

Semiparametric efficiency in nonlinear LATE models

HAN HONG

Departments of Economics, Stanford University

DENIS NEKIPELOV

Departments of Economics, University of California at Berkeley

In this paper we study semiparametric efficiency for the estimation of a finite-dimensional parameter defined by generalized moment conditions under the local instrumental variable assumptions. These parameters identify treatment effects on the set of compliers under the monotonicity assumption. The distributions of covariates, the treatment dummy, and the binary instrument are not specified in a parametric form, making the model semiparametric. We derive the semiparametric efficiency bounds for both conditional models and unconditional models. We also develop multistep semiparametric efficient estimators that achieve the semiparametric efficiency bound. To illustrate the efficiency gains from using the optimal semiparametric weights, we design a Monte Carlo study. It demonstrates that our semiparametric estimator performs well in nonlinear models.

KEYWORDS. Semiparametric efficiency bound, local treatment effect, FTP, child achievement, unemployment benefits.

JEL CLASSIFICATION. C25, C26, C31.

1. INTRODUCTION

Semiparametric efficiency is an important issue in the estimation of treatment effect models and models with endogenous regressors; see, for example, Chernozhukov and Hansen (2005) and Newey (1990a), among others. Under the strong ignorability assumption, Hahn (1998) and Hirano, Imbens, and Ridder (2003) derived the semiparametric efficiency bound and developed semiparametric efficient estimators for the averaged treatment effect and the averaged treatment effect on the treated. Firpo (2003) extended their analyses to quantile treatment effects.

Han Hong: hanhong@stanford.edu

Denis Nekipelov: nekiyelov@econ.berkeley.edu

The authors acknowledge generous research supports from the NSF and excellent research assistance from Tim Armstrong. The usual disclaimer applies. We would like to thank Alberto Abadie, Shakeeb Khan, Guido Imbens, Bernard Salanié, Ed Vytlačil, conference participants in the 2007 NASM meeting in Duke University, and seminar participants at Columbia University, University of Pennsylvania, Harvard, and MIT for helpful comments. The authors have made use of the FTP data without representing positions by the State of Florida and MDRC.

Copyright © 2010 Han Hong and Denis Nekipelov. Licensed under the [Creative Commons Attribution-NonCommercial License 3.0](http://creativecommons.org/licenses/by-nc/3.0/). Available at <http://www.qeconomics.org>.

DOI: 10.3982/QE43

An alternative approach to address the endogeneity problem is based on the local instrumental variable (LIV) method. The baseline model for this method has a dummy endogenous regressor and a dummy instrument variable. Under the LIV assumption, the instrumental variable weakly changes the endogenous regressor in one direction. [Abadie \(2003\)](#) showed that the entire distributional causal effect is identified for the *complier* population where the endogenous regressor changes from 0 to 1 as the instrumental variable changes from 0 to 1, and proposed linear and nonlinear conditional local average treatment effects (LATE) models as an extension of [Imbens and Angrist \(1994\)](#) and [Angrist, Imbens, and Rubin \(1996\)](#) for the case when the instrument is valid conditional on a vector of covariates, X . In the context of quantile regression, conditional LATE models were first applied in [Abadie, Angrist, and Imbens \(2002\)](#). In contrast to the strong ignorability assumption, semiparametric efficiency under the LIV assumption has not been subject to careful studies. An exception is [Frolich \(2007\)](#), who derived the efficiency bound for the average treatment effect for compliers and showed that the propensity score, properly defined in the LIV context, does not affect the efficiency bound. [Henderson, Millimet, Parmeter, and Wang \(2006\)](#) applied the estimator for the averaged treatment effect on compliers to a fertility analysis. We emphasize that in this paper, we do not develop new models of treatment effects for compliers. The generality that we consider is solely aimed at encompassing the existing conditional and unconditional models of treatment effects for compliers.

We make several theoretical contributions in this paper. First we derive the semiparametric efficiency bound for both unconditional and conditional versions of the nonlinear treatment effect parameters, particularly in the context of a general nonlinear conditional mean treatment effect model developed in [Abadie \(2003\)](#). We illustrate the specialization of the efficiency bounds to the quantile and linear treatment effect parameters of [Abadie, Angrist, and Imbens \(2002\)](#) and [Abadie \(2003\)](#). Our semiparametric efficiency calculations include both conditional models and unconditional models, which characterize different treatment effect parameters. The unconditional efficiency bounds include as a special case the mean parameter of [Frolich \(2007\)](#), and also include the treatment effect on the treated compliers, which is related to the average treatment effect of the treated (ATT) when endogeneity is absent. Our results also generalize the efficiency calculations in [Hahn \(1998\)](#) and [Chen, Hong, and Tarozzi \(2008\)](#). Second, we show that the semiparametric efficiency bounds for the treatment effect of treated compliers are different when the propensity score is unknown, is known, or is correctly specified parametrically. We also simplify the structure of the efficiency analysis compared to the existing literature.

In addition, we develop semiparametric estimators that achieve the theoretical efficiency bounds. Efficient estimators are developed for both conditional and unconditional models. In the conditional case, we identify a member among a class of estimators admissible under the structure given in [Abadie \(2003\)](#) that achieves the efficiency bound. The structure of the model allows us to make use of the binary instrument feature of a conditional moment model and to reduce the problem of finding semiparametric efficiency bound to the moment-based framework as in [Newey \(1990b\)](#), [Bickel, Klaassen, Ritov, and Wellner \(1993\)](#), and [Robins and Rotnitzky \(1995\)](#). For unconditional

models, we described efficient estimators for both the treatment effect of compliers and the treatment effect of treated compliers for cases when the propensity score is unknown, known, and parametrically specified. In general, efficiency can be achieved by choosing the instrument functions optimally in the propensity weighting framework of [Abadie \(2003\)](#) or in a conditional expectation projection framework. We demonstrate the efficiency gain from using the optimal weights in a set of Monte Carlo experiments.

Section 2 develops the semiparametric efficiency results for the complier treatment effect model in [Abadie \(2003\)](#). Section 3 develops efficient estimators that achieve the semiparametric efficiency bound, and explicitly quantifies the amount of efficiency improvement over existing methods. This section also gives regularity assumptions that validate the proposed semiparametric efficient estimator. Each of these two sections also discusses extensions to parameters that are defined unconditionally. Section 4 reports the results from a simulation exercise. Finally Section 5 concludes. The Appendix is provided as Supplemental Material ([Hong and Nekipelov \(2010\)](#)): it contains the mathematical proofs and also an application of the efficient estimator to the Florida Transition Program which was offered as an alternative to the existing state welfare system in Florida.

2. SEMIPARAMETRIC EFFICIENCY BOUND

2.1 Local treatment effect parameters

The local (complier) treatment effect model (see [Imbens and Angrist \(1994\)](#) and [Abadie \(2003\)](#), for example) is defined through a random vector $(Y_1, Y_0)' \in \mathbb{R}^2$, a vector of binary variables $(D_1, D_0)' \in \{0, 1\} \times \{0, 1\}$, a binary instrument $Z \in \{0, 1\}$, and a vector of covariates $X \in \mathcal{X} \subset \mathbb{R}^k$. The following assumptions are used by these authors to describe the distributions of the variables under consideration:

ASSUMPTION 1. *Almost everywhere in $\mathcal{X}$,*

- (i) $(Y_1, Y_0, D_1, D_0)' \perp Z|X$,
- (ii) $E[D_1|X] \neq E[D_0|X]$,
- (iii) $\Pr(Z = 1|X) \in (0, 1)$,
- (iv) $\Pr(D_1 \geq D_0|X) = 1$.

Under these four assumptions, in particular the last assumption (monotonicity), the data directly identify the differences between the cohort that would have been treated for both values of the instrument (always-takers) and the cohort that would not have been treated under any circumstances (never-takers). The combination of the always-taker cohort and the never-taker cohort indirectly recovers the compliers, which is the cohort that change behavior when the instrument changes. The variables in the model, Y_1 , Y_0 , D_1 , and D_0 are not always completely observable. Only the following transformed variables are observed:

$$Y = Y_1 D + Y_0 (1 - D) \quad \text{and} \quad D = D_0 + Z(D_1 - D_0).$$

Note that in this setup, covariates X can be endogenous. The binary treatment D is endogenous by construction. On the other hand, as we will see later, the presence of the “switching” dummy Z allows us to recover the effect of the treatment for some subpopulation of the sample. In this way, provided the assumption of conditional independence of Z , and Y_i and D_i ($i = 0, 1$) given X , we can use variable Z as an instrument for the endogenous treatment.

Due to Assumption 1(i), the conditional probabilities of the observable binary variable D can be written as $\mathcal{P}_i(X) = P(D = 1|Z = i, X) = E[D_i|X]$, $i = 0, 1$, where the second equalities follow from the conditional independence Assumption 1(i).

Also define $\mathcal{Q}(X) = E[Z|X]$. Consequently, the conditional probability of the binary treatment d given the instrument in terms of the probabilities of treatment dummies can be expressed as

$$P(D = 1|Z = z, X = x) = \mathcal{F}(z, x) = \mathcal{P}_1(x)z + \mathcal{P}_0(x)(1 - z).$$

Taking expectation over Z given X produces the conditional probability of d given only X : $\mathbf{P}(x) = \mathcal{P}_1(x)\mathcal{Q}(x) + \mathcal{P}_0(x)(1 - \mathcal{Q}(x))$.

