{\tau}}$ varies. The first-order condition for maximization over $\tilde{\tau}$ is

$$\frac{d\tilde{Q}_\tau(\theta, \gamma_{\tilde{\tau}})}{d\tilde{\tau}}\bigg|_{\tilde{\tau}=\tau} = 0.$$

This equation holds identically in θ, so that we can differentiate both sides of the equality with respect to θ, evaluate at $\theta = \theta_0$ and $\tau = 0$, and interchange the order of differentiation to obtain

$$\frac{\partial^2 \tilde{Q}(\theta_0, \gamma_\tau)}{\partial\tau\partial\theta} = 0.$$

Then it follows by the chain rule that

$$\frac{\partial^2 \tilde{Q}_\tau(\theta_0, \gamma_\tau)}{\partial\tau\partial\theta} = \frac{\partial^2 \tilde{Q}_\tau(\theta_0, \gamma_0)}{\partial\tau\partial\theta} + \frac{\partial^2 \tilde{Q}(\theta_0, \gamma_\tau)}{\partial\tau\partial\theta} = \frac{\partial^2 \tilde{Q}_\tau(\theta_0, \gamma_0)}{\partial\tau\partial\theta}. \tag{7}$$

That is, the influence function can be obtained by treating the limit γ_τ as if it were equal to the true value γ_0.

Equation (7) generalizes Proposition 2 of Newey (1994) and Theorem 3.4 of Ichimura and Lee (2010) to objective functions that are not necessarily a sample average of a function of θ and γ. There are many important estimators included in this generalization. One of those is NPIV where the residual includes both parametric and nonparametric components. The result implies that estimation of the function of the nonparametric component γ can be ignored in calculating the influence function of θ. Another interesting estimator is partially linear regression with generated regressors. There the estimation of the nonparametric component can also be ignored in deriving the influence function, just as in Robinson (1988), though the presence of generated regressors will often affect the influence function, as in Hahn and Ridder (2013, 2016) and Mammen, Rothe, and Schienle (2012).

Appendix C: Endogenous orthogonality conditions with misspecification

In this Appendix, we derive the FSIF for endogenous orthgonality conditions under overidentification and misspecification where

$$\bar{\pi}(X) = \pi\big(\rho(W, \gamma_0)|X\big) \neq 0.$$

The first-order conditions for $\gamma_\tau = \arg\min_\gamma E_\tau[\pi_\tau(\rho(W, \gamma_\tau)|X)^2]$ give

$$\begin{aligned} 0 &= E_\tau\big[\pi_\tau\big(\rho(W, \gamma_\tau)|X\big)\pi_\tau\big(v_{\rho\tau}(W)\Delta(Z)|X\big)\big] \\ &= E_\tau\big[\pi_\tau\big(\rho(W, \gamma_\tau)|X\big)v_{\rho\tau}(W)\Delta(Z)\big] \text{ for all } \Delta \in \Gamma, \end{aligned}$$

identically in τ. Define $\alpha(X, \Delta) := \pi(v_\rho(W)\Delta(Z)|X)$ for $\Delta \in \Gamma$. Differentiating the previous identity with respect to τ gives for all $\Delta \in \Gamma$,

$$\begin{aligned} 0 &= \frac{\partial}{\partial\tau}E\big[\pi_\tau\big(\rho(W, \gamma_\tau)|X\big)\alpha(X, \Delta)\big] + \int \phi_1(w, \Delta)H(dw) + T_{v_\rho}(\Delta), \\ \phi_1(w, \Delta) &:= \bar{\pi}(X)v_\rho(W)\Delta(Z) - E\big[\bar{\pi}(X)v_\rho(W)\Delta(Z)\big], \\ T_{v_\rho}(\Delta) &:= \frac{\partial}{\partial\tau}E\big[\bar{\pi}(X)v_{\rho\tau}(W)\Delta(W)\big]. \end{aligned}$$

where $v_\rho(W) = v_{\rho 0}(W)$. Solving gives

$$\frac{\partial}{\partial \tau} E\big[\pi_\tau\big(\rho(W, \gamma_\tau)|X\big)\alpha(X, \Delta)\big] = -\int \phi_1(w, \Delta)H(dw) - T_{v_\rho}(\Delta) \tag{8}$$

for all $\Delta \in \Gamma$.

