Uniformly Semiparametric Efficient Estimation of Treatment Effects with a Continuous Treatment*

Antonio F. Galvao† Liang Wang‡

October 15, 2013

Abstract

This paper 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 which are possibly non-smooth. 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.

Keywords: Dose-response function, continuous treatment, ignorability, semiparametric estimation and inference, birthweight

*The authors would like to express their appreciation to Carlos Flores, Alfonso Flores-Lagunes, Zhengyuan Gao, Chuan Goh, Roger Koenker, Carlos Lamarche, Ying-Ying Lee, Cristine Pinto, Alexandre Poirier, Suyong Song, Zhijie Xiao, Ting Zhang, and participants in the seminars at the University of Iowa, University of Illinois at Urbana-Champaign, Central Bank of Brazil, and University of Wisconsin-Milwaukee for useful comments and discussions regarding this paper. Computer programs to replicate the numerical analyses are available from the authors. All the remaining errors are ours.

†Department of Economics, University of Iowa, W210 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242. E-mail: antonio-galvao@uiowa.edu

‡Department of Economics, University of Wisconsin-Milwaukee, 2025 E. Newport Ave., Milwaukee, WI 53201, E-mail: wang42@uwm.edu
1 Introduction

There is a large and 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 studies in economics and statistics, where these methods help to analyze how treatments or social programs affect the outcome distributions of interest.

This paper studies identification, estimation, and inference of general unconditional treatment effect (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 non-smooth functions. In this general formulation, the range of moment models for which the methods are applicable is very broad, for instance, the framework includes average and quantile DRF as special cases, and consequently average treatment effects (ATE) and quantile treatment effects (QTE) are direct applications of the methods developed in the paper. QTE are especially important in empirical research because they uncover the effects of policy variables on the entire distribution of outcomes.

Consistent estimation requires identification of the parameters of interest. In this paper, 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 TE literature, see e.g. Rubin (1977), Barnow, Cain, and Goldberger (1980), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), Firpo (2007), Flores (2007), Angrist and Pischke (2009), and Cattaneo (2010). The moment condition model together with the selection on observables allow identification of the parameters of interest.

Based on the identification condition, we construct a two-step estimation procedure. The implementation of the estimator in practice is simple. In the first step, one estimates a ratio of conditional distributions, similar to a propensity score. In the second step, an optimization problem is solved. It is important to notice that, once the identification is achieved and a DRF is estimated, other parameters of interest based on these functions can be estimated with little additional effort. For example, one can easily estimate TE, which are defined as differences of the DRF evaluated at different levels of treatment. In addition, one could
estimate the entire curve of potential outcomes.

Mild sufficient conditions are provided for the two-step estimator to have desired asymptotic properties, namely, consistency, weak convergence, and semiparametric efficiency. In particular, we show that the two-step estimator of the DRF is uniformly consistent over the set of treatment. Differently from the binary or multi-valued treatment models, in which case pointwise results are equivalent to uniform results (because the numbers of treatment levels are finite), when treatment levels are an interval $T$, the uniform results are stronger than pointwise results, and consequently, only pointwise results are often not adequate for inference. In addition, we show that the estimator converges weakly to a Gaussian process, and that it is uniformly semiparametric efficient. For the latter result we use the method of Bickel, Klaassen, Ritov, and Wellner (1993). From a technical point of view, the derivations of the limiting properties of the proposed estimator are of independent interest.

In addition to identification and estimation, we develop statistical inference procedures. In particular, we conduct inference on a DRF uniformly over the treatment levels. We propose testing procedures for the hypothesis of the equality of the DRF and any given function. The test statistics used are Kolmogorov and Cramér-von Mises types, which detect deviations from the null hypothesis. Since the parameter of interest is infinite dimensional and the weak limit of these statistics are not standard, we compute critical values using a bootstrap method. We provide sufficient conditions under which the bootstrap is valid, and discuss an algorithm for practical implementation. The proof of the validity of the bootstrap also extends that of Chen, Linton, and Van Keilegom (2003).

We conduct Monte Carlo simulations to evaluate the finite sample performance of the proposed methods. The simulations show that the estimators are approximately unbiased and the Cramér-von Mises type test statistic has good empirical size and high power against a few alternatives. In addition, the result is improved when the sample size increases, and is not sensitive to the selected numbers of bootstrap.

To illustrate the proposed methods, we consider an empirical application to a birthweight study using the National Vital Statistics System of Centers for Disease Control and Prevention. We estimate the unconditional average and quantile dose-birthweight functions for both mothers’ weight gain during pregnancy as well as mother’s age. The empirical results document important heterogeneity in the dose-birthweight functions for weight gain and age across quantiles. The estimates of the dose-birthweight functions regarding the mother’s weight gain during pregnancy reveal interesting heterogeneity across quantiles and weight
The findings provide evidence that, in general, more weight gain during pregnancy leads to higher birthweight. However, the treatment effects differ at different levels of weight gain. For a given quantile of interest, the positive impacts are larger for low and high weight gains while relatively lower in the middle range of weight gain. The quantile dose-birthweight functions of the mother’s age on birthweight is downward-sloping. In particular, the impact of the mother’s age on birthweight remains negative, for all quantiles, as age increases. In addition, for a given age, this impact becomes more severe for lower parts of the distribution of birthweight. Although intuitive, these results complement the existing results in the literature.¹

1.1 Literature and outline

There has been a large and growing literature on unconditional TE, most of which focuses on models with discrete (usually binary) treatment levels. Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), and Imbens, Newey, and Ridder (2006) study efficient estimation of the average treatment effects (ATE) nonparametrically. To estimate the ATE, Hirano, Imbens, and Ridder (2003) estimate the propensity score nonparametrically in the first step while Abadie and Imbens (2006) applies matching methods. In addition, Li, Racine, and Wooldridge (2009) propose efficient estimation of ATE with mixed categorical and continuous data. The study of unconditional ATE has been extended to the quantile framework by Firpo (2007) with a two-step estimator that is semiparametric efficient for estimation of quantile treatment effects (QTE). This method, as that of Hirano, Imbens, and Ridder (2003), is based on nonparametric estimation of propensity score in the first step. There is also literature on multi-valued treatment effect models. Imbens (2000) shows that the multi-valued counterpart of the propensity score theorem of Rosenbaum and Rubin (1983) still holds. Imbens (2000) and Lechner (2001) discuss the unconditional ATE. Cattaneo (2010) extends the literature and proposes semiparametric efficient estimation of a family of multi-valued DRF which are implicitly defined by sets of possibly over-identified non-smooth moment conditions under the ignorability condition. The TE include mean and quantile as special cases.

¹Previous quantile estimation approaches to estimating birthweight outcome regressions have used reduced form models and, therefore, cannot be interpreted as causal effects. For instance, Abrevaya (2001) (see also Koenker and Hallock (2001) and Chernozhukov and Fernandez-Val (2011)) used federal natality data and found that various observables have significantly stronger associations with birthweight at lower quantiles of the birthweight distribution.
However, the literature on the study of continuous TE is relatively sparse. Among others, Hirano and Imbens (2004) and Imai and van Dyk (2004) develop the generalized propensity score for continuous average treatment models, and Flores (2007) develops nonparametric estimators for average dose-response functions (ADRF), its maximizer, and its global maximum under the ignorability assumption. Also, Florens, Heckman, Meghir, and Vytlacil (2008) consider the identification of ATE using control functions. More recently, Lee (2012) studies unconditional distribution of potential outcomes with continuous treatments as a partial mean process with generated regressors. Despite this sparsity, many questions of interest in applied research involve continuous treatments. For example, in the study of TE of mothers’ weight gain during pregnancy as well as mother’s age on birthweight, the weight gain in pounds and age are continuous variables.

