The Asymptotic Variance of Semi-parametric Estimators with Generated Regressors*

Jinyong Hahn
Department of Economics, UCLA

Geert Ridder
Department of Economics, USC

June 30, 2011

Abstract

We study the asymptotic distribution of three-step estimators of a finite dimensional parameter vector where the second step consists of one or more non-parametric regressions on a regressor that is estimated in the first step. The first-step estimator is either parametric or non-parametric. Using Newey's (1994) path-derivative method we derive the contribution of the first-step estimator to the influence function. In this derivation it is important to account for the dual role that the first-step estimator plays in the second-step non-parametric regression, i.e., that of conditioning variable and that of argument.

JEL Classification: C01, C14.

Keywords: Semi-parametric estimation, generated regressors, asymptotic variance.

*Financial support for this research was generously provided through NSF SES 0819612 and 0819638. We thank Guido Imbens and seminar participants at UC Riverside, the Tinbergen Institute, Yale, FGV-Rio de Janeiro, FGV-São Paulo, Harvard/MIT, UCLA, Mannheim, Sciences Po and CEMFI for comments. Geert Ridder thanks the Department of Economics, PUC, Rio de Janeiro for their hospitality. Addresses: Jinyong Hahn, Department of Economics, Bunche Hall, UCLA, Los Angeles, CA 90095, hahn@econ.ucla.edu; Geert Ridder, Department of Economics, Kaprilian Hall, USC, Los Angeles, CA 90089, ridder@usc.edu.

1 Introduction

We study the asymptotic distribution of estimators of a finite dimensional parameter in a semi-parametric three (or more) step estimation problem. This topic has become quite important especially due to the recent developments in the econometric analysis of treatment effects and in the identification and estimation of non-linear models with endogenous covariates using control variables. We undertake a theoretical investigation of such estimators, and illustrate the usefulness of our result by examining the asymptotic variance of the estimator of the Average Treatment Effect (ATE) proposed by Heckman, Ichimura and Todd (1998) that is based on non-parametric regressions on the estimated propensity score.

The estimators under consideration are all characterized by three steps. In the first step we estimate a regressor. In the second step we estimate a non-parametric regression with the "generated regressor" as one of the independent variables. In the third step we estimate a finite dimensional parameter (without loss of generality we consider the scalar case) that satisfies a moment condition that depends on the non-parametric regression estimated in the second step.

Pagan (1984), who considered regression estimators involving generated regressors in the parametric context, is an intellectual predecessor. We heavily use Newey's (1994) path-derivative based characterization of the asymptotic variance of semi-parametric estimators. We extend his result to three-step estimators, where the second step is a non-parametric regression on a variable estimated in the first step.

This paper has the following structure. Our main result is in Section 2. In Section 3 we consider estimators that involve partial means with an application to regression on the estimated propensity score in Section 4.

2 The Influence Function of Semi-parametric Three-Step Estimators

We now present our two main results on semi-parametric three-step estimators. We distinguish between two cases: (i) the first step is parametric, (ii) the first step is non-parametric. Moreover, we first consider estimators that can be expressed as a sample average of a function of the second-step non-parametric regression only. Next, we allow the third-step estimator to depend on other variables besides the second-step non-parametric regression. In both cases the estimators are, in Newey's (1994) terminology, full means, because they average over all arguments of the second-step non-parametric regression. In Section 3 we consider estimators that average over most but not all independent variables in the second-step non-parametric regression, i.e. the estimator is a partial mean. This makes the influence function more complicated, which is the reason that we start with the full mean case.

2.1 Parametric First Step, Non-parametric Second Step

We assume that we observe i.i.d. observations $s_i = (y_i, x_i, z_i,)\,, i = 1, \ldots, n$. In the first step, we compute an estimator $\widehat{\alpha}$ such that $\sqrt{n}\,(\widehat{\alpha} - \alpha_*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\,(x_i, z_i) + o_p\,(1)$ with $\mathbb{E}\,[\psi\,(x_i, z_i)] = 0$. The parameter vector α indexes a relation between a dependent variable that is a component of

x (and that we later denote by u) and independent variables that are some or all of the other variables in x and those in z. Either the predicted value or the residual of this relationship is an independent variable in the second-step non-parametric regression. The notation $\varphi(x, z, \alpha)$ for the generated regressor covers both cases.

In the second step, our goal is to estimate

$$\mu(x, v_*) = \mathbb{E}[y \mid x, v_*]$$

where $v_* = \varphi(x, z, \alpha_*)$. Because we do not observe α_*, we use $\widehat{v}_i = \varphi(x_i, z_i, \widehat{\alpha})$ in the non-parametric regression. The non-parametric regression estimator of y on $x, v = \varphi(x, z, \alpha)$ is denoted by $\widehat{\gamma}$. (The $\widehat{\gamma}$ is distinct from the nonparametric regression $\widehat{\mu}$ of y on $x, v_* = \varphi(x, z, \alpha_*)$.)

Our goal is to characterize the first order asymptotic properties of

$$\widehat{\beta} = \frac{1}{n} \sum_{i=1}^{n} h(\widehat{\gamma}(x_i, \varphi(x_i, z_i, \widehat{\alpha})))$$

We can consider $\widehat{\beta}$ as the solution of a sample moment equation that is derived from a population moment equation that depends on β and $\mu(x, \varphi(x, z, \alpha_*))$.

Using Newey's (1994) path-derivative approach, it can be shown that we have the approximation

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta} - \beta_*\right) = & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\mu(x_i, \varphi(x_i, z_i, \alpha_*))) - \beta_*\right) && \text{(i)} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\widehat{\gamma}(x_i, \varphi(x_i, z_i, \alpha_*))) - h(\mu(x_i, \varphi(x_i, z_i, \alpha_*)))\right) && \text{(ii)} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\gamma(x_i, \varphi(x_i, z_i, \widehat{\alpha}))) - h(\mu(x_i, \varphi(x_i, z_i, \alpha_*)))\right) + o_p(1) && \text{(iii)}
\end{aligned}$$

The first term (i) is the main term, the second term (ii) is the adjustment for the estimation of $\widehat{\gamma}$, and the third term (iii) is the adjustment related to the estimation of $\widehat{\alpha}$. The decomposition here is based on the fact that Newey's approach can be used "term-by-term". Therefore, we may without loss of generality assume that α is a scalar.

The second component (ii) in the decomposition can be analyzed as in Newey (1994, pp. 1360 – 61), and we therefore focus on the analysis of the third component

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(\gamma(x_i, \varphi(x_i, z_i, \widehat{\alpha}))) - h(\mu(x_i, \varphi(x_i, z_i, \alpha_*)))\right)$$

We define

$$\begin{aligned}
\gamma(x, v^*; \alpha) &= \mathbb{E}[y \mid x, \varphi(x, z, \alpha) = v^*] \\
g(s, \alpha_1, \alpha_2) &= h(\gamma(x, \varphi(x, z, \alpha_1); \alpha_2))
\end{aligned}$$

Note that the two roles that α plays are made explicit in $g(s, \alpha_1, \alpha_2)$ that is obtained by substituting $v^* = \varphi(x, z, \alpha_1)$ in $\gamma(x, v^*; \alpha_2)$. Note also that $\mu(x, v_*) = \gamma(x, v_*; \alpha_*)$, where $\mu(x, v_*) = \mathbb{E}[y \mid x, \varphi(x, z, \alpha_*) = v_*]$. The notation α_1, α_2 is just an expositional device, since $\alpha_1 = \alpha_2 = \alpha$.