Next, we use the orthogonality condition for the projection that for all $b \in \mathcal{B}$,

$$E_\tau\big[\rho(W, \gamma_\tau)b(X)\big] = E_\tau\big[\pi_\tau\big(\rho(W, \gamma_\tau)|X\big)b(X)\big].$$

Because $\mathcal{A}$ is a subset of $\mathcal{B}$, it follows that

$$E_\tau\big[\rho(W, \gamma_\tau)\alpha(X, \Delta)\big] = E_\tau\big[\pi_\tau\big(\rho(W, \gamma_\tau)|X\big)\alpha(X, \Delta)\big] \quad \text{for all } \Delta \in \Gamma,$$

identically in τ. Differentiating both sides of this identify with respect to τ and applying the chain rule gives

$$\begin{aligned}
\frac{\partial}{\partial \tau} E_\tau\big[\rho(W, \gamma_\tau)\alpha(X, \Delta)\big] &= \frac{\partial}{\partial \tau} E_\tau\big[\bar{\pi}(X)\alpha(X, \Delta)\big] + \frac{\partial}{\partial \tau} E\big[\pi_\tau\big(\rho(W, \gamma_\tau)|X\big)\alpha(X, \Delta)\big] \\
&= \frac{\partial}{\partial \tau} E_\tau\big[\bar{\pi}(X)\alpha(X, \Delta)\big] - \int \phi_1(w, \Delta)H(dw) - T_{v_\rho}(\Delta) \\
&= -\int \phi_\Gamma(w, \Delta)H(dw) - T_{v_\rho}(\Delta), \\
\phi_\Gamma(w, \Delta) &= \bar{\pi}(X)\big\{v_\rho(X)\Delta(Z) - \alpha(X, \Delta)\big\},
\end{aligned}$$

for all $\Delta \in \Gamma$ where the second equality follows by equation (8) and the third equality follows by $E[\bar{\pi}(X)\alpha(X, \Delta)] = E[\bar{\pi}(X)v_\rho(W)\Delta(Z)]$. Applying the chain rule to the left-hand side and solving then gives

$$\begin{aligned}
-\frac{\partial}{\partial \tau} E\big[\rho(W, \gamma_\tau)\alpha(X, \Delta)\big] &= \frac{\partial}{\partial \tau} E_\tau\big[\rho(W, \gamma_0)\alpha(X, \Delta)\big] + \int \phi_\Gamma(w, \Delta)H(dw) + T_{v_\rho}(\Delta) \\
&= \int \big\{\rho(w, \gamma_0)\alpha(x, \Delta) + \phi_\Gamma(w, \Delta)\big\}H(dw) + T_{v_\rho}(\Delta), \qquad (9)
\end{aligned}$$

for all $\Delta \in \Gamma$, where the last equality follows by the first-order condition at $\tau = 0$ that implies $E[\rho(W, \gamma_0)\alpha(X, \Delta)] = 0$ for all Δ. Suppose that there exists b_m such that the projection of b_m on $\mathcal{A}$ is $\alpha(X, \Delta_m)$ for some $\Delta_m \in \Gamma$ and

$$\Pi\big(v_m(Z)|Z\big) = -\Pi\big(v_\rho(W)b_m(X)|Z\big).$$

Then by $\gamma_\tau(Z) \in \Gamma$,

$$\begin{aligned}
E\big[v_m(Z)\gamma_\tau(Z)\big] &= E\big[\Pi\big(v_m(Z)|Z\big)\gamma_\tau(Z)\big] \\
&= -E\big[\Pi\big(v_\rho(W)b_m(X)|Z\big)\gamma_\tau(Z)\big] \\
&= -E\big[v_\rho(W)b_m(X)\gamma_\tau(Z)\big] \\
&= -E\big[b_m(X)\pi\big(v_\rho(W)\gamma_\tau(Z)|X\big)\big]
\end{aligned}$$

$$= -E\big[\alpha(X,\Delta_m)\pi\big(v_\rho(W)\gamma_\tau(Z)|X\big)\big]$$
$$= -E\big[\alpha(X,\Delta_m)v_\rho(W)\gamma_\tau(Z)\big]. \tag{10}$$

Then differentiating gives

$$\begin{aligned}
\frac{\partial}{\partial\tau}E\big[m(W,\gamma_\tau)\big] &= \frac{\partial}{\partial\tau}E\big[v_m(Z)\gamma_\tau(Z)\big] \\
&= -\frac{\partial}{\partial\tau}E\big[\alpha(X,\Delta_m)v_\rho(W)\gamma_\tau(Z)\big] \\
&= -\frac{\partial}{\partial\tau}E\big[\alpha(X,\Delta_m)\rho(W,\gamma_\tau)\big] \\
&= \int\big\{\rho(w,\gamma_0)\alpha(x,\Delta_m)+\phi_\Gamma(w,\Delta_m)\big\}H(dw)+T_{v_\rho}(\Delta_m)
\end{aligned}$$

where the first equality follows by Assumption 3, the second equality by equation (10), the third equality by Assumption 4, and the fourth equality by equation (9). Combining this last equation with the conditions on which it depends gives the following result.