This paper contributes to the existing TE literature by studying continuous treatments considering general forms of dose response for TE models, which include both ATE and QTE as special cases. Thus, this paper extends the literature on ATE and QTE for binary and multi-value doses of treatment (see e.g. Heckman and Vytlacil (2005), Firpo (2007), Cattaneo (2010)) to a continuous dose of treatment. We point out that the extension from the discrete to continuous treatment levels is non-trivial. In fact, since the parameters of interest are now infinite dimensional, the results need to be uniform on the set of treatment levels. In addition, we extend the literature on continuous treatments, which, to our knowledge, only allows for ADRF and ATE (see e.g. Hirano and Imbens (2004) and Flores (2007)), to general (possibly non-smooth) DRF, with quantile dose-response functions (QDRF) and QTE being important examples. This is an important innovation because the extension to the non-smooth cases are important in practice and technically challenging.

The remaining of the paper is organized as follows. Section 2 provides the identification conditions of the continuous treatment model and proposes a two-step estimator. Section 3 studies the asymptotic properties of the two-step estimator. Section 4 provides Monte Carlo simulation results and Section 5 illustrates the methods with an application to the estimation of dose-birthweight functions. Section 6 concludes the paper. The proofs of the main results are collected in the Appendix.

**Notations:** Let E and \( \mathbb{E} \) denote the expectation and sample average, respectively. Let \( \rightsquigarrow \), \( p \rightarrow \), and \( p^* \rightarrow \) denote weak convergence, convergence in probability, and convergence in outer probability, respectively. Finally, let \( |g(x)|_\infty \) denote \( \sup_x |g(x)| \) for \( x \in \mathcal{X} \subset \mathbb{R}^d \), where \( d \) is a positive integer.
The model, identification, and estimation

In this paper, we assume that a random sample of size $n$ is available. The objective is to learn how an outcome variable of an agent 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 variable 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, one cannot observe the random process $Y(t)$ 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 $T$. Therefore, the observed outcome is the random variable

$$Y = Y(T) = \int_{t \in T} Y(t) \, d1\{t \geq T\},$$

where $1\{\cdot\}$ is the indicator function.

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 be introduced by direct use of the sample counterparts to calculate treatment effects. To illustrate this point, we consider the estimation of average treatment effects as an example. For any $t_1 < t < t_2$, since

$$E[Y|T = t_2] - E[Y|T = t_1] = \frac{E[Y(t_2) - Y(t_1)|T = t]}{\text{Average treatment effect on the treated}} + \frac{E[Y(t_2)|T = t_2] - E[Y(t_2)|T = t]}{\text{Selection bias 1}}$$

$$+ \frac{E[Y(t_1)|T = t] - E[Y(t_1)|T = t_1]}{\text{Selection bias 2}},$$

we have

$$E[Y|T = t_2] - E[Y|T = t_1] = \frac{E[Y(t_2) - Y(t_1)\mid T = t]}{\text{Average treatment effect on the treated}} + \frac{E_t[E[Y(t_2)|T = t_2] - E[Y(t_2)|T = t]]}{\text{Average of selection bias 1}}$$

$$+ \frac{E_t[E[Y(t_1)|T = t] - E[Y(t_1)|T = t_1]]}{\text{Average of selection bias 2}}.$$

This simple example indicates that, due to the existence of averages of the selection biases 1 and 2, it is impossible to directly use the sample counterparts to calculate treatment effects. 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,
\[ E[Y|X, T = t_2] - E[Y|X, T = t_1] = E[Y(t_2)|X, T = t_2] - E[Y(t_1)|X, T = t_1] \]
\[ = E[Y(t_2)|X] - E[Y(t_1)|X] \]
\[ = E[Y(t_2) - Y(t_1)|X], \]
which has a causal interpretation. This is the ignorability condition and it is discussed in more detail below. Finally, we need to combine the results for each \( X \) and obtain an unconditional treatment effect. In this case, using the law of iterated expectation, this unconditional expectation can be recovered.

The objective of this paper is to study average and quantile dose-response functions, ADRF and QDRF respectively. From the corresponding DRF it is straightforward to recover the average treatment effect (ATE) and quantile treatment effect (QTE) respectively. To accomplish this aim we develop a general framework for generic moment restriction estimators (\( Z \)-estimators) with possibly non-smooth functions. For each \( t \in \mathcal{T} \), 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, which we discuss in more details in condition \( I.I \) stated below. Then the DRF is defined as the parameters of interest \( \beta(t) \) that solve the moment condition. Henceforth, we discuss the important examples of ADRF and QDRF as applications of the general theory. The following two examples show that ADRF and QDRF are special cases of \( \beta(t) \) in condition \( I.I \) which result from choosing specific forms of \( m(\cdot) \).

**Example (Average).** We first discuss the ADRF example. Setting \( m(Y(t); \mu(t)) = Y(t) - \mu(t) \) and letting the first moment to equal to zero, we can 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 (Quantile).** QDRF is another special case of our general model. Let \( m(Y(t); q_\tau(t)) = \tau - 1\{Y(t) < q_\tau(t)\} \), we obtain \( q_{\tau_0}(t) \in \inf\{q : F_{Y(t)}(q) \geq \tau\} \), the unconditional \( \tau \)th QDRF, where \( F_{Y(t)} \) is the distribution function of \( Y(t) \). From the QDRF, one can estimate the QTE.
as $QTE(t, t') = q_{\tau_0}(t) - q_{\tau_0}(t')$. Note that, as is in Firpo (2007) and Cattaneo (2010), in this paper the QTE is defined as the difference of the $\tau$th quantile at different levels of treatment. This definition does not require the assumption of rank preservation, and it is regarded as a convenient way to summarize interesting aspects of marginal distributions of the potential outcomes. However, if rank preservation holds, then the simple differences in quantiles turn out to be the QTE. We refer the reader to Firpo (2007) for a detailed discussion on rank preservation in QTE.

Now we state assumptions on the general model to achieve identification of the parameters of interest.

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

I.II For all $t \in \mathcal{T}$, we have

1. $Y(t) \perp T | X$;

2. $f_{0T|X,Y}(t|x, y) > 0$ for $t \in \mathcal{T}$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

I.III Assume that

1. There exists a function $e(y)$ with $\int e(y) dy < \infty$ such that

$$|m(y; \beta(t_0))f_{T,Y|X}(t_0 + \Delta t, y|x)| \leq e(y);$$

2. $E[m(Y; \beta(t_0))|X, T = t_0] = \lim_{\Delta t \downarrow 0} E[m(Y; \beta(t_0))|X, T \in [t_0, t_0 + \Delta t]]$. Also the interval $\mathcal{T}$ is right open.

Condition I.I is an identification condition provided that $Y(t)$ are 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 $Y(t)$ are not observable for all $t \in \mathcal{T}$. Therefore, condition I.II.1, the assumption of ignorability, is fundamental. According to condition 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, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), and Hirano and Imbens (2004). Condition I.II.2 states that for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the density of treatment levels is positive. Thus in our model the triple $(X, Y, T)$
is observable, and a random sample of size $n$ can be obtained. Condition I.III allows us to change orders of limits and integral. Also, the set $\mathcal{T}$ is right open without loss of generality. If we would like to have $\mathcal{T}$ to be right closed, we shrink the interval $[t_0 - \Delta t, t_0]$ to obtain $t_0$.

