Uniformly Semiparametric Efficient Estimation of Treatment Effects With a Continuous Treatment

Antonio F. GALVAO and Liang WANG

This article studies identification, estimation, and inference of general unconditional treatment effects models with continuous treatment under the ignorability assumption. We show identification of the parameters of interest, the dose–response functions, under the assumption that selection to treatment is based on observables. We propose a semiparametric two-step estimator, and consider estimation of the dose–response functions through moment restriction models with generalized residual functions that are possibly nonsmooth. This general formulation includes average and quantile treatment effects as special cases. The asymptotic properties of the estimator are derived, namely, uniform consistency, weak convergence, and semiparametric efficiency. We also develop statistical inference procedures and establish the validity of a bootstrap approach to implement these methods in practice. Monte Carlo simulations show that the proposed methods have good finite sample properties. Finally, we apply the proposed methods to estimate the unconditional average and quantile effects of mothers’ weight gain and age on birthweight. Supplementary materials for this article are available online.

KEY WORDS: Birthweight; Dose–response function; Ignorability; Semiparametric estimation and inference

1. INTRODUCTION

There is a growing literature on program evaluation studies. Estimation of treatment effects has provided a valuable method of statistical analysis of the effects of policy variables. This is especially true for program evaluation in economics and statistics, where these methods help to analyze how treatments or social programs affect the outcome distributions of interest.

The literature on unconditional treatment effects (TE) concentrates almost exclusively on the special case of discrete treatment, as binary and multi–valued treatment assignments. On the binary TE models, Hahn (1998), Heckman et al. (1998), Hirano, Imbens, and Ridder (2003), Abadie and Imbens (2006), Imbens, Newey, and Ridder (2006), and Li, Racine, and Wooldridge (2009) studied efficient estimation of the average treatment effects (ATE). The study of ATE has been extended to the quantile treatment effects (QTE) by Firpo (2007). There is also literature on estimation of ATE and QTE for multi-valued TE, see, for example, Imbens (2000), Lechner (2001), and Cattaneo (2010).

However, the literature on continuous TE is relatively sparse. Among others, Hirano and Imbens (2004) and Imai and van Dyk (2004) developed a generalized propensity score for continuous average treatment models. Flores (2007) proposed nonparametric estimators for average dose–response functions (ADRF). Florens et al. (2008) considered identification of ATE using control functions. Lee (2012) studied unconditional distribution of potential outcomes with continuous treatments. Despite this sparsity, many empirical questions of interest in applied research involve continuous treatments. For example, in the study of the effects of mothers’ weight gain during pregnancy on birthweight, the weight gain in pounds is a continuous variable.

This article studies identification, estimation, and inference of general unconditional TE models with a continuous dose of the treatment. We consider estimating the parameters of interest, the dose–response functions (DRF), through moment restriction models (Z-estimators) in which the generalized residual functions are possibly nonsmooth. In this general formulation, the range of models for which the methods are applicable is very broad. For instance, the framework includes average and quantile DRF, ADRF, and QDRF, as special cases, and consequently ATE and QTE are direct applications of the proposed methods. Throughout the article, we use the examples of ADRF and QDRF as applications of the general theory. Thus, we extend the literature on ATE and QTE from discrete to continuous dose of treatment. In addition, we extend the literature on continuous treatments, which to our knowledge only allows for ADRF and ATE, to general DRF, with QDRF and QTE being important examples.

Consistent estimation of the DRF requires identification of the parameters of interest. In this article, following Rosenbaum and Rubin (1983), the relevant restriction for identification is the ignorability assumption, that is, the selection to treatment is based on observable variables. The ignorability assumption states that given a set of observed covariates, each individual is randomly assigned to either the treatment group or the control group. This condition has been largely employed in the literature, see, for example, Rubin (1977), Heckman et al. (1998), Dehejia and Wahba (1999), Firpo (2007), Flores (2007), and Cattaneo (2010).

Based on the identification, we propose a two-step estimation procedure. The practical implementation of the estimator

Antonio F. Galvao is Associate Professor, Department of Economics, University of Iowa, Iowa City, IA 52242 (E-mail: antonio-galvao@uiowa.edu). Liang Wang is PhD Student, Department of Economics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201 (E-mail: wang42@uwm.edu). The authors express their appreciation to Sergio Firpo, Carlos Flores, Alfonso Flores-Lagunes, Zhengyuan Gao, Chuan Goh, Roger Koenker, Carlos Lamarche, Ying-Ying Lee, Marcelo Medeiros, Marcelo Moreira, Cristine Pinto, Alexandre Poitier, Suyong Song, Elie Tamer, Zhijie Xiao, Ting Zhang, and participants in the seminars at the University of Wisconsin Madison, University of Iowa, University of Illinois at Urbana-Champaign, University of Wisconsin-Milwaukee, PUC-Rio, Central Bank of Brazil, and the 2013 Latin American Workshop in Econometrics for useful comments and discussions regarding this article. The authors also thank the editor and three anonymous referees for their careful reading and comments to improve the article. Computer programs to replicate the numerical analyses are available from the authors. All the remaining errors are ours.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jasa.
Estimation of TEs with a Continuous Treatment

2. THE MODEL, IDENTIFICATION, AND ESTIMATION

In this article, we assume that a random sample of size \( n \) is available. The objective is to learn how an outcome variable changes as the dose of some treatment variable varies. The dose is denoted by \( t \), where \( t \in T \), an interval in \( \mathbb{R} \), and the outcome is denoted by \( Y(t) \). More specifically, for each \( t \in T \), \( Y(t) \) is the outcome when the dose of treatment is \( t \). When \( t \) varies in \( T \), a random process \( Y(t) \) is defined. The random process \( Y(t) \) indexed by \( t \in T \) denotes potential outcomes under treatment levels in \( T \). However, in practice, one cannot observe \( Y(t) \) for all \( t \in T \). Rather, only a single \( Y(t_0) \) can be observed, where \( t_0 \) is the realization of a random variable \( T \). Thus, the observed outcome is the random variable

\[
Y = Y(T) = \int_{t \in T} Y(t) d\mathbb{1}[t \geq T].
\]

Ideally we would like to estimate the value of the DRF at \( t_0 \) using the sample with \( T = t_0 \). However, in general, due to the self-selection problem, bias can arise by direct use of sample counterparts. To illustrate this point, we consider the estimation of ATE, from the DRF, as an example. For any \( t_1 < t < t_2 \),

\[
\mathbb{E}[Y|T = t_2] - \mathbb{E}[Y|T = t_1] = \mathbb{E}[Y(t_2) - Y(t_1)|T = t] + \mathbb{E}[Y(t_2)|T = t_2] - \mathbb{E}[Y(t_1)|T = t_1].
\]

\( \mathbb{E}[Y(t_2)|T = t_2] - \mathbb{E}[Y(t_1)|T = t_1] \) is observed difference in birthweight, \( \mathbb{E}[Y(t_2) - Y(t_1)|T = t] \) is average treatment effect on the treated, \( \mathbb{E}[Y(t_1)|T = t_1] \) is selection bias 1 and \( \mathbb{E}[Y(t_2)|T = t_2] - \mathbb{E}[Y(t_1)|T = t_1] \) is selection bias 2.

This simple calculation indicates that, due to the existence of the selection biases 1 and 2, it is impossible to directly use the sample counterparts to compute TE. To solve this problem, it is common in the literature to assume the existence of a set of random variables \( X \) conditional on which \( Y(t) \) is independent from \( T \) for all \( t \in T \). In such case,

\[
\mathbb{E}[Y|X, T = t_2] - \mathbb{E}[Y|X, T = t_1] = \mathbb{E}[Y(t_2)|X, T = t_2] - \mathbb{E}[Y(t_1)|X, T = t_1] = \mathbb{E}[Y(t_2)|X] - \mathbb{E}[Y(t_1)|X] = \mathbb{E}[Y(t_2) - Y(t_1)|X],
\]

which has a causal interpretation. This is the ignorability condition, which is discussed in more detail below. Finally, we need to combine the results for \( X \) to obtain an unconditional TE. By the law of iterated expectations, this unconditional expectation can be recovered.

The objective of this article is to study continuous treatments considering general forms of dose–response for TE models. To accomplish this aim, we develop a general framework for generic moment restriction estimators (Z-estimators) with
possibly nonsmooth functions. For each \( t \in T \), the parameter of interest \( \beta(t) \in B \subset \mathbb{R} \) is assumed to uniquely solve the identifying conditions as
\[
E[m(Y(t); \beta(t))] = 0,
\]
where \( m(\cdot) \) is defined as the generalized residual function. Then for each \( t \), the DRF is defined as the value \( \beta(t) \) that solves the moment condition. We will use the important examples of average and quantile dose–response functions (ADRF and QDRF) as applications of the general theory. In addition, we use the ADRF and QDRF estimators in the empirical application to a birthweight study. The following examples show that ADRF and QDRF are special cases of \( \beta(\cdot) \), which result from choosing specific forms of \( m(\cdot) \).

**Example 1 (Average).** The ADRF example. Setting \( m(Y(t); \mu(t)) = Y(t) - \mu(t) \) and letting for each \( t \) the first moment equal to zero, we obtain \( \mu_0(t) = E[Y(t)] \), the unconditional ADRF. From this, it is easy to recover the ATE, which is given by \( \text{ATE}(t, t') = \mu_0(t) - \mu_0(t') \).

**Example 2 (Quantile).** QDRF is another special case of the general model. Let \( m(Y(t); q_0(t)) = \tau - 1\{Y(t) < q_0(t)\} \) for each \( t \), we obtain \( q_0(t) \in \text{inf}\{q : F(Y(t)) \geq \tau\} \), the unconditional \( r \)th QDRF, where \( F_Y(t) \) is the distribution function of \( Y(t) \). From the QDRF, one can estimate the QTE as \( \text{QTE}(t, t') = q_0(t) - q_0(t') \). In this article, QTE is defined as the difference of the \( \tau \)th quantile at different levels of treatment.

We state assumptions on the general model for identification of the parameters of interest.