With these definitions, we can now write

$$\frac{1}{n}\sum_{i=1}^{n} h\left(\gamma\left(x_i, \varphi\left(x_i, z_i, \widehat{\alpha}\right); \widehat{\alpha}\right)\right) = \frac{1}{n}\sum_{i=1}^{n} g\left(s_i, \widehat{\alpha}_1, \widehat{\alpha}_2\right)$$

where $\widehat{\alpha}_1 = \widehat{\alpha}_2 = \widehat{\alpha}$, but we keep them separate to emphasize the two roles of $\widehat{\alpha}$. This forces us to deal with the two roles that $\widehat{\alpha}$ plays in the linearization that involves partial derivatives:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h\left(\gamma\left(x_i, \varphi\left(x_i, z_i, \widehat{\alpha}\right); \widehat{\alpha}\right)\right) - h\left(\gamma\left(x_i, \varphi\left(x_i, z_i, \alpha_*\right); \alpha_*\right)\right)\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(g\left(s_i, \widehat{\alpha}_1, \widehat{\alpha}_2\right) - g\left(s_i, \alpha_*, \alpha_*\right)\right) \\
&= \left(\mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_1}\right] + \mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_2}\right]\right)\sqrt{n}(\widehat{\alpha} - \alpha_*) + o_p\left(1\right)
\end{aligned}$$

Therefore we must compute $\mathbb{E}\left[\frac{\partial g(s,\alpha_*,\alpha_*)}{\partial \alpha_1}\right]$ and $\mathbb{E}\left[\frac{\partial g(s,\alpha_*,\alpha_*)}{\partial \alpha_2}\right]$. The computation of the first expectation is easy; it is straightforward to show that

$$\mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_1}\right] = \mathbb{E}\left[\frac{\partial h\left(\mu\left(x, \varphi\left(x, z, \alpha_*\right)\right)\right)}{\partial \mu}\frac{\partial \mu\left(x, \varphi\left(x, z, \alpha_*\right)\right)}{\partial v}\frac{\partial \varphi\left(x, z, \alpha_*\right)}{\partial \alpha}\right]$$

The headache is to compute the second expectation. By the chain rule

$$\mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_2}\right] = \mathbb{E}\left[\frac{\partial h\left(\mu\left(x, \varphi\left(x, z, \alpha_*\right)\right)\right)}{\partial \mu}\frac{\partial \gamma\left(x, \varphi\left(x, z, \alpha_*\right); \alpha_*\right)}{\partial \alpha_2}\right] \tag{1}$$

Unfortunately, it is not obvious how to differentiate $\gamma\left(x, \varphi\left(x, z, \alpha_*\right); \alpha\right)$ with respect to α. After all, $\gamma\left(x, \varphi\left(x, z, \alpha_*\right); \alpha\right)$ has the functional form of $\mathbb{E}\left[y \mid x, \varphi\left(x, z, \alpha\right) = v^*\right]$ that depends on α. The next lemma gives the solution in a generic case.[1]

Lemma 1 *Let $t\left(x, \varphi(x, z, \alpha_*)\right)$ denote an arbitrary mean square integrable function that is continuously differentiable in the second argument. Then,*

$$\begin{aligned}
&\frac{\partial}{\partial \alpha}\mathbb{E}\left[t(x, \varphi(x, z, \alpha_*))\gamma(x, \varphi(x, z, \alpha_*); \alpha)\right]\Big|_{\alpha=\alpha_*} \\
&= -\mathbb{E}\left[t(x, v_*)\frac{\partial \mu(x, v_*)}{\partial v}\frac{\partial \varphi(x, z, \alpha_*)}{\partial \alpha}\right] + \mathbb{E}\left[\left(\mu(x, z) - \mu(x, v_*)\right)\frac{\partial t(x, v_*)}{\partial v}\frac{\partial \varphi(x, z, \alpha_*)}{\partial \alpha}\right]
\end{aligned} \tag{2}$$

with $v_* = \varphi(x, z, \alpha_*)$ and $\mu(x, z) = \mathbb{E}(y|x, z)$.

[1] Our analysis is predicated on the assumption that the derivative exists on the left of (1) exists. Analysis of the case when the derivative does not exist is beyond the scope of this paper. A note that analyzes and gives sufficient conditions for the existence of the derivative is available on request.

Proof. Because $\gamma(x, \varphi(x, z, \alpha); \alpha)$ is the solution to

$$\min_p \mathbb{E}\left[(y - p(x, \varphi(x, z, \alpha)))^2\right]$$

we have that for all α

$$\mathbb{E}\left[(y - \gamma(x, \varphi(x, z, \alpha); \alpha))\, t\left(x, \varphi\left(x, z, \alpha\right)\right)\right] = 0$$

Differentiating with respect to α and evaluating the result at $\alpha = \alpha_*$, we find after rearranging (2). ■

This key lemma is used repeatedly in the sequel, beginning with the proof of the following theorem.

Theorem 1 (Contribution parametric first-step estimator) *The adjustment to the influence function that accounts for the first-stage estimation error is*

$$\begin{aligned}
&\left(\mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_1}\right] + \mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_2}\right]\right)\sqrt{n}\left(\widehat{\alpha} - \alpha_*\right) \\
&= \mathbb{E}\left[\left(\mu\left(x, z\right) - \mu\left(x, v_*\right)\right)\frac{\partial^2 h\left(\mu\left(x, v_*\right)\right)}{\partial \mu^2}\frac{\partial \mu\left(x, v_*\right)}{\partial v}\frac{\partial \varphi(x, z, \alpha_*)}{\partial \alpha}\right]\sqrt{n}(\widehat{\alpha} - \alpha_*) \qquad (3)
\end{aligned}$$

with $v_* = \varphi(x, z, \alpha_*)$.

Proof. We compute the right hand side of (1) that by Lemma 1 is equal to

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial h\left(\mu\left(x, \varphi\left(x, z, \alpha_*\right)\right)\right)}{\partial \mu}\frac{\partial \gamma\left(x, \varphi\left(x, z, \alpha_*\right); \alpha_*\right)}{\partial \alpha}\right] \\
&= -\,\mathbb{E}\left[\frac{\partial h(x, \mu(x, v_*))}{\partial \mu}\frac{\partial \mu(x, v_*)}{\partial v}\frac{\partial \varphi(x, z, \alpha_*)}{\partial \alpha}\right] \\
&\quad + \mathbb{E}\left[(\mu(x, z) - \mu(x, v_*))\frac{\partial^2 h(x, \mu(x, v_*))}{\partial \mu^2}\frac{\partial \mu(x, v_*)}{\partial v}\frac{\partial \varphi(x, z, \alpha_*)}{\partial \alpha}\right]
\end{aligned}$$

Adding $\mathbb{E}\left[\frac{\partial g(s, \alpha_*, \alpha_*)}{\partial \alpha_1}\right]$ that is equal to the opposite of the first term on the right-hand side, we find the desired result. ■

Remark 1 *From Theorem 1, it can be easily deduced that the first-stage estimate has no contribution to the influence function if* h *is linear in* μ. *Although straightforward ex post, this simplification does not seem to have been recognized in the past. The literature focused instead on the simplification that occurs if the index restriction* $\mathbb{E}\left[y|\, x, z\right] = \mathbb{E}\left[y|\, x, \varphi(x, z, \alpha_*)\right]$ *holds. See, e.g., Newey (1994) and Klein, Shen and Vella (2010).*

Suppose now that the μ is multidimensional, i.e., y is a J-dimensional random vector. The estimator is now

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} h\left(\widehat{\gamma}_1\left(x_i, \varphi\left(x_i, z_i, \widehat{\alpha}\right)\right), \ldots, \widehat{\gamma}_J\left(x_i, \varphi\left(x_i, z_i, \widehat{\alpha}\right)\right)\right)$$

The product rule of calculus suggests that we can tackle this problem by adding the derivatives. This is formalized in the next theorem.

