Flexible Estimation of Treatment Effect Parameters

Thomas MaCurdy[a] and Xiaohong Chen[b] and Han Hong[c] [1]

1 Introduction

Many empirical studies of program evaluations are complicated by the presence of both observed and unobserved heterogeneity in measuring the effects of treatments on outcomes. Challenges in estimation arise when a researcher never observes individual outcomes with and without treatment. This missing information requires the introduction of identifying assumptions to infer treatment impacts, assumptions that impose semiparametric constraints on the underlying data generating process. This paper discusses how a variety of treatment effect parameters, both conditional and unconditional measures, can be flexibly estimated using conventional regression techniques and conditional moment restrictions. This includes the parameters estimated by the popular approaches relying on propensity score matching and other weighting methods. Moreover, except in situations where propensity scores or weighting factors are known a priori, the regression and conditional moment methods posed here attain statistically efficiency bounds.

The following discussion initially outlines the basic statistical problem underlying the estimation of treatment effects. Section 3 next discusses the estimation of models that rely on the common identification strategy based on conditional independence. The discussion then addresses the important distinctions and relevant issues in estimating conditional verse unconditional treatment effect parameters, with Section 4 covering the linear case and Section 5 describing nonlinear extensions. Section 6 considers the case of endogenous treatments, and, finally, Section 7 presents concluding remarks.

2 Background

In a typical treatment effect model, the researchers have access to a data set with information on the observed covariates X, the treatment status variable D and the treatment outcome variable Y. The treated outcome (when $D = 1$) is denoted Y_1, while the untreated outcome (when $D = 0$) is denoted Y_0. For each individual in the sample, only Y_1 or Y_0, but not both, is observed. They are related to the observed outcome variable by the relation:

$$Y = DY_1 + (1 - D)Y_0.$$

In the presence of observed heterogeneity defined by covariates X that partition the data into *cells*, many identifying assumptions, e.g. conditional independence, monotonicity, difference in difference, regression discontinuity, are often made conditional on each *cell*.

With a sufficient amount of data, treatment effect analysis can be estimated cell-by-cell. However, it is often the case that available samples do not provide enough observations to conduct precise statistical inference for each of the individual cells. The literature implements two alternative empirical strategies in this situation.

[1a,c] Stanford University and [b] Yale University.

The first approach is to change the parameter of interest to an overall unconditional treatment effect parameter which is defined as the average over either the entire population or subpopulation consisting of treated and untreated observations. These parameters can usually be estimated at the usual asymptotic rate with a normal limiting distribution. However, the cost of doing this is that the parameter of interest becomes different.

A second approach defines weighted unconditional treatment effect parameters over the population or subpopulation with weights formulated to measure a parameter of economic interest. In such applications, it is not uncommon for researchers to select weights for the sake of statistical significance instead of economic importance.

Regardless of which of these estimation approach one considers, the underlying nonparametric and semiparametric procedures can be essentially interpreted as partitioning the sample into cells defined by the observed variables, with treatment effects estimated by comparing average outcomes for the treated and untreated observations within a cell. Different averaging across cells yields estimates of treatment effects. Smoothing across cells is often done to compensate for insufficient data; and, to satisfy the conditions nonparametric theory, an analyst must shrink the coverage of each cell as the sample size increases. The discussion below ignores these factors. Understanding the concepts in this discussion, however, does continually return to the theme that estimating treatment effects amounts to computing averages of estimates across cells which, in turn, merely represents a flexible regression method.

3 Conditional Independence Models

A common identification strategy for controlling observed treatment effect heterogeneity is to impose the assumption that the missing outcome information for the counterfactual treatment status for a given individual can be recovered using individuals with similar observed characteristics. This is usually stated formally as conditional independence, or unconfoundedness, or strong ignorability:

Assumption 1 $Y_1, Y_0 \perp D \mid X$.