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

**Theorem 1.** Under conditions I.I–I.III, we have

$$E[m(Y(t); \beta(t))] = E[m(Y; \beta(t))\pi_0(U; t)]$$

(1)

for each $t \in \mathcal{T}$, where $\pi_0(\mathbf{u}; t) := \frac{f_{U|X,Y}(t|\mathbf{x},y)}{f_{U|X}(t|\mathbf{x})}$. Consequently,

$$E[m(Y; \beta(t))\pi_0(\mathbf{u}; t)] = 0$$

(2)

if and only if $\beta(t) = \beta_0(t)$.

**Proof.** See Appendix C.

The result in equation (1) allows identification of the DRF. The left hand side of equation (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 equation (1) is expressed in terms of observable $(\mathbf{X}, Y, T)$, and therefore, 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 $\mathbf{X}$ delivers identification the parameter of interest. That is, conditional on observed covariates $\mathbf{X}$, each individual is randomly assigned to a treatment level.

**Remark 1.** The result in Theorem 1 indeed has a similar format as the equation in (2) of Cattaneo (2010) after we transform the latter. We begin with

$$E \left[ \mathbf{1}\{T = t\}m(Y; \beta(t)) \right] = 0.$$ 

By the law of iterated expectation, the left hand side of the previous equation equals

$$E \left[ \frac{m(Y; \beta(t))}{p_t(X)}E\left[ \mathbf{1}\{T = t\} | X, Y \right] \right].$$

(3)
Noting that

\[ E[\{T = t\} | \bm{X}, \bm{Y}] = P(\bm{T} = t | \bm{X}, \bm{Y}) \]

and, by definition, \( p_t(\bm{X}) = P(\bm{T} = t | \bm{X}) \), equation (3) equals

\[ E \left[ m(\bm{Y}; \beta) \frac{P(\bm{T} = t | \bm{X}, \bm{Y})}{P(\bm{T} = t | \bm{X})} \right]. \]

Thus, our result simply “replaces” the conditional probabilities by conditional densities.

**Remark 2.** The result in Theorem 1 is also related to Hirano and Imbens (2004). They extend the propensity score method in a setting with continuous treatment for ADRF. This paper 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.

Given the identification condition in equation (2) of Theorem 1, we are able to estimate the parameters of interest. We propose a Z-estimator that involves two steps estimation as follows.\(^2\)

**Step 1** Estimate \( \pi_0(\bm{U}; t) \) parametrically or nonparametrically and obtain an estimator \( \hat{\pi} \).

**Step 2** For each \( t \in \mathcal{T} \), find \( \hat{\beta}(t) \) as a zero of the following condition

\[ \frac{1}{n} \sum_{i=1}^{n} m(Y_i; \beta(t)) \hat{\pi}(\bm{U}_i; t) = 0. \quad (4) \]

The estimator \( \hat{\beta}(t) \) is defined as the zero of the equation above.\(^3\)

The identification conditions and the estimator are illustrated below using the previous two examples.

\(^2\)We work directly with the estimating equations. However, estimation could be carried with GMM methods.

\(^3\)We state the exact zero root for simplicity. Nevertheless, the solution does not need to be characterized by an exact zero. Technically, we only need a solution that approximately solves the estimating equation, which is common condition in the literature; see, e.g., He and Shao (1996) and He and Shao (2000).
**Example (Average, continued).** The identification condition for $\mu_0(t)$ is

$$E[(Y - \mu_0(t))\pi_0(U; t)] = 0.$$  

An estimator of $\mu_0(t)$ is

$$\hat{\mu}(t) = \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\pi}(U_i; t)\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}(U_i; t) Y_i.$$  

(5)

**Example (Quantile, continued).** The identification condition for $q_{\tau_0}(t)$ is

$$E[(\tau - 1\{Y < q_{\tau_0}(t)\})\pi_0(U; t)] = 0.$$  

An estimator of $q_{\tau_0}(t)$ is

$$\hat{q}_{\tau}(t) = \arg \min_q \frac{1}{n} \sum_{i=1}^{n} \hat{\pi}_0(U_i; t) \rho_{\tau} (Y_i - q),$$  

(6)

where $\rho_{\tau}(u) := u(\tau - 1\{u < 0\})$ is the check function as in Koenker and Bassett (1978).

### 3 Asymptotic properties

In this section, we first derive the asymptotic properties of the general two-step estimator described above. In particular, we establish the uniform consistency, the weak limit, and the uniform semiparametric efficiency. Second, we apply the general results and derive the corresponding asymptotic properties of ADRF and QDRF examples. In addition, we discuss estimation of the nuisance parameter, $\pi_0$, and inference based on the two-step estimator.

To establish these results, in **Lemmas 1 and 2** in Appendix A we first provide sufficient conditions to ensure consistency and weak convergence of the infinite dimensional parameters obtained from generic moment restriction estimators (Z-estimators) with possibly non-smooth criterion that depends on a preliminary infinite dimensional (nuisance) parameter estimator. The results allow for the case where the nonparametric estimator is profiled, i.e., is allowed to depend on the parameters. These results are similar to those in van der Vaart (2002) and van der Vaart and Wellner (2007), and extend Chen, Linton, and Van Keilegom (2003) in that the parameter of interest is not in a Euclidean space but
in a generic Banach space. Moreover, the results extend Theorem 3.3.1 of van der Vaart and Wellner (1996) in that a possibly infinite dimensional nuisance parameter needs to be estimated in the first step.

We then proceed by applying Lemmas 1 and 2 to show uniform consistency and weak convergence of the estimator of DRF, \( \hat{\beta}(t) \), in \( \ell^\infty(T) \). Applications of these results to the specific cases of ADRF and QDRF are also provided. The uniform semiparametric efficiency is based on the pointwise semiparametric efficiency and the weak convergence of the estimator to a tight random process. We then provide ways of estimating the nuisance parameter \( \pi_0 \). As for the inference, we focus on hypothesis testing based on a Kolmogorov and a Cramér-von Mises statistic.

3.1 Consistency

Consistency is a desired property for most estimators. In this paper, different from the discrete or multi-valued treatment models, the treatment levels take values on an interval \( T \). Therefore, the consistency results are established uniformly over the set \( T \). For the general two-step estimator given in equation (4) to be uniformly consistent, we state the following sufficient conditions.

\begin{align*}
\text{C.I} & \quad |\mathbb{E}[m(Y; \hat{\beta}(t)) \pi(U; t)]|_\infty = o_p(1). \\
\text{C.II} & \quad |\mathbb{E}[m(Y; \beta_n(t)) \pi_0(U; t)]|_\infty \to 0 \text{ implies } |\beta_n(t) - \beta_0(t)|_\infty \to 0 \text{ for any sequence } \beta_n(t). \\
\text{C.III} & \quad \sup_{\beta \in B} |\mathbb{E}[m(Y; \beta(t))]|_\infty < M < \infty \text{ for some } M > 0. \\
\text{C.IV} & \quad |\hat{\pi} - \pi_0|_\infty = o_p(1). \\
\text{C.V} & \quad \sup_{\beta(t) \in \ell^\infty(T), \pi \in \Pi_{\delta_n}} |\mathbb{E}[m(Y; \beta(t)) \pi(U; t)] - \mathbb{E}[m(Y; \beta(t)) \pi(U; t)]|_\infty = o_{p^*}(1) \text{ for } \delta_n \downarrow 0, \\
or \quad \text{C.V'} & \quad \{\psi_{1,\beta,t} : \beta \in \ell^\infty(T), t \in T\} \text{ and } \{\psi_{2,\pi,t} : \pi \in \Pi_{\delta_n}, t \in T\} \text{ are Glivenko-Cantelli with respective envelopes } F_1 \text{ and } F_2 \text{ such that } EF_1 F_2 < \infty, \text{ where } \psi_{1,\beta,t} = m(Y; \beta(t)) \text{ and } \psi_{2,\pi,t} = \pi(U; t). 
\end{align*}