Theorem 2 (Contribution parametric first-step estimators) *The adjustment to the influence function that accounts for the first-stage estimation error is*

$$\sum_j \mathbb{E}\left[\frac{\partial^2 h\left(\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial \mu_j^2}\left(y_j-\mu_j\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)\frac{\partial \mu_j\left(x,\varphi\left(x,z,\alpha_*\right)\right)}{\partial v}\frac{\partial \varphi\left(x,z,\alpha_*\right)}{\partial \alpha}\right]\sqrt{n}\left(\widehat{\alpha}-\alpha_*\right).$$

Proof. See Appendix. ■

2.2 Non-parametric First Step, Non-parametric Second Step

We now assume that the first step is non-parametric. Again we have a random sample $s_i = (y_i, x_i, z_i), i = 1, \ldots, n$. The first-step projection of one of the components of x, that we denote by u, on some or all of the other components of x and z is denoted by $v_* = \varphi_*(x,z) = \mathbb{E}[u \mid x, z]$. The first step is to estimate this projection by non-parametric regression. In the second step we estimate $\mu(x, v_*) = \mathbb{E}[y \mid x, v_*]$ by non-parametric regression of y on $x, \widehat{v} = \widehat{\varphi}(x,z)$. Our interest is to characterize the first-order asymptotic properties of

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} h\left(\widehat{\gamma}\left(x_i, \widehat{\varphi}\left(x_i, z_i\right)\right)\right)$$

where $\widehat{\gamma}(x_i, \widehat{v}_i)$ is the non-parametric regression estimate.

We define

$$\begin{aligned}\gamma\left(x, v_1; v_2\right) &= \mathbb{E}\left[y \mid\ x, \varphi\left(x,z\right) = v_1\right] \\ g\left(w, v_1, v_2\right) &= h\left(\gamma\left(x, v_1; v_2\right)\right)\end{aligned}$$

with $v_2 \equiv \varphi(x,z)$ and conditioning on $\varphi(x,z) = v_1$. Note that v_1 and v_2 play the roles of α_1 and α_2.

With these definitions, we can now write

$$\frac{1}{n}\sum_{i=1}^{n} h\left(\gamma\left(x_i, \widehat{v}_1; \widehat{v}_2\right)\right) = \frac{1}{n}\sum_{i=1}^{n} g\left(s_i, \widehat{v}_1, \widehat{v}_2\right)$$

where $\widehat{v}_1 = \widehat{v}_2 = \widehat{v}$. We keep them separate to emphasize their different roles. Our objective is to approximate

$$\frac{1}{n}\sum_{i=1}^{n} g\left(s_i, \widehat{v}_1, \widehat{v}_2\right) - \frac{1}{n}\sum_{i=1}^{n} g\left(s_i, v_1, v_2\right)$$

To find the contribution of the sampling variation in $\hat{v}$ we take γ as known, i.e., the sampling variation in the second-step non-parametric regression is accounted for in a separate term that has a well-known form, since it follows directly from Newey (1994). As in Newey (1994) we consider a path v_α indexed by $\alpha \in \mathbb{R}$ such that $v_{\alpha_*} = v_*$. First, using the calculation in the previous section we obtain that

$$\left.\frac{\partial \mathbb{E}\left[h\left(\gamma\left(x, v_\alpha; v_\alpha\right)\right)\right]}{\partial \alpha}\right|_{\alpha=\alpha_*} = \frac{\partial \mathbb{E}\left[D\left(s, v_\alpha\right)\right]}{\partial \alpha}$$

for

$$D\left(s, v_{\alpha}\right)=\frac{\partial^{2} h\left(\mu\left(x, v_{*}\right)\right)}{\partial \mu^{2}}\left(y-\mu\left(x, v_{*}\right)\right) \frac{\partial \mu\left(x, v_{*}\right)}{\partial v} v_{\alpha}.$$

which is linear in v_α. Second, we have that for any $v = \varphi(x, z)$,

$$\mathbb{E}\left[D\left(s, v\right)\right]=\mathbb{E}\left[\delta_{1}\left(x, z\right) \varphi(x, z)\right]$$

with

$$\delta_{1}\left(x, z\right)=\mathbb{E}\left[\left.\frac{\partial^{2} h\left(\mu\left(x, \varphi_{*}(x, z)\right)\right)}{\partial \mu^{2}}\left(y-\mu\left(x, \varphi_{*}(x, z)\right)\right) \frac{\partial \mu\left(x, \varphi_{*}(x, z)\right)}{\partial v}\right| x, z\right] \tag{4}$$

where it is understood that we condition on the variables that are in φ_* so that we average over all x that are not in the generated regressor.

By Newey (1994, Proposition 4), these two facts imply that the adjustment to the influence function is equal to

$$\delta_{1}\left(x_{i}, z_{i}\right)\left(u_{i}-\mathbb{E}\left[u \mid x_{i}, z_{i}\right]\right)=\delta_{1}\left(x_{i}, z_{i}\right)\left(u_{i}-\varphi_{*}(x_{i}, z_{i})\right)$$

with u the component of x that is projected on x, z.

We summarize the result in a theorem:

Theorem 3 (Contribution non-parametric first-step estimator) *The adjustment to the influence function that accounts for the first-stage estimation error is*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{1}\left(x_{i}, z_{i}\right)\left(u_{i}-\varphi_{*}(x_{i}, z_{i})\right)$$

with $\varphi_*(x, z) = \mathbb{E}[u|x, z]$ *and* δ_1 *as in (4).*

2.3 Additional Variables in the Third Step

So far, we have assumed that the parameter of interest is

$$\beta_{*}=\mathbb{E}[h(\mu(x, v_{*}))]$$

where h depends only on $\mu(x, v_*)$. We now consider the extension to

$$\beta_{*}=\mathbb{E}[h(w, \mu(x, v_{*}))]$$

where w is a vector of other variables that may have x, z as subvectors. We consider both the case that φ is parametric and the case that this function is non-parametric. Because as before the main term and the contribution of the estimation of $\mathbb{E}\left[y|x, v_*\right]$ do not raise new issues, the next two theorems only give the contribution of the first-stage estimator. In these theorems we use the function

$$\kappa\left(x, v\right)=\mathbb{E}\left[\left.\frac{\partial h\left(w, \mu\left(x, v_{*}\right)\right)}{\partial \mu}\right| x, \varphi\left(x, z, \alpha_{*}\right)=v\right]$$

with $\varphi_*(x, z)$ substituted in the non-parametric case.

Theorem 4 (Contribution parametric first-step estimator) *The adjustment to the influence function that accounts for the first-stage estimation error is*

$$\left(\mathbb{E}\left[\left(\frac{\partial h\left(w,\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu}-\kappa\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)\frac{\partial\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)}{\partial v}\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]\right.$$
$$\left.+\mathbb{E}\left[\frac{\partial\kappa\left(x,\varphi\left(x,z,\alpha_*\right)\right)}{\partial v}\left(\mu(x,z)-\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]\right)\sqrt{n}\left(\widehat{\alpha}-\alpha_*\right)$$

with $v_*=\varphi(x,z,\alpha_*)$.

Proof. See Appendix. ■

Now, we consider the case where the first step is non-parametric. The discussion preceding Theorem 3 implies that

Theorem 5 (Contribution non-parametric first-step estimator) *The adjustment to the influence function that accounts for the first-stage estimation error is*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_2\left(x_i,z_i\right)\left(u_i-\varphi_*(x_i,z_i)\right)$$

with $\varphi_*(x,z)=\mathbb{E}[u|x,z]$ *and*

$$\delta_2\left(x,z\right)=\mathbb{E}\left[\left(\frac{\partial h\left(w,\mu\left(x,\varphi_*\left(x,z\right)\right)\right)}{\partial\mu}-\kappa\left(x,\varphi_*\left(x,z\right)\right)\right)\frac{\partial\mu\left(x,\varphi_*\left(x,z\right)\right)}{\partial v}\right|x,z\Bigg]$$
$$+\mathbb{E}\left[\frac{\partial\kappa\left(x,\varphi_*\left(x,z\right)\right)}{\partial v}\left(\mu(x,z)-\mu\left(x,\varphi_*\left(x,z\right)\right)\right)\right|x,z\Bigg]$$

Remark 2 *Theorems 4 and 5 are easily generalized to the case of multidimensional* μ.