The key implication of this identifying assumption is that for a group of individuals with similar observed characteristics, some of which are treated and some of which are not treated, the effects of treatment can be calculated in an average sense.

Two parameters of interest in this model are the conditional average treatment effect (CATE), defined as $E(Y_1 - Y_0|X)$ and the conditional average treatment effect on the treated (CATT), defined as $E(Y_1 - Y_0|D = 1, X)$ Under the conditional independence assumption, both CATE and CATT are given by

$$\beta(X) \equiv E(Y_1 - Y_0|X) = E(Y_1 - Y_0|D, X) = E(Y|D = 1, X) - E(Y|D = 0, X).$$

To estimate such parameters, consider a flexible nonparametric regression of Y on D and X: $E(Y|D, X)$. Since D is a dummy variable taking values in 0 and 1, this regression function can be written as

$$E(Y|D, X) = \alpha(X) + \beta(X) D,$$

where $\beta(X) = E(Y|D=1, X) - E(Y|D=0, X)$.

Hence we see that with a flexible regression of Y on D and X, the coefficients appearing in the function of interacting X with the dummy variable D gives the desired conditional treatment effect parameter.

The propensity score is the probability of treatment given the observed covariates X:

$$p(X) = p(D=1|X).$$

A celebrated result due to Rosenbaum and Rubin (1983) shows that

$$D \perp X|p(X), \tag{1}$$

and, for any *balancing score* $b(X)$ such that $D \perp X|b(X)$, it is necessarily the case that $p(X)$ is a function of $b(X)$. Assumption 1 and result (1) together imply that

$$E(Y|D=1, p(X)=p) - E(Y|D=0, p(X)=p) = E(Y_1 - Y_0|p(X)=p).$$

Relying on this finding, Rosenbaum and Rubin (1983) suggest matching propensity scores, or on any balancing score for that matter, enables one to estimate averaged treatment effect among the set of individuals with the same propensity score of p.

While this statement is true, it is often not the case that a researcher happens to be interested in the treatment effect of a group of individuals who share the same value of $p(X)$. In most economic applications, one is not interested in the quantity

$$E(Y_1 - Y_0|p(X)=p),$$

because X is multi-dimensional and the group of individuals with identical $p(X)$ might include very heterogenous values of X and heterogenous treatment effects $E(Y_1 - Y_0|X)$. Moreover, a researcher can *choose* $p(X)$. Through different choices of $p(X)$, which is used to "randomly" assign individuals to the treatment and control groups conditional on their observed characteristics, the average effects of treatment can be arbitrarily defined.

4 Relating Unconditional and Conditional Treatment Effects

Often researchers are interested in estimating *unconditional parameters* which predominately includes the average treatment effect (ATE), the average treatment effect of the treated (ATT), or weighted versions of both. These quantities are defined as $ATE = E(Y_1 - Y_0)$ and $ATT = E(Y_1 - Y_0|D=1)$. Identification follows from integrating $\beta(X)$ over either the marginal distribution of X or the conditional distribution of X given $D=1$. In particular,

$$ATE = E(\beta(X)) \quad \text{and} \quad ATT = E(\beta(X)|D=1).$$

Given a flexible nonparametric estimate $\hat{\beta}(X)$ of $\beta(X)$, these parameters can be estimated by integration over the relevant samples:

$$\widehat{ATE} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}(X_i) \quad \text{and} \quad \widehat{ATT} = \sum_{i=1}^{n}\hat{\beta}(X_i)D_i \Big/ \sum_{i=1}^{n}D_i.$$

Alternative estimators can be developed based on the use of propensity score matching or weighting methods. The discussion below argues that unless the propensity score is known, implementing these methods using the estimated propensity score does not bring either identification or efficiency benefits over the nonparametric estimate provided above.