The objects of interest that can be identified under Assumption 1 are the distributions of the outcomes Y_1 and Y_0 given $D_1 > D_0$ (implying that $D_1 = 1$ and $D_0 = 0$): for $j = 0, 1$, $f(y_j|D_1 > D_0, X = x)$. The subpopulation for which $D_1 > D_0$ is usually referred to as *compliers*, for whom random selection into treatment affects the treatment dummy monotonically. Under the monotonicity Assumption 1(iv), the distributions of compliers can be expressed in terms of the observed conditional distributions:

$$\begin{aligned} f_{**}(y|x, d = 1) \\ &\equiv f(y_1 = y|D_1 > D_0, x) \\ &= \frac{\mathcal{P}_1(x)}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} f(y|d = 1, z = 1, x) - \frac{\mathcal{P}_0(x)}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} f(y|d = 1, z = 0, x). \end{aligned} \quad (1)$$

To see this relation, note that under the monotonicity assumption, $\mathcal{P}_0(x)$ is the proportion of always-takers ($D_0 = D_1 = 1$) conditional on x while $\mathcal{P}_1(x)$ is the sum of always-takers and compliers. $f(y_1|d = 1, z = 1, x)$ gives the distribution of y_1 conditional on being either an always-taker or a complier and the covariate x . $f(y_1|d = 1, z = 0, x)$ gives the distribution of y_1 conditional on being just an always-taker and x . Therefore, $\mathcal{P}_1(x)f(y_1|d = 1, z = 1, x)$ can be written as a linear combination for the known distribution of always-takers and the unknown distribution for compliers. Similarly, one can write the joint distribution of y_0 and the event of being either a never-taker or a complier, which is $(1 - \mathcal{P}_0(x))f(y_0|d = 0, z = 0, x)$, as a linear combination of the distributions for never-takers and compliers. Hence

$$\begin{aligned} f_{**}(y|X = x, d = 0) \\ &\equiv f(Y_0 = y|D_1 > D_0, x) \\ &= \frac{1 - \mathcal{P}_0(x)}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} f(y|d = 0, Z = 0, x) - \frac{1 - \mathcal{P}_1(x)}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} f(y|d = 0, Z = 1, x). \end{aligned} \quad (2)$$

The semiparametric model that we consider incorporates the linear quantile regression model of [Abadie, Angrist, and Imbens \(2002\)](#) to a parameter vector β determined by a conditional moment equation, $\forall x$ and $\forall d$,

$$\varphi(\beta, x, d) = E[g(y, d, x, \beta)|x, d, D_1 > D_0] = 0 \quad (3)$$

for some parametric function $g(\cdot)$.¹ The conditional expectation for $d = 1, 0$ is taken with respect to the corresponding conditional density $f_{**}(y|x, d)$.

Two direct applications of this general definition are the mean treatment effect of [Imbens and Angrist \(1994\)](#) and the quantile treatment effect of [Abadie, Angrist, and Imbens \(2002\)](#). The mean treatment effect model corresponds to a moment condition

$$g(y, d, x, \beta) = y - \beta_1 d - (1 - d)\beta_0 - \beta'_2 x.$$

The quantile treatment effect model characterizes the difference in conditional distributions of potential outcomes y_1 and y_0 for compliers through a linear specification of the conditional quantile functions: $Q_\tau(y|x, d, D_1 > D_0) = \beta_0 d + \beta'_1 x$. The corresponding moment function that defines the quantile treatment effect (QTE) parameter is, therefore,

$$g(y, d, x, \beta) = 1(y \leq \beta_0 d + \beta'_1 x) - \tau.$$

These models can be extended to allow for a semiparametric component in the conditional moment function. For $\mu(x)$ being a nonparametric function of x , we may consider estimating $\mu(x)$ and β simultaneously in the moment function: $g(y, d, x, \mu(x), \beta)$. For example, the parametric mean treatment effect model can be generalized to a semiparametric partial linear model:

$$g(y, d, x, \mu(x), \beta) = y - \beta_1 d - (1 - d)\beta_0 - \mu(x).$$

In the rest of the paper, we derive semiparametric efficiency bounds for the parameter vector β and develop a semiparametric procedure that achieves the efficiency bound. This framework can be extended to derive the semiparametric efficiency bound for a nonparametric component in the specification of the conditional moment equations.

2.2 Efficiency bound for treatment effect parameters

We will use the arguments of [Newey \(1990a\)](#) and [Severini and Tripathi \(2001\)](#) to construct the efficiency bounds for the system of conditional moments. More specifically, given a set of instrument functions of the covariates x , the conditional moments are first transformed into a system of unconditional moments. Then choosing the instrument functions optimally will produce the semiparametric efficiency bound of the conditional moment model.²

¹The class of conditional models of treatment effects for compliers was developed in [Abadie \(2003\)](#).

²We note that the finiteness of the efficiency bound may involve some strong conditions on the absolute integrability of the inverse propensity score function, noted in [Khan and Tamer \(2009\)](#). When these conditions are not satisfied, one cannot estimate the treatment effect estimator at a parametric rate and the efficiency bound becomes meaningless. One, however, may use the efficient procedure outlined in [Khan and Nekipelov \(2010\)](#) even in such a nonregular case.

THEOREM 1. *Under Assumption 1, the semiparametric efficiency bound for a k -dimensional parameter β that characterizes the subsample of compliers in (3) can be expressed as:*

$$V(\beta) = E \left((\mathcal{P}_1(x) - \mathcal{P}_0(x))^2 E \left[\frac{\partial \varphi(\beta, d, x)}{\partial \beta} \zeta(x, d)' \middle| x \right] \right. \\ \left. \times \bar{\Omega}(x)^{-1} E \left[\zeta(x, d) \frac{\partial \varphi(\beta, d, x)'}{\partial \beta} \middle| x \right] \right)^{-1}.$$

Denote $\omega_{d,z}(x) = V(g(y, d, x, \beta) | d, z, x)$ and $\gamma_{d,z}(x) = E(g(y, d, x, \beta) | d, z, x)$. We can then express the elements of the matrix $\bar{\Omega}(x)$ in the manner

$$\bar{\Omega}_{11}(x) = \left(\frac{\mathcal{P}_1(x)\omega_{11}(x)}{\mathcal{Q}(x)} + \frac{\mathcal{P}_0(x)\omega_{10}(x)}{1 - \mathcal{Q}(x)} \right. \\ \left. + \frac{\gamma_{11}^2(x)\mathcal{P}_1(x)\mathbf{P}(x)}{\mathcal{P}_0(x)\mathcal{Q}(x)(1 - \mathcal{Q}(x))} \left[1 - \frac{\mathcal{P}_1(x)\mathcal{P}_0(x)}{\mathbf{P}(x)} \right] \right), \\ \bar{\Omega}_{22}(x) = \left(\frac{(1 - \mathcal{P}_1(x))\omega_{01}(x)}{\mathcal{Q}(x)} + \frac{(1 - \mathcal{P}_0(x))\omega_{00}(x)}{1 - \mathcal{Q}(x)} \right. \\ \left. + \frac{\gamma_{00}^2(x)(1 - \mathcal{P}_0(x))(1 - \mathbf{P}(x))}{\mathcal{Q}(x)(1 - \mathcal{Q}(x))(1 - \mathcal{P}_1(x))} \left[1 - \frac{(1 - \mathcal{P}_0(x))(1 - \mathcal{P}_1(x))}{1 - \mathbf{P}(x)} \right] \right),$$

and

$$\bar{\Omega}_{21}(x) = \bar{\Omega}_{12}(x) = \left(\frac{\mathcal{P}_1(x)(1 - \mathcal{P}_0(x))}{\mathcal{Q}(x)(1 - \mathcal{Q}(x))} \gamma_{11}(x)\gamma_{00}(x) \right).$$

In this theorem, we have also used the notation

$$\zeta(d, x) = \left(\frac{d}{\mathbf{P}(x)}, \frac{1 - d}{1 - \mathbf{P}(x)} \right)'.$$

This structure of the variance bound shows several visible features. First of all, the semi-parametric efficiency bound will grow if the fraction of compliers $\mathcal{P}_1(x) - \mathcal{P}_0(x)$ in the sample decreases. Moreover, the efficiency bound will be higher if the binary instrument is taking one of the values most of the time, in which case $\mathcal{Q}(x)$ is closer to 0 or 1. In addition, the proof and the estimation section show that the structure of the variance reflects the optimal instrument function as $\mathcal{M}(x)\zeta(x, d)$, where

$$\mathcal{M}(x) = E \left[\frac{\partial \varphi(d, x, \beta)}{\partial \beta} \zeta(x, d)' \middle| x \right] \bar{\Omega}(x)^{-1} \text{diag} \left\{ \frac{\mathbf{P}(x)}{\mathcal{Q}(x)}, \frac{1 - \mathbf{P}(x)}{1 - \mathcal{Q}(x)} \right\}.$$

2.3 Unconditional parameters

Often times researchers can be mainly interested in parameters that are defined unconditionally. For example, under the unconfoundedness assumption where the latent outcome is conditionally independent of the treatment status given exogenous covariates,

the semiparametric efficiency literature has focused on the average treatment effect and the average treatment effect on the treated, both of which are defined unconditionally with respect to the exogenous covariates X .

Under the unconfoundedness assumption, one can also specify a model where the average treatment effect or effect on the treated conditional on each covariate is constant or a known parametric function of the covariates, similar to the analysis in the previous section and in [Abadie, Angrist, and Imbens \(2002\)](#). However, most of the literature has focused on analyzing the average treatment effect or effect on the treated without requiring that this effect is a constant conditional on every value of the exogenous covariate.

When X is not a constant, the conditional model and the unconditional model imply very different parameters of interest. For example, the semiparametric efficiency bound for an average treatment effect that is assumed to be constant across all possible values of covariate X is tighter than that for the average treatment effect defined unconditionally with respect to the covariates X . This section investigates efficient estimators for unconditionally defined treatment effect parameters under the LIV monotonicity assumption.