I.I For each \( t \in T \), \( \beta_0(t) \) uniquely solves \( E[m(Y(t); \beta(t))] = 0 \), where \( m : \mathbb{R} \times B \mapsto \mathbb{R} \) is measurable.

I.II For all \( t \in T \), we have:
1. \( Y(t) \perp T \mid X \);
2. \( f_{Y|X}(y|x) > 0 \) for \( t \in T, x \in X \) and \( y \in Y \).

I.III Assume that
1. There exists a function \( e(y) \) with \( \int e(y) \, dy < \infty \) such that \( |m(y; \beta(t_0))| f_{Y|X}(t_0 + \Delta t, y|X) \leq e(y) \);
2. \( E[m(Y; \beta(t_0)) X, T = t_0] \rightarrow \lim_{\Delta t \to 0} E[m(Y; \beta(t_0)) X, T \in [t_0, t_0 + \Delta t]] \). Also the interval \( T \) is right open.

Assumption I.I is an identification condition provided that \( Y(t) \) is observable. The parameter of interest, \( \beta(t) \), is defined by the moment condition. However, this condition cannot be used directly to estimate \( \beta(t) \) because our data are not experimental and \( Y(t) \) is not observable for all \( t \in T \). Therefore, condition I.II.1, the ignorability assumption, is fundamental. According to I.II.1, although the assignment of the treatment level is not random, it is random within subpopulations characterized by \( X \). This assumption has been used, among others, by Heckman et al. (1998), Dehejia and Wahba (1999), and Hirano and Imbens (2004). I.II.2 states that the density of treatment levels is positive. Thus, the triple \( (X, Y, T) \) is observable, and a random sample of size \( n \) can be obtained. I.III allows for changing the orders of limits and integral. The set \( T \) is right open without loss of generality.

The identification result is presented in the following theorem. For notational convenience, denote \( u := (X^\top, y)^\top \) and \( U := (X^\top, Y)^\top \).

**Theorem 1.** Under conditions I.I–I.III, we have
\[
E[m(Y(t); \beta(t))] = E[m(Y; \beta(t))\pi_0(U; t)]
\]
for each \( t \in T \), where \( \pi_0(u; t) := \int_{X \times Y} f_{X|Y}(x|y) \cdot f_{Y|X}^u(y|x) \). Consequently,
\[
E[m(Y; \beta(t))\pi_0(u; t)] = 0
\]
if and only if \( \beta(t) = \beta_0(t) \).

**Proof.** See the Appendix. \( \square \)

The result in Equation (1) allows identification of the DRF. The left-hand side of (1) is used to define \( \beta(t) \), which involves the unobservable \( Y(t) \). Consequently, it cannot be used to estimate \( \beta(t) \). Nevertheless, the right-hand side of (1) is expressed in terms of the observables \( (X, Y, T) \), and hence, can be used to estimate \( \beta(t) \). Note that \( Y(t) \) is not observable while \( Y \) is. The intuition behind the result is that the existence of \( X \) delivers identification of the parameter of interest. That is, conditional on observed covariates \( X \), each individual is randomly assigned to a treatment level.

**Remark 1.** The result in Theorem 1 has a similar format as eq. (2) of Cattaneo (2010), after we transform the latter. To see this, we begin with \( E \left[ \left( T = t \mid m(Y; \beta(t)) \right) \pi_0(U; t) \right] = 0 \). By the law of iterated expectation, the left-hand side of the previous equation equals to \( E \left[ m(Y; \beta(t)) \pi_0(U; t) \mid T = t \right] E \left[ 1 \{ T = t \} \mid X, Y \right] \). Noting that \( E \left[ 1 \{ T = t \} \mid X, Y \right] = P(T = t \mid X, Y) \) and, by definition, \( p_t(X) = P(T = t \mid X) \), the last equation equals to \( E \left[ m(Y; \beta) \pi_0(U; t) \mid T = t \right] E \left[ 1 \{ T = t \} \mid X, Y \right] \). Thus, our result “replaces” the conditional probabilities by conditional densities.

**Remark 2.** The result in Theorem 1 is also related to Hirano and Imbens (2004). They extended the propensity score method to a setting with continuous treatment for ADRF. This article complements their results by providing a more general model for estimating DRF. This general formulation includes ADRF as a special case. In addition, we generalize the results in Hirano and Imbens (2004) by establishing uniform asymptotic results of the two-step estimator. In particular, in the next section we show uniform consistency, weak convergence, and semiparametric efficiency of the proposed estimator. We also develop practical inference.

Given the identification condition in Equation (2) of Theorem 1, we are able to estimate the parameters of interest. We propose a two-step estimator as follows.

**Step 1.** Estimate \( \pi_0(U; t) = \frac{f_{X|Y}(x|y) \cdot f_{Y|X}^u(y|x)}{f_{X|Y}^u(x|y)} \) and obtain an estimator \( \hat{\pi}_0 \).

**Step 2.** For each \( t \in T \), find \( \hat{\beta}(t) \) as a zero of the following condition
\[
\frac{1}{n} \sum_{i=1}^n m(Y_i; \beta(t))\hat{\pi}(U_i; t) = 0.
\]
The estimator \( \hat{\beta}(t) \) is defined as the zero of the equation above. The identification conditions and the estimator are illustrated in the following ADRF and QDRF examples.

**Example 3 (Average).** The identification condition for \( \mu_0(t) \) is \( \mathbb{E}[(Y - \mu_0(t))\pi_0(U; t)] = 0 \). An estimator of \( \mu_0(t) \) is
\[
\hat{\mu}(t) = \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}(U_i; t) \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}(U_i; t) Y_i.
\]

**Example 4 (Quantile).** The identification condition for \( q_{01}(t) \) is \( \mathbb{E}[(\tau - 1|Y < q_{01}(t))\pi_0(U; t)] = 0 \). An estimator of \( q_{01}(t) \) is
\[
\hat{q}_1(t) = \arg \min_q \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}_0(U_i; t) \rho_q(Y_i - q),
\]
where \( \rho_q(u) := u(\tau - 1[u < 0]) \) is the check function as in Koenker and Bassett (1978).

### 3. ASYMPTOTIC PROPERTIES

In this section, we derive the asymptotic properties of the two-step estimator. We establish the uniform consistency and the weak limit of the DRF, \( \hat{\beta}(\cdot) \), in \( \ell^\infty(T) \). In addition, as important applications and examples, we specialize the general discussion of this type of identification assumption, see van der Vaart (1998). C.III only requires the moment of the estimation equation to be finite. This is analogous to 4(b) in Cattaneo (2010). C.IV requires consistent estimation of the nuisance parameter. This is a usual requirement corresponding to (1.4) of Theorem 1 of Chen, Linton, and Van Keilegom (2003). We discuss estimation of the nuisance parameter in Section 3.3. C.V implies a uniform law of large numbers, this is standard; see, for example, Newey and McFadden (1994). We provide more primitive conditions to establish consistency for the specific cases of ADRF and QDRF. Now we state the uniform consistency result for the DRF estimator.

**Theorem 2.** Suppose that \( \mathbb{E}[m(Y, \beta_0(t))\pi_0(U; t)] = 0 \), and that conditions C.I–C.V are satisfied. Then, as \( n \to \infty \)
\[
\sup_{t \in T} |\hat{\beta}(t) - \beta_0(t)| = o_p(1).
\]

**Proof.** See the Appendix.

As illustration of the general result, we discuss uniform consistency over \( t \in T \) of the two-step estimators ADRF and QDRF given in (4) and (5), respectively. In these specific cases, the functional form of \( m(\cdot) \) is given and, hence, we can check the high level conditions of Theorem 2 (C.I–C.V). We show that both ADRF and QDRF satisfy C.I–C.III. They are formally verified in Corollaries 1 and 2. Assumptions C.IV–C.V refer to the estimation of the nuisance parameter \( \pi_0 \) in the first step, and examples satisfying these conditions are given in Section 3.3. Consider the assumptions for the ADRF.

AC.I There exists \( 0 < M_Y < \infty \) such that \( \mathbb{E}[Y(t)] < M_Y \). Also, the parameter space for \( \mu \) is a bounded sub-Banach space \( \mathcal{M} \) of \( \ell^\infty(T) \).

AC.II The function class \( \{\psi_{2,\pi,t} : \pi \in \Pi_\delta, t \in T\} \) is Glivenko-Cantelli, and has an envelope \( F_2(y) \) such that \( yF_2(y) \) is integrable.

AC.I requires the expectation of \( Y(t) \) to be finite. The diameter of the parameter space is finite, which is common for \( \hat{\mu} \)-estimators; see, for example, Chen, Linton, and Van Keilegum (2003). Hirano, Imbens, and Ridder (2003) assumed the second moment of \( Y(1) \) and \( Y(0) \) to be finite, which is slightly stronger. AC.II is a high level condition on the nuisance parameter, parallel to C.V, and will be discussed in more detail below. Nevertheless, there are many functional classes that satisfy this condition. Examples include the smooth function class in Example 19.9 of van der Vaart (1998) for sufficiently smooth functions and sufficiently small tail probabilities. Uniform consistency of the ADRF is summarized in the following corollary. The intuition of this result is that under AC.I and AC.II we are able to check conditions C.I–C.V.

**Corollary 1 (Average).** The two-step estimator of ADRF is uniformly consistent, that is, \( |\hat{\beta}(\cdot) - \mu_0(\cdot)|_\infty = o_p(1) \), provided AC.I–AC.II and C.IV are satisfied.

**Proof.** See the Appendix.

For the uniform consistency of the QDRF over \( t \in T \), consider the following conditions.

**QC.1** Uniformly in \( t \), the densities \( f_{Y_0}(y) \) is bounded above and \( f_{Y_0}(q_{01}(t)) > 0 \). Also, for any \( \delta > 0 \), \( \inf_{|y-q_{01}(t)| > \delta} |\mathbb{E}[(\tau - 1|Y < q)|\pi_0(U; t)|_\infty > \epsilon_\delta \).