As demonstrated in the literature, the unconditional treatment effects parameters can be rewritten as,

$$
\begin{aligned}
ATE &= & EE\left(\beta\left(X\right)|p\left(X\right)\right) = E\left[E\left(Y|D=1,p(X)\right) - E\left(Y|D=0,p(X)\right)\right] & \quad (2) \\
ATT &= & E\left[E\left(\beta\left(X\right)|p\left(X\right)\right)|D=1\right] = E\left[\left[E\left(Y|D=1,p(X)\right) - E\left(Y|D=0,p(X)\right)\right]|D=1\right]. & \quad (3)
\end{aligned}
$$

These fomulations suggest that the ATE and ATT parameters can be estimated by averaging the propensity score matched estimates over the desired population. In particular, for an initial consistent estimate $\hat{p}\left(X\right)$ of $p\left(X\right)$, one can form

$$
\widetilde{ATE} = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{E}\left(Y|\hat{p}\left(X_i\right),D=1\right) - \hat{E}\left(Y|\hat{p}\left(X_i\right),D=0\right)\right] \quad (4)
$$

$$
\widetilde{ATT} = \sum_{i=1}^{n}\left[\hat{E}\left(Y|\hat{p}\left(X_i\right),D=1\right) - \hat{E}\left(Y|\hat{p}\left(X_i\right),D=0\right)\right]D_i \Big/ \sum_{i=1}^{n}D_i. \quad (5)
$$

Furthermore, the unconditional parameters can also be rewritten as:

$$
\begin{aligned}
ATE &= & E\left(\beta\left(X\right)\right) = \int\left[E\left(Y|D=1,X\right) - E\left(Y|D=0,X\right)\right]f\left(X\right)dX \\
&= & E\left(Y\frac{p}{p(X)}|D=1\right) - E\left(Y\frac{1-p}{1-p(X)}|D=0\right),
\end{aligned}
$$

and

$$
ATT = \quad E\left(Y|D=1\right) - E\left(Y\frac{p}{1-p}1-p\left(X\right)p\left(X\right)|D=0\right).
$$

These relations form the basis of the inverse propensity score weighting estimators:

$$
\overline{ATE} = \sum_{i=1}^{n}Y_i\frac{\hat{p}}{\hat{p}\left(X_i\right)}D_i\Big/\sum_{i=1}^{n}D_i - \sum_{i=1}^{n}Y_i\frac{1-\hat{p}}{1-\hat{p}\left(X_i\right)}\left(1-D_i\right)\Big/\sum_{i=1}^{n}\left(1-D_i\right). \quad (6)
$$

Essentially all three estimation methods (2), (4) and (6) identify the same population parameters without any additional assumptions. Various papers (e.g. Hahn (1998), Hirano, Imbens, and Ridder (2003) and Chen, Hong, and Tarozzi (2008)) have shown that nonparametric variants of these different estimators all converge at the same rate and have the same limiting distribution. They also all achieve the semiparametric efficiency bound exploiting general nonparametric specifications.

When the propensity score is known ex ante, or if a research assumes that the propensity score has a correctly specified parametric functional form, efficient estimators require different modifications to each of the three estimation methods. See for example Hirano, Imbens, and Ridder (2003), Hahn (1998) and Chen, Hong, and Tarozzi (2008). It is also important to note that

under Assumption 1, $E(Y|D=1) - E(Y|D=0)$ does not estimate either ATE or ATT. On the other hand, when treatment assignment is completely random, it estimates both.

A researcher interested in estimating the effect of the treatment D sometimes specifies a model in which the treatment effect is either: (i) homogenous $\beta(X) \equiv \beta$, or (ii) takes a parametric specification such as a linear functional form given by $\beta(X) = \beta_0 + \beta_1'X$. These structures introduce additional identifying and efficiency information. We interpret these as *conditional* parameters.