2.4 Semiparametric efficiency of unconditional mean treatment effects

This section will restrict attention to mean effect parameters to illustrate the ideas. However, the results are readily extendible to general moment conditions in Section 3.6. Specifically, we consider the average treatment effect on compliers (ATEC) $\beta \equiv \beta_1 - \beta_0 = E(Y_1 - Y_0 | D_1 > D_0)$ and the average treatment effect on the treated compliers (ATTC) $\gamma \equiv \gamma_1 - \gamma_0 = E(Y_1 - Y_0 | d = 1, D_1 > D_0)$. These parameters reduce to the usual notation of average treatment effect (ATE) and effect on the treated (ATT) under strong ignorability when $P(D_1 > D_0) = 1$. The efficiency bound for ATEC was derived by [Frolich \(2007\)](#), although we develop a simplified derivation. Our results for ATTC are new and are applicable when the propensity score $Q(x)$ is unknown, known, or parametrically specified. The first theorem considers unknown propensity scores.

THEOREM 2. *The semiparametric efficient bound for β is given by the variance of the efficient influence function*

$$\begin{aligned} & \frac{1}{P(D_1 > D_0)} \left\{ \frac{z}{Q(x)} (y - E(Y|Z = 1, x)) + E(Y|Z = 1, x) \right. \\ & \quad - \frac{1 - z}{1 - Q(x)} (y - E(Y|Z = 0, x)) - E(Y|Z = 0, x) \\ & \quad - \left(\frac{z}{Q(x)} (d - E(D|Z = 1, x)) + E(D|Z = 1, x) \right. \\ & \quad \left. \left. - \frac{1 - z}{1 - Q(x)} (d - E(D|Z = 0, x)) - E(D|Z = 0, x) \right) \beta \right\}, \end{aligned}$$

while the semiparametric efficiency bound for γ is given by the variance of the efficient influence function

$$\frac{1}{P(d=1, D_1 > D_0)} \left\{ y - \frac{1-z}{1-Q(x)} (y - E(Y|Z=0, x)) - E(Y|Z=0, x) \right. \\ \left. - \left(d - \frac{1-z}{1-Q(x)} (d - E(D|Z=0, x)) - E(D|Z=0, x) \right) \gamma \right\}.$$

Obviously, under the strong ignorability assumption when $Z = D$, $P(D_1 > D_0) = 1$, both of these reduce to the corresponding influence functions derived in Hahn (1998). In fact, the only difference (other than the factor outside the bracket) is in the coefficient in front of β and γ , which become 1 and z under strong ignorability.

The literature has also been concerned with the semiparametric efficiency when the so-called propensity score, in our case $Q(x)$, is either known or parametrically specified. We will still leave $\mathcal{P}_1(x) - \mathcal{P}_0(x)$ nonparametrically specified, even though cases when this is known or parametrically specified can be analyzed too.

From the proof of Theorem 2, it is clear that $Q(x)$ does not even enter the moment conditions that define the parameters β . (See equations (15) and (17) in Appendix C). Consequently, any knowledge of $Q(x)$ will have no impact on the efficiency bound for β .

Such knowledge, however, will improve the efficiency bound for γ , as described in the following theorem.

THEOREM 3. *When the propensity score $Q(x; \alpha)$ is correctly specified up to a finite-dimensional parameter α , the semiparametric efficiency bound for γ is the variance of the efficient influence function*

$$\frac{1}{P(D=1, D_1 > D_0)} \left\{ z(y - E(Y|Z_1=1, x)) + Q(x)E(Y|Z_1=1, x) \right. \\ \left. - \frac{1-z}{1-Q(x)} Q(x)[y - E(Y|Z=0, x)] - Q(x)E(Y|Z=0, x) \right. \\ \left. - \left\{ z(d - E(D|Z_1=1, x)) + Q(x)E(D|Z_1=1, x) \right. \right. \\ \left. \left. - \frac{1-z}{1-Q(x)} Q(x)[d - E(D|Z=0, x)] - Q(x)E(D|Z=0, x) \right\} \gamma \right. \\ \left. + \text{Proj}[(z - Q(x))\kappa(x)|S_\alpha(z; x)] \right\}.$$

In the above expression we have used the definition

$$\kappa(x) = E(Y|Z=1, x) - E(Y|Z=0, x) \\ - (E(D=1|Z=1, x) - E(D=1|Z=0, x))\gamma,$$

and the efficient influence function of the parametric propensity score model

$$S_\alpha(z; x) = \frac{z - Q(x)}{Q(x)(1 - Q(x))} \frac{\partial Q}{\partial \alpha}(x, \alpha).$$

In the same equations, Proj denotes the linear projection operator

$$\text{Proj}[(z - Q(x))\kappa(x)|S_\alpha(z; x)] = S_\alpha(z; x) \text{Var}(S_\alpha(z; x))^{-1} \text{Cov}(S_\alpha(z; x), \kappa(x)).$$

In fact, Theorem 2 can be considered a special case of this influence function when

$$\text{Proj}[(z - Q(x))\kappa(x)|S_\alpha(z; x)]$$

is replaced by just $(z - Q(x))\kappa(x)$. In another special case, the efficient influence function when the propensity score $Q(x)$ is known is the same as in Theorem 3, except that the last term $\text{Proj}[(z - Q(x))\kappa(x)|S_\alpha(z; x)]$ is replaced by 0.

3. EFFICIENT ESTIMATION

In this section, we describe an estimator that achieves the semiparametric efficiency bound that makes use of the knowledge of the efficiency variance bound and the efficient score function of the model. The connection between the efficient estimators and the structure of the efficient influence function is exploited in Bickel et al. (1993) and Murphy and van der Vaart (1997). In particular, the linear quantile treatment effect estimator of Abadie, Angrist, and Imbens (2002) has a limiting variance that is strictly larger than the semiparametric variance bound.

3.1 Efficiency improvement over existing methods

We have seen that the parameters of the treated and nontreated distributions form a conditional moment equation:

$$\int g(y, d, x, \beta) f(y|d, D_1 > D_0, x) = 0.$$

The idea of the estimator is closely related to the identification argument. First of all, any given set of instrument functions, denoted $A(d, x) = \mathcal{M}(x)\zeta(x, d)$ and

$$\mathcal{A}(d, x) = (\mathcal{P}_1(x) - \mathcal{P}_0(x)) \left(\frac{Q(x)d}{\mathbf{P}(x)} + \frac{(1 - Q(x))(1 - d)}{1 - \mathbf{P}(x)} \right) A(d, x), \quad (4)$$

can be used to transform the conditional moment equations (3) into unconditional ones $E[E[\mathcal{A}(x, d)g(y, d, x, \beta)|D_1 > D_0, x, d]] = 0$, where the outer expectation is taken with respect to the marginal distribution of d and x . For a given $A(x, d)$, we conjecture the form of the efficient estimator from the identification arguments. It is then shown that efficiency bound is achieved when the optimal $A(x, d)$ is estimated consistently. We note that this approach to finding a regular estimator of the finite-dimensional parameter in the conditional moment model is a particular case of the general approach to estimating regular conditional moment models outlined in Newey (1993). As noted in Newey (1993), regular estimation can be performed by constructing a nonlinear instrument based on the conditioning set. We follow this approach to provide a regular efficient estimator in our case.

The identification condition in (2) translates into the following implications on the conditional moment functions for $\tilde{g} = A(x, d)g(y, d, x, \beta)$ and $\hat{g} = \mathcal{A}(x, d)g(y, d, x, \beta)$:

$$\begin{aligned} E(\hat{g}|d, D_1 > D_0, x) &= \frac{\mathcal{P}_1(x)d}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E(\hat{g}|d = 1, z = 1, x) \\ &\quad - \frac{\mathcal{P}_0(x)d}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E(\hat{g}|d = 1, z = 0, x) \\ &\quad + \frac{(1 - \mathcal{P}_0(x))(1 - d)}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E(\hat{g}|d = 0, z = 0, x) \\ &\quad - \frac{(1 - \mathcal{P}_1(x))(1 - d)}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E(\hat{g}|d = 0, z = 1, x). \end{aligned} \quad (5)$$

Using (4), this can be reexpressed in terms of $\tilde{g}$:

$$\begin{aligned} E(\tilde{g}|d, d_1 > d_0, x) &= \frac{\mathcal{P}_1(x)\mathcal{Q}(x)d}{\mathbf{P}(x)} E(\hat{g}|d = 1, z = 1, x) - \frac{\mathcal{P}_0(x)\mathcal{Q}(x)d}{\mathbf{P}(x)} E(\hat{g}|d = 1, z = 0, x) \\ &\quad + \frac{(1 - \mathcal{P}_0(x))(1 - \mathcal{Q}(x)(1 - d))}{1 - \mathbf{P}(x)} E(\hat{g}|d = 0, z = 0, x) \\ &\quad - \frac{(1 - \mathcal{P}_1(x))(1 - \mathcal{Q}(x))(1 - d)}{1 - \mathbf{P}(x)} E(\hat{g}|d = 0, z = 1, x). \end{aligned} \quad (6)$$

Using the Bayes rule and the law of iterated expectation, one can further write

$$\begin{aligned} E[E(\tilde{g}|d, d_1 > d_0, x)] &= E\left\{\left(dZ - \frac{\mathcal{Q}(x)}{1 - \mathcal{Q}(x)}d(1 - Z) + (1 - d)(1 - Z) - \frac{(1 - \mathcal{Q}(x))}{\mathcal{Q}(x)}(1 - d)Z\right)\hat{g}\right\} \\ &= E\{\kappa(Z, d, x)\hat{g}\}, \end{aligned}$$

where

$$\kappa(d, z, x) = 1 - \frac{d(1 - z)}{1 - \mathcal{Q}(x)} - \frac{(1 - d)z}{\mathcal{Q}(x)}$$

is the weight function defined in Abadie (2003).