Conditions C.I defines the Z-estimator and C.II is an identification condition for the Z-estimator. Pakes and Pollard (1989) and Chen, Linton, and Van Keilegom (2003) have similar assumptions. For a detailed discussion of this type of identification condition, see p.
45 of van der Vaart (1998). Condition C.III only requires the moment of the estimating equation to be finite. This is a standard assumption and analogous to condition 4 (b) in Cattaneo (2010). Condition C.IV requires consistent estimation of the nuisance parameter. This is also a usual requirement, which corresponds to condition (1.4) of Theorem 1 of Chen, Linton, and Van Keilegom (2003). We will discuss estimation of the nuisance parameter in Section 3.4 below. Condition C.V is a uniform law of large numbers, which is implied by condition C.V’. These conditions are standard; see Newey and McFadden (1994). We provide more primitive conditions for specific cases, as average and quantile. Now we state the consistency result for the estimator of the DRF.

**Theorem 2.** Suppose that \( E[m(Y, \beta_0(t))\pi_0(U; t)] = 0 \), 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 Appendix C.

Now we discuss the consistency of the two-step estimators of ADRF and QDRF given in the examples in equations (5) and (6), respectively. To establish the result for the ADRF, the following conditions are imposed.

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

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

Condition AC.I requires that the expectation of \( Y(t) \) be finite. Also the diameter of the parameter space is finite, which is a common condition for M-estimators, see e.g., Theorem 3 of Chen, Linton, and Van Keilegom (2003) or Wooldridge (2010). Hirano, Imbens, and Ridder (2003) assume the second moment of \( Y(1) \) and \( Y(0) \) to be finite, which is slightly stronger than our condition. This is mainly because they do not explicitly describe conditions for the consistency of their estimator. Cattaneo (2010) has the same second moment restriction as well. Condition AC.II is a high level condition on the nuisance parameters, 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 for the two-step estimator of the ADRF is summarized in the following corollary.

**Corollary 1** (Average). The two-step estimator of ADRF is consistent, i.e., $|\hat{\mu}(t) - \mu_0(t)|_\infty = o_p(1)$, provided conditions **AC.I**–**AC.II** and **C.IV** are satisfied.

*Proof.* See Appendix C.

For the uniform consistency of the QDRF estimator over $t \in \mathcal{T}$, the following conditions are imposed.

**QC.I** Uniformly in $t$, the densities $f_{Y(t)}(y)$ is bounded above and $f_{Y(t)}(q_{\tau_0}(t)) > 0$. Also, for any $\delta > 0$, $\inf_{y-q_{\tau_0}} > \delta |E[(\tau - 1\{Y < q\})\pi_0(U; t)]_\infty > \epsilon_0$.

**QC.II** There exists $0 < M_\pi < \infty$ such that $\pi_0(u; t) < M_\pi$.

**QC.III** The function class $\{\psi_{2,\pi,t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is Glivenko-Cantelli, and have an envelope $F_2(y)$ such that $F_2(y)$ that is integrable.

Condition **QC.I** is an identification condition on the parameter of interest. This condition is the analogue to the general condition **C.I**, and it is similar to A.2 and A.3 of Angrist, Chernozhukov, and Fernandez-Val (2006), and corresponds to Assumption 2 of Firpo (2007). Cattaneo (2010) uses a similar assumption for the quantile estimation. Condition **QC.II** is a boundedness condition of the joint density of $(U, T)$, and is the continuous treatment version of Assumption 1 (ii) of Firpo (2007). Condition **QC.III** is weaker than condition **AC.II** since only $F_2(y)$ is required to be integrable. This is because $\tau - 1\{\cdot\}$ is uniformly bounded. Those two conditions are imposed for the estimation of the nuisance parameter. The consistency result for the estimator of QDRF is provided in the following corollary.

**Corollary 2** (Quantile). For a given quantile of interest, the two-step estimator of the QDRF is consistent, i.e., $|\bar{q}_\tau(t) - q_{\tau_0}(t)|_\infty = o_p(1)$, provided conditions **QC.I**–**QC.III** and **C.IV** are satisfied.

*Proof.* See Appendix C.
3.2 Weak convergence

In this section, we apply the results of Lemma 2 to derive the limiting distribution of the general two-step estimator in (4). Later, we demonstrate the results for the ADRF and QDRF estimators. To this end, we impose the following sufficient conditions.

\textbf{G.I} \ |E[m(Y; \hat{\beta}(t))\hat{\pi}(U; t)]|_\infty = o_p(1/\sqrt{n}).

\textbf{G.II} The map $\beta \mapsto E[m(Y; \beta)\pi_0(U; \cdot)]$ is Fréchet differentiable at $\beta_0$ with a continuously invertible derivative $Z_1(\beta_0, \pi_0)$.

\textbf{G.III} $E[m(Y; \beta(t))]$ is Lipschitz continuous at $\beta_0(t)$. In addition, $\sup_{\beta \in \mathbb{B}} |E[m(Y; \beta(t))^2]|_\infty < M < \infty$ for some $M > 0$.

\textbf{G.IV} $|\hat{\pi} - \pi_0|_\infty = o_p(n^{-1/4}).$

\textbf{G.V} The functional classes $\{\psi_{1, \beta, t} : \beta \in \ell^\infty_\delta(T), t \in \mathcal{T}\}$ and $\{\psi_{2, \pi, t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ are uniformly bounded Donsker classes.

\textbf{G.V'} The functional classes $\{\psi_{1, \beta, t}\psi_{2, \pi, t} : \beta \in \ell^\infty_\delta(T), \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is a Donsker classes.

\textbf{G.VI} $\sqrt{n}(E[m(Y; \beta_0(t))(\pi(U; t) - \pi_0(U; t))]_{|\pi = \hat{\pi}} + E[m(Y; \beta_0(t))\pi_0(U; t)])$ converges weakly to a tight random element $G(t)$ in $\ell^\infty(\mathcal{T})$.

Condition \textbf{G.I} defines the Z-estimator, which is slightly stronger than condition \textbf{C.I} but still allows the right hand size to be zero only approximately. This type of $o_p(n^{-1/2})$ condition is also assumed in (i) of Theorem 3.3 of Pakes and Pollard (1989) and (2.1) of Theorem 2 of Chen, Linton, and Van Keilegom (2003). Condition \textbf{G.II} requires the model to be differentiable in $\beta$ and the derivative is invertible, and corresponds to (ii) of Theorem 3.3 of Pakes and Pollard (1989) and (2.2) of Theorem 2 of Chen, Linton, and Van Keilegom (2003). Condition \textbf{G.III} corresponds to assumption 6 (b) of Cattaneo (2010). Condition \textbf{G.IV} strengthens condition \textbf{C.IV} such that the estimator of the nuisance parameter converges at a rate faster than $n^{-1/4}$. A similar condition appears in condition (2.4) in Theorem 2 of Chen, Linton, and Van Keilegom (2003). Conditions \textbf{G.V} (\textbf{G.V'}) and \textbf{G.VI} are high level and 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

$$\sqrt{n}(\hat{\beta}(t) - \beta_0(t)) \rightsquigarrow Z_1^{-1}(\beta_0(t), \pi_0(U; t))G(t)$$

in $\ell^\infty(\mathcal{T})$.