Remark 3 *Suppose that* $\kappa\left(x,\varphi\left(x,z,\alpha_*\right)\right)=0$ *in Theorem 4. The adjustment is then equal to the derivative with respect to* α_1, *i.e., the naive derivative that only accounts for* α *as an argument (see equation (15) in the proof of Theorem 4). Therefore, it may be useful to check whether* $\kappa\left(x,\varphi\left(x,z,\alpha_*\right)\right)=0$ *in specific models. If it is the case, we need not worry about the effect of first-step estimation on the second-stage non-parametric regression. Such a characterization turns out to be useful for the partially linear regression model*

$$y_i=x_i\beta_*+m\left(\varphi(x_i,z_i,\alpha_*)\right)+\varepsilon_i,$$

where m *is non-parametric and* $\mathbb{E}\left[\varepsilon_i|\,x_i,v_{*i}\right]=0$.[2]

Remark 4 *The theorems can be applied to general semi-parametric GMM estimators. If we consider the moment condition*

$$\mathbb{E}[m(w,\mu(x,v_*),\beta_*)]=0$$

and we linearize the corresponding sample moment condition to obtain

$$\sqrt{n}(\widehat{\beta}-\beta_*)=\left(\mathbb{E}\left[\frac{\partial m(w,\mu(x,v_*),\beta_*)}{\partial\beta'}\right]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}m\left(w_i,\widehat{\gamma}(x_i,\widehat{\varphi}(x_i,z_i)),\beta_*\right)+o_p(1)$$

Therefore, the contribution of the first-stage estimate to the asymptotic distribution of $\widehat{\beta}$ *can be found by applying Theorem 5 to* $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}m\left(w_i,\widehat{\gamma}(x_i,\widehat{\varphi}(x_i,z_i)),\beta_*\right)$.

[2] A detailed discussion can be found in a previous version of the paper, which is available on request.

3 The Influence Function of Semi-parametric Three-step Estimators: Partial Means

In Section 2, the estimator $\widehat{\beta}$ averaged over all arguments of the second-step non-parametric regression function. In Newey's (1994) terminology the estimator is a full mean. In this section we consider the case that the discrete independent variable d is fixed, i.e., $\widehat{\beta}$ does not average over this variable. Hence the estimator is a partial mean. Because the variable that is fixed in the partial mean is discrete, the parametric rate of convergence applies. (If that variable were continuous we would have a slower rate.) Let the discrete variable d take the values $d_{(1)}, \ldots, d_{(K)}$. In the second step we estimate

$$\mu\left(x, v_*, d\right) = \mathbb{E}\left[y \mid x, v_*, d\right]$$

As in Section 2 we use $\widehat{v}_i$ in the non-parametric regression. Define $\widehat{\gamma}_k(x_i, \widehat{v}_i) = \widehat{\gamma}(x_i, \widehat{v}_i, d_{(k)})$, i.e., $\widehat{\gamma}_k$ is the non-parametric regression function if we set x and $\widehat{v}$ to the observed values for i, but fix d at value $d_{(k)}$ which may not be its value for i. The K vector $\widehat{\gamma}(x_i, \widehat{v}_i)$ stacks the $\widehat{\gamma}_k(x_i, \widehat{v}_i)$. We also define

$$\mu_k\left(x, v_*\right) = \mathbb{E}\left[y \mid x, v_*, d_{(k)}\right]$$

and $\mu(x, v_*)$ as the K vector with components $\mu_k\left(x, v_*\right)$. Our goal is to characterize the first order asymptotic properties of

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} h\left(w_i, \widehat{\gamma}\left(x_i, \widehat{v}_i\right)\right)$$

with $\widehat{\gamma}$ the K vector of non-parametric regression functions where the discrete variable d is fixed at its K distinct values. As in Section 2.3, we allow for a vector of additional variables w in h.

3.1 Partial Means: Parametric First Step, Non-parametric Second Step

We assume that the first step is parametric, and $\widehat{v}_i = \varphi\left(x_i, z_i, \widehat{\alpha}\right)$. As in Section 2, we can use Newey's (1994) approach, and express the influence function of $\widehat{\beta}$ as a sum of three terms: (i) the main term, (ii) a term that adjusts for the estimation of $\widehat{\gamma}$, and (iii) an adjustment related to the estimation of $\widehat{\alpha}$. Define

$$\pi_k(x, \varphi(x, z, \alpha)) = \Pr(d = d_{(k)} | x, \varphi(x, z, \alpha))$$

$$\pi_k(x, z) = \Pr(d = d_{(k)} | x, z)$$

and

$$\kappa_k\left(x, v\right) = \mathbb{E}\left[\left.\frac{\partial h\left(w, \mu\left(x, v_*\right)\right)}{\partial \mu_k}\right| x, \varphi\left(x, z, \alpha_*\right) = v\right]$$

The second component in the decomposition can be analyzed as in Newey (1994, pp. 1360 – 61) and is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{K} 1(d_i = d_{(k)})(y_i - \mu_k(x_i, v_{*i}))\frac{\kappa_k(x_i, v_{*i})}{\pi_k(x_i, v_{*i})} + o_p\left(1\right) \tag{5}$$

As in Section 2 we therefore focus on the analysis of the third component

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h\left(w_i,\gamma\left(x_i,\varphi\left(x_i,z_i,\widehat{\alpha}\right);\widehat{\alpha}\right)\right)-h\left(w_i,\gamma\left(x_i,\varphi\left(x_i,z_i,\alpha_*\right)\right);\alpha_*\right)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(g\left(s_i,\widehat{\alpha}_1,\widehat{\alpha}_2\right)-g\left(s_i,\alpha_*,\alpha_*\right)\right)$$

with

$$\begin{aligned}\gamma_k\left(x,v^*;\alpha\right)&=\mathbb{E}\left[y \mid\ x,\varphi\left(x,z,\alpha\right)=v^*,d=d_{(k)}\right]\\ \gamma\left(x,v^*;\alpha\right)&=\left(\gamma_1\left(x,v^*;\alpha\right)\cdots\gamma_K\left(x,v^*;\alpha\right)\right)'\\ g\left(w,\alpha_1,\alpha_2\right)&=h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_1\right);\alpha_2\right)\right)\end{aligned}$$

Lemma 1 can be generalized to the partial means case:

Lemma 2 *For partial means we have for all $k=1,\ldots,K$*

$$\begin{aligned}&\frac{\partial}{\partial\alpha}\mathbb{E}\left[t(x,\varphi(x,z,\alpha_*))\gamma_k(x,\varphi(x,z,\alpha_*);\alpha)\right]\Big|_{\alpha=\alpha_*}\\ =&-\mathbb{E}\left[\frac{\pi_k(x,z)}{\pi_k(x,v_*)}t(x,v_*)\frac{\partial\mu_k(x,v_*)}{\partial v}\frac{\partial\varphi(x,z,\alpha_*)}{\partial\alpha}\right]\\ &+\mathbb{E}\left[\frac{\pi_k(x,z)}{\pi_k(x,v_*)^2}\left(\mu_k(x,z)-\mu_k(x,v_*)\right)\left(\pi_k(x,v_*)\frac{\partial t(x,v_*)}{\partial v}-t(x,v_*)\frac{\partial\pi_k(x,v_*)}{\partial v}\right)\frac{\partial\varphi(x,z,\alpha_*)}{\partial\alpha}\right]\end{aligned}\tag{6}$$

with $v_*=\varphi(x,z,\alpha_*)$.