**QC.2** There exists \( 0 < M_\pi < \infty \) such that \( \pi_0(u; t) < M_\pi \).
QC.I The function class \( \{ \psi_{2,\pi,t} : \pi \in \Pi_t, t \in T \} \) is Glivenko-Cantelli, and have an envelope \( F_2(y) \) such that \( F_2(y) \) that is integrable.

QC.I is a standard identification condition in the quantile regression literature. It allows one to verify C.I. It is similar to A.2–A.3 of Angrist, Chernozhukov, and Fernandez-Val (2006), and corresponds to Assumption 2 of Firpo (2007). QC.II imposes boundedness on the joint density of \((U, T)\) analog to Assumption 1(ii) of Firpo (2007). QC.III is weaker than AC.II, since \( \tau = 1\{ \} \) is uniformly bounded. Those two conditions are imposed for estimation of the nuisance parameter. Under QC.I–QC.III, we can check the general conditions C.I–C.V. Now we provide uniform consistency of QDRF.

**Corollary 2 (Quantile).** The two-step estimator of the QDRF is uniformly consistent, that is, \( |\hat{q}_t(\cdot) - q_{0,t}(\cdot)|_\infty = o_p(1) \), provided QC.I–QC.III and C.IV are satisfied.

**Proof.** See the Appendix. □

### 3.2 Weak Convergence

Now we derive the limiting distribution of the general two-step estimator in (3). We impose the following sufficient conditions. As application of the general results, later, we establish the results for the ADRF and QDRF and check the general conditions in these examples.

**G.I** \[ |E[m(Y; \beta(t))|\tau(T; \cdot)|_\infty = o_p(1/\sqrt{n_r_n}), \] where \( n_r_n = o(n^{-1}) \) or \( n_r_n = 1 \).

**G.II** The parameter space for \( \mu_0 \) is a bounded sub-Banach space of \( \ell^\infty(T) \). In addition, \( |E[Y(t)|_\infty < M_T \) for some \( 0 < M_T < \infty \).

**G.III** The function class \( \{ \psi_{3,\pi,t} : \pi \in \Pi_t, t \in T \} \) is Donsker, where \( \psi_{3,\pi,t}(u) = y\tau(u; t) \).

**G.IV** \[ \sqrt{n_r_n} E[(\hat{m}(Y; \beta(t))\pi_0(U; t) - \pi_0(U; t))]_{\pi = \hat{x}} \] converges weakly to a tight random element \( \tau(t) \) in \( \ell^\infty(T) \).

In G.I, \( n_r_n \) is a random sequence referring to a possible nonparametric estimation of \( \pi_0 \) in the first step. For example, a bandwidth parameter in kernel estimation. Note that when \( \pi_0 \) is estimated parametrically in the first step, we have \( n_r_n = 1 \). When \( \pi_0 \) is estimated nonparametrically, \( n_r_n \) converges to 0 and, therefore, the second term disappears. Assumption G.I defines the Z-estimator. This type of \( o_p(n^{-1/2}) \) condition is slightly stronger than C.I but still allows the right-hand side to be zero only approximately. G.II requires the model to be differentiable in \( \beta \) and the derivative to be invertible. G.I and G.II are also assumed in theorem 3.3 of Pakes and Pollard (1989). G.III corresponds to assumption 6(b) of Cattaneo (2010). G.I–G.II are high level and we provide concrete examples using the ADRF and QDRF applications. G.IV strengthens C.IV such that the estimator of the nuisance parameter converges at a rate faster than \( n^{-1/4} \). A similar assumption appears in Chen, Linton, and Van Keilegom (2003). G.V and G.VI are high level conditions, similar to Cattaneo (2010), and will be discussed below in the estimation of \( \pi_0 \). Now we present the weak convergence result.

**Theorem 3.** Suppose that \( |E[m(Y; \beta_0(t))\pi_0(U; t)]|_\infty = 0 \), that \( |\beta - \beta_0|_\infty = o_p(1) \), and that conditions G.I–G.VI are satisfied. Then, in \( \ell^\infty(T) \),

\[ \sqrt{n_r_n}(\hat{\beta}(t) - \beta(t)) \approx Z^{-1}\beta_0(t), \pi_0(U; t))G(t). \]

**Proof.** See the Appendix. □

The result in Theorem 3 shows that the limiting distribution of the two-step DRF estimator is nonstandard (Gaussian process). This result is due to the presence of the set of continuous treatments. However, if one fixes the treatment at \( \tilde{t} \), then the limiting distribution collapses to a simple normal distribution. In spite of this, below we provide inference methods for DRF over the set of treatments that are simple to implement in practice. Nevertheless, Theorem 3 is an important step to rigorously establish the inference procedure.

This result has significant practical applications, for instance in the ADRF and QDRF examples. In these cases, the functional forms of \( m(\cdot) \) are given and, hence, we can check the conditions in Theorem 3. We show that both ADRF and QDRF satisfy G.I–G.III, they are verified in Corollaries 3 and 4. G.IV–G.VI refer to the estimation of the nuisance parameter \( \pi_0 \). For the weak convergence of the ADRF, we impose the following assumptions.

**AG.I** The parameter space for \( \mu_0 \) is a bounded sub-Banach space of \( \ell^\infty(T) \). In addition, \( |E[Y(t)]|_\infty < M_T \) for some \( 0 < M_T < \infty \).

**AG.II** The function class \( \{ \psi_{3,\pi,t} : \pi \in \Pi_t, t \in T \} \) is Donsker, where \( \psi_{3,\pi,t}(u) = y\tau(u; t) \).

**AG.III** \[ \sqrt{n_r_n} E[(Y - \mu_0(t))(\pi(U; t) - \pi_0(U; t))]_{\pi = \hat{x}} \] converges weakly.

**AG.I** is standard and requires the parameter space to be bounded. Also the second moment of \( Y(t) \) is bounded, which is used in Hirano, Imbens, and Ridder (2003) and Cattaneo (2010). Many functional classes satisfy AG.II, for example, the smooth class discussed above. AG.III is a high level condition on the nuisance parameter, and we provide estimators that satisfies this condition below.

**Corollary 3 (Average).** The two-step estimator of the ADRF is \( \sqrt{n_r_n} \)-consistent and converges weakly in \( \ell^\infty(T) \), provided conditions AG.I–AG.III and G.IV.

**Proof.** See the Appendix. □

To obtain the weak convergence of the QDRF, we impose the following assumptions.

**QG.I** The function class \( \{ \psi_{2,\pi,t} : \pi \in \Pi_t, t \in T \} \) is Donsker with a uniform bound.

**QG.II** \[ \sqrt{n_r_n} E[(\tau \in \mu_0(t)](\pi(U; t) - \pi_0(U; t))]_{\pi = \hat{x}} \] converges weakly.

Examples satisfying QG.II include smooth function classes. QG.I is a high level condition and will be discussed in the section
of the estimation of \( \pi_0 \). This assumption is similar to AG.III and a version of G.VI. Now we state the weak convergence of QDRF.

**Corollary 4 (Quantile).** The two-step estimator of QDRF is \( \sqrt{n} \)-consistent and converges weakly in \( \ell^\infty(T) \), provided conditions QC.I–QC.II, QG.I–QG.II, and G.IV.

**Proof.** See the Appendix.

### 3.3 Estimation of \( \pi_0 \)

We have been assuming that the estimator \( \hat{\pi} \) of the nuisance parameter \( \pi_0 \) satisfies various conditions. In this section, we discuss the estimation of \( \pi_0 \), that is, \( f_{r|X,Y}(t|x,y)/f_{r|X}(t|x) \). The estimation of the nuisance parameter in the first step is very important for practical implementation of the proposed methods.

#### 3.3.1 Nonparametric Estimation of \( \pi_0 \)

A potential nonparametric estimator for the nuisance parameter is the kernel estimator, as \( \hat{\pi}_n = n^{-1}h_n^d \sum_{i=1}^n K((W_i - W_i^0)/h_n) \). Although \( \pi_0 \) is the ratio of two densities, in fact the estimation is about the density of \( w = x, y, \) and \( t \). Under some assumptions described below, Gine and Nickl (2008), in their Proposition 4, showed that the kernel estimator converges weakly. Consider the following assumptions.

- **NP.I** The density function \( \rho_0 \in C^t(\mathbb{R}^d) \) for some \( t \geq 0 \).
- **NP.II** The kernel \( K \) of order \( r = t + 1 - k \) for some positive \( k < t + 1 \).
- **NP.III** Let \( C \) be: (a) any translation invariant \( \mathbb{P} \) Donker class of convex sets of diameter bounded by a fixed constant, or (b) all \( C^\infty \) domains with \( a \geq 2, a > (d - 1)/2 \) and with both diameter and \( H^{d-1}(aC) \) bounded by a constant (in this case, \( \rho_0 \) needs to have compact support).
- **NP.IV** The bandwidth is such that \( h_n^{k+1-t}n^{1/2} \rightarrow 0 \).

Under these assumptions, it is possible to verify the stated high level conditions (C.IV–C.V and G.IV–G.V) as a general theory.

**Proposition 1.** Under NP.I–NP.IV, conditions C.IV–C.V and G.IV–G.V hold.

**Proof.** See the Appendix.

Although the nonparametric estimation of \( \pi_0 \) seems persuasive from the theoretical point of view, there are issues with its practical implementation. First, in most empirical applications the number of variables in \( X \) used to satisfy the ignorability condition is relatively large. However, the dimension of \( X \) has an adverse effect on nonparametric methods due to the curse of dimensionality. Hence, practical estimation might be infeasible. Second, as Theorem 3 shows, when using nonparametric estimation in the first step, the convergence rate of the estimator is slower relative to parametric estimation. In addition, in the TE examples, for a formal verification of G.VI, parametric estimation in the first step is requested. Therefore, there are compelling reasons to use flexible parametric models to estimate the ratio of the conditional density functions, \( \pi_0 \). In the next subsection, we describe such estimators.