Proof. See Appendix C.

The result given in Theorem 3 shows that the limiting distribution of the two-step DRF estimator is nonstandard. This result is due to the presence of the set of continuous treatments. However, if one fixes the treatment at $\bar{t}$, 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. In addition, this result has important applications to the two leading examples of ADRF and QDRF. For the weak convergence of the two-step estimator of the ADRF, we impose the following conditions.

AG.I The parameter space for $\mu_0$ is a bounded sub-Banach space $\mathcal{M}$ of $\ell^\infty(\mathcal{T})$. In addition, $|E[Y(t)^2]|_\infty < M_Y$ for some $0 < M_Y < \infty$.

AG.II The function class $\{\psi_{3,\pi,t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is Donsker, where $\psi_{3,\pi,t}(u) = y\pi(u; t)$.

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

Condition 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 function classes satisfy condition AG.II, i.e., the smooth function class discussed above. Condition AG.III is a high level condition on the nuisance parameter, and we provide an estimator that satisfies this condition.

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

Proof. See Appendix C.

To obtain the weak convergence of the QDRF estimator, equation (6), we impose the following conditions.
The function class \( \{ \psi_{2, \pi, t} : \pi \in \mathcal{\Pi}, t \in T \} \) is Donsker with a uniform bound.

\[ \sqrt{n} \mathbb{E}[(\tau - 1\{Y < q_{r0}(t)\})(\pi(U; t) - \pi_0(U; t))] \big|_{\pi = \hat{\pi}} \text{ converges weakly.} \]

Examples satisfying condition QG.II include smooth function classes. Condition QG.I is a high level condition and will be discussed in the section of the estimation of the nuisance parameter. This assumptions is similar to AG.III. and a version of condition G.VI. Now we state the weak convergence result for the estimator of QDRF.

**Corollary 4** (Quantile, continued). 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 Appendix C. \( \square \)

### 3.3 Semiparametric efficiency of the two-step estimator

In this subsection, we first calculate the efficient influence function of the parameter \( \beta(t) \) in the following semiparametric model

\[ \mathcal{F} = \{ F_{\beta, \pi} : \beta \in \ell^\infty(T), \pi \in \Pi \}, \]

where \( F_{\beta_0, \pi_0} \) is the distribution function of the observed data. Then, we provide sufficient conditions under which the proposed two-step estimator is uniformly semiparametric efficient.

**Proposition 1.** Suppose \( \Gamma_0(t) := \frac{\partial \mathbb{E}[m(Y(t); \beta_0(t))]}{\partial \beta(t)} \) exists. For each \( t \in T \), the efficient influence function of the parameter \( \beta(t) \) is

\[ \Psi_\beta(y, t, x) = -\Gamma_0^{-1}(t)\psi(y, x, t, \beta_0, \pi_0, e_0), \]

where \( \psi(y, x, t, \beta_0, \pi_0, e_0) = m(y; \beta_0(t))\pi_0(u; t) - e_0(x, \beta_0(t))(\pi_0(u; t) - 1) \) with \( e_0(x, \beta(t)) = \mathbb{E}[m(Y; \beta(t)) | X = x] \).

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

Based on the efficient influence function of \( \beta(t) \), we show that the two-step estimator is uniformly semiparametric efficient provided the following condition

**E.** \( \sqrt{n} \mathbb{E}[m(Y; \hat{\beta}_0(t))\hat{\pi}(U; t)] = \sqrt{n} \mathbb{E}[\psi(Y, X, t, \beta_0, \pi_0, e_0)] + o_p(1). \)
Condition \( E \) is critical to the efficiency of the two-step estimator, and it is similar to its corresponding condition for the multi-valued model is condition (4.2) of Cattaneo (2010).

**Theorem 4.** Assume that the conditions of Theorem 3 and condition \( E \) hold. Then the two-step estimator is uniformly semiparametric efficient.

**Proof.** See Appendix C.

This result guarantees that the two-step estimator is uniformly semiparametric efficient. Hypothesis testings based on this estimator are expected to be optimal.

### 3.4 Estimation of \( \pi_0 \)

We have been assuming that the estimator \( \hat{\pi} \) of the nuisance parameter \( \pi_0 \) satisfies various conditions. In this subsection we discuss the estimation of the nuisance parameter \( \pi_0 \), i.e., \( \frac{f_{T|X,Y}(t|x,y)}{f_{T|X}(t|x)} \). The estimation of the nuisance parameter is a very important step for implementation of the proposed estimator in practice.

It is common in the literature to estimate nuisance parameters in two-step estimators by using parametric models, see, e.g., Newey (1984), 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 use a flexible parametric approach. In addition, there is persuasive reason to use flexible parametric models to estimate the ratio of the conditional density functions, \( \pi_0 \). 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 convergence rates of nonparametric estimation. Hence practical nonparametric estimation might be infeasible due to the curse of dimensionality problem.

For estimators of \( \pi_0 \) to have the desirable properties, we impose the following assumptions.

**N.I** Assume \( \pi = \pi(u; t; \vartheta) \), where \( \vartheta \in \mathbb{R}^{d_\vartheta} \) with \( d_\vartheta \) being a positive integer. \( \pi(u; t; \vartheta) \) is a smooth function of \( \vartheta \) with uniformly continuous, bounded, and square integrable first derivative, \( \pi'(u; t; \vartheta) \), with respect to \( \vartheta \).

**N.II** There exists \( \hat{\vartheta} \) such that \( \sqrt{n} (\hat{\vartheta} - \vartheta) \xrightarrow{d} N(0, \Im^{-1}) \) with \( \Im \) is nonsingular.

Condition **N.I** is a smoothness and boundedness condition for the function to be estimated, and condition **N.II** assumes that there is an estimator of the parameter that is asymptotically normal. Examples which satisfy condition **N.II** include maximum likelihood estimator.

Proof. See Appendix C.

Proposition 2 can be applied to verify the conditions on the nuisance parameters for the average and quantile examples. Now we provide a few examples to demonstrate the estimation of $\pi_0$ in practice.

Example. We estimate $f_{Y|X,T}(y|x,t)$ and $f_Y(y|x)$ separately. For $f_{Y|X,T}(y|x,t)$, assume the relationship

$$Y = g(X, T; b) + \epsilon,$$

and $\epsilon|X, T \sim N(0, \sigma^2_{\epsilon})$, where $g(\cdot)$ is some known function and $b$ is an unknown parameter to be estimated. Using Nonlinear Least Square Estimation (see e.g. Davidson and MacKinnon (1993)), we obtain estimators of the conditional mean and variance, and therefore, the conditional density of $Y$ given $X$ and $T$. Similarly, we estimate $f_Y(x|y)$.

Example. To estimate the conditional density $f_{Y|X,T}(y|x,t)$, we can also assume the following model

$$\Lambda (Y, \lambda) = \Lambda (g(X, T), \lambda) + \epsilon,$$

where $\epsilon|X, T \sim N(0, \sigma^2_{\epsilon})$, $g(\cdot)$ is some known function, and $\Lambda(\cdot)$ is the Box-Cox transformation function, which is defined as $\Lambda(Z, \lambda) = \log Z$ if $\lambda = 0$ and $= \frac{Z^{\lambda-1}}{\lambda}$ otherwise. Using Maximum Likelihood Estimation, we obtain the unknown parameter $\lambda$ and therefore the conditional density $f_{Y|X,T}(y|x,t)$. Similarly, we estimate $f_Y(x|y)$.