Proof. Because $\gamma_k(x,\varphi(x,z,\alpha);\alpha)$ is the solution to

$$\min_p\mathbb{E}\left[1(d=d_{(k)})(y-p(x,\varphi(x,z,\alpha)))^2\right]$$

we have that for all α

$$\mathbb{E}\left[1(d=d_{(k)})\left(y-\gamma_k(x,\varphi(x,z,\alpha);\alpha)\right)\frac{t(x,\varphi(x,z,\alpha))}{\pi_k(x,\varphi(x,z,\alpha)}\right]=0$$

Differentiating with respect to α and evaluating the result at $\alpha=\alpha_*$, we find after rearranging (6). ■

The next theorem generalizes Theorem 4:

Theorem 6 (Contribution parametric first-step estimator) *The adjustment to the influence function that accounts for the first-stage estimation error is*

$$\begin{aligned}&\left(\mathbb{E}\left[\left(\sum_{k=1}^{K}\left(\frac{\partial h(w,\mu(x,v_*))}{\partial\mu_k}-\frac{\pi_k(x,z)}{\pi_k(x,v_*)}\kappa_k(x,v_*)\right)\frac{\partial\mu_k(x,v_*)}{\partial v}\right)\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]\right.\\ &+\mathbb{E}\left[\left(\sum_{k=1}^{K}\frac{\pi_k(x,z)}{\pi_k(x,v_*)}(\mu_k(x,z)-\mu_k(x,v_*))\frac{\partial\kappa_k(x,v_*)}{\partial v}\right)\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]\\ &\left.-\mathbb{E}\left[\left(\sum_{k=1}^{K}\frac{\pi_k(x,z)}{\pi_k(x,v_*)^2}(\mu_k(x,z)-\mu_k(x,v_*))\kappa_k(x,v_*)\frac{\partial\pi_k(v,v_*)}{\partial v}\right)\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]\right)\sqrt{n}\left(\widehat{\alpha}-\alpha_*\right)\end{aligned}$$

Proof. See Appendix. ■

If h only depends on μ, then $\kappa_k(x,v_*) = \frac{\partial h(\mu(x,v_*))}{\partial \mu_k}$ so that the contribution of the first stage is

$$\left(\mathbb{E}\left[\left(\sum_{k=1}^{K} \frac{\pi_k(x,z) - \pi_k(x,v_*)}{\pi_k(x,v_*)} \frac{\partial h(\mu(x,v_*))}{\partial \mu_k} \frac{\partial \mu_k(x,v_*)}{\partial v} \right) \frac{\partial \varphi(x,z,\alpha_*)}{\partial \alpha} \right] \right.$$
$$+ \mathbb{E}\left[\left(\sum_{k=1}^{K} \sum_{k'=1}^{K} \frac{\pi_k(x,z)}{\pi_k(x,v_*)} (\mu_k(x,z) - \mu_k(x,v_*)) \frac{\partial^2 h(\mu(x,v_*))}{\partial \mu_k \partial \mu_{k'}} \frac{\partial \mu_{k'}(x,v_*)}{\partial v} \right) \frac{\partial \varphi(x,z,\alpha_*)}{\partial \alpha} \right]$$
$$\left. - \mathbb{E}\left[\left(\sum_{k=1}^{K} \frac{\pi_k(x,z)}{\pi_k(x,v_*)^2} (\mu_k(x,z) - \mu_k(x,v_*)) \frac{\partial h(\mu(x,v_*))}{\partial \mu_k} \frac{\partial \pi_k(x,v_*)}{\partial v} \right) \frac{\partial \varphi(x,z,\alpha_*)}{\partial \alpha} \right] \right) \sqrt{n}(\widehat{\alpha} - \alpha_*) \tag{7}$$

3.2 Partial Means: Nonparametric First Step, Non-parametric Second Step

Now assume that the first step consists of a non-parametric regression estimate of $v_* = \varphi_*(x,z) = \mathbb{E}[u \mid x,z]$. We can obtain the adjustment by replicating the arguments in Section 2.3 leading to Theorem 5. Letting

$$\begin{aligned} \delta_3(x,z) = {} & \mathbb{E}\left[\sum_{k=1}^{K} \left(\frac{\partial h(w,\mu(x,v_*))}{\partial \mu_k} - \frac{\pi_k(x,z)}{\pi_k(x,v_*)} \kappa_k(x,v_*) \right) \frac{\partial \mu_k(x,v_*)}{\partial v} \middle| x,z \right] + \\ & \mathbb{E}\left[\sum_{k=1}^{K} \frac{\pi_k(x,z)}{\pi_k(x,v_*)} (\mu_k(x,z) - \mu_k(x,v_*)) \frac{\partial \kappa_k(x,v_*)}{\partial v} \middle| x,z \right] \\ & - \mathbb{E}\left[\sum_{k=1}^{K} \frac{\pi_k(x,z)}{\pi_k(x,v_*)^2} (\mu_k(x,z) - \mu_k(x,v_*)) \kappa_k(x,v_*) \frac{\partial \pi_k(v,v_*)}{\partial v} \middle| x,z \right] \end{aligned} \tag{8}$$

we obtain an analog of Theorem 3:

Theorem 7 (Contribution non-parametric first-step estimator) *The adjustment to the influence function that accounts for the first-stage estimation error is*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_3(x_i,z_i)(u_i - \varphi_*(x_i,z_i))$$

with $\delta_3(x,z)$ given in (8).

If h depends on μ only we replace δ_3 above by

$$\begin{aligned} \delta_3(x,z) = \mathbb{E}\Bigg[& \sum_{k=1}^{K} \frac{\pi_k(x,z) - \pi_k(x,\varphi_*(x,z))}{\pi_k(x,v_*)} \frac{\partial h(\mu(x,\varphi_*(x,z)))}{\partial \mu_k} \frac{\partial \mu_k(x,\varphi_*(x,z))}{\partial v} \\ & + \sum_{k=1}^{K} \sum_{k'=1}^{K} \frac{\pi_k(x,z)}{\pi_k(x,\varphi_*(x,z))} (\mu_k(x,z) - \mu_k(x,\varphi_*(x,z))) \frac{\partial^2 h(\mu(x,\varphi_*(x,z)))}{\partial \mu_k \partial \mu_{k'}} \frac{\partial \mu_{k'}(x,\varphi_*(x,z))}{\partial v} \\ & - \sum_{k=1}^{K} \frac{\pi_k(x,z)}{\pi_k(x,\varphi_*(x,z))^2} (\mu_k(x,z) - \mu_k(x,\varphi_*(x,z))) \frac{\partial h(\mu(x,\varphi_*(x,z)))}{\partial \mu_k} \frac{\partial \pi_k(x,\varphi_*(x,z))}{\partial v} \Bigg| x,z \Bigg] \end{aligned}$$

4 Application: Regression on the Estimated Propensity Score

We consider an intervention with potential outcomes y_0, y_1 that are the control and treated outcome, respectively. The treatment indicator is d and $y = dy_1 + (1-d)y_0$ is the observed outcome. The vector x contains covariates that are not affected by the intervention. As shown by Rosenbaum and Rubin (1983) unconfounded assignment, i.e., the assumption that $y_1, y_0 \perp d|x$, implies $y_1, y_0 \perp d|\varphi(x)$ with $\varphi(x) = \Pr(d=1|x)$ the probability of selection or propensity score. This observation has led to a large number of estimators. The asymptotic variance of the estimators can be compared to the semi-parametric efficiency bound for the ATE derived by Hahn (1998).