The conditional probability in this moment condition, $\mathcal{Q}(x)$, is not observed. However, it can be consistently estimated in a first step, using, for example, either kernel regression or a sieve-based estimator. This estimate, $\hat{\mathcal{Q}}(x)$, can then be used to form a sample analog of the above moment conditions given any estimated instrument function $\hat{A}(x, d)$ or $\hat{\mathcal{M}}(x)$:

$$\frac{1}{N} \sum_{k=1}^N \hat{\psi}_k(\beta) = \frac{1}{N} \sum_{k=1}^N \psi_k(\beta, \hat{\mathcal{Q}}, \hat{\mathcal{M}}) = \frac{1}{N} \sum_{k=1}^N \hat{\kappa}(d_k, z_k, x_k) \tilde{g}_k,$$

where

$$\hat{\kappa}(d, z, x) = 1 - \frac{d(1-z)}{1 - \hat{Q}(x)} - \frac{(1-d)z}{\hat{Q}(x)}$$

is the estimated version of the weight function of [Abadie \(2003\)](#).

The proof of Theorem 5 shows that this estimator achieves the efficiency bound under suitable regularity conditions, and the asymptotic representation

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{k=1}^N \hat{\psi}_k(\beta) &= \frac{1}{\sqrt{N}} \sum_{k=1}^N \mathcal{M}(x_k) \left\{ \chi_k(\beta) + E \left[\frac{\partial \chi_k(\beta)}{\partial Q} \middle| x_k \right] (z_k - Q(x_k)) \right\} \\ &\quad + o_p(1) \end{aligned}$$

holds. This influence function falls into the framework of [Abadie \(2003\)](#). In the above equality, we have used the following notation to isolate the instrument matrix:

$$\chi_k(\beta) = \mathcal{M}(x_k)^{-1} \psi_k(\beta, Q, \mathcal{M}). \quad (7)$$

We now compare the efficient variance to the variance of the estimator obtained using the approach in [Abadie \(2003\)](#). To describe his estimator, we start with a distance function $\rho(\cdot)$ whose first order condition can produce a moment condition that is implied by the conditional moment model. For some consistent estimate of the probability $Q(x)$, the estimator for β will solve

$$\tilde{\beta} = \arg \min_{\beta \in \mathcal{B}} \left\{ \frac{1}{N} \sum_{k=1}^N \hat{\kappa}(d_k, z_k, x_k) \rho(y_k, d_k, x_k, \beta) \right\}.$$

This optimization problem usually leads to a moment equation in the form

$$\psi(\beta) = \kappa(d, z, x) h(d, x, \beta) g(y, d, x, \beta).$$

In the above equality, $h(d, x, \beta)$ is an instrument function that can also depend on β . In estimation, we replace the functions under consideration with their empirical analogs. In this case,

$$\frac{1}{\sqrt{N}} \sum_{k=1}^N \hat{\psi}_k(\beta) = \frac{1}{\sqrt{N}} \sum_{k=1}^N \left\{ \psi_k(\beta) + E \left[\frac{\partial \psi_k(\beta)}{\partial Q} \middle| x_k \right] (z_k - Q(x_k)) \right\} + o_p(1),$$

where ψ is similar to $\hat{\psi}$ with $\hat{Q}(x)$ replaced by $Q(x)$. Note that we can write

$$E \left[\frac{\partial \psi_k(\beta)}{\partial Q} \middle| x \right] = \tilde{\theta}(x)' \tilde{D}(x)^{-1} E \left[\frac{\partial \chi_k(\beta)}{\partial Q} \middle| x \right],$$

where χ is defined implicitly in (7) and

$$\tilde{D}(x) = \text{diag}\{\mathbf{P}(x), 1 - \mathbf{P}(x)\}, \quad \tilde{\theta}(x) = \begin{pmatrix} h(1, x, \beta_0)' \\ h(0, x, \beta_0)' \end{pmatrix}.$$

To compute asymptotic variance associated with the empirical moment equation, note that

$$\begin{aligned} & V\left(E\left[\frac{\partial\psi_k(\beta)}{\partial Q}\middle|x_k\right](z_k - Q(x_k))\right) \\ &= V\left\{\tilde{\theta}(x_k)' \tilde{D}(x_k)^{-1} E\left[\frac{\partial\chi_k(\beta)}{\partial Q}\middle|x_k\right](z_k - Q(x_k))\right\}. \end{aligned}$$

Moreover,

$$\begin{aligned} & \text{cov}\left(\psi_k(\beta), E\left[\frac{\partial\psi_k(\beta)}{\partial Q}\middle|x_k\right](z_k - Q(x_k))\right) \\ &= -V\left(E\left[\frac{\partial\psi_k(\beta)}{\partial Q}\middle|x_k\right](z_k - Q(x_k))\right). \end{aligned}$$

Finally,

$$V(\psi_k(\beta)) = E\{\tilde{\theta}(x)' \tilde{D}(x)^{-1} V(\chi_k(\beta)|x) \tilde{D}(x)^{-1} \tilde{\theta}(x)\},$$

which suggests that $V(\hat{\psi}_k(\beta)) = V(\tilde{\theta}(x) \tilde{D}(x)^{-1} \hat{\chi}_k(\beta))$. We can express the Jacobi matrix for this model as $J = E\{\tilde{\theta}(x)' \tilde{D}(x)^{-1} \theta(x)\}$, where $\theta(x) = E[\zeta(x, d) \frac{\partial\varphi(\beta, d, x)'}{\partial\beta} | x]$. This gives the expression for the asymptotic variance:

$$\begin{aligned} V(\tilde{\beta}) &= E\{\tilde{\theta}(x)' \tilde{D}(x)^{-1} \theta(x)\}^{-1} E\{\tilde{\theta}(x)' \tilde{D}(x)^{-1} \tilde{\Omega}(x) \tilde{D}(x)^{-1} \tilde{\theta}(x)\} \\ &\quad \times E\{\tilde{\theta}(x)' \tilde{D}(x)^{-1} \theta(x)\}^{-1}. \end{aligned}$$

Next we note that $V(\hat{\beta})^{-1} - V(\tilde{\beta})^{-1}$ can be written as the variance-covariance of the residual vector of the set of regression where the dependent variables are $\tilde{\Omega}(x)^{-1/2} \theta(x)$ and the regressors are $\tilde{\theta}(x)' \tilde{D}(x)^{-1} \tilde{\Omega}(x)^{1/2}$. This result implies that $V(\tilde{\beta}) - V(\hat{\beta})$ is a positive semidefinite matrix and thus the variance in [Abadie \(2003\)](#) is larger than that for the efficient estimator.

3.2 Efficient propensity score weighting estimator

The following multistep procedure summarizes a semiparametric efficient estimator under suitable regularity conditions. In step 1, we first use a kernel-based or sieve-based nonparametric estimator to obtain estimates $\hat{\mathcal{P}}_1(x)$, $\hat{\mathcal{P}}_0(x)$, and $\hat{Q}(x)$ of the conditional probabilities $\mathcal{P}_1(x)$, $\mathcal{P}_0(x)$, and $Q(x)$. In step 2, using an initial choice of an instrument matrix $\tilde{A}(x, d)$ of dimension $d_\beta \times d_g$, construct an initial estimate $\tilde{\beta}$ such that

$$\frac{1}{N} \sum_{k=1}^N \hat{\kappa}(d_k, z_k, x_k) \tilde{A}(x_k, d_k) g(y_k, d_k, x_k, \tilde{\beta}) = 0. \quad (8)$$

In step 3, $\tilde{\beta}$ is used to estimate the optimal instrument nonparametrically. For this purpose, we need to estimate

$$\hat{\omega}_{d,z}(x, \tilde{\beta}) = \hat{V}(g(y, d, x, \tilde{\beta})|d, z, x)$$

and

$$\hat{\gamma}_{d,z}(x, \bar{\beta}) = \hat{E}(g(y, d, x, \bar{\beta})|d, z, x)$$

for $d = 0, 1$ and $z = 0, 1$. Then an estimate of $\bar{\Omega}(x)$ and $\mathcal{M}(x)$ can be analytically computed as

$$\hat{\mathcal{M}}(x) = \left(\frac{\partial \varphi(1, x, \bar{\beta})}{\partial \beta'}, \frac{\partial \varphi(0, x, \bar{\beta})}{\partial \beta'} \right)' \hat{\bar{\Omega}}(x; \bar{\beta})^{-1} \text{diag} \left\{ \frac{\hat{\mathbf{P}}(x)}{\hat{\mathbf{Q}}(x)}, \frac{1 - \hat{\mathbf{P}}(x)}{1 - \hat{\mathbf{Q}}(x)} \right\}.$$

Finally, the efficient $\hat{\beta}$ is obtained through a sample moment condition similar to the one that leads to $\bar{\beta}$, except that we replace $\tilde{A}(x, d)$ by $\hat{\mathcal{M}}(x)\zeta(d, x)$. The particular form of $\varphi(d, x, \beta)$ is model specific. For example, for quantile treatment effect parameters,

$$\frac{\partial \varphi(d, x, \beta)}{\partial \beta} = f_{**}(d\beta_0 + x'\beta_1|d, x)(d, x)'.$$

This is analogous to the efficiency improvement using density weighting in [Newey and Powell \(1990\)](#) over the nonweighted quantile regression estimator of [Koenker and Bassett \(1978\)](#). Section 3.3 formally provides regularity condition for the asymptotic distribution. The structure of the efficient estimator also shows that while the inverse propensity weighting method (the $\kappa(d, z, x)$ weight function) in [Abadie \(2003\)](#) is an efficient method to construct unconditional moment conditions for compliers, efficient estimation with conditional moment restrictions also requires the optimal choice of moment conditions for compliers.