Example. 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, Maximum Likelihood Estimation can be applied and the estimator $\hat{f}_{T,X,Y}(t, x, y)$ calculated, and then $\hat{\pi}$ can be obtained.

3.5 Inference on the DRF

In this section, we turn our attention to inference on the DRF. Inference is carried uniformly over the set of treatment levels, $\mathcal{T}$. Given the formulation for inference of DRF, inference
for the treatment effects is straightforward. In particular, it is possible to formulate a wide
variety of tests using variants of the proposed tests.

General hypotheses on the vector $\beta(t)$ can be accommodated by Kolmogorov and Cramér-
von Mises type tests. These statistics and their associated limiting theory provide a natural
foundation for testing the null hypothesis $H_{01}: \beta(t) = r(t)$ when $r \in \ell^\infty(\mathcal{T})$ is known. Thus,
from the result in Theorem 3, under the null hypothesis $H_{01},$

$$V_n(t) := \sqrt{n}(\hat{\beta}(t) - r(t)) \rightsquigarrow \mathbb{G}(t).$$

More precisely, we propose the following test statistics:

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

They are a Kolmogorov type and a Cramér-von Mises type statistic, respectively. Now we
present the limiting distributions of the test statistics under the null hypothesis.

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

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

**Proof.** The assertion holds by Theorem 3 and the continuous mapping theorem. \qed

In addition to testing the hypothesis $\beta_0(t) = r(t)$ with known $r \in C(\mathcal{T})$, we could also
test the hypothesis with unknown $r$, in which case, the estimation of $r$ is needed. Often, a
$\sqrt{n}$-consistent estimator $\hat{r}$ is available, and under the null hypothesis $H_{02}: \beta_0(t) = r(t)$, the
test statistic becomes

$$\bar{V}_n(t) := \sqrt{n}(\hat{\beta}(t) - \hat{r}(t)) \rightsquigarrow \mathbb{G} - \mathbb{G}_r,$$

where $\mathbb{G}_r$ is the weak limit of $\sqrt{n}(\hat{r}(t) - r(t))$. Therefore, due to the estimation of $r(t)$, a
drift component is introduced.
We propose the following test statistics:

\[
\bar{T}_1^n := \sup_{t \in T} |\bar{V}_n(t)|, \\
\bar{T}_2^n := \int_{t \in T} |\bar{V}_n(t)| \, dt.
\]

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

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

\[
\bar{T}_1^n \rightsquigarrow \sup_{t \in T} |G(t) - G_r|, \quad \bar{T}_2^n \rightsquigarrow \int_{t \in T} |G(t) - G_r| \, dt.
\]

**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. A formal justification for our simulation method, discussed below, is stated in **Lemma 3**, in Appendix A. This result provides a proof of the validity of the bootstrap for general Z-estimator, and it is also an extension of that in Chen, Linton, and Van Keilegom (2003). A simple application of Corollaries 5 and 6 produces the bootstrap procedure for the ADRF or QDRF.

Given the above framework, inference for the treatment effects is simple. Using the inference of ATE from treatment level \( t_1 \) to \( t_2 \) as an example, the point estimate of the ATE is \( \hat{\mu}(t_2) - \hat{\mu}(t_1) \), which has an asymptotic normal distribution with mean \( \mu(t_2) - \mu(t_1) \), and its variance is computable from the covariance kernel of the weak limit of \( \hat{\mu}(t) \). Therefore, the inference can be done using standard methods.

### 3.5.1 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 \( H_{01} \) 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 (4). For special cases of DRF, as ADRF and QDRF, one replaces equation (4) with (5) and (6) respectively. Second, \( T_{1n} \) or \( T_{2n} \) are calculated by centralizing \( \hat{\beta}(t) \) at \( r(t) \) and taking the
maximum over \( t \) (for \( T_{1n} \)) or summing over \( t \) (for \( T_{2n} \)). For the general case, \( H_{02} \) with unknown \( r(t) \), the tests can be implemented in the same fashion. The only adjustment is after estimating \( \beta(t) \) one uses \( \hat{r}(t) \) to compute \( \hat{T}_{1n} \) or \( \hat{T}_{2n} \).

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 \( \hat{T}_{1n} \) 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 \( \{(Y_{i}^{b}, U_{i}^{b}), i = 1, \ldots, n\} \).

(ii) Estimate the DRF \( \hat{\beta}^{b}(t) \) and set \( V_{n}^{b}(t) := \sqrt{n}(\hat{\beta}^{b}(t) - r(t)) \)

(iii) Compute the quantity statistic of test of interest

\[
\hat{T}_{1n}^{b} = \max_{t \in \mathcal{T}} |V_{n}^{b}(t)|
\]

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

4 Monte Carlo

In this section we report the results of small Monte Carlo simulations. The data generating process has treatment level \( t \in [0, 1] \). A sample of \( n \) i.i.d. random elements \((X_{i}, \epsilon_{i}(t), v_{i}(t))\) whose components are mutually independent are generated, where the independent white Gaussian noises \( \epsilon_{i}(t) \) and \( v_{i}(t) \) are represented by \((\epsilon_{i}(0), \epsilon_{i}(0.01), \ldots, \epsilon_{i}(0.99), \epsilon_{i}(1))\) and \((v_{i}(0), v_{i}(0.01), \cdots, v_{i}(0.99), v_{i}(1))\), respectively. The observed individual characteristics \( X_{i} \sim Unif[-0.5, 0.5] \) and \( Y_{i}(t) = 0.5 - |0.5 - t| + X_{i} + v_{i}(t) \) where independent innovations \( v_{i}(0), v_{i}(0.1), \cdots, v_{i}(1) \sim N(0, 1) \). The treatment assignment is generated by

\[
T_{i} = \arg\max_{t \in \{0.0, 0.01, \ldots, 0.99\}} H_{t,i},
\]

where \( H_{t,i} = \sin(2\pi t)X_{i} + \epsilon_{i}(t) \) where independent innovations \( \epsilon_{i}(0), \epsilon_{i}(0.1), \cdots, \epsilon_{i}(1) \sim N(0, 1) \). We generate the data in such a way that the mean and median functions are \( 0.5 - |0.5 - t| \), an up-side-down symmetric check function. The level is highest in the middle range and decreases as \( t \) deviates from the middle. The number of replications is 2,000.
Our null and alternative hypotheses are summarized below.

\begin{align*}
H_{m0} & : \mu_0(t) = 0.5 - |0.5 - t| \text{ for } t \in [0.2, 0.8] \\
H_{q0} & : q_{0.5,0}(t) = 0.5 - |0.5 - t| \text{ for } t \in [0.2, 0.8] \\
H_{m1} & : \mu_0(t) = t \text{ for } t \in [0.2, 0.8] \\
H_{q1} & : q_{0.5,0}(t) = t \text{ for } t \in [0.2, 0.8] \\
H_{m2} & : \mu_0(t) = t^2 \text{ for } t \in [0.2, 0.8] \\
H_{q2} & : q_{0.5,0}(t) = t^2 \text{ for } t \in [0.2, 0.8] \\
H_{m3} & : \mu_0(t) = 0.25 - (t - 0.5)^2 \text{ for } t \in [0.2, 0.8] \\
H_{q3} & : q_{0.5,0}(t) = 0.25 - (t - 0.5)^2 \text{ for } t \in [0.2, 0.8]
\end{align*}

On the one hand, the first alternative is a linear function while the second is an asymmetric quadratic function, which are quite different from the null. On the other hand, the third alternative is a quadratic function symmetric around 0.5 and attains its maximum at 0.5.