Consider, first, the specification that $\beta(X) \equiv \beta$. Obviously β can be estimated by ATE or ATT, or $\beta(X)$ can be estimated for any X. In particular, an optimal estimate of β should be given by a weighted average of $\beta(X)$ across the covariates X, where the optimal weights are determined through the conditional variance structure of Y given X. Similarly, for the specification $\beta(X) = \beta_0 + \beta_1'X$, optimal estimates of β_0 and β_1 are given by weighted least square regressions where the optimal weights are also determined through the conditional variance structure of Y given X.

Alternatively, one can interpret $\beta(X) \equiv \beta$ as a semiparametric partial linear model that includes both a finite dimensional parameter β and an infinite dimensional function $\alpha(X)$:

$$E(Y|D,X) = \alpha(X) + \beta D.$$

Similarly, the linear specification of $\beta(X) = \beta_0 + \beta_1'X$ translates into

$$E(Y|D,X) = \alpha(X) + \beta_0 D + \beta_1 DX.$$

These partial-linear models can be estimated efficiently and flexibly using a variety of methods, such as the procedures discussed in Robinson (1988) and Ai and Chen (2003). Comparing Theorem 1 in Hahn (1998) and the partial linear model in Robinson (1988) using Cauchy-Schwartz inequality shows that for a homoscedastic model, in which $Var(\epsilon|D,X) \equiv \sigma^2$, the average treatment effect estimate $\widehat{ATE}$ has a strictly larger variance than the semiparametrially efficient partial-linear least square estimator. This is not surprising given that the constraint $\beta(X) \equiv \beta$ imposes additional structure in the model. The variances are the same only when the propensity score is a constant, or when the assignment is completely random.

5   Extensions to Nonlinear Conditional and Unconditional Effects

The recent literature has become increasingly focused on estimating nonlinear treatment effects, such as quantile parameters. See for example Firpo (2007), Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005). In these models the parameter of interest is generally a nonlinear functional of a conditional distribution of the outcome variables Y_1 and Y_0. Unconditional parameters (such as quantile) are defined as functionals of $f(Y_1)$ and $f(Y_0)$, or $f(Y_1|D=1)$ and $f(Y_0|D=1)$ for the "effects on the treated".

Under the conditional independence Assumption 1, the effects for the entire population are measured by comparing the quantities

$$\begin{aligned} f(Y_1) = & \quad f_1(Y) \equiv \int f(Y|D=1,X) f(X) dX, \\ f(Y_0) = & \quad f_0(Y) \equiv \int f(Y|D=0,X) f(X) dX. \end{aligned}$$

The unconditional effects on the "treated population" are analogously defined as

$$f(Y_1|D = 1) = f(Y|D = 1),$$
$$f(Y_0|D = 1) = \int f(Y|D = 0, X) f(X|D = 1) \, dX.$$

Each of these unconditional densities of Y_1 and Y_0 can be estimated using nonparametric analogs of any one of the three methods described in Section 4.

For example, an estimate of the unconditional effects $f(Y_0|D = 1)$ can be obtained either by

$$\sum_{i=1}^{n} \hat{f}(Y|D = 0, X_i) D_i / \sum_{i=1}^{n} D_i,$$

where

$$\hat{f}(Y|D = 0, x) = \sum_{i=1}^{n_0} \frac{1}{h} k\left(\frac{y_i - y}{h}\right) k\left(\frac{x_i - x}{h}\right) / \sum_{i=1}^{n_0} k\left(\frac{x_i - x}{h}\right);$$

or by

$$\frac{1}{n_0} \sum_{i=1}^{n_0} \frac{1}{h} k\left(\frac{y_i - y}{h}\right) \frac{\hat{p}(X_i)}{p} \frac{1 - p}{1 - \hat{p}(X_i)}.$$