The most popular estimators are the matching estimators that estimate the ATE given x or given $\varphi(x)$ by averaging outcomes over units with a 'similar' value of x or $\varphi(x)$. Abadie and Imbens (2009a, 2009b) are recent contributions. They show that matching estimators that have an asymptotic distribution that is notoriously difficult to analyze, are not asymptotically efficient. The second class of estimators do not estimate the ATE given x or $\varphi(x)$ but use the propensity scores as weights. Hahn's (1998) estimator and the estimator of Hirano, Imbens and Ridder (2003) are examples of such estimators. These estimators are asymptotically efficient. The third class of estimators use non-parametric regression to estimate $\mathbb{E}[y|d=1,x]$, $\mathbb{E}[y|d=0,x]$ or $\mathbb{E}[y|d=1,\varphi(x)]$, $\mathbb{E}[y|d=0,\varphi(x)]$. Of these estimators the estimator based on $\mathbb{E}[y|d=1,x]$, $\mathbb{E}[y|d=0,x]$, the imputation estimator, is known to be asymptotically efficient, which suggests that there is no role for the propensity score. The missing result is that for the estimator that uses the non-parametric regression on the propensity score that is estimated in a preliminary step. This estimator that was suggested and analyzed by Heckman, Ichimura, and Todd (HIT) (1998) fits into our framework and is analyzed here.[3]

Our conclusion is that the HIT estimator has the same asymptotic variance as the imputation estimator, so that there is no efficiency gain in using the propensity score. This should settle the issue whether there is a role for the propensity score in achieving semi-parametric efficiency.[4]

4.1 Parametric First Step, Non-parametric Second Step

We have a random sample $s_i = (y_i, x_i, d_i), i = 1, \ldots, n$. The propensity score $\Pr(d=1|x) = \varphi(x,\alpha)$ is parametric and its parameters α are estimated in the first step, by e.g. Maximum Likelihood or OLS (Linear Probability model) or any other method, such that

$$\sqrt{n}\left(\widehat{\alpha} - \alpha_*\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\left(d_i, x_i\right) + o_p(1)$$

with $\mathbb{E}[\psi(d_i, x_i)] = 0$. In the second step, we estimate

$$\mu\left(\varphi(x,\alpha_*)\right) = \left(\mathbb{E}\left[y \mid \varphi(x,\alpha_*), d=1\right], \mathbb{E}\left[y \mid \varphi(x,\alpha_*), d=0\right]\right)',$$

[3]Heckman, Ichimura, and Todd actually consider an estimator of the Average Treatment Effect on the Treated (ATT) that we also analyze.

[4]That does not mean that there is no role for the propensity score in assessing the identification or in improved small sample performance of ATE estimators.

Because we do not observe α_*, we use $\varphi(x_i, \widehat{\alpha})$ in the non-parametric regression.

Our interest is to characterize the first order asymptotic properties of

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\gamma}_1\left(\varphi\left(x_i, \widehat{\alpha}\right)\right) - \widehat{\gamma}_2\left(\varphi\left(x_i, \widehat{\alpha}\right)\right)\right)$$

This estimator can be handled by applying Theorem 6 for the special case that h only depends on μ.

The vector $\mu\left(\varphi(x, \alpha_*)\right)$ is a 2-vector of partial means and d is either $d_{(1)} = 1$ or $d_{(2)} = 0$. Further φ, π_k depend on x only and $\pi_1(x) = \varphi(x, \alpha_*)$ and $\pi_1(\varphi(x, \alpha_*)) = \Pr(d = 1|\varphi(x, \alpha_*)) = \varphi(x, \alpha_*)$. Also $h(\mu) = \mu_1 - \mu_2$ so that the second derivatives are 0. Upon substitution the first two terms on the right-hand side of (7) are 0. Because $\frac{\partial h}{\partial \mu_k}\frac{\partial \pi_k}{\partial v} = 1$ for $k = 1, 2$ we find that the contribution of the first-stage estimator to the influence function is

$$\left(\mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_1}\right] + \mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_2}\right]\right)\sqrt{n}(\hat{\alpha} - \alpha_*) =$$

$$-\mathbb{E}\left[\left(\frac{\mathbb{E}\left[y|\, x, d = 1\right] - \mu_1\left(\varphi\left(x, \alpha_*\right)\right)}{\varphi\left(x, \alpha_*\right)} + \frac{\mathbb{E}\left[y|\, x, d = 0\right] - \mu_2\left(\varphi\left(x, \alpha_*\right)\right)}{1 - \varphi\left(x, \alpha_*\right)}\right)\frac{\partial \varphi\left(x, \alpha_*\right)}{\partial \alpha}\right]\sqrt{n}(\hat{\alpha} - \alpha_*)$$

The contribution of $\widehat{\gamma}$ can be derived using Newey (1994), and is given in (10) below.

We also consider the HIT estimator of the Average Treatment Effect on the Treated (ATT)

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{d_i}{p}\left(\widehat{\gamma}_1\left(\varphi(x_i, \widehat{\alpha})\right) - \widehat{\gamma}_2\left(\varphi(x_i, \widehat{\alpha})\right)\right)$$

with $p = \Pr(d = 1)$. This estimator is a special case of that considered in Theorem 6 with $h(w, \mu_1, \mu_2) = \frac{d}{p}(\mu_1 - \mu_2)$, so that h not only depends on μ_1, μ_2, so that the simplification in (7) does not apply.

Substitution of

$$\frac{\partial h(w, \mu_1, \mu_2)}{\partial \mu_1} = \frac{d}{p} \qquad\qquad \frac{\partial h(w, \mu_1, \mu_2)}{\partial \mu_2} = -\frac{d}{p}$$

and (κ_1, κ_2 are functions of v only)

$$\kappa_1(v) = \frac{v}{p} \qquad\qquad \kappa_2(v) = -\frac{v}{p}$$

give that the first term in Theorem 6 is 0, and the second term is

$$\frac{1}{p}\mathbb{E}\left[\left(\left(\mu_1(x) - \mu_1(\varphi(x, \alpha_*))\right) - \left(\mu_2(x) - \mu_2(\varphi(x, \alpha_*))\right)\right)\frac{\partial \varphi\left(x, \alpha_*\right)}{\partial \alpha}\right]$$

and the third

$$\frac{1}{p}\mathbb{E}\left[\left(\left(\mu_1(x) - \mu_1(\varphi(x, \alpha_*))\right) - \frac{\varphi(x, \alpha_*)}{1 - \varphi(x, \alpha_*)}\left(\mu_2(x) - \mu_2(\varphi(x, \alpha_*))\right)\right)\frac{\partial \varphi\left(x, \alpha_*\right)}{\partial \alpha}\right]$$

so that by taking the difference of the last two terms we find that the contribution is

$$\left(\mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_1}\right] + \mathbb{E}\left[\frac{\partial g\left(s, \alpha_*, \alpha_*\right)}{\partial \alpha_2}\right]\right)\sqrt{n}(\hat{\alpha} - \alpha_*) =$$

$$-\mathbb{E}\left[\frac{\mathbb{E}[y|x, d = 0] - \mu_2\left(\varphi\left(x, \alpha_*\right)\right)}{p\left(1 - \varphi\left(x, \alpha_*\right)\right)}\frac{\partial \varphi(x, \alpha_*)}{\partial \alpha}\right]\sqrt{n}(\hat{\alpha} - \alpha_*)$$

4.2 Non-parametric First Step, Non-parametric Second Step

The analysis in the previous section combined with the results in Newey (1994) shows that in the case that the first stage is non-parametric the contribution of the first-stage estimation to the influence function of the ATE estimator is

$$-\left(\frac{\mathbb{E}\left[y\right|x,d=1\right]-\mu_1\left(\varphi_*\left(x\right)\right)}{\varphi_*\left(x\right)}+\frac{\mathbb{E}\left[y\right|x,d=0\right]-\mu_2\left(\varphi_*\left(x\right)\right)}{1-\varphi_*\left(x\right)}\right)\left(d-\varphi_*(x)\right)$$

which can be alternatively written as

$$\begin{aligned}&-\frac{\mathbb{E}\left[y\right|x,d=1\right]-\mu_1\left(\varphi_*\left(x\right)\right)}{\varphi_*\left(x\right)}d+\left(\mathbb{E}\left[y\right|x,d=1\right]-\mu_1\left(\varphi_*\left(x\right)\right)\right)\\&+\frac{\mathbb{E}\left[y\right|x,d=0\right]-\mu_2\left(\varphi_*\left(x\right)\right)}{1-\varphi_*\left(x\right)}\left(1-d\right)-\left(\mathbb{E}\left[y\right|x,d=0\right]-\mu_2\left(\varphi_*\left(x\right)\right)\right)\end{aligned}\tag{9}$$