3.3 Regularity conditions and asymptotic distribution

In this section, we state a set of sufficient regularity conditions for the semiparametric efficient estimator. We will focus mainly on the reweighting estimator described in (the previous) Section 3.2. Similar conditions can be given for the conditional expectation projection estimator described in (the next) Section 3.4 and for the unconditional parameters estimators described in Section 2.3. We follow much of the recent literature and describe regularity conditions in terms of sieve nonparametric estimators for conditional probabilities and conditional expectations. Most of these conditions are well understood in the recent literature (e.g., [Ai and Chen \(2003\)](#), [Chen, Linton, and Van Keilegom \(2003\)](#), and [Newey \(1994\)](#)). Therefore, we only highlight the essential elements.

Let $\{q_l(X), l = 1, 2, \dots\}$ denote a sequence of known basis functions that can approximate any square-measurable function of X arbitrarily well. Also let

$$q^{k(n)}(X) = (q_1(X), \dots, q_{k(n)}(X))'$$

and

$$Q = (q^{k(n)}(X_1), \dots, q^{k(n)}(X_n))'$$

for some integer $k(n)$, with $k(n) \rightarrow \infty$ and $k(n)/n \rightarrow 0$ when $n \rightarrow \infty$. A first stage nonparametric estimator for $\mathcal{Q}(x)$ is then defined as

$$\hat{\mathcal{Q}}(X) = \sum_{j=1}^n Z_j q^{k(n)}(X_j) (Q'Q)^{-1} q^{k(n)}(X).$$

An estimator of the instrument function $\hat{\mathcal{M}}(x)$ depends on a preliminary parameter estimate $\bar{\beta}$ and nonparametric estimates of the quantities that define $\bar{\Omega}(x)$, which include $\hat{\omega}_{kl}(x, \bar{\beta})$, $k, l = 0, 1$, $\hat{\gamma}_{kl}(x, \bar{\beta})$, $k, l = 0, 1$, $\hat{P}_k(x)$, $k = 0, 1$, $\hat{\mathbf{P}}(x)$, and the Jacobi matrix term

$$\hat{E} \left[\frac{\partial \hat{\varphi}(d, x, \bar{\beta})}{\partial \beta} \zeta(x, d)' | x \right],$$

which can be nonparametrically estimated by

$$\hat{E} \left[\frac{\partial \hat{\varphi}(d, x, \bar{\beta})}{\partial \beta} \zeta(x, d)' | x \right] = \sum_{j=1}^n \hat{\mathcal{W}}_j(\bar{\beta}) q^{k(n)}(X_j) (Q'Q)^{-1} q^{k(n)}(x),$$

where

$$\hat{\mathcal{W}}_j(\bar{\beta}) = \frac{g(y_j, d_j, x_j, \bar{\beta} + h) - g(y_j, d_j, x_j, \bar{\beta} - h)}{2h} \zeta(x_j, d_j).$$

For the sake of clarity, we collect all the regularity conditions in Appendix D. The following two theorems are the direct consequences of these assumptions.

THEOREM 4. *Under Assumption 2, $\hat{\beta} - \beta_0 = o_p(1)$.*

THEOREM 5. *Under Assumptions 2, 3, and 4, the obtained M -estimates are consistent, are asymptotically normal, and achieve the variance lower bound. In other words,*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V(\beta))$$

for $V(\beta)$ given in Theorem 1.

The proofs of the theorems and propositions in this section are provided in the Appendix and follow immediately from the assumptions.

3.4 Conditional expectation projection estimator

The estimation method described in (the previous) Sections 3.1 and 3.2 is based on a sample average of the properly reweighted moment conditions, where the weights are related to the conditional probabilities $\mathcal{Q}(x)$, $\mathcal{P}_1(x)$, and $\mathcal{P}_0(x)$, all of which need to be estimated nonparametrically. Borrowing from the terminology from treatment effect estimation under the unconfoundedness (i.e., strong ignorability) assumption, we will call this the inverse propensity score weighting estimator. In fact, in the exogenous case when $\mathcal{P}_1(x) = 1$ and $\mathcal{P}_0(x) = 0$, this is identical to the inverse probability weighting estimator for strongly ignorable conditional treatment effect models.

There also exists an alternative estimator that relies on direct estimation of the conditional expectation $E[g(Y, D, X, \beta)|D, X = x, D_1 > D_0]$ for each candidate parameter β instead of on reweighting the moment conditions using the inverse of $\hat{Q}(x)$. To describe this estimator, begin with rewriting the identification condition (5) as

$$\begin{aligned} E(\tilde{g}|D = d, D_1 > D_0, X = x) \\ = \frac{d}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E(D\tilde{g}|Z = 1, x) - \frac{d}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E(D\tilde{g}|Z = 0, x) \\ + \frac{(1-d)}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E((1-D)\tilde{g}|Z = 0, x) \\ - \frac{(1-\mathcal{P}_1(x))}{\mathcal{P}_1(x) - \mathcal{P}_0(x)} E((1-D)\tilde{g}|Z = 1, x). \end{aligned}$$

For a given instrument matrix $\mathcal{M}(x)$, this suggests estimating β by equating to zero the sample analog

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^N \phi_k(\beta) = \frac{1}{N} \sum_{k=1}^N \left\{ \frac{d_k}{\hat{\mathcal{P}}_1(x_k) - \hat{\mathcal{P}}_0(x_k)} \hat{E}(d_k \tilde{g}|Z = 1, x_k) \right. \\ - \frac{d_k}{\hat{\mathcal{P}}_1(x_k) - \hat{\mathcal{P}}_0(x_k)} \hat{E}(d_k \tilde{g}|Z = 0, x_k) \\ + \frac{1-d_k}{\hat{\mathcal{P}}_1(x_k) - \hat{\mathcal{P}}_0(x_k)} \hat{E}((1-d_k)\tilde{g}|Z = 0, x_k) \\ \left. - \frac{1-d_k}{\hat{\mathcal{P}}_1(x_k) - \hat{\mathcal{P}}_0(x_k)} \hat{E}((1-d_k)\tilde{g}|Z = 1, x_k) \right\}, \end{aligned} \quad (9)$$

where each of the conditional expectation terms are estimated nonparametrically at every given parameter value β . For example,

$$\hat{E}(d_k \tilde{g}|Z = 1, x_k) = \hat{Q}(x_k)^{-1} \hat{E}(d_k z_k \tilde{g}|x_k). \quad (10)$$

Both conditional expectations can be estimated using a variety of nonparametric regression methods such as sieve expansion or kernel smoothing.

It is easy to show that the asymptotic linear influence function that corresponds to the moment condition $\frac{1}{N} \sum_{k=1}^N \phi_k(\beta)$ for a given $\mathcal{M}(x)$ including the optimal one coincides with the semiparametric efficient function. First of all, similar to before, estimating $\mathcal{P}_1(x) - \mathcal{P}_0(x)$ has no impact on the asymptotic variance due to the conditional nature of the moment restrictions. Using the representation theorem of Newey (1994), we can, for example, expand the first component as

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{k=1}^N \frac{d_k}{\hat{\mathcal{P}}_1(x_k) - \hat{\mathcal{P}}_0(x_k)} \frac{\hat{E}(d_k z_k \tilde{g}|x_k)}{\hat{Q}(x_k)} \\ = \frac{1}{\sqrt{N}} \sum_{k=1}^N \left\{ \frac{\mathbf{P}(x_k) d_k z_k \tilde{g}}{(\mathcal{P}_1(x_k) - \mathcal{P}_1(x_k)) \mathcal{Q}(x_k)} \right\} \end{aligned}$$

$$- \frac{\mathbf{P}(x_k)\mathcal{P}_1(x_k)E(\tilde{g}|d_k=1, z_k=1, x_k)}{(\mathcal{P}_1(x_k) - \mathcal{P}_1(x_k))\mathcal{Q}(x_k)}(z_k - \mathcal{Q}(x_k)) \\ + \frac{d_k - \mathbf{P}(x)}{(\mathcal{P}_1(x) - \mathcal{P}_0(x))}E(d_k\tilde{g}|z_k=1, x_k) \Big\} + o_p(1).$$

Similar calculations can be applied to the other three terms.

When summing these four components, we note that the last terms in each of the components cancel out due to the implications of the conditional moment restrictions that $E(d_k\tilde{g}|z_k=1, x_k) = E(d_k\tilde{g}|z_k=0, x_k)$ and $E((1-d_k)\tilde{g}|z_k=1, x_k) = E((1-d_k)\tilde{g}|z_k=0, x_k)$. Therefore, it is easy to check that the sum of the four influence functions is identical to the semiparametric efficient influence function when the instrument is chosen optimally. For the sake of brevity, we omit the regularity conditions for the conditional expectation projection estimator.

To summarize, the implementation of this estimation method is a two step procedure, each step of which involves a profiled semiparametric estimator. In the first step, for an initial arbitrary choice of the instrument matrix $\mathcal{M}(x)$ and for each trial parameter β , the moment condition in each of the terms in (10) in the moment condition (9) is estimated nonparametrically to form the moment condition (9). The near zero of this moment gives an initial estimate of β . In second step, the same procedure is repeated using a consistent estimate of the efficient instrument matrix $\mathcal{M}(x)$ which depends on the initial estimate of β following the procedure outlined in Section 3.2.