Turning to an example of a nonlinear conditional parameter, nonparametric quantile regressions define *conditional quantile treatment effects* $\beta(X)$ through conditional quantiles:

$$E1(Y \le \beta_0(X) + \beta(X) D|D, X) = \tau,$$

with $\beta(X) = \beta_1(X) - \beta_0(X)$, $\beta_1(X) = Q_\tau(Y_1|X)$ and $\beta_0(X) = Q_\tau(Y_0|X)$. When $\beta(X)$ is heterogenous, one may be interested in learning about its entire distribution over X measured by its mean and percentiles. Alternatively, Firpo (2007) defines the notion of a quantile treatment effect as the difference between the quantiles of the marginal distributions of Y_1 and Y_0:

$$\beta \equiv \beta_1 - \beta_0 = Q_\tau(Y_1) - Q_\tau(Y_0).$$

These marginal quantiles can be defined and estimated through the regression principle:

$$E_X E[1(Y \le \beta_1)|D = 1, X] = E_X E[1(Y \le \beta_0)|D = 0, X] = \tau;$$

or, as in Firpo (2007) through the propensity score weighting principle:

$$E1(Y \le \beta_1) \frac{p}{p(D = 1|X)} = E1(Y \le \beta_0) \frac{1 - p}{1 - p(D = 1|X)} = \tau.$$

As yet another alternative, a researcher may be interested in efficiently estimating a model under the assumption of homogeneous treatment effects $\beta(X) = \beta$, or by introducing additional parametric or semiparametric assumptions on the functions $\beta_{0,\theta}(X)$ and $\beta_\theta(X)$ as in Chernozhukov and Hansen (2005) and Abadie, Angrist, and Imbens (2002). The *parameters θ* appearing in these functions, being either finite or infinite dimensional, can be estimated by parametric, semiparametric, or nonparametric quantile regressions of the form:

$$E1(Y \le \beta_{0,\theta}(X) + \beta_\theta(X) D|D, X) = \tau.$$

6 Allowing for Endogenous Treatment Effects

The above estimation concepts can be extended to instances where the treatment status is endogenous, regardless of whether one considers the linear or nonlinear effects.

6.1 Linear Treatment Parameters

With endogenous treatments D, Imbens and Angrist (1994) studied a model with a binary instrument. Under their independence and monotonicity assumptions,

Assumption 2 $Y_1, Y_0, D_1, D_0 \perp Z, and \quad P(D_1 \geq D_0) = 1,$

for $\tilde{Z} = (1, Z)$, the instrumental variable parameter β defined by

$$E\tilde{Z}(Y - \alpha - \beta D) = 0 \tag{7}$$

estimates the local average treatment effect (LATE) parameter: $\beta = E(Y_1 - Y_0|D_1 > D_0)$. Their paper did not consider exogenous covariates X.

The LATE framework with endogenous D is generalized to allow for conditioning on exogenous covariates X in a sequence of papers by Abadie, Angrist, and Imbens (2002) and Abadie (2003). In the conditional version, the independence assumption takes the form of

Assumption 3 *Conditional on each* X, $Y_1, Y_0, D_1, D_0 \perp Z, and \quad P(D_1 \geq D_0|X) = 1.$

Under these conditions, a parameter $\beta(x)$ defined by the conditional moment restriction:

$$E\left[\tilde{Z}(Y - \alpha(X) - \beta(X)D)|X\right] = 0, \tag{8}$$

which can also be written analytically as

$$\beta(X) = \frac{E(Y|Z = 1, X) - E(Y|Z = 0, X)}{E(D|Z = 1, X) - E(D|Z = 0, X)},$$

estimates a conditional LATE parameter:

$$\beta(X) = E(Y_1 - Y_0|D_1 > D_0, X). \tag{9}$$

Note that Assumption 1 is a special case of Assumption 8 with $D = Z$.