#### 3.3.2 Parametric Estimation of \( \pi_0 \)

It is common in the literature to estimate nuisance parameters in two-step estimators using parametric models, see, for example, Murphy and Topel (1985), Newey and McFadden (1994), Chernozhukov and Hong (2002), Hirano and Imbens (2004), and Wei and Carroll (2009). We follow this literature and also propose a flexible parametric approach. For estimators of \( \pi_0 \) to have the desirable properties, we impose the following assumptions.

- **PI** Assume \( \pi = \pi(u; t; \theta) \), where \( \theta \in \mathbb{R}^{d_\theta} \) with \( d_\theta \) being a positive integer. \( \pi(u; t; \theta) \) is a smooth function of \( \theta \) with uniformly continuous, bounded, and square integrable first derivative, \( \pi'(u; t; \theta) \), with respect to \( \theta \).
- **PII** There exists \( \tilde{\theta} \) such that \( \sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N(0, \Sigma_\theta^{-1}) \) with \( \Sigma_\theta \) nonsingular.

Condition PI is a smoothness and boundedness condition on the function to be estimated. P.II assumes there is an estimator of the parameter that is asymptotically normal.

**Proposition 2.** Under PI and P.II, conditions C.IV–C.V and G.IV–G.VI hold.

**Proof.** See the Appendix.

We provide examples to illustrate the estimation of \( \pi_0 \) in practice. These examples satisfy PI–P.II and are a direct application of Proposition 2 such that the high level conditions C.IV–C.V and G.IV–G.VI are automatically satisfied.

**Example 5.** We estimate \( f_{r|X,Y}(t|x,y) \) and \( f_{r|X}(t|x) \) separately. For \( f_{r|X,Y}(t|x,y) \), assume that \( T = g(X, Y; b) + \epsilon \), where \( \epsilon \) is some known function, and \( b \) is an unknown parameter to be estimated. Using nonlinear least squares we obtain estimates of the conditional mean and variance, and therefore, the conditional density of \( T \) given \( X \) and \( Y \). Similarly, we estimate \( f_{r|X}(t|x) \).

**Example 6.** To estimate the conditional density \( f_{r|X,Y}(t|x,y) \) assume the following model \( \Lambda(T, \lambda) = \Lambda(g(X, Y), \lambda) + \epsilon \), where \( \epsilon \) is some known function, and \( \lambda \) is a Box-Cox transformation function, which is defined as \( \Lambda(T, \lambda) = \log T \) if \( \lambda = 0 \) and \( \Lambda(T, \lambda) = T^{\lambda-1} \lambda \) otherwise. Using maximum likelihood estimation (MLE), we obtain the unknown parameter \( \lambda \) and therefore the conditional density \( f_{r|X,Y}(t|x,y) \). Similarly, we estimate \( f_{r|X}(t|x) \).

**Example 7.** A simple approach to estimate \( \pi_0 \) is to assume that \( (t, x, y) \) follow a known multivariate distribution, as a Normal distribution for instance. Then, MLE can be applied, \( f_{T,X,Y}(t, x, y) \) calculated, and \( \hat{\pi} \) obtained.

### 4. Inference on the DRF

In this section, we turn our attention to inference on the DRF. Important questions posed in the econometric and statistical literatures concern the nature of the impact of a policy intervention or treatment on the outcome distributions of interest; such as, for example, whether a policy exerts a significant effect, a constant versus heterogeneous effect, or a nondecreasing effect. Thus, we consider the following general null hypothesis

\[ \beta(t) - r(t) = 0, \quad t \in T, \]

uniformly, where \( r(t) \) is assumed to be known, continuous in \( t \) over \( T \), and \( r \in \ell^\infty(T) \). Inference is carried uniformly over the
set of treatment levels, \( T \). The basic inference process is

\[
V_n(t) := \hat{\beta}(t) - r(t), \quad t \in T.
\]

General hypotheses on the vector \( \beta(t) \) can be accommodated through functions of \( V_n(\cdot) \). We consider the Kolmogorov and Cramér–von Mises-type test statistics, \( T_n = f(V_n(\cdot)) \), where \( f(\cdot) \) represents the functionals for those two test statistics, as

\[
T_{1n} := \sqrt{n} \sup_{t \in T} |V_n(t)|, \quad T_{2n} := \sqrt{n} \int_{t \in T} |V_n(t)| \, dt.
\]

These statistics and their associated limiting theory provide a natural foundation for testing the null hypothesis. It is possible to formulate a wide variety of tests using variants of the proposed tests. The following are examples of hypotheses that may be considered.

**Example 8 (The hypothesis of a significant effect).** A basic hypothesis is that the treatment impact summarized by \( \beta(t) \) is ineffective for all doses, that is, statistically equal to zero for all \( t \in T \). The alternative is that the treatment differs from zero at least for some \( t \in T \). In this case, \( r(t) = 0 \).

**Example 9 (The hypothesis monotone nondecreasing).** The test of nondecreasing dose–response hypothesis involves the composite null hypothesis \( \beta(t) \leq 0 \), for all \( t \in T \), versus the alternative of \( \beta(t) < 0 \), for some \( t \in T \). In this case, the least favorable null involves \( r(t) = 0 \).

Now we present the limiting distributions of the test statistics under the null hypothesis. From Theorem 3 and under the null hypothesis \( H_0 : \beta(t) = r(t) \), it follows \( \sqrt{n} \sup_{t \in T} (\hat{\beta}(t) - r(t)) \rightsquigarrow \mathcal{G}(t) \). Thus, the following corollary summarizes the limiting distributions.

**Corollary 5.** Assume the conditions of Theorem 3. Under \( H_{01} : \beta_0(t) = r(t) \), as \( n \to \infty \),

\[
T_{1n} \rightsquigarrow \sup_{t \in T} |\mathcal{G}(t)|, \quad T_{2n} \rightsquigarrow \int_{t \in T} |\mathcal{G}(t)| \, dt.
\]

**Proof.** The assertion holds by Theorem 3 and the continuous mapping theorem. \( \square \)

In addition to testing the hypothesis \( \beta(t) = r(t) \) with known \( r \in C(T) \), we could also test the hypothesis with unknown \( r \), in which case, the estimation of \( r \) is needed.

**Example 10 (The hypothesis of a constant effect vs. heterogenous effects).** An important hypothesis is whether the treatment impact does not vary across the dose, that is, \( \beta(t) = \beta \) for some \( \beta \) for all \( t \). In this case, \( r(t) = \beta \), and one can estimate this component by \( \int_{T} \beta(t) \). The alternative is the hypothesis of heterogenous effect, that is, \( \beta(t) \) varies across \( t \). This can also be interpreted as a test for asymmetry of the dose–response function.

In this framework, given a \( \sqrt{n} \sup_{t \in T} |V_n(t)| \), \( \sqrt{n} \sup_{t \in T} (\hat{\beta}(t) - \bar{\beta}(t)) \rightsquigarrow \mathcal{G} - \mathcal{G}_r \), where \( \mathcal{G}_r \) is the weak limit of \( \sqrt{n} \sup_{t \in T} |V_n(t)| \). Thus, due to the estimation of \( r(t) \), a drift component is introduced. Let \( \tilde{V}_n(t) := (\hat{\beta}(t) - \bar{\beta}(t)) \). Under the null hypothesis \( H_{02} : \beta_0(t) = r(t) \), the test statistics become:

\[
\tilde{T}_{1n} := \sqrt{n} \sup_{t \in T} |\tilde{V}_n(t)|, \quad \tilde{T}_{2n} := \sqrt{n} \int_{t \in T} |\tilde{V}_n(t)| \, dt.
\]

Now we display the limiting distributions of the test statistics under the null hypothesis.

**Corollary 6.** Assume the conditions of Theorem 3. Under \( H_{02} : \beta_0(t) = r(t) \), as \( n \to \infty \),

\[
\bar{T}_{1n} \rightsquigarrow \sup_{t \in T} |\mathcal{G}(t) - \mathcal{G}_r|, \quad \bar{T}_{2n} \rightsquigarrow \int_{t \in T} |\mathcal{G}(t) - \mathcal{G}_r| \, dt.
\]

**Proof.** The assertion holds by Theorem 3 and the continuous mapping theorem. \( \square \)

The weak limits in Corollaries 5 and 6 are not standard. Therefore, to make practical inference we suggest the use of bootstrap techniques to approximate the limiting distribution.

### 4.1 Implementation of Testing Procedures

Implementation of the proposed tests in practice is simple. To test \( H_{01} \) with known \( r(t) \), one needs to compute the statistics of test \( T_{1n} \) or \( T_{2n} \). Analogously, to test \( H_{02} \) one computes \( \bar{T}_{1n} \) or \( \bar{T}_{2n} \). The steps for implementing the tests are as following.

First, the estimates of \( \beta(t) \) are computed by solving the problem in Equation (3). For special cases of DRF, as ADRF and QDRF, one replaces (3) with (4) and (5), respectively. Second, \( V_n(t) \) is calculated by centralizing \( \hat{\beta}(t) \) at \( r(t) \), and \( T_{1n} \) or \( T_{2n} \) is computed by taking the maximum over \( t \) (\( T_{1n} \)) or summing over \( t \) (\( T_{2n} \)). For the general case, \( H_{02} \) with unknown \( r(t) \), the tests are computed in the same fashion. The only adjustment is the use of \( \bar{r}(t) \) to compute \( \tilde{V}_n(t) \). Third, after obtaining the statistic of test, it is necessary to compute the critical values. We propose the following scheme. We use the statistic of test \( T_{1n} \) as example, but the procedure is the same for the other cases. Take \( B \) as a large integer. For each \( b = 1, \ldots, B \):

(i) Obtain the resampled data \((Y^b_i, X^b_i, T^b_i), i = 1, \ldots, n\).

(ii) Estimate the DRF \( \hat{\beta}^b(t) \) and set \( V_n^b(\cdot) := (\hat{\beta}^b(t) - \hat{\beta}(t)) \).

(iii) Compute the statistic of test of interest \( \bar{T}_{1n}^b = \max_{t \in \bar{T}} \sqrt{n} \sup_{t \in T} |V_n^b(t)| \).