3.5 Efficient estimation of unconditional parameters

It is easy to show that an efficient estimator can be derived from the principle of conditional expectation projection that follows the identification condition. Consider first the case of the average treatment effect on compliers (ATEC) $\beta = E[Y_1 - Y_0|D_1 > D_0]$. Combining equations for the means of the distributions of treated and nontreated observations for compliers, we obtain the unconditional moment equation

$$E\{\beta(\mathcal{P}_1(x) - \mathcal{P}_0(x)) - (E[y|z=1, x] - E[y|z=0, x])\} = 0.$$

The efficient semiparametric estimator is obtained from the sample analog of this moment equation and takes the form

$$\hat{\beta} = \left(\frac{1}{N} \sum_{k=1}^N (\hat{\mathcal{P}}_1(x_k) - \hat{\mathcal{P}}_0(x_k)) \right)^{-1} \frac{1}{N} \sum_{k=1}^N (\hat{E}[y_k|z_k=1, x_k] - \hat{E}[y_k|z_k=0, x_k]).$$

Conditional expectations in this expression can be estimated nonparametrically by kernel- or sieve-based methods. Semiparametric efficiency of this estimator can be established by the same projection arguments that we used before to establish efficiency of the estimator for the conditional moment-based model.

Similarly to the ATEC, we can estimate the average treatment effect for the treated (ATTTC) as $\gamma = E[Y_1 - Y_0|d=1, D_1 > D_0]$. The ATTTC can be written in terms of the unconditional moment equation

$$E\{\gamma(\mathcal{P}_1(x) - \mathcal{P}_0(x)) - (E[Y|Z=1, x] - E[Y|Z=0, x])\mathcal{Q}(x)\} = 0.$$

By the same principle as the ATEC, we express the efficient estimator as an empirical analog

$$\hat{\gamma} = \left(\frac{1}{N} \sum_{k=1}^N \hat{Q}(x_k) (\hat{P}_1(x_k) - \hat{P}_0(x_k)) \right)^{-1} \\ \times \frac{1}{N} \sum_{k=1}^N \hat{Q}(x_k) (\hat{E}[y_k | z_k = 1, x_k] - \hat{E}[y_k | z_k = 0, x_k]).$$

Using the projection argument of [Newey \(1994\)](#), we can easily verify that this estimator achieves the semiparametric efficiency bound when each of the conditional expectations and conditional probabilities above are estimated nonparametrically using either kernel- or sieve-based methods.

If the $Q(x)$ is specified as a parametric function or is a known function $Q_\alpha(x)$, then the efficient estimator for γ becomes

$$\hat{\gamma} = \left(\frac{1}{N} \sum_{k=1}^N \hat{Q}_{\hat{\alpha}}(x_k) (\hat{P}_1(x_k) - \hat{P}_0(x_k)) \right)^{-1} \\ \times \frac{1}{N} \sum_{k=1}^N Q_{\hat{\alpha}}(x_k) (\hat{E}[y_k | z_k = 1, x_k] - \hat{E}[y_k | z_k = 0, x_k]),$$

where $\hat{\alpha}$ is the parametric maximum likelihood estimator (MLE), or the known α_0 if $Q(x)$ is fully known.

When the *propensity score* $Q(x)$ is entirely unknown, an alternative efficient estimator can be developed using the inverse propensity score weighting approach of [Abadie \(2003\)](#). When $Q(x)$ is known or parametrically specified, however, [Hahn \(1998\)](#) and [Hirano, Imbens, and Ridder \(2003\)](#) showed that efficient estimators based on inverse propensity score weighting typically require combining a nonparametric estimate of $\hat{Q}(x)$ with the known or parametrically estimated $Q(x)$. A detailed comparison between the conditional expectation projection approach and the inverse propensity weighting approach is provided in [Chen, Hong, and Tarozzi \(2008\)](#), but they maintained the unconfoundedness assumption and did not investigate endogeneity.

To summarize, a recipe for empirically implementing the efficient estimators for β and γ only requires summing over the data a function of the nonparametrically estimated choice probabilities $\hat{Q}(x)$, $\hat{P}_1(x)$, and $\hat{P}_0(x)$, and nonparametric estimates of the conditional expectations $\hat{E}[y_k | z_k = 1, x_k]$ and $\hat{E}[y_k | z_k = 0, x_k]$. This is a very straightforward procedure that does not involve any nonlinear or numerical optimization procedures.

3.6 General separable unconditional model for compliers

The treatment effect models considered in this section have a straightforward generalization to the separable conditional moment restrictions expressed in terms of unobservable outcome variables. Consider a problem where a finite-dimensional parameter

$\beta \in \mathbb{R}^k$ is given by the following unconditional moment equation described in terms of unobservable variables Y_1 and Y_0 :

$$\varphi(\beta) = E[g_1(Y_1, x, \beta) - g_0(Y_0, x, \beta) | D_1 > D_0] = 0. \quad (11)$$

In particular, when $g_1(Y_1, \beta) = Y_1 - \beta$ and $g_0(Y_0, \beta) = Y_0 + \beta$, parameter β defines the average treatment effect for compliers. On the other hand, $g_1(Y_1, \beta) = 1(Y_1 \leq \beta_1) - \tau$ and $g_0(Y_0, \beta) = 1(Y_0 \leq \beta_0) + \tau$ define a complier analog of the average quantile treatment effect parameter proposed in [Firpo \(2003\)](#).

Note that we can represent this moment equation for compliers in terms of distributions for the entire population. Using the Bayes's rule, we find that this equation is equivalent to

$$E[(\mathcal{P}_1(x) - \mathcal{P}_0(x))(\mathcal{Q}(x)E[g_1(Y_1, x, \beta) | d = 1, D_1 > D_0, x] - (1 - \mathcal{Q}(x))E[g_0(Y_0, x, \beta) | d = 0, D_1 > D_0, x])] = 0,$$

which can be redefined in terms of only observable variables in the form

$$E\left[(\mathcal{P}_1(x) - \mathcal{P}_0(x))\left(\frac{\mathcal{Q}(x)d}{\mathbf{P}(x)} + \frac{(1 - \mathcal{Q}(x))(1 - d)}{1 - \mathbf{P}(x)}\right) \times E[dg_1(y, x, \beta) - (1 - d)g_0(y, x, \beta) | d, x, D_1 > D_0]\right] = 0.$$

This equation in general defines an overidentified system of moments for β . Using a constant matrix A (which we can then choose optimally), we can transform this vector of moments into an exactly identified system. The Jacobi matrix J for this system given A is computed in the standard way. The following theorem describes the structure of the efficient influence function for this model.

THEOREM 6. *In the model given by the general moment condition (11), the efficient influence function, which corresponds to finite-dimensional parameter β , can be expressed as*

$$\begin{aligned} \Phi(y, d, x, z) &= -J^{-1} A \frac{z - \mathcal{Q}(x)}{1 - \mathcal{Q}(x)} \left\{ dg_1(y, x, \beta) + (1 - d)g_0(y, x, \beta) \right. \\ &\quad - \frac{1}{\mathcal{Q}(x)} [(1 - \mathcal{Q}(x))E[(1 - d)g_0(y, x, \beta) | z = 1, x] \\ &\quad \left. + \mathcal{Q}(x)E[dg_1(y, x, \beta) | z = 0, x]] \right\} \\ &= -J^{-1} A \phi(y, d, x, z). \end{aligned}$$

The structure of the efficient influence function in this case is similar to that in the ATE model which we considered earlier in this section. We can further choose the ma-

trix A such that it minimizes the variance of the efficient influence function. In particular, given that the Jacobi matrix can be expressed as

$$J = A \frac{\partial \varphi(\beta)}{\partial \beta'},$$

the semiparametric efficiency bound for this model when A is chosen optimally takes the form $V(\beta) = (\frac{\partial \varphi(\beta)}{\partial \beta'} E[\phi(y, d, x, z) \phi(y, d, x, z)'] \frac{\partial \varphi(\beta)}{\partial \beta'})^{-1}$. An optimally weighted generalized method of moments (GMM) estimator based on the nonparametrically estimated moment condition

$$\begin{aligned} & \left(\frac{1}{N} \sum_{k=1}^N (\hat{P}_1(x_k) - \hat{P}_0(x_k)) \right)^{-1} \\ & \times \frac{1}{N} \sum_{k=1}^N (\hat{E}[g_1(y_k, x_k, \beta) | z_k = 1, x_k] - \hat{E}[g_0(y_k, x_k, \beta) | z_k = 0, x_k]) = 0 \end{aligned}$$

can easily be shown to achieve the efficiency bound derived in Theorem 6.

Similarly, it is immediate to develop semiparametric efficiency bounds for a nonlinear treatment effect parameter for *treated compliers*, defined as

$$\varphi(\gamma) = E[g_1(Y_1, x, \gamma) - g_0(Y_0, x, \gamma) | d = 1, D_1 > D_0] = 0.$$

In addition, an optimally weighted GMM estimator based on the nonparametric estimated moment condition

$$\begin{aligned} & \left(\frac{1}{N} \sum_{k=1}^N \hat{Q}_{\hat{\alpha}}(x_k) (\hat{P}_1(x_k) - \hat{P}_0(x_k)) \right)^{-1} \\ & \times \frac{1}{N} \sum_{k=1}^N \hat{Q}_{\hat{\alpha}}(x_k) \\ & \times (\hat{E}[g_1(y_k, x_k, \gamma) | z_k = 1, x_k] - \hat{E}[g_0(y_k, x_k, \gamma) | z_k = 0, x_k]) = 0, \end{aligned}$$

where $\hat{Q}_{\hat{\alpha}}(x_k)$ can be nonparametrically estimated, parametrically estimated, or the known propensity, can easily be shown to achieve the required corresponding semiparametric efficiency bound when the propensity score is unknown, parametrically specified, or known.