In the presence of treatment heterogeneity where $\beta(X)$ can vary substantially across X, a parameter of interest is the "unconditional average treatment effect" defined by $\beta = E(Y_1 - Y_0|D_1 > D_0)$. This parameter can be estimated by averaging $\beta(X)$ over the subpopulation of $f(X|D_1 > D_0)$:

$$\beta = \int \beta(X) f(X|D_1 > D_0) \, dX,$$

where

$$f(X|D_1 > D_0) = \frac{P(D = 1|Z, X) - P(D = 1|Z = 0, X)}{P(D_1 > D_0)} f(X)$$

7

and

$$P\left(D_1 > D_0\right) = \int \left[P\left(D = 1|Z, X\right) - P\left(D = 1|Z = 0, X\right)\right] f\left(X\right) dX.$$

Hence β can be estimated by

$$\hat{\beta} = \frac{\frac{1}{n} \sum_{i=1}^n \left(\hat{E}\left(Y|Z = 1, X_i\right) - \hat{E}\left(Y|Z = 0, X_i\right)\right)}{\frac{1}{n} \sum_{i=1}^n \left(\hat{E}\left(D|Z = 1, X_i\right) - \hat{E}\left(D|Z = 0, X_i\right)\right)}.$$

With endogenous treatment status, Abadie (2003) shows that a proper notion of the *propensity score* is defined by $p_z\left(X\right) = p\left(Z = 1|X\right)$. The same unconditional parameter of interest can also be expressed in terms of propensity score weighting schemes:

$$\beta = \frac{\int E\left(Y|Z = 1, X\right) \frac{p}{p_z(X)} f\left(X|Z = 1\right) dX - \int E\left(Y|Z = 0, X\right) \frac{1-p}{1-p_z(X)} f\left(X|Z = 0\right) dX}{P\left(D_1 > D_0\right)}.$$

Hence, an alternative estimate is given by

$$\tilde{\beta} = \frac{\frac{1}{n_1} \sum_{i=1}^{n_1} Y_i \frac{\hat{p}}{\hat{p}_z(X_i)} - \frac{1}{n_0} \sum_{j=1}^{n_0} Y_j \frac{1-\hat{p}}{1-\hat{p}_z(X_j)}}{\frac{1}{n_1} \sum_{i=1}^{n_1} D_i \frac{\hat{p}}{\hat{p}_z(X_i)} - \frac{1}{n_0} \sum_{j=1}^{n_0} D_j \frac{1-\hat{p}}{1-\hat{p}_z(X_j)}}.$$

In this expression, n_1 and n_0 denote the subsamples for which $Z = 1$ and $Z = 0$. Hong and Nekipelov (2007) and Frolich (2006) show that $\hat{\beta}$ and $\tilde{\beta}$ achieve the semiparametric effiency bound for estimating β.

If a researcher is willing to assume the treatment effect is homogeneous (i.e., $\beta\left(X\right) \equiv \beta$), then the set of moment conditions

$$E\left[\tilde{Z}\left(\alpha\left(X\right) + \beta D\right)|X\right] = 0,$$

again defines a semiparametric conditional moment model with both a finite dimensional parameter β and an infinite dimensional parameter $\alpha\left(X\right)$. These quantities can be estimated using procedures described in Newey and Powell (2003) and Ai and Chen (2003).

Instead of assuming that the treatment effect is purely homogeneous, it is common to adopt a parametric or semiparametric specification of $\beta\left(X\right)$, such as a linear specification $\beta\left(X\right) = \beta_0 + \beta_1 X$. These models can also be estimated efficiently as a partial-linear semiparametric conditional moment model:

$$E\left[\tilde{Z}\left(\alpha\left(X\right) + \beta_0 D + \beta_1 DX\right)|X\right] = 0.$$