To obtain the complete influence function of $\widehat{\beta}$ we need the contribution of the estimation error in $\widehat{\gamma}$. This contribution is equal to[5]

$$\begin{aligned}&\left(\mu_1\left(\varphi_*\left(x\right)\right)-\mu_2\left(\varphi_*\left(x\right)\right)-\beta_*\right)\\&+\frac{d}{\varphi_*\left(x\right)}\left(y-\mu_1\left(\varphi_*\left(x\right)\right)\right)-\frac{1-d}{1-\varphi_*\left(x\right)}\left(y-\mu_2\left(\varphi_*\left(x\right)\right)\right)\end{aligned}\tag{10}$$

Adding (9) and (10), we obtain the influence function of the estimator based on regressions on the estimated propensity score:

$$\left(\mathbb{E}\left[y\right|x,d=1\right]-\mathbb{E}\left[y\right|x,d=0\right]-\beta_*\right)+\frac{d}{\varphi_*\left(x\right)}\left(y-\mathbb{E}\left[y\right|x,d=1\right]\right)-\frac{1-d}{1-\varphi_*\left(x\right)}\left(y-\mathbb{E}\left[y\right|x,d=0\right]\right),$$

which is the influence function of the efficient estimator and also that of the imputation estimator

$$\widehat{\beta}_I=\frac{1}{n}\sum_{i=1}^{n}(\widehat{\lambda}_1(x_i)-\widehat{\lambda}_2(x_i))$$

with $\lambda_1\left(x\right)=\mathbb{E}[y|x,d=1],\lambda_2\left(x\right)=\mathbb{E}[y|x,d=0]$. The imputation estimator involves non-parametric regressions on x and not on the estimated propensity score. However these two estimators have the same influence function which shows that regressing on the non-parametrically estimated propensity score does not result in an efficiency gain. The infeasible estimator that depends on non-parametric regressions on the population propensity score is less efficient than the estimator that uses the estimated propensity score.

For the estimator of the ATT the contribution of the first stage is

$$-\frac{\mathbb{E}[y|x,d=0]-\mu_2\left(\varphi_*(x)\right)}{p\left(1-\varphi_*(x)\right)}\left(d-\varphi_*(x)\right)$$

[5]Derivation is straightforward, and is contained in a previous version of the paper, which is available on request.

The main term and the contribution of the estimation of the (infeasible) non-parametric regressions is[6]

$$\frac{d}{p}\left(y-\mu_1\left(\varphi_*\left(x\right)\right)\right)-\frac{(1-d)\varphi_*\left(x\right)}{p\left(1-\varphi_*\left(x\right)\right)}\left(y-\mu_2\left(\varphi_*\left(x\right)\right)\right)+\frac{d}{p}\left(\mu_1\left(\varphi_*\left(x\right)\right)-\mu_2\left(\varphi_*\left(x\right)\right)-\beta_*\right).$$

Adding these expressions we obtain the full influence function

$$\frac{d}{p}\left(y-\mathbb{E}[y|x,d=1]\right)-\frac{(1-d)\varphi_*(x)}{p(1-\varphi_*(x))}\left(y-\mathbb{E}[y|x,d=0]\right)+\frac{d}{p}\left(\mathbb{E}[y|x,d=1]-\mathbb{E}[y|x,d=0]-\beta_*\right)$$

As in the case of the ATE the influence function is the same as that for the estimator that involves non-parametric regressions on x and not on the estimated propensity score, so that again there is no first-order asymptotic efficiency gain if we use the estimated propensity score in the non-parametric regressions.

It should be noted that the influence functions derived in this section are different from those found in the literature. Recently, Mammen, Rothe, and Schienle (2010) derived the influence function for the ATE estimator considered in this section. They concluded that it is identical to that of the infeasible estimator that regresses on the population propensity score. This is because they imposed the index assumption $E\left[y|\,d,x\right]=E\left[y|\,d,\varphi(x)\right]$, which is not made in the standard program evaluation literature, because it restricts the distribution of the potential outcomes. For instance, in a linear selection (on observables) model the index restriction implies that the regression coefficients in the outcome equations are proportional to those in the selection equation. HIT derived the influence function for the ATT estimator that is also different from ours. In this case, the difference appears to be due to an error in their derivation, which fails to account for the effect of the first-stage estimation on the conditional expectation in the second stage.

5 Summary

We studied the asymptotic distribution of three-step estimators of a finite dimensional parameter vector where the second step consists of one or more non-parametric regressions on a regressor that is estimated in the first step. The first step estimator is either parametric or non-parametric. We showed that Newey's (1994) method can be used to determine the contribution of the first-step estimation error on the influence function. In doing so it is essential to recognize that the first-stage estimate has two effects on the sampling distribution of the finite-dimensional parameter vector. First, the first-stage estimate enters the argument at which the conditional expectation is evaluated, second, the first-stage estimate changes the conditional expectation itself. In the literature the second contribution of the first-stage estimate to the influence function is sometimes forgotten. Our contribution is that we show how to derive this contribution so that we obtain the correct influence function for three- or more stage estimators.

[6] This can be derived using an argument that leads to (10).

Appendix

Proof of Theorem 2 We write

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h\left(\gamma\left(x_i,\varphi\left(x_i,z_i,\widehat{\alpha}\right);\widehat{\alpha}\right)\right)-h\left(\gamma\left(x_i,\varphi\left(x_i,z_i,\alpha_*\right);\alpha_*\right)\right)\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(g\left(s_i,\widehat{\alpha}_1,\widehat{\alpha}_2\right)-g\left(s_i,\alpha_*,\alpha_*\right)\right)$$
$$=\left(\mathbb{E}\left[\frac{\partial g\left(s,\alpha_*,\alpha_*\right)}{\partial\alpha_1}\right]+\mathbb{E}\left[\frac{\partial g\left(s,\alpha_*,\alpha_*\right)}{\partial\alpha_2}\right]\right)\sqrt{n}(\widehat{\alpha}-\alpha_*)+o_p\left(1\right)$$

Therefore we must compute $\mathbb{E}\left[\frac{\partial g(s,\alpha_*,\alpha_*)}{\partial\alpha_1}\right]$ and $\mathbb{E}\left[\frac{\partial g(s,\alpha_*,\alpha_*)}{\partial\alpha_2}\right]$. The computation of the first expectation is easy. Because $\gamma\left(x,\varphi\left(x,z,\alpha\right);\alpha_*\right)=\mu\left(x,\varphi\left(x,z,\alpha\right)\right)$, we have

$$\mathbb{E}\left[\frac{\partial g\left(s,\alpha_*,\alpha_*\right)}{\partial\alpha_1}\right]=\sum_{j=1}^{J}\mathbb{E}\left[\frac{\partial h\left(\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu_j}\frac{\partial\mu_j\left(x,\varphi\left(x,z,\alpha_*\right)\right)}{\partial v}\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]$$

where μ_j denotes the j-th component of μ, etc. We now tackle the second expectation. By the chain rule

$$\mathbb{E}\left[\frac{\partial g\left(s,\alpha_*,\alpha_*\right)}{\partial\alpha_2}\right]=\sum_{j=1}^{J}\mathbb{E}\left[\frac{\partial h\left(\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu_j}\frac{\partial\gamma_j\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)}{\partial\alpha_2}\right]\tag{11}$$

We compute the right hand side of (11) using Lemma 1.

$$\mathbb{E}\left[\frac{\partial g\left(s,\alpha_*,\alpha_*\right)}{\partial\alpha_2}\right]=\sum_{j}\mathbb{E}\left[\frac{\partial h\left(\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu_j}\frac{\partial\gamma_j(x,\varphi(x,z,\alpha_*);\alpha_*)}{\partial\alpha_2}\right]=$$

$$-\sum_{j=1}^{J}\mathbb{E}\left[\frac{\partial h\left(\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu_j}\frac{\partial\mu_j(x,\varphi(x,z,\alpha_*))}{\partial v}\frac{\partial\varphi(x,z,\alpha_*)}{\partial\alpha}\right]+$$

$$\sum_{j=1}^{J}\mathbb{E}\left[\left(y_j-\mu_j\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)\frac{\partial^2 h\left(\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu_j^2}\frac{\partial\mu_j\left(x,\varphi\left(x,z,\alpha_*\right)\right)}{\partial v}\frac{\partial\varphi(x,z,\alpha_*)}{\partial\alpha}\right]$$