Let \( \bar{Z}_n^b \) denote the empirical \((1 - \alpha)\)-quantile of the simulated sample \( \{\bar{T}_{1n}^1, \ldots, \bar{T}_{1n}^B\} \), where \( \alpha \in (0, 1) \) is the nominal size. We reject the null hypothesis if \( \bar{T}_{1n} \) is larger than \( \bar{Z}_n^B \). In practice, the maximum in Step (iii) is taken over a discretized subset of \( T \).

A formal justification the simulation method is stated as follows. Let \( Z = E[|m(Y; \beta(t)|\pi_0(u); t)] \), and consider the following conditions.

G4.B With \( P^* \)-probability tending to one, \( \hat{\pi}^* \in \Pi \) and \( ||\hat{\pi}^* - \pi||_p = o_P(n^{-\gamma}) \).

G5.B For any \( \delta_n \downarrow 0 \), \( \sup_{||\beta - \beta_0|| \leq \delta_n, \pi - \pi_0 \leq \delta_n} \| \sqrt{n} \sup_{t \in T} |Z_n(t) - Z_0(t)| \|_p = o_P(1) \).

G6.B If \( \sqrt{n} \sup_{t \in T} |Z_n(\beta, \pi) - \pi^*| + (Z_n^* - Z_0(\beta, \pi)) \) converges weakly to a tight random element \( \mathcal{G} \) in \( L \) in \( P^* \)-probability.

**Theorem 4.** Suppose \( \beta_0 \in \text{int}(B) \), the internal of the parameter space, and \( \hat{\beta} \to \beta_0 \). Assume that conditions G.I, G.IV , G.V, and G.VI are satisfied with “in probability” replaced by “almost surely.” Let G.II and G.III hold with \( \pi_0 \) replaced by \( \pi \in \Pi_{\beta_0} \). Also, assume that \( Z_b(\beta, \pi) \) is continuous in \( \pi \) at \( \beta = \beta_0 \) and \( \pi = \pi_0 \). Then, under G4.B–G6.B (with \( \theta \) replaced by \( \beta, h \) by
π, and \( m^1(\cdot) \) by \( m(\cdot) \), \( \sqrt{\hat{m}_n}(\hat{\beta} - \hat{\beta}) \sim Z_n^{-1}(\beta_0, \pi)G \) in \( P^* \)-probability.

**Proof.** This theorem is a restatement of the Lemma 3 in the supplemental appendix.

Theorem 4 establishes the consistency of the bootstrap procedure. It is important to highlight the connection between this result and the previous section. In fact, Theorem 4 shows that the limiting distribution of the bootstrap estimator is the same as that of Theorem 3, and hence the above resample scheme is able to mimic the asymptotic distribution of interest. Thus, computation of critical values and practical inference are feasible.

Now we provide examples of applications of the previous bootstrap result in Theorem 4 to the ADRF and QDRF cases. Consider the following conditions for ADRF.

**AG.IB** The parameter space for \( \mu_0 \) is a bounded sub-Banach space \( M \) of \( \ell^\infty(T) \) and \( \mu_n \xrightarrow{a.s.} \mu_0 \). In addition, \( |E[Y(t)^2]|_\infty < M_T \) for some \( 0 < M_T < \infty \).

**AG.IIB** For any \( \delta_n \downarrow 0 \), \( \sup_{|\pi - \pi_n|_n \leq \delta_n} \{ \frac{1}{n} \sum_{t=1}^{n} \pi_n(nu; t) - E[\pi_0(nu; t)] \} = o_p(1/\sqrt{n}) \).

**Cattaneo** \( \sqrt{\hat{m}_n} \frac{1}{n} \sum_{t=1}^{n} [(\tau - 1)(Y_t < q_0(t))]E[\pi_0(U_i; t) - \hat{\pi}(U_i; t)] \) converges weakly to a tight random element \( G \) in \( L \) in \( P^* \)-probability.

**Corollary 8 (Quantile).** Under QC.I–QC.II, AG.IB–AG.IIB, and G.IV with “in probability” replaced by “almost surely,” the bootstrap estimator of the QDRF is \( \sqrt{\hat{m}_n} \)-consistent and converges weakly in \( \ell^\infty(T) \).

**Proof.** It is an application of Theorem 4 and parallel to that of weak convergence of \( \hat{q}_i(t) \).

As a remark, note that given the above framework, inference on the TE is simple. For example, consider the inference of QTE from treatment levels \( t_1 \) to \( t_2 \). The point estimate is \( \hat{q}(t_2) - \hat{q}(t_1) \), which has an asymptotic normal distribution with mean \( q(t_2) - q(t_1) \), and its variance is computable from the covariance kernel of the weak limit of \( \hat{q} \). In a related work, He, Hsu, and Hu (2010) developed a testing procedure to test the hypothesis of no TE against a class of alternatives where the two outcome distributions differ only or mainly in the right tail.

# 5. MONTE CARLO

We conduct numerical experiments to assess the bias, size, and power properties of the proposed methods in finite samples. The data-generating process has treatment level \( t \in \{0,1\} \), with \( Y_i(t) = (1 + \gamma) + X_i + v_i(t) \), where the observed individual characteristics are \( X_i \sim \text{Unif}[-0.5, 0.5] \), and the independent innovations \( (v_i(0), v_i(1), \ldots, v_i(1)) \sim \text{N}(0, 1) \). The treatment assignment is generated by \( T_i = \text{argmax}_{t \in \{0,1\}} H_{i,t} \), where \( H_{i,t} = \sin(2\pi t)X_i + \epsilon_i(t) \) with independent innovations \( \epsilon_i(0), \epsilon_i(0.1), \ldots, \epsilon_i(1) \sim \text{N}(0, 1) \). Thus, a sample of \( n \) iid random elements \( (X_i, \epsilon_i(t), v_i(t)) \) whose components are mutually independent is generated. We set \( \gamma = 0 \) under the null hypothesis. We examine bias, and the empirical rejection frequencies for 5% nominal level tests for different choices of sample size \( n = \{250, 500, 1000\} \), number of bootstraps \( 250, 500 \), and parameter \( \gamma = \{0, 0.05, 0.10\} \). We use the Cramér–von Mises test for the simulations. We report results for both average and median estimates, and also parametric and nonparametric estimation of \( \pi_0 \), for completeness. In the former case, we use a normal distribution to estimate the densities. In the latter case, we employ Hall, Racine, and Li (2004), a nonparametric method to estimate conditional densities. The number of replications is 2000. Moreover, for comparison we report results for Cattaneo (2010) estimators. Although Cattaneo’s (2010) methods are

| \( B = 250 \) | \( B = 500 \) |
|---|---|---|
| \( \gamma = 0 \) | 0.051 | 0.056 |
| \( \gamma = 0.05 \) | 0.076 | 0.0973 |
| \( \gamma = 0.10 \) | 0.992 | 1.000 |
| 0.056 | 0.053 | 1.000 |
| 0.965 | 1.000 | 1.000 |
| 1.000 | 0.053 | 1.000 |
| 0.049 | 0.055 | 0.053 |
| 1.000 | 0.992 | 1.000 |
| 0.767 | 0.867 | 1.000 |
| 0.996 | 0.995 | 0.995 |
| 0.944 | 0.995 | 0.995 |
| 0.880 | 0.867 | 1.000 |
| 0.809 | 0.867 | 1.000 |
designed to categorical treatments, not continuous, in the experiments, we categorize the continuous treatment in three groups to study the effects of categorizing a continuous variable.

We first report results on bias in Table 1. The bias of the proposed estimators are defined as the supreme of the pointwise biases. As expected, the results show the bias for mean and median are approximately zero for both parametric and nonparametric estimation of the nuisance parameter in the first step. In addition, the bias decreases as sample size increases. On the other hand, Cattaneo’s (2010) methods show some bias for small samples.

Now we present the empirical size and power. Table 2 collects the results for mean and median, for the parametric estimation of the nuisance parameter, and Table 3 for the nonparametric estimation. For comparison, the results for Cattaneo (2010) are in Table 4.

In Table 2, we observe that the empirical sizes ($\gamma = 0$) are close to the nominal, 5%. The size improves with the sample size as well. The power is already high for small deviations from the null ($\gamma = 0.05$), and it increases with sample size, but it is not very sensitive to the number of bootstraps. The simulations show numerical evidence that, in practice, although one is able to use nonparametric estimation in the first step, parametric methods are preferable and perform very well in finite sample.

Finally, Table 4 collects the results for Cattaneo (2010) methods. The results show there are large size distortions when using Cattaneo (2010) to analyze continuous data. In addition, our proposed methods substantively outperform Cattaneo’s in terms of power. There is a significant loss of power when using a methodology that is not designed to handle continuous treatments. These results are in fact expected, since it is known from the literature that categorizing continuous treatments generally leads to a number of serious problems. Discretizing continuous outcomes leads to loss of power in testing, misclassification (which is associated with potential bias), problems for prediction, and even interpretation of the results and coefficients of interest (see, e.g., Cox 1957; Cohen 1983; van Belle 2008; Fedorov, Mannino, and Zhang 2009).

Overall the results suggest the proposed methods have good finite sample performance. Our main proposal, the uniform test, in addition to having a considerably better power than other methods, makes the bootstrap method a practical inference procedure.