We use the Cramér-von Mises test for the simulations. Also, we use the method of Hall, Racine, and Li (2004), a nonparametric method to estimate the conditional densities. We first show the biasedness of the estimator when sample sizes are 150 and 300. The bias is defined as the supreme of the pointwise biases and presented below.

<table>
<thead>
<tr>
<th></th>
<th>n=150</th>
<th>n=300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.071</td>
<td>0.059</td>
</tr>
<tr>
<td>Median</td>
<td>0.068</td>
<td>0.056</td>
</tr>
</tbody>
</table>

As expected, the bias is approximately zero and decreases as sample size increases. Now we present the empirical size and power below.

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>Power for $H_1$</th>
<th>Power for $H_2$</th>
<th>Power for $H_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=150</td>
<td>Mean</td>
<td>0.03</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>n=300</td>
<td>Mean</td>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.02</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>n=300</td>
<td>Mean</td>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>n=300</td>
<td>Mean</td>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>n=500</td>
<td>Mean</td>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
In the simulations, we evaluate the empirical size and power for a variety of sample size and number of bootstrap. We observe that the sizes are close to the level of significance, 5%, and the power is high for the alternative hypotheses $H_1$ and $H_2$. To study the impact of sample size and number of bootstrap on the power of the test, we test for $H_3$, which is closer to the null hypothesis. It turns out that there is power gain from increasing the sample size. However, the power is not sensitive to the number of bootstrap, implying that smaller number of bootstrap is satisfactory and using larger number of bootstrap is not necessary.

The simulations show numerical evidence that, although we have not theoretically shown the weak convergence of the Z-estimator under kernel estimation in the first step, in practice, one nevertheless is able use the kernel estimation as an alternative method in the first step.

5 Application to the estimation of dose-birthweight functions

In this section, 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, and corresponding treatment effects, for both mothers’ weight gain during pregnancy as well as mother’s age. Recently birthweight has been shown to be the foremost telltale of infant health. 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 at birth. There is empirical evidence showing that the direct medical costs of LBW are quite high. Almond, Chay, and Lee (2005) document that the hospital costs for newborns are elevated: the expected costs of delivery and initial care of a baby weighing one kilogram at birth can exceed $100,000 (in year 2000 dollars). The costs remain elevated even among babies weighing 2–2.1 kilograms; an additional pound (454 grams) of weight is still associated with a $10,000 difference in hospital charges for inpatient services.\textsuperscript{4} The infant mortality rate also increases at lower birthweights.

On the other hand, problems associated with high birthweight have become more recognizable. For instance, babies weighing more than 4 kilograms (defined as high birthweight

\textsuperscript{4}Appurtenant expenditures, such as radiological, pharmaceutical, respiratory, and laboratory fees, greatly extend the costs of intensive care for LBW infants. (See, e.g., Behrman, Butler, and Committee on Understanding Premature Birth and Assuring Healthy Outcomes (2007)).
(HBW)) and especially those weighing more than 4.5 kilograms (classified as very high birth-weight) are more likely to require cesarean-section births, have higher infant mortality rates, and develop health problems later in life. Recent research also suggests giving birth to infants over 4.5 kilograms carries significant risks to both the infant and the mother; see, e.g., Cesur and Kelly (2010) and Webb (2011) for more detailed discussions. 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.

Some other contributions have been made on the study of the distributional effects of the treatment. The literature on QTE includes, among others, Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005), which explicitly deal with the fact that treatment effects may be nonmonotonic along the outcome distribution. Chernozhukov and Hansen (2006, 2008) introduce a class of instrumental quantile regression methods for structural and treatment effect models. Imbens and Newey (2009) extend Newey, Powell, and Vella (1999), Pinkse (2000), and Blundell and Powell (2003) to identification and estimation of a class of parameters, termed structural quantile functions, and apply to the case of continuous treatments. Moreover, there is an emerging literature on average treatment effects for continuous variables, which includes Hirano and Imbens (2004) and Flores (2007).

5.1 Data

The United States natality data from the National Vital Statistics System (NVSS) of Centers for Disease Control and Prevention (CDC) document nearly all births in registered areas. The data in this study are the 2004 public use natality data of Wisconsin.

For this study, we consider only live, singleton births (without missing values of any characteristics used in the study) to new, white mothers that are not older than 45, with less than five years of college, whose counties of occurrence (birthing) and residence are the same. Birthweights have been found to differ across different ethnicities, numbers of babies at delivery, and so on. By using a more homogeneous sample, we can focus more on the effects of birth inputs on the birthweights. This results in a sample of 13,581 births. We emphasize that all the inferences done using the sample should be applicable only for the subpopulation represented by the sample choice.

Now we summarize the variables of interest. Out of 13,581 births, there are 6,508 females (proportion 0.4792), and 7,732 mothers are married (proportion 0.5693). Table 1 displays statistics for birthweight (measured in kilograms), 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. And Figure 1 shows the distribution of the month of first time prenatal care visits.

Our dataset is similar to “1st child” Washington and Arizona datasets of Abrevaya and Dahl (2008) for the variables that are directly comparable. The number of observations are 45,067 and 56,201 for Washington and Arizona, respectively. For example, the averages of the birthweights of Washington and Arizona data are 3.44 and 3.34 kilograms, respectively, while the average of the birthweights in Wisconsin is 3.35 kilograms. The averages of the ages of Washington and Arizona are 25.27 and 25.23, respectively, while that of Wisconsin is 24.88. The averages of number of prenatal care visits and education in Wisconsin are slightly lower than those of Washington and Arizona. The proportion of male infants are similar for all the births in the three states, but the proportion of married mothers in Wisconsin is much lower than those of Washington and Arizona.

[Table 1 and Figure 1 about here]

5.2 Estimation of nuisance parameter \( \pi_0 \)

The estimation strategy of nuisance parameters \( \pi_0(u; t) := \frac{f_{0|Y|X}(t|x,y)}{f_{0|Y|X}(t|x)} \) in (5) and (6) is the same for both the treatment effects of the mother’s age and weight gain during pregnancy. In this section, we describe the details of the estimation procedure using the first treatment effect model, i.e., 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 distribution, we assume the relationship \( \log \left( \frac{T_i - 13}{46 - T_i} \right) = X_i^\top \theta_0 + \epsilon_i \), where \( \epsilon_i \) is independent of \( X_i \) and has density \( N(0, \sigma_0^2) \). The choice of \( X \) is discussed in the following subsections. The log-ratio form of the dependent variable makes mothers’ age to be limited to [13, 46]. This strategy is similar to the one in the logit model where the probability is confined to [0, 1]. Therefore, \( \frac{T_i - 13}{46 - T_i} =: \eta_i \) follows log-normal distribution \( \log-N(X_i^\top \theta_0, \sigma_0^2) \).
The density of $T|X$ is obtained by calculating the distribution function,

$$F_{0T|X}(t|x) = P(T \leq t | X = x) = \Phi \left( \frac{\log \frac{t-13}{46-t} - x^T \theta_0}{\sigma_0} \right),$$

where $\Phi$ and $\phi$ are distribution and density functions of a standard normal random variable. For the conditional density of $T|X, Y$, we assume the relationship $\log \left( \frac{T_i-13}{46-T_i} \right) = U_i^T \vartheta_0 + \varepsilon_i$, where $\varepsilon_i$ is independent of $U_i$ and has density $N(0, \varsigma_0)$. Therefore, the conditional distribution function $f_{0T|U}(t|u) = \phi \left( \frac{\log \frac{t-13}{46-t} - u^T \vartheta_0}{\varsigma_0} \right) \frac{1}{\varsigma_0} \left( \frac{1}{t-13} + \frac{1}{46-t} \right)$.