4. NUMERICAL SIMULATIONS

In this section we report the results from a Monte Carlo study to illustrate the finite sample properties of the proposed estimators and the numerical efficiency comparisons with existing estimators. The design of the Monte Carlo study is motivated by the empirical illustration in Appendix E.

4.1 The structure of the data-generating process

To analyze the performance of our semiparametric estimator, we designed an experiment where the outcome variable Y depends on the endogenous regressor X in a non-linear fashion. We construct the data in which a binary instrumental variable Z depends on X , but is conditional on X potential outcomes and the treatments are independent from Z . The observable outcome is generated from these variables. The data-generating mechanism for the simulation is characterized by the vector of potential outcomes (Y_1, Y_0) and the vector of potential treatments (D_1, D_0) . Their distributions depend on the vector of covariates X . In light of the empirical illustration, the observable outcome for compliers has a Poisson distribution whose mean depends on the covariates and the parameters. We first consider the benchmark model. In Monte Carlo simulation experiments, we consider alternative parameter values and measure parameter differences in relation to the benchmark values. The sequential sampling scheme in the benchmark model is given by the following steps.

First, we generate potential treatments as $D_1 = 1(\gamma_0 + x'\gamma_1 + \delta + v \geq 0)$ and $D_0 = 1(\gamma_0 + x'\gamma_1 + v \geq 0)$, where $\gamma_0 = -0.5$, $\gamma_1 = 1$, $\delta = 1$, x is generated from the uniform distribution, and v is standard normal. Second, we generate potential outcomes based on Poisson distributions. We first generate four independent Poisson random variables $\xi_1 \sim \text{Poisson}(\exp(\alpha + x'\beta))$, $\xi_2 \sim \text{Poisson}(\exp(x'\beta))$, $\xi_3 \sim \text{Poisson}(\lambda_{11})$, and $\xi_4 \sim \text{Poisson}(\lambda_{00})$, where $\alpha = 1$, $\beta = 0.5$, $\lambda_{11} = 2$, and $\lambda_{00} = 1$. Denote $e = (1, 1)'$. Then we construct the potential outcomes as

$$\begin{pmatrix} Y_1 \\ Y_0 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \xi_3 e \mathbf{1}\{D_1 = 1, D_0 = 1\} + \xi_4 e \mathbf{1}\{D_1 = 0, D_0 = 0\}.$$

Note that this structure assures that for compliers ($D_1 = 1$ and $D_0 = 0$), two potential outcomes are independent and $Y_i \sim \text{Poisson}(\exp(\alpha i + x'\beta))$ for $i = 0, 1$. For always-takers ($D_1 = D_0 = 1$), the potential outcomes have covariance λ_{11} , and for never-takers ($D_1 = D_0 = 0$), the potential outcomes have covariance λ_{00} . Finally, the instrument is generated as an independent Bernoulli random variable $Z \sim \text{Bernoulli}(\Phi(x))$.

Given the latent variables generated above, we compute the observable treatments and outcomes as $D = D_1 Z + D_0(1 - Z)$ and $Y = Y_1 D + Y_0(1 - D)$. This structure of the data-generating process guarantees that for compliers, the outcome will be independent from the observable treatment D and the mean of the treatment outcome can be computed from the pair of Poisson random variables (ξ_1, ξ_2) . As a result, the data-generating process for the Monte Carlo experiment is characterized by conditional moments

$$E[Y - \exp(\alpha D + X'\beta) | D_1 > D_0, D = d, X = x] = 0.$$

This model fits into our general conditional moment framework in the LATE context. To analyze the performance of our estimation procedure for parameters α and β , we designed a series of experiments.

4.2 Experiment 1: Basic comparison with alternative procedures

The analysis of our estimation procedure begins with a comparison between our efficient two-stage estimator and the alternative existing estimator. We estimate this model using both the original method of Abadie (2003) and our more efficient estimator. Three parameters are estimated: $\beta_0 = 0$ is the coefficient on the constant term of the covariate, $\beta_1 = 1$ is the slope coefficient on the uniform regressor, and $\alpha = 0.5$ is the coefficient on the treatment dummy. The simulation results across 1000 simulations are summarized in Table 1.

We separately investigate the impact of the error in Monte Carlo sampling on our results by considering different Monte Carlo sample sizes. We analyze the differences in the mean-squared error of the estimated treatment effect α for different numbers of Monte Carlo samples. We study the impact of the Monte Carlo sampling error by repeating the simulation results for 1000 Monte Carlo replications (which we use throughout our analysis) 100 times for the baseline parameters of the model. By decomposing the mean-squared error into a between Monte Carlo component and a within Monte Carlo

TABLE 1. Simulation Summary.

Parameter	Mean Bias	Median Bias	Std Deviation	Mean Squared Errors
Sample Size 1000				
Abadie et al. estimator				
β_0	0.0956	0.1120	0.2176	0.0565
β_1	0.0559	0.0527	0.1880	0.0385
α	-0.3825	-0.3186	0.3254	0.2504
Semiparametric efficient estimator				
β_0	0.0053	-0.1045	0.1587	0.0254
β_1	0.0493	0.0540	0.1605	0.0282
α	-0.0365	-0.0286	0.1672	0.0288
Sample Size 2000				
Abadie et al. estimator				
β_0	0.1126	0.1209	0.1435	0.0332
β_1	0.0452	0.0384	0.1271	0.0182
α	-0.3727	-0.3510	0.2205	0.1875
Semiparametric efficient estimator				
β_0	0.0113	0.079	0.1132	0.0129
β_1	0.0434	0.0384	0.1141	0.0149
α	-0.0186	-0.0254	0.1188	0.0144
Sample Size 4000				
Abadie et al. estimator				
β_0	0.0974	0.1050	0.0998	0.0194
β_1	0.0534	0.0502	0.0893	0.0108
α	-0.3441	-0.3325	0.1353	0.1367
Semiparametric efficient estimator				
β_0	-0.0009	0.0035	0.0787	0.0062
β_1	0.0464	0.0453	0.0787	0.0083
α	-0.0029	-0.0069	0.0824	0.0068

component, we find that the mean-squared error for the estimated treatment effect associated with sampling error constitutes only 16.3% of the overall mean-squared error across a total of 100,000 Monte Carlo simulations. This leads us to the conclusion that even though the impact of the sampling error is visible, it does not substantially interfere with the results of the Monte Carlo experiment.

4.3 Experiment 2: Proportional reduction of noncompliers

In this experiment, we study the robustness of the estimator against the endogeneity of the dependent variable. We proportionally decrease the variances of ξ_3 and ξ_4 , which are responsible for the dependence between the binary regressors and the outcome. In the limiting case where these variances λ_{00} and λ_{11} are equal to zero, the latent outcomes are completely independent. In Table 2, we document its effect on the mean-squared error of the estimate for the coefficient of the endogenous dummy variable. The tabulated data are obtained from 1000 Monte Carlo replications. The table shows that a reduction in the correlation between the treatment outcomes (Y_1 and Y_0) leads to a smaller variance of the estimated treatment effect. Moreover, an increase in the sample size results in a decrease of the mean-squared error. The numbers in the table are computed from the Monte Carlo sample where we trimmed away the top and bottom 1% of observations to avoid including the cases where the distance minimization algorithm did not converge.

TABLE 2. Simulation Summary for Decreasing Correlation Between Treatment Outcomes.

$\lambda_{11} = 2K, \lambda_{00} = K$	Sample Sizes				
$K =$	250	350	450	600	1000
1	0.2716	0.2307	0.1865	0.1782	0.1282
0.9474	0.2774	0.2242	0.1873	0.1640	0.1306
0.8947	0.2791	0.2155	0.1792	0.1642	0.1249
0.8421	0.2578	0.2065	0.1900	0.1667	0.1207
0.7895	0.2509	0.2174	0.1829	0.1519	0.1138
0.7368	0.2453	0.2079	0.1808	0.1577	0.1209
0.6842	0.2528	0.2201	0.1883	0.1544	0.1171
0.6316	0.2504	0.1941	0.1732	0.1621	0.1139
0.5789	0.2463	0.1938	0.1663	0.1566	0.1137
0.5263	0.2506	0.1938	0.1684	0.1457	0.1124
0.4737	0.2305	0.2054	0.1746	0.1554	0.1140
0.4211	0.2169	0.1914	0.1600	0.1463	0.1173
0.3684	0.2187	0.1868	0.1589	0.1495	0.1151
0.3158	0.2135	0.1847	0.1625	0.1431	0.1096
0.2632	0.2116	0.1762	0.1514	0.1408	0.1133
0.2105	0.2181	0.1804	0.1596	0.1383	0.1125
0.1579	0.2149	0.1839	0.1551	0.1372	0.1067
0.1053	0.2074	0.1729	0.1642	0.1385	0.1035
0.0526	0.2137	0.1730	0.1660	0.1305	0.1030
0.0000	0.2050	0.1714	0.1528	0.1312	0.1053

4.4 Experiment 3: Smaller treatment effect

In this experiment we study the sensitivity of our estimator with respect to the value of the treatment effect. We vary the coefficient α of the dummy endogenous variable, keeping the remaining components of the model the same. This exercise illustrates the robustness of our estimation method with respect to the magnitude of the treatment effect parameter of interest relative to the remaining components of the model. The results in Table 3 show that, in principle, our procedure gives stable mean-squared errors across different choices of the treatment effect parameters and the sample sizes. Similar to the previous experiment, the numbers in the table are computed from the trimmed Monte Carlo sample (removing top and bottom 1% quantiles) to avoid including the cases where the distance minimization algorithm did not converge. One can see, however, from Table 3 that reduction of the actual treatment effect leads to an increase in the distribution range for the estimated treatment effect.