In addition, parametric and semiparametric specifications can also be imposed on $\alpha\left(X\right)$, e.g., $\alpha\left(X\right) \equiv \alpha$ or $\alpha\left(X\right) = \alpha_0 + \alpha_1 X$. Because Z is a dummy variable, the second set of moment condition $E\left[Z\left(\alpha_\theta\left(X\right) + \beta_\theta\left(X\right) D\right)|X\right] = 0$ is equivalent to

$$E\left[A\left(X\right)\left(\alpha_\theta\left(X\right) + \beta_\theta\left(X\right) D\right)|Z = 1\right] = 0$$

for any instrument function $A(X)$. Abadie, Angrist, and Imbens (2002) and Abadie (2003) show that parameters can be estimated using a transformation of it based on *propensity score weighting*:

$$E\left[A\left(X\right)\left(\alpha_{\theta}\left(X\right)+\beta_{\theta}\left(X\right)D\right)|Z=1\right]=EA\left(X\right)\left(\alpha_{\theta}\left(X\right)+\beta_{\theta}\left(X\right)D\right)\frac{p_{z}}{p_{z}\left(X\right)}.$$

The instrument functions $A(X)$ can be chosen efficiently to achieve the semiparametric variance bound (Hong and Nekipelov (2007)).

6.2 Nonlinear Treatment Parameters

The finding that the parameters $\beta(X)$ estimated using the linear instrumental variable condition given by (8) coincides with the LATE parameter (9) of Imbens and Angrist (1994) appears to be specific to the linear model and does not generalize to nonlinear specifications. For example, the quantile treatment effect model under rank invariance conditions in Chernozhukov and Hansen (2005) and under LATE conditions in Abadie, Angrist, and Imbens (2002) depend on different moment conditions. In the moment Condition 8, while $\beta(X)$ estimates (9), $\alpha(X)$ and $\alpha(X)+\beta(X)$ do not estimate $E(Y_0|D_1 > D_0)$ and $E(Y_1|D_1 > D_0)$.

Under the LATE assumption, Abadie (2003) shows that for any function $g(\cdot)$ of Y_1 and Y_0:

$$\begin{aligned} E\left(g\left(Y_1\right)|D_1 > D_0, X\right) &= \frac{E(g(Y_1)D_1|X)-E(g(Y_1)D_0|X)}{P(D_1>D_0|X)} \\ &= \frac{E(g(Y)D|Z=1,X)-E(g(Y)(1-D)|Z=0,X)}{E(D|Z=1,X)-E(D|Z=0,X)}, \end{aligned}$$

and

$$\begin{aligned} E\left(g\left(Y_0\right)|D_1 > D_0, X\right) &= \frac{E(g(Y_0)(1-D_0)|X)-E(g(Y_0)(1-D_1)|X)}{P(D_1>D_0|X)} \\ &= \frac{E(g(Y)(1-D)|Z=0,X)-E(g(Y)(1-D)|Z=1,X)}{E(D|Z=1,X)-E(D|Z=0,X)}. \end{aligned}$$

Hence a conditional quantile difference parameter can be defined as $\beta(X) = \beta_1(X) - \beta_0(X)$, where

$$\frac{E\left(1\left(Y \leq \beta_1\left(X\right)\right)D|Z=1,X\right)-E\left(1\left(Y \leq \beta_1\left(X\right)\right)D|Z=0,X\right)}{E\left(D|Z=1,X\right)-E\left(D|Z=0,X\right)}=\tau,$$

and

$$\frac{E\left(1\left(Y \leq \beta_0\left(X\right)\right)\left(1-D\right)|Z=0,X\right)-E\left(1\left(Y \leq \beta_0\left(X\right)\right)\left(1-D\right)|Z=1,X\right)}{E\left(D|Z=1,X\right)-E\left(D|Z=0,X\right)}=\tau.$$

Unconditional parameters can be defined by functionals of the distribution of $\beta(X)$, or in terms of the unconditional quantiles in the spirit of Firpo (2007), Frolich (2006) and Hong and Nekipelov (2007).