### Table 3. Size and power: Nonparametric estimation of nuisance parameter

|         | $B = 250$ | $B = 500$ |
|---------|-----------|-----------|
| $\gamma = 0$ | $\gamma = 0.05$ | $\gamma = 0.10$ | $\gamma = 0$ | $\gamma = 0.05$ | $\gamma = 0.10$ |
| $n = 250$ | 0.021 | 0.410 | 0.973 | 0.025 | 0.446 | 0.988 |
| $n = 500$ | 0.023 | 0.772 | 1.000 | 0.031 | 0.800 | 1.000 |
| $n = 1000$ | 0.031 | 0.955 | 1.000 | 0.038 | 0.973 | 1.000 |
| A. Mean |           |           |           |           |           |           |
| $n = 250$ | 0.022 | 0.113 | 0.626 | 0.023 | 0.156 | 0.654 |
| $n = 500$ | 0.023 | 0.358 | 0.879 | 0.034 | 0.389 | 0.985 |
| $n = 1000$ | 0.033 | 0.711 | 1.000 | 0.037 | 0.729 | 1.000 |
| B. Median |           |           |           |           |           |           |

### Table 4. Size and power: Cattaneo

|         | $B = 250$ | $B = 500$ |
|---------|-----------|-----------|
| $\gamma = 0$ | $\gamma = 0.05$ | $\gamma = 0.10$ | $\gamma = 0$ | $\gamma = 0.05$ | $\gamma = 0.10$ |
| $n = 250$ | 0.101 | 0.111 | 0.284 | 0.150 | 0.125 | 0.427 |
| $n = 500$ | 0.120 | 0.129 | 0.625 | 0.101 | 0.113 | 0.705 |
| $n = 1000$ | 0.118 | 0.195 | 0.718 | 0.095 | 0.178 | 0.844 |
| A. Mean |           |           |           |           |           |           |
| $n = 250$ | 0.030 | 0.035 | 0.053 | 0.070 | 0.061 | 0.145 |
| $n = 500$ | 0.046 | 0.055 | 0.175 | 0.030 | 0.036 | 0.201 |
| $n = 1000$ | 0.036 | 0.088 | 0.201 | 0.079 | 0.087 | 0.345 |
| B. Median |           |           |           |           |           |           |
the 90%, 75%, 50%, 25%, and 10% quantiles of birthweights. The solid curve is the average and the dashed curves are the high and low birthweight, respectively. The horizontal lines represent the high and low birthweight.

6. APPLICATION: DOSE-BIRTHWEIGHT FUNCTIONS

We illustrate the use of the two-step estimator with a study of dose-birthweight functions. We estimate the unconditional average and quantile dose-birthweight functions, for both mothers’ weight gain during pregnancy as well as mothers’ age. Recently, birthweight has been shown to be the foremost indicator of infant health. In addition, unhealthy births have large economic costs in both immediate medical costs and longer care costs.

Infants are classified as low birthweight (LBW) when weighing less than 2.5 kilograms (kg) at birth. There is empirical evidence showing that LBW is associated with several health problems for the baby and the direct medical costs of LBW are quite high. Almond, Chay, and Lee (2005) documented that the birthweight has been shown to be the foremost indicator of infant health in addition, unhealthy births have large economic costs in both immediate medical costs and longer care costs.

Recent research also suggests giving birth to infants over 4.5 kg carries significant risks to both the infant and the mother; see, for example, Cesur and Kelly (2010). Fetal disorders such as shoulder dystocia, stillbirth, Erb’s palsy, jaundice, and respiratory distress have been found to be more common in HBW infants in addition to greater levels of obesity later.

6.1 Data

The data in this study are from the 2004 public use natality data of Wisconsin from the National Vital Statistics System of Centers for Disease Control and Prevention. We consider only live, singleton births (without missing values of any used characteristics) to new, white mothers that are not older than 45, with less than 5 years of college, whose counties of occurrence (birthing) and residence are the same. By using a more homogenous sample, we focus on the effects of birth inputs on the birthweights. This results in a sample of 13,581 births. We emphasize that the inferences using the sample should be applicable only for the subpopulation represented by the sample choice. Table 5 displays the descriptive statistics for birthweight (measured in kg), the mother’s age, the mother’s weight gain during pregnancy (WG), number of cigarettes per day (Cigarettes), number of prenatal care visits (No. care), and the mother’s years of education for the sample. Out of 13,581 births, there are 6508 females (proportion 0.4792), and 7732 mothers are married (proportion 0.5693).

6.2 Estimation of Nuisance Parameter \( \pi_0 \)

The estimation strategy of nuisance parameters \( \pi_0(u; 1) := \frac{f(u | x, y)}{f(x | y)} \) in (4) and (5) is the same for both the effects of the mother’s age and weight gain during pregnancy. We describe the details of the estimation procedure using the mother’s age, as an example.

The mothers’ ages in our sample range from 14 to 45 years old. Therefore, it is natural to treat the mother’s age as a continuous variable in the interval [13,46]. For the estimation of...
conditional density, we assume that \( \log(T_{-13}) = X^\top \theta_0 + \epsilon_i \), where \( \epsilon_i \) is independent of \( X_i \) and has density \( N(0, \sigma^2_0) \). The choice of \( X \) is discussed below. The log-ratio form of the dependent variable makes mothers’ age to be limited to [13, 46]. Hence, \( T_{-13} = -u_i \) follows log-normal as \( \log(N(X_i \theta_0, \sigma^2_0)) \). The density of \( T \mid X \) is obtained by calculating the distribution function, as \( F_{0T \mid X}(t\mid x) = F_{0U}(\frac{\log(\frac{t + u_i - x \theta_0}{\theta_0})}{\sigma_0}) \) and \( f_{0T \mid X}(t\mid x) = \phi(\log(\frac{t + u_i - x \theta_0}{\theta_0})) \frac{1}{\sigma_0} \frac{1}{\phi(\log(\frac{t + u_i - x \theta_0}{\theta_0}))} \), where \( \phi \) and \( \Phi \) are the distribution and density of a standard normal. For the conditional density of \( T \mid X, Y \), we assume that \( \log(f_{-13}) = U^\top \theta_0 + \epsilon_i \), where \( \epsilon_i \) is independent of \( U_i \) and has density \( N(0, \xi_0) \). Thus, the conditional density \( f_{0T \mid U}(u\mid t) = \phi(\frac{\log(\frac{t - u \theta_0}{\theta_0})}{\xi_0}) \frac{1}{\phi(\log(\frac{t - u \theta_0}{\theta_0}))} \), and \( f_0(u, t) = f_{0T \mid X}(t\mid x) f_{0T \mid U}(u\mid t) \).

6.3 Empirical Results

6.3.1 Mother’s Weight Gain During Pregnancy. In the study of the mother’s weight gain and age, we control for characteristics of mothers, that is, marital status, years of education, and number of cigarettes per day during pregnancy. It is important to note that although we are controlling for some characteristics of mothers, we are estimating the unconditional average of birthweight. In particular, the impact is very prominent for ages from 14 to 45. The impact of the age on birthweight is negative for all the quantiles as well. However, for a given age, the impact becomes more severe for lower parts of the distribution of birthweight. In particular, the impact is very prominent for the 10% quantile of mothers after 40 years old. The estimated DRF. The results regarding the mother’s weight gain during pregnancy show evidence that, after controlling for a mother’s characteristics, in general, larger weight gain during pregnancy leads to higher birthweight. Just as the estimation of the unconditional treatment effects (TE) of the mother’s age, we are estimating the unconditional TE of the mother’s weight gain during pregnancy. Figure 1 reports the estimates of the average and selected quantiles of the birthweight for different levels of the mother’s weight gain during pregnancy. From the figure, we see that the slopes are relatively larger for low or high weight gain. The shape of the curves resembles a simple cubic function with steeper slopes at the extremes. This implies weight gaining generates higher birthweights at low and high levels of weight gain. For low weight gains, the impact on the birthweight is higher for upper quantiles and relatively mild for low quantiles. However, for the middle range of weight gain, all the curves are relatively parallel. The disaggregated plots with 90% confidence bands are shown in Figure 2. The confidence bands in general are relatively wider at the extremes of weight gain due to the sparsity of the data at the extremes. In addition, Table 6 describes TE for selected weight. It contains 20 pound effects. The results show that the impact of gaining weight is positive and statistically significant.

6.3.2 Mother’s Age. The empirical results display negative effects of the mother’s age on birthweight, after controlling for a mother’s characteristics (i.e., age, marital status, years of education, number of cigarettes per day, and the month of first prenatal care visit). As expected, the birthweight decreases as age increases. Figure 3 plots the point estimates for the mean, 10%, 90%, and the three quartiles of birthweights for mothers’ ages from 14 to 45. The impact of the age on birthweight is negative for all the quantiles as well. However, for a given age, the impact becomes more severe for lower parts of the distribution of birthweight. In particular, the impact is very prominent for the 10% quantile of mothers after 40 years old. The estimated

| Mother’s Weight Gain During Pregnancy | 10% Qt. | 25% Qt. | 50% Qt. | 75% Qt. | 90% Qt. | Average |
|--------------------------------------|--------|--------|--------|--------|--------|---------|
| 0–20                                 | 2.23   | 2.52   | 2.68   | 2.77   | 2.72   | 2.53    |
| SD                                   | 0.05   | 0.07   | 0.05   | 0.08   | 0.07   | 0.05    |
| 20–40                                | 0.37   | 0.28   | 0.23   | 0.25   | 0.25   | 0.30    |
| SD                                   | 0.04   | 0.07   | 0.06   | 0.08   | 0.08   | 0.06    |
| 40–60                                | 0.21   | 0.20   | 0.20   | 0.20   | 0.23   | 0.22    |
| SD                                   | 0.05   | 0.07   | 0.06   | 0.09   | 0.09   | 0.07    |
| 60–80                                | 0.25   | 0.25   | 0.28   | 0.31   | 0.37   | 0.30    |
| SD                                   | 0.05   | 0.08   | 0.07   | 0.10   | 0.11   | 0.08    |

Table 6. Selected treatment effects of the mother’s weight gain during pregnancy.
average birthweight is downward sloping, and more negative at high ages, which is different from the median and is probably capturing the effect of the low quantile. On the other hand, the median birthweight is robust to this feature. Although intuitive, these results complement the existing results in the literature with three advantages. First, our results can be interpreted as causal effects. Second, we estimate the unconditional quantile and mean of the birthweight for a range of the mother’s age. Third, unlike the regression framework, our results show that the confidence interval for 10% quantile at the age of 45 is relatively wide. Table 7 describes the TE for selected ages, containing 5 year effects. Most of them are statistically significant, and negative values show evidence that aging is negatively related to birthweight.

The results also show evidence of certain risks for birthweight when the mother’s age is at the sample boundary. Although on average the birthweight is within the “healthy range,” between 2.5 and 4 kg, estimates show that mothers younger than 20 years are likely to have LBW infants, while mothers older than 44 years are likely to have LBW infants.