Proposition C1. (*i*) *If Assumptions* 3–4 *are satisfied*; (*ii*) *there exists* $b_m(X)$ *and* $\Delta_m \in \Gamma$ *such that* $\alpha(X,\Delta_m)$ *is the projection of* $b_m(X)$ *on* $\mathcal{A}$ *and* $\Pi(v_m(Z)|Z) = \Pi(v_\rho(W)b_m(X)|Z)$; *and* (*iii*) *there is* $\phi_\rho(w)$ *such that* $\partial E[\bar{\pi}(X)v_{\rho\tau}(W)\Delta_m(W)]/\partial\tau = \int\phi_\rho(w)H(dw)$ *then the FSIF is*

$$\phi(w,\gamma,\alpha) = \alpha(x,\Delta_m)\rho(w,\gamma)+\bar{\pi}(x)\big\{v_\rho(x)\Delta_m(z)-\pi\big(v_\rho(X)\Delta_m(Z)|X=x\big)\big\}+\phi_\rho(w).$$

This expression for the FSIF contains the term $\phi_\rho(w)$, which is the influence function of $E[\bar{\pi}(X)v_{\rho\tau}(W)\Delta_m(Z)]$. This $\phi_\rho(w)$ need not exist. In particular, for quantile orthogonality conditions where $v_{\rho\tau}(W)$ depends on the conditional pdf of Y given Z and X evaluated at the point $Y=\gamma_0(Z)$ it seems that this $\phi_\rho(w)$ generally does not exist. In that case, the NPIV estimator may not be root-n consistent under misspecification. This problem does not appear to be present for expectiles, where $E[\bar{\pi}(X)v_{\rho\tau}(W)\Delta_m(Z)]$ can be shown to have an influence function.

Ai and Chen (2007, p. 40) gave an influence function for a function of the solution to a conditional moment restriction under misspecification. In this case, the expression given in Proposition 3 is analogous to that in Ai and Chen (2007). Proposition C1 generalizes that expression to orthogonality conditions.

References

Ai, C. and X. Chen (2007), "Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables." *Journal of Econometrics*, 141, 5–43. [10]

Bickel, P., C. Klaasen, Y. Ritov, and J. Wellner (1993), "Efficient and adaptive estimation for semiparametric models." Johns Hopkins, Washington. [5]

Hahn, J. and G. Ridder (2013), "The asymptotic variance of semi-parametric estimators with generated regressors." *Econometrica*, 81, 315–340. [8]

Hahn, J. and G. Ridder (2016), "Three-stage semi-parametric inference: Control variables and differentiability." USC-INET Research Paper No. 16–17. [8]

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators." *Econometrica*, 50, 1029–1054. [6]

Ichimura, H. and S. Lee (2010), "Characterization of the asymptotic distribution of semiparametric M-estimators." *Journal of Econometrics*, 159, 252–266. [8]

Mammen, E., C. Rothe, and M. Schienle (2012), "Nonparametric regression with nonparametrically generated covariates." *Annals of Statistics*, 40, 1132–1170. [8]

Newey, W. K. (1994), "The asymptotic variance of semiparametric estimators." *Econometrica*, 62, 1349–1382. [6, 7, 8]

Ritov, Y. and P. J. Bickel (1990), "Achieving information bounds in non and semiparametric models." *Annals of Statistics*, 18, 925–938. [6]

Robinson, P. M. (1988), "Root-N-consistent semiparametric regression." *Econometrica*, 56, 931–954. [8]

Sherman, R. (1993), "The limiting distribution of the maximum rank correlation estimator." *Econometrica*, 61, 123–137. [7]

Van der Vaart, A. W. (1991), "On differentiable functionals." *Annals of Statistics*, 19, 178–204. [5, 6]

Van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press, Cambridge. [4]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 24 January, 2017; final version accepted 19 May, 2021; available online 28 July, 2021.