4.5 Experiment 4: Choice of the bandwidth

In this experiment, we study the sensitivity of the estimator with respect to the choice of the bandwidth parameter. We use the same structure of the bandwidth for the estimation of all nonparametric components of the model, including the conditional probability of treatment selection $Z = 0$ and the conditional probability of treatment $D = 0$. We choose the bandwidth as $h_n = 4\sigma \exp(K)n^{-1/3}$, where σ is the unconditional variance of

TABLE 3. Simulation Summary for Reduction of the Treatment Effect.

$\lambda_{11} = 0.5K$	Sample Sizes				
$K =$	250	350	450	600	1000
1	0.2716	0.2276	0.1903	0.1617	0.1347
0.9474	0.2728	0.2287	0.1943	0.1690	0.1319
0.8947	0.2871	0.2383	0.1999	0.1681	0.1266
0.8421	0.2899	0.2332	0.2064	0.1661	0.1288
0.7895	0.2793	0.2358	0.1888	0.1679	0.1216
0.7368	0.2845	0.2315	0.2060	0.1688	0.1223
0.6842	0.2924	0.2371	0.2067	0.1745	0.1251
0.6316	0.2922	0.2172	0.2040	0.1743	0.1271
0.5789	0.2929	0.2314	0.1900	0.1756	0.1322
0.5263	0.2921	0.2321	0.1959	0.1621	0.1299
0.4737	0.2931	0.2291	0.2027	0.1684	0.1313
0.4211	0.2820	0.2118	0.1988	0.1573	0.1234
0.3684	0.2894	0.2224	0.1904	0.1695	0.1326
0.3158	0.2481	0.2101	0.1881	0.1538	0.1259
0.2632	0.2558	0.2200	0.1782	0.1637	0.1255
0.2105	0.2700	0.1825	0.1550	0.1250	0.1117
0.1579	0.2480	0.1886	0.1513	0.1127	0.0970
0.1053	0.2577	0.1613	0.1305	0.1155	0.0779
0.0526	0.2596	0.1687	0.1309	0.1109	0.0796
0.0000	0.2013	0.1532	0.1339	0.1110	0.0868

TABLE 4. Simulation Summary for Various Bandwidth Choices.

$h = 4\sigma \exp(K)n^{-1/3}$	Sample Sizes				
$K =$	250	350	450	600	1000
1	0.2860	0.2488	0.2210	0.1854	0.1565
0.9474	0.3022	0.2436	0.2210	0.1900	0.1528
0.8947	0.3002	0.2392	0.2152	0.2000	0.1570
0.8421	0.2669	0.2392	0.2202	0.2016	0.1535
0.7895	0.3215	0.2492	0.2266	0.1961	0.1540
0.7368	0.2868	0.2609	0.2183	0.2045	0.1492
0.6842	0.3027	0.2403	0.2236	0.1922	0.1581
0.6316	0.2926	0.2482	0.2275	0.1857	0.1601
0.5789	0.2825	0.2457	0.2194	0.1950	0.1582
0.5263	0.2749	0.2475	0.2154	0.1972	0.1502
0.4737	0.2863	0.2437	0.2172	0.1972	0.1555
0.4211	0.2895	0.2524	0.2231	0.1990	0.1553
0.3684	0.3074	0.2566	0.2043	0.1795	0.1650
0.3158	0.2802	0.2478	0.2221	0.1857	0.1540
0.2632	0.2809	0.2362	0.2189	0.1969	0.1571
0.2105	0.2841	0.2419	0.2135	0.2008	0.1533
0.1579	0.2857	0.2430	0.2179	0.1876	0.1542
0.1053	0.2937	0.2331	0.2209	0.1824	0.1520
0.0526	0.2759	0.2315	0.2069	0.1996	0.1528
0.0000	0.3066	0.2393	0.2245	0.1875	0.1495

the binary variable (Z or D), n is the sample size, and K is the constant of choice which we vary from 0 to 1. The results in Table 4 demonstrate the mean squared errors across the Monte Carlo simulations. As one can see from the table, the mean-squared error remains stable across all different bandwidth choices. It is especially visible for sample size 1000. This confirms our theoretical results that if the regularity conditions are satisfied, the choice of the estimation procedure for non-parametric components of the model should not have a large impact on the estimated treatment effect parameter.

5. CONCLUSION

In this paper, we derive the semiparametric efficiency bound for the estimation of a finite-dimensional parameter defined by generalized moment conditions under the local instrumental variable assumptions of Imbens and Angrist (1994) and Abadie, Angrist, and Imbens (2002). These parameters identify the treatment effect on the set of compliers under the monotonicity assumption. The moment equation characterizes the parametrized moment of the outcome distribution given a set of covariates and the treatment dummy. The distributions of covariates, the treatment dummy, and the binary instrument are not specified in a parametric form, making the model semiparametric. We also develop multistep semiparametric efficient estimators that achieve the semiparametric efficiency bound. The results of the Monte Carlo simulations demonstrate good performance of the semiparametric efficient estimator for finite samples.

REFERENCES

- Abadie, A. (2003), "Semiparametric instrumental variable estimation of treatment response models." *Journal of Econometrics*, 113, 231–263. [280, 281, 283, 288, 289, 290, 291, 295, 299]
- Abadie, A., J. Angrist, and G. Imbens (2002), "Instrumental variables estimates of the effects of subsidized training on the quantiles of trainee earnings." *Econometrica*, 70, 91–117. [280, 283, 285, 287, 302]
- Ai, C., and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71 (6), 1795–1843. [291]
- Angrist, J., G. Imbens, and D. Rubin (1996), "Identification of causal effects using instrumental variables." *Journal of the American Statistical Association*, 91 (434). [280]
- Bickel, P. J., C. A. Klaassen, Y. Ritov, and J. A. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, New York. [280, 287]
- Chen, X., H. Hong, and A. Tarozi (2008), "Semiparametric efficiency in GMM models with auxiliary data." *The Annals of Statistics*, 36 (2), 808–843. [280, 295]
- Chen, X., O. Linton, and I. Van Keilegom (2003), "Estimation of semiparametric models when the criterion function is not smooth." *Econometrica*, 71 (5), 1591–1608. [291]
- Chernozhukov, V., and C. Hansen (2005), "An IV model of quantile treatment effects." *Econometrica*, 73 (1), 245–261. [279]
- Firpo, S. (2003), "Efficient semiparametric estimation of quantile treatment effects." Working paper, UC Berkeley, Department of Economics. [279, 296]
- Frolich, M. (2007), "Nonparametric IV estimation of local average treatment effects with covariates." *Journal of Econometrics*, 139 (1), 35–75. [280, 285]
- Hahn, J. (1998), "On the role of propensity score in efficient semiparametric estimation of average treatment effects." *Econometrica*, 66 (2), 315–331. [279, 280, 286, 295]
- Henderson, D., D. Millimet, C. Parmeter, and L. Wang (2006), "Fertility and the health of children: A nonparametric investigation." Working paper, Virginia Polytechnic Institute and State University. [280]
- Hirano, K., G. Imbens, and G. Ridder (2003), "Efficient estimation of average treatment effects using the estimated propensity score." *Econometrica*, 71 (4), 1161–1189. [279, 295]
- Hong, H., and D. Nekipelov (2010), "Supplement to 'Semiparametric Efficiency in Nonlinear LATE Models'." *Quantitative Economics Supplemental Material*, <http://www.qeconomics.org/supp/43/supplement.pdf>. [281]
- Imbens, G., and J. Angrist (1994), "Identification and estimation of local average treatment effects." *Econometrica*, 62 (2), 467–475. [280, 281, 283, 302]
- Khan, S., and D. Nekipelov (2010), "Semiparametric efficiency in irregularly identified models." Mimeo, Duke University and UC Berkeley. [283]

- Khan, S., and E. Tamer (2009), “Irregular identification, support conditions and inverse weight estimation.” Unpublished manuscript, Northwestern University. [283]
- Koenker, R., and G. S. Bassett (1978), “Regression quantiles.” *Econometrica*, 46, 33–50. [291]
- Murphy, S., and A. van der Vaart (1997), “Semiparametric likelihood ratio inference.” *The Annals of Statistics*, 25 (4), 1471–1509. [287]
- Newey, W. (1990a), “Efficient instrumental variables estimation of nonlinear models.” *Econometrica*, 58 (4), 809–837. [279, 283]
- (1990b), “Semiparametric efficiency bounds.” *Journal of Applied Econometrics*, 5 (2), 99–135. [280]
- Newey, W. (1993), “Efficient estimation of models with conditional moment restrictions.” In *Handbook of Statistics*, Volume 11, 419–454, Elsevier. [287]
- Newey, W. (1994), “The asymptotic variance of semiparametric estimators.” *Econometrica*, 62, 1349–1382. [291, 293, 295]
- Newey, W., and J. Powell (1990), “Efficient estimation of linear and type I censored regression models under conditional quantile restrictions.” *Econometric Theory*, 6, 295–317. [291]
- Robins, J. M., and A. Rotnitzky (1995), “Semiparametric efficiency in multivariate regression models with missing data.” *Journal of the American Statistical Association*, 90 (429), 122–129. [280]
- Severini, T., and G. Tripathi (2001), “A simplified approach to computing efficiency bounds in semiparametric models.” *Journal of Econometrics*, 102, 23–66. [283]