7. CONCLUSION

In this article, we first study the identification of the DRF with continuous treatment levels. In empirical studies, we usually have observational data. Agents can choose the levels of treatment they desire. Under the ignorability assumption, we derive moment conditions that are identified by observational data. Based on the moment conditions, we propose a two-step estimator. Throughout the article, we use the important examples of average and quantile DRF as applications of the general theory. Sufficient conditions are provided for establishing the asymptotic properties of the estimator. We develop uniform inference for the two-step estimator. More specifically, we are interested in testing the null hypothesis that a DRF \( \beta(t) = r(t) \) with \( t \in T \) for known or unknown \( r(t) \). Because the parameters are of infinite dimension and the weak limits of test statistics are not standard, we use a bootstrap method when conducting practical inference. Finally, we apply the methods to the study of unconditional effects of mother’s age and weight gain during pregnancy on infants’ birthweight, illustrating the usefulness of the new approach. An important topic for future research includes further analysis of potential nonparametric estimation in the first-step for the continuous treatment effects models.

APPENDIX: PROOFS

This appendix collects the proofs of the results given in the text.

Proof. Fixing \( t = t_0 \), by law of iterated expectations, 
\[ E[m(Y(t_0); \beta(t_0))] = E[E[m(Y(t_0); \beta(t_0))|X]]. \]
For the conditional expectation, 
\[ E[m(Y(t_0); \beta(t_0))|X, T \in [t_0, t_0 + \Delta t]] = E[I(T \in [t_0, t_0 + \Delta t]) m(Y(t_0); \beta(t_0))|X], \]
the right-hand side of Equation (A.1) equals 
\[ E[I(T \in [t_0, t_0 + \Delta t]) m(Y(t_0); \beta(t_0))|X] = \frac{E(I(T \leq t_0 + \Delta t) - 1[T < t_0]) m(Y; \beta(t_0))|X)}{F_{T,X}(t_0 + \Delta t|x) - F_{T,X}(t_0|x)}, \]
where \( F_{T,X} \) denotes the conditional distribution function of \( T \) given \( X \). Noting that 
\[ E[I(T \leq t_0 + \Delta t) - 1[T \leq t_0]) m(Y; \beta(t_0))|X = x] = \int (1[t \leq t_0 + \Delta t] - 1[t \leq t_0]) m(y; \beta(t_0)) f_{T,Y|X}(t, y|x) dt dy = \int_{t_0}^{t_0 + \Delta t} m(y; \beta(t_0)) f_{T,Y|X}(t, y|x) dt dy, \]
it follows that 
\[ \lim_{\Delta t \to 0} E[m(Y; \beta(t_0))|X = x, T \in [t_0, t_0 + \Delta t]] = \lim_{\Delta t \to 0} \int_{t_0}^{t_0 + \Delta t} m(y; \beta(t_0)) f_{T,Y|X}(t, y|x) dt dy \]
\[ = \lim_{\Delta t \to 0} \int_{t_0}^{t_0 + \Delta t} \frac{f_{T,Y|X}(t + \Delta t|x) - f_{T,Y|X}(t|x)}{f_{T,Y|X}(t|x)} \]
\[ = \lim_{\Delta t \to 0} \int m(y; \beta(t_0)) f_{T,Y|X}(t, y|x) \frac{f_{T,Y|X}(t + \Delta t|x) - f_{T,Y|X}(t|x)}{f_{T,Y|X}(t|x)} dy \]
\[ = \lim_{\Delta t \to 0} \int m(y; \beta(t_0)) f_{T,Y|X}(t, y|x) \frac{f_{T,Y|X}(t + \Delta t|x) - f_{T,Y|X}(t|x)}{f_{T,Y|X}(t|x)} dy \]
\[ = \lim_{\Delta t \to 0} \int m(y; \beta(t_0)) f_{T,Y|X}(t, y|x) f_{T,Y|X}(t|x) dy \]
\[ = E[m(Y; \beta(t_0)) f_{T,Y|X}(t, y|x) f_{T,Y|X}(t|x)|X = x]] \]
where $\epsilon_1$ and $\epsilon_2$ are fixed numbers in $[0, 1]$. The second equality is by mean value theorems for differentiation and integration. And the third equality is by condition III.1 and dominated convergence theorem. Hence, $E[m(Y(t_0); \beta(t_0))] = E[m(Y; \beta(t_0))]_{Z(t_0) = \beta(t_0)}$.

To demonstrate Theorems 2, 3, and 4, we make use of Lemmas 1, 2, and 3, respectively, given in the online supplemental appendix. These lemmas establish, respectively, uniform consistency, weak convergence, and validity of the bootstrap for generic Z-estimators with possibly nonsmooth functions and a nuisance parameter, when both the parameter of interest and the nuisance parameter are possibly infinitely dimensional. The results allow for the case of profiled nonparametric estimator, that is, it depends on the parameters.

**Proof of Theorem 2.** The general result for consistency of Z-estimator is given in Lemma 1 in the online supplemental appendix. To prove the result we apply the lemma to our continuous treatment model with $\theta_0 = \beta_0(\cdot)$, $h_0 = \pi(\cdot)$, $Z(\theta, h)(t) = E\psi_{\theta, h}(t)$, and $\hat{Z}(\theta, h)(t) = E\psi_{\hat{\theta}, \hat{h}}(t)$, where $\psi_{\theta, h}(t) = m(y; \beta(t))\pi(u(t))$. In this case, $\hat{Z} = \hat{Z} = \hat{Z}$, and $E\|\hat{Z} - Z\|_2 = \sup_{t \in \mathcal{T}} \sup_{u(t) \in \mathcal{U}} \|\hat{Z}(\theta, h)(t) - Z(\theta, h)(t)\|$. For any $\delta > 0$, $\Pi_{\delta} = \{\pi \in \Pi : |\pi - \pi_0|_2 < \delta\}$.

First, we show that condition C.3 (online supplemental appendix), the continuity of $E[m(Y; \beta(t))\pi(U; t)]$ at $\pi_0$ uniformly in $\beta(t)$ in $\mathcal{E} = \mathcal{E}$, is satisfied. For any $|\pi - \pi_0|_2 \leq \delta$, which is equivalent to $\sup_{t \in \mathcal{T}} \sup_{\pi(U; t) \in \mathcal{U}} |\pi(U; t) - \pi_0(U; t)| \leq \delta$, we have

$$\sup_{t \in \mathcal{T}} \sup_{\pi(U; t) \in \mathcal{U}} \left| E[m(Y; \beta(t))\pi(U; t)] - E[m(Y; \beta(t))\pi_0(U; t)] \right| \leq \sup_{t \in \mathcal{T}} \left| E[m(Y; \beta(t))] - E[m(Y; \beta(t))] \right|.$$

Therefore, condition C.3 is implied by condition C.III.

Condition C.IV and Corollary 9.27 of Kosorok (2008) imply condition CSS, which, in turn implies C.5.

**Proof of Corollary 1. Consistency of $\hat{\mu}(\cdot)$.** To show consistency, we verify the conditions of Theorem 2. Note that in this case we have a closed form solution $\mu(t) = E[Y - E(Y|U(t))] = E[Y - E(Y|U(t))]$. Condition C.I is verified by the fact that $\hat{\mu} = E[Y - E(Y|U(t))]$ is the exact zero of the estimating equation. For condition C.II, if there is a sequence $\mu_n(\cdot)$ such that $E[Y - \mu_n(\cdot)]\pi_0(U; t) = 0$, then $\mu_n(\cdot) \rightarrow \mu_0(\cdot)$ uniformly. To see this, note that $E[Y - \mu_n(\cdot)]\pi_0(U; t) = (o(1))$ implies $\mu_n(\cdot) \rightarrow \mu_0(\cdot)$ uniformly. Hence, condition C.III is verified by the direct calculation

$$E[Y - \mu(\cdot)] = E[Y - \mu(\cdot)] = E[Y] + |\mu(\cdot)|,$$

both of which are finite by condition AC.I. As for condition C.V, note that $\psi_{\beta, t} = y - \mu(t)$, automatically $\psi_{\beta, t} : \beta \in \mathcal{E} = \mathcal{E}$ is Glivenko-Cantelli. To see this, note

$$\sup_{t \in \mathcal{T}} \left| E[Y - \mu(t)] - E[Y - \mu(t)] \right| = E[Y] - E[Y] \rightarrow 0,$$

by Khintchine’s law of large numbers. It has envelope $F_t(y) = y + |\mu(t)|$ by condition AC.I. Condition AC.II completes the verification of condition C.V. Hence, all the conditions of Theorem 2 are satisfied.

**Proof of Corollary 2. Consistency of $\hat{\sigma}_2(\cdot)$.** To establish the result we verify the conditions of Theorem 2. Condition C.I is satisfied by the computational properties of quantile regression estimator of Theorem 3.3 of Koenker and Bassett (1978) and conditions C.IV and C.Q.II such that we have

$$\left| E[(\tau - 1)(y < q_2(t))]\hat{\sigma}_2(U; t) \right| \leq \text{const} \sup_{t \in \mathcal{T}} \sup_{u(t) \in \mathcal{U}} |\hat{Z}(\theta, h)(t) - Z(\theta, h)(t)| = O_p(1/n),$$

Condition C.II holds by condition QC.I. Condition C.III is satisfied because $\tau - 1(y < q_2(t))$ is a bounded function. Condition C.V is implied by the fact that the function class $\{\psi_{\beta, t} : \beta \in \mathcal{E} = \mathcal{E}\} = \text{Glivenko-Cantelli}$ because it is a Vapnik-Cervonenkis class and by condition QC.III. Hence, all the conditions of Theorem 2 are satisfied.