Adding $\mathbb{E}\left[\frac{\partial g(s,\alpha_*,\alpha_*)}{\partial\alpha_1}\right]$ we find the desired result.$\blacksquare$

Proof of Theorem 4 The contribution of $\widehat{\alpha}$ is the sum of

$$\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)\right)}{\partial\alpha_1}\right]\sqrt{n}\left(\widehat{\alpha}-\alpha_*\right)\tag{12}$$

and

$$\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)\right)}{\partial\alpha_2}\right]\sqrt{n}\left(\widehat{\alpha}-\alpha_*\right)\tag{13}$$

Note that

$$\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_1\right);\alpha_*\right)\right)}{\partial\alpha_1}\right]\bigg|_{\alpha_1=\alpha_*}=\mathbb{E}\left[\frac{\partial h\left(w,\mu\left(x,v_*\right)\right)}{\partial\mu}\frac{\partial\mu(x,v_*)}{\partial v}\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right] \tag{14}$$

Because $\frac{\partial\gamma(x,\varphi(x,z,\alpha_*);\alpha_*)}{\partial\alpha}$ is a function of $(x,\varphi\left(x,z,\alpha_*\right))$, we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)\right)}{\partial\alpha_2}\right]\\
&=\mathbb{E}\left[\frac{\partial h\left(w,\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu}\frac{\partial\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)}{\partial\alpha_2}\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\frac{\partial h\left(w,\mu\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right)}{\partial\mu}\bigg|\,x,\varphi\left(x,z,\alpha_*\right)\right]\frac{\partial\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)}{\partial\alpha_2}\right]\\
&=\mathbb{E}\left[\kappa\left(x,\varphi\left(x,z,\alpha_*\right)\right)\frac{\partial\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)}{\partial\alpha_2}\right]
\end{aligned} \tag{15}$$

By Lemma 1

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)}{\partial\alpha_2}\kappa\left(x,\varphi\left(x,z,\alpha_*\right)\right)\right]\\
&=\mathbb{E}\left[\left(\mu(x,z)-\mu\left(x,v_*\right)\right)\frac{\partial\kappa\left(x,v_*\right)}{\partial v}\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]\\
&-\mathbb{E}\left[\kappa\left(x,v_*\right)\frac{\partial\mu\left(x,v_*\right)}{\partial v}\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]
\end{aligned} \tag{16}$$

Combining (14) - (16), we conclude that

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)\right)}{\partial\alpha_1}\right]+\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)\right)}{\partial\alpha_2}\right]\\
&=\mathbb{E}\left[\left(\frac{\partial h\left(w,\mu\left(x,v_*\right)\right)}{\partial\mu}-\kappa\left(x,v_*\right)\right)\frac{\partial\mu\left(x,v_*\right)}{\partial v}\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]\\
&+\mathbb{E}\left[\frac{\partial\kappa\left(x,v_*\right)}{\partial v}\left(\mu(x,z)-\mu\left(x,v_*\right)\right)\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right]
\end{aligned}$$

which gives us the desired result.■

Proof of Theorem 6 Note that

$$\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_1\right);\alpha_*\right)\right)}{\partial\alpha_1}\right]\bigg|_{\alpha_1=\alpha_*}=\mathbb{E}\left[\left(\sum_{k=1}^{K}\frac{\partial h\left(w,\mu\left(x,v_*\right)\right)}{\partial\mu_k}\frac{\partial\mu_k(x,v_*)}{\partial v}\right)\frac{\partial\varphi\left(x,z,\alpha_*\right)}{\partial\alpha}\right] \tag{17}$$

Because $\frac{\partial \gamma_k(x,v_*;\alpha_*)}{\partial \alpha}$ is a function of (x, v_*), we have

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial h\left(w,\gamma\left(x,\varphi\left(x,z,\alpha_*\right);\alpha_*\right)\right)}{\partial \alpha_2}\right] &= \mathbb{E}\left[\sum_{k=1}^{K}\mathbb{E}\left[\left.\frac{\partial h\left(w,\mu\left(x,v_*\right)\right)}{\partial \mu_k}\right| x,v_*\right]\frac{\partial \gamma_k(x,v_*;\alpha_*)}{\partial \alpha}\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{K}\kappa_k\left(x,v_*\right)\frac{\partial \gamma_k(x,v_*;\alpha_*)}{\partial \alpha}\right] \qquad (18)
\end{aligned}$$

By Lemma 2

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k=1}^{K}\kappa_k\left(x,v_*\right)\frac{\partial \gamma_k\left(x,v_*;\alpha_*\right)}{\partial \alpha}\right] \\
&= -\mathbb{E}\left[\sum_{k}\frac{\pi_k(x,z)}{\pi_k(x,v_*)}\kappa_k\left(x,v_*\right)\frac{\partial \mu_k(x,v_*)}{\partial v}\frac{\partial \varphi(x,z,\alpha_*)}{\partial \alpha}\right] \\
&+ \mathbb{E}\left[\sum_{k=1}^{K}\frac{\pi_k(x,z)}{\pi_k(x,v_*)}\left(\mu_k\left(x,z\right)-\mu_k\left(x,v_*\right)\right)\frac{\partial \kappa_k(x,v_*)}{\partial v}\frac{\partial \varphi(x,z,\alpha_*)}{\partial \alpha}\right] \\
&- \mathbb{E}\left[\sum_{k=1}^{K}\frac{\pi_k(x,z)}{\pi_k(x,v_*)^2}\left(\mu_k\left(x,z\right)-\mu_k\left(x,v_*\right)\right)\kappa_k\left(x,v_*\right)\frac{\partial \pi_k\left(x,v_*\right)}{\partial v}\frac{\partial \varphi(x,z,\alpha_*)}{\partial \alpha}\right] \qquad (19)
\end{aligned}$$

Combining (17) - (19), we obtain the desired conclusion. ■

References

[1] Abadie, A. and G. Imbens (2009a),“A Martingale Representation for Matching Estimators,” unpublished working paper.

[2] Abadie, A. and G. Imbens (2009b),“Matching on the Estimated Propensity Score, ” unpublished working paper.

[3] Hahn, J. (1998): “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects, ” *Econometrica*, 66,pp. 315–331.

[4] Heckman, J.J., H. Ichimura, and P. Todd (1998): “Matching as an Econometric Evaluation Estimator,” *Review of Economic Studies* 65, pp. 261–294.

[5] Hirano, K., G. Imbens, and G. Ridder (2003): “Efficient Estimation of the Average Treatment Effect Using the Estimated Propensity Score, ” *Econometrica*, 71, pp. 1161–1189.

[6] Klein, R., C. Shen, and F. Vella (2010): “Triangular Semiparametric Models Featuring Two Dependent Endogenous Binary Outcomes,” unpublished working paper.

[7] Mammen, E., C. Rothe, and M. Schienle (2010): “Nonparametric Regression with Nonparametrically Generated Covariates, ” unpublished working paper.

[8] Murphy, K. M. and R. H. Topel (1985): “Estimation and Inference in Two-Step Econometric Models,” *Journal of Business and Economic Statistics* 3, pp. 370 – 379.

[9] Newey, W.K. (1994): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica* 62, pp. 1349–1382.

[10] Pagan, A. (1984): “Econometric Issues in the Analysis of Regressions with Generated Regressors,” *International Economic Review* 25, pp. 221–247.

[11] Rosenbaum, P., and D. Rubin (1983): “The Central Role of the Propensity Score in Observational Studies for Causal effects, ” *Biometrika* 70, pp. 41–55.