If a researcher is willing to impose parametric or semiparametric restrictions on $\beta_0(X)$ and $\beta_1(X)$, such as $\beta_1(X) \equiv \theta_1$ and $\beta_0(X) = \theta_0$, or more generally $\beta(X)$'s take a parametric or semiparametric functional form $\beta_1^\theta(X)$ and $\beta_0^\theta(X)$, then θ can be estimated by the system of moment equations:

$$E\left(1\left(Y \leq \beta_1^\theta\left(X\right)\right)D-\tau D|Z=1,X\right)=E\left(1\left(Y \leq \beta_1^\theta\left(X\right)\right)D-\tau D|Z=0,X\right),$$

and

$$E\left(1\left(Y \le \beta_0^\theta\left(X\right)\right)\left(1-D\right)+\tau D|Z=0,X\right) = E\left(1\left(Y \le \beta_0^\theta\left(X\right)\right)\left(1-D\right)-\tau D|Z=1,X\right).$$

The above system of moment conditions has different conditioning sets. However, it can be transformed into a system with the same conditioning parameters as in Ai and Chen (2003), by introducing auxiliary variables η_0 and η_1:

$$E\left(Z\left(1\left(Y \le \beta_1^\theta\left(X\right)\right)D-\tau D\right)-\eta_1 Z|X\right)=0,$$
$$E\left(\left(1-Z\right)\left(1\left(Y \le \beta_1^\theta\left(X\right)\right)D-\tau D\right)-\eta_1\left(1-Z\right)|X\right)=0,$$

and

$$E\left(\left(1-Z\right)\left(1\left(Y \le \beta_0^\theta\left(X\right)\right)\left(1-D\right)+\tau D\right)-\eta_0\left(1-Z\right)|X\right)=0,$$
$$E\left(Z\left(1\left(Y \le \beta_0^\theta\left(X\right)\right)D-\tau D\right)-\eta_0 Z|X\right)=0.$$

The parameters θ, η_0 and η_1 can be jointly efficiently estimated by the above system of conditional moment restrictions.

7   Concluding Remarks

The efficiency properties and efficient estimators under the unconfoundness assumption extend generally not only to local average treatment effect models, but also to a variety of other identification strategies, including difference in difference methods, panel data fixed effect methods and regression discontinuity methods. Each of these models have a common cell structure, in which the observed heterogeneity of the regression defines a partition of the sample into cells, while the treatment and instrumental like variables further partition each cell into subcells. Efficiency results under the unconfoundness assumption can be viewed as special cases of the efficient results under a general cell structure, and in particular as a special case of the local average treatment effect model when the instrumental variable coincides with the treatment status variable. Typically in the presence of exogenous covariates that defines the cell structure, identification assumptions are imposed conditional on each value of the covariate, or cell by cell. Treatment effects across cells are typically heterogeneous. Researchers might be interested in unconditional parameters which are the averaged treatment effects across the cells. Alternatively, treatment effects can be estimated more efficiently if researchers are willing to impose additional parametric and semiparametric structures on the heterogeneous treatment effects across cells.

References

ABADIE, A. (2003): "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 113(2), 231–263.

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.

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

——— (2004): "Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions containing unknown functions," working paper, Yale University and the University of Florida.

CHEN, X., H. HONG, AND A. TAROZZI (2008): "Semiparametric efficiency in GMM models with auxiliary data," *Annals of Statistics*, 36(2), 808–843.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

FIRPO, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259–276.

FROLICH, M. (2006): "Nonparametric IV estimation of local average treatment effects with covariates," *Journal of Econometrics*, 139(1), 35–75.

HAHN, J. (1998): "On the Role of Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66(2), 315–332.

HIRANO, K., G. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica*, 71(4), 1161–1189.

HONG, H., AND D. NEKIPELOV (2007): "Semiparametric Efficiency in Nonlinear LATE Models," working paper, Stanford University and UC Berkeley, accept at QE.

IMBENS, G., AND J. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

NEWEY, W., AND J. POWELL (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71(5), 1565–1578.

ROBINSON, P. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

ROSENBAUM, P., AND D. RUBIN (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41.