**Proof of Theorem 3.** We apply Lemma 2 in the online supplemental appendix. We first verify condition G.3. To find the pathwise derivative of $Z(\beta, \pi_0)$ with respect to $\pi$, we conduct the following calculations. For any $\pi$ such that $|\pi_0 + \alpha(\pi - \pi_0)|_2 \leq 0$, $\alpha \in [0, 1] \subset \Pi$, and $E[m(Y; \beta)(\pi_0 + \alpha(\pi - \pi_0))] = E[m(Y; \beta)\pi_0]$

$$E[m(Y; \beta)\pi_0] = E[m(Y; \beta)(\pi - \pi_0)].$$

and has the limit $E[m(Y; \beta)(\pi - \pi_0)]$ as $\alpha \rightarrow 0$. Therefore $Z(\beta, \pi_0)[\pi - \pi_0] = E[m(Y; \beta)(\pi - \pi_0)]$ in all directions $[\pi - \pi_0] \in \Pi$. Condition G.3.1 is satisfied by noting that

$$E[m(Y; \beta)(\pi_0(U; t))] - E[m(Y; \beta)(\pi(U; t))] \leq (E[m(Y; \beta)(\pi(U; t))] - E[m(Y; \beta)(\pi(U; t))]|_{\Pi = \Pi_{\delta}} = 0.$$
Condition G.III is satisfied because the distribution function of $Y$ is continuous. Condition G.V was verified in the proof of Corollary 2. Condition G.VI holds by G.G.II. Hence, all the conditions of Theorem 3 are satisfied.

Proof of Proposition 1. Verification of conditions C.IV and G.IV are standard, as stated on p. 1599 of Chen, Linton, and Van Keilegom (2003). Note that conditions C.V and G.V are satisfied by condition NP.III.

Proof of Proposition 2. By the mean value theorem, $\pi(u; \cdot \theta) = \pi(u; \cdot \theta^*) + \pi'(u; \cdot \theta^*) (\hat{\theta} - \theta)$, where $\theta^*$ is a convex combination of $\theta$ and $\theta_0$. Therefore,

$$|\pi(u; \cdot \theta) - \pi(u; \cdot \theta_0)| = |\pi'(u; \cdot \theta^*)||\hat{\theta} - \theta_0|| = O_p(n^{-1/2})$$

since $|\pi'(u; \cdot \theta^*)|_\infty$ is bounded and $|\hat{\theta} - \theta_0|| = O_p(n^{-1/2})$. There, conditions C.IV and G.IV are verified.

For condition C.V, by Theorem 19.7 of van der Vaart (1998), this verification of conditions C.IV and G.IV is verified.

To verify the weak convergence of condition C.VI, we need to use the functional delta method, which involves Hadamard differentiability of a map between norm spaces. A map $\phi: \mathbb{D} \mapsto \mathbb{E}$ is Hadamard differentiable at $\theta$ if there exists a continuous linear map $\phi': \mathbb{D} \mapsto \mathbb{E}$ such that $\lim_{n \to \infty} \phi'(h_n) = \phi(h)$, as $n \to \infty$, for all converging sequences $h_n \to h$ in $\mathbb{D}$, with $h_n \in \mathbb{D}$ and $\theta + h_n \in \mathbb{D}$ for all $n \geq 1$ sufficiently large; see p. 22 of Kosorok (2008). We first verify that $\eta$ is Hadamard differentiable at $\theta$ tangentially to $\mathbb{D}$. For any $h_n \to 0$ and $h_n \to h \in \mathbb{D}$,

$$\frac{\eta(h + l_n h_n) - \eta(h)}{l_n} = \frac{\pi(u; \cdot \theta + l_n h_n) - \pi(u; \cdot \theta)}{l_n} = \frac{\pi'(u; \cdot \theta^*) l_n h_n}{l_n} \to \pi'(u; \cdot \theta) h.$$ 

Using the functional delta method, since $\pi(u; \cdot \theta)$ is uniformly bounded,

$$\sqrt{n} (\pi(u; \cdot \hat{\theta}) - \pi(u; \cdot \theta)) \sim \pi'(u; \cdot \theta) Z_\delta \text{ in } \ell^\infty(U \times T),$$

where $Z_\delta \sim N(0, \delta^{-1})$.

**SUPPLEMENTARY MATERIALS**

The supplementary appendix presents results for the asymptotic theory for the generic $Z$-estimator, and the semiparametric efficiency of the two-step estimator.

[Received October 2013. Revised August 2014.]

**REFERENCES**

Abadie, A., and Imbens, G. (2006), “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, 74, 235–267. [1528]

Almond, D., Chay, K. Y., and Lee, D. S. (2005), “The Costs of Low Birth Weight,” *Quarterly Journal of Economics*, 120, 1031–1083. [1537]

Angrist, J., Chernozhukov, V., and Fernandez-Val, I. (2006), “Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure,” *Econometrica*, 74, 539–563. [1532]

Cattaneo, M. (2010), “Efficient Semiparametric Estimation of Multi-Valued Treatment Effects under Ignorability,” *Journal of Econometrics*, 155, 138–154. [1528,1530,1531,1532,1535,1536]

Cesar, R., and Kelly, I. (2010), “From Cradle to Classroom: High Birth Weight and Cognitive Outcomes,” *Health Economics*, 13, 1–24. [1537]

Chen, X., Linton, O., and Van Keilegom, I. (2003), “Estimation of Semiparametric Models When the Criterion Function is not Smooth,” *Econometrica*, 71, 1591–1608. [1531,1532,1541]

Chernozhukov, V., and Hong, H. (2002), “Three-Step Censored Quantile Regression and Extramartial Affairs,” *Journal of the American Statistical Association*, 97, 872–882. [1533]

Cohen, J. (1983), “The Cost of Dichotomization,” *Applied Psychological Measurement*, 7, 249–253. [1536]

Cox, D. R. (1957), “Note on Grouping,” *Journal of the American Statistical Association*, 52, 543–547. [1536]

Dehejia, R., and Wahba, S. (1999), “Causal Effects in Nonexperimental Studies: Evaluating the Evaluation of Training Programs,” *Journal of the American Statistical Association*, 94, 1053–1062. [1528,1530]

Fedorov, V., Mannino, F., and Zhang, R. (2009), “Consequences of Dichotomization,” *Pharmaceutical Statistics*, 8, 50–61. [1536]

Firpo, S. (2007), “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, 75, 259–276. [1528,1530,1532]

Flores, J. P., Heckman, J. J., Meghir, C., and Vytlacil, E. (2008), “Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects,” *Econometrica*, 76, 1191–1206. [1528]

Flores, C. A. (2007), “Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment,” *Mimeo, Department of Economics, University of Miami*. [1528]

Gine, E., and Nickl, R. (2008), “Uniform Central Limit Theorems for Kernel Density Estimators,” *Probability Theory and Related Fields*, 141, 333–387. [1533]

Hahn, J. (1998), “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects,” *Econometrica*, 66, 315–331. [1528]

Hall, P., Racine, J., and Li, Q. (2004), “Cross-Validation and the Estimation of Conditional Probability Densities,” *Journal of the American Statistical Association*, 99, 1015–1026. [1535]

He, X., Hsu, Y.-H., and Hu, M. (2010), “Detection of Treatment Effects by Covariate-Adjusted Expected Shortfall,” *Annals of Applied Statistics*, 4, 2114–2125. [1535]

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998), “Characterizing Selection Bias Using Experimental Data,” *Econometrica*, 66, 1017–1098. [1528,1530]

Hirano, K., Imbens, G., and Ridder, G. (2003), “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, 71, 1161–1189. [1528,1531,1532]

Hirano, K., and Imbens, G. W. (2004), “The Propensity Score with Continuous Treatment,” in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, eds. A. Gelman and X.-L. Meng, West Sussex, UK: Wiley. [1528,1530,1533]

Imai, K., and van Dyk, D. (2004), “Causal Inference with General Treatment Regimes: Generalizing the Propensity Score,” *Journal of the American Statistical Association*, 99, 854–866. [1528]

Imbens, G. (2000), “The Role of the Propensity Score in Estimating Dose-Response Function,” *Biometrika*, 87, 706–710. [1528]

Imbens, G., Newey, W., and Ridder, D. (2006), “Mean-Squared-Squared Error Calculations for Average Treatment Effects” *Mimeo, Department of Economics, Harvard University*. [1528]

Koenker, R., and Bassett, G. W. (1978), “Regression Quantiles,” *Econometrica*, 46, 33–49. [1531,1540]

Kosorok, M. (2008), *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer. [1540,1541]

Lechner, M. (2001), “Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption,” in *Econometric Evaluation of Labour Market Policies*, eds. M. Lechner and F. Pfeiffer, Heidelberg: Physica Springer. [1528]

Lee, Y. (2012), “Partial Mean Processes with Generated Regressors: Continuous Treatment Effects and Nonseparable Models,” *Mimeo, Department of Economics, University of Oxford*. [1528]

Li, Q., Racine, J., and Wooldridge, J. (2009), “Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data,” *Journal of Business and Economic Statistics*, 27, 206–222. [1528]

Murphy, K., and Topel, R. (1985), “Estimation and Inference in Two-Step Econometric Models,” *Journal of Business and Economic Statistics*, 3, 88–97. [1533]

Newey, W., and McFadden, D. L. (1994), “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics* (Vol. 4), eds. R. F. Engle and D. L. McFadden, Amsterdam: North Holland, Elsevier. [1531,1533]

Pakes, A., and Pollard, D. (1989), “Simulations and the Asymptotics of Optimization Estimators,” *Econometrica*, 57, 1027–1057. [1531,1532]
Rosenbaum, P., and Rubin, D. (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41–55. [1528]
Rubin, D. (1977), “Assignment to Treatment Group on the Basis of a Covariate,” *Journal of Educational Statistics*, 2, 1–28. [1528]
van Belle, G. (2008), *Statistical Rules of Thumb* (2nd ed.), Hoboken, NJ: Wiley-Interscience. [1536]

van der Vaart, A. (1998), *Asymptotic Statistics*. New York: Cambridge University Press. [1531,1541]
van der Vaart, A., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer-Verlag Press. [1540]
Wei, Y., and Carroll, R. J. (2009), “Quantile Regression with Measurement Error,” *Journal of the American Statistical Association*, 104, 1129–1143. [1533]