**5.3 Empirical results**

**5.3.1 Mother’s weight gain during pregnancy**

The results regarding the mothers’ weight gain during pregnancy show evidence that, after controlling for a mother’s characteristics chosen (i.e., age, marital status, years of education, number of cigarettes per day, and the month of first prenatal care visit), in general, greater weight gain during pregnancy leads to higher birthweight. Just as the estimation of the unconditional treatment effects of the mother’s age, we are estimating the unconditional treatment effects of mother’s weight gain during pregnancy. Figure 2 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 3.
at the extremes of weight gain due to the sparsity of the data at the extremes.

Table 2 describes treatment effects for selected weight treatment effects. It is divided according to the weight gain interval effects. The first part contains 20 pound effects. The second contains 40 pound effects, and so on until an 80 pound interval. The results show that the impact of gaining weight is positive.

[Figures 2 and 3 and Table 2 about here]

In summary, to produce an infant with healthy birthweight, mothers should gain weight between approximately 20 to 40 pounds. The average birthweight is below 2.5 kilograms for mothers with weight gain less than around 10 pounds and is above 4 kilograms for mothers with weight gain more than around 80 pounds, in contrast to the study of mothers’ ages, in which the average birthweight is always between 2.5 and 4 kilograms. It seems optimal for pregnant women to gain between 20 to 40 pounds to lower the chances of having LBW or HBW infants.

5.3.2 Mothers’ age

The QDRF of the mother’s age on birthweight is negative. For a given age, this negative impact becomes more severe for lower parts of the distribution of birthweight. Although intuitive, this result complements 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 mothers’ age. Third, unlike using regression framework, our results show that the treatment effects are not confined to be constant or a linear function of ages.

In the current study of the mothers’ age, we control for characteristics of mothers, e.g., 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 treatment effects. The empirical results for treatment effects of the mother’s age on birthweight reveal that the treatment effect is negative; that is, as expected, the birthweight decreases as the mother’s age increases. Figure 4 plots the point estimates for the mean, 10%, 90%, and the three quartiles of birthweights for mothers’ ages from 14 to 45. This impact of the mother’s age on birthweight is negative for all the quantiles. However, for a given age, this 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 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.

From the disaggregated figures (Figure 5) one can see that the 90% confidence bands are narrower in the middle ages because there are more data for that age range. In contrast, we can see that the confidence interval for 10% quantile at the age of 45 is relatively wide.

Table 3 describes the treatment effects for selected age treatment effects. The table is divided according to the age interval effects. The first part contains 5 year effects. The second part contains 10 year effects, and so on, until a 30 year interval. Most of them are statistically significant, and negative values show evidence that aging is negatively related to birthweight. Finally, the effect is larger (in absolute values) for the low part of the distribution of birthweights; for example, for a mother aged 25 to 35 years the treatment effect is -0.08 at 10% and -0.05 at 90%.

In general, there are certain risks of having a baby when the mother is too young or too old. Although on average the birthweight is within the “healthy range” between 2.5 and 4 kilograms, our estimates show that mothers younger than 20 years are likely to have HBW infants, while mothers older than 44 years are likely to have LBW infants. Therefore, it may be prudent for women who plan to have a baby to do so approximately between 20 and 44 years of age. To prevent female teenagers from having unexpected babies, more education and other forms of help may be needed.

6 Conclusion

In this paper, we first study the identification of the dose-response function 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. Sufficient conditions are provided for the estimator to be consistent, converge weakly, and to be semiparametric efficient. We study hypothesis testing procedures based on 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 infinite dimensional and the weak limits of test statistics are not standard, we use the bootstrap method when conducting inferences. Finally, we apply our estimation methods to the study of unconditional effects of mothers’ age and weight gain during pregnancy on infants’ birthweight, illustrating the usefulness of the new estimator.
7 Appendix

In Appendix A, we provide asymptotic properties of a generic Z-estimator. More specifically, we describe the model, the regularity conditions, and state the asymptotic results. The proofs of these results are collected in Appendix B. Finally, Appendix C collects the proofs of the results given in the text.

In Lemmas 1 and 2 below, we provide verifiable sufficient conditions for consistency and weak convergence of generic moment restriction estimators (Z-estimators) with possibly non-smooth functions and a nuisance parameter, when both the parameter of interest and the nuisance parameter are possibly infinite dimensional. Lemma 3 establishes the validity of the bootstrap. These general results are used to prove the asymptotic properties of the two-step estimator discussed above. In this general setting, the data need not be independent and identically distributed (i.i.d.). These approaches and results are similar to those in van der Vaart (2002) and van der Vaart and Wellner (2007). While these later works provide high level conditions, we describe simpler verifiable conditions for Z-estimators. The results presented here also extend those of Chen, Linton, and Van Keilegom (2003) in that the parameter of interest is not in a Euclidean space but in a generic Banach space. Moreover, the results extend Theorem 3.3.1 of van der Vaart and Wellner (1996) in that a possibly infinite dimensional nuisance parameter needs to be estimated in the first step.

7.1 Appendix A

Let $\Theta$ and $\mathcal{L}$ denote Banach spaces, and $\mathcal{H}$ a norm space, with norms $\|\cdot\|_\Theta$, $\|\cdot\|_\mathcal{H}$, and $\|\cdot\|_\mathcal{L}$, respectively. Let $Z_n : \Theta \times \mathcal{H} \mapsto \mathcal{L}$, $Z : \Theta \times \mathcal{H} \mapsto \mathcal{L}$ be random maps and a deterministic map, respectively. We suppress the dependence of $Z$ on $n$ for simplicity. The Z-estimator $\hat{\theta}$ is defined as the root of

$$Z(\theta, \hat{h}) = 0,$$

where $\hat{h}$ is a first step estimation of a possibly infinite dimensional nuisance parameter. This general theory is an extension of Theorem 1 of Chen, Linton, and Van Keilegom (2003) in that the parameter of interest is a Banach valued quantity instead of a Euclidean vector, and of Theorem 3.3.1 of van der Vaart and Wellner (1996) to the model with a nuisance parameter.
7.1.1 Consistency

We first derive a general consistency result for a Z-estimator in a Banach space. To obtain the consistency of the generic Z-estimator, we impose the following conditions.

C.1 \[ ||Z(\hat{\theta}, \hat{h})||_L = o_p^*(1). \]

C.2 \[ ||Z(\theta_n, h_0)||_L \rightarrow 0 \text{ implies } \theta_n \rightarrow \theta_0 \text{ for any sequences } \theta_n \in \Theta. \]

C.3 Uniformly in \( \theta \in \Theta \), \( Z(\theta, h) \) is continuous at \( h_0 \).

C.4 \[ ||\hat{h} - h_0||_H = o_p^*(1). \]

C.5 For all sequences \( \delta_n \downarrow 0 \),

\[ \sup_{\theta \in \Theta : ||h - h_0||_H \leq \delta_n} \frac{||Z(\theta, h) - Z(\theta, h)||_L}{1 + ||Z(\theta, h)||_L + ||Z(\theta, h)||_L} = o_p^*(1). \]

Condition C.1 requires that \( \hat{\theta} \) solves the estimating equation \( ||Z(\theta, \hat{h})||_L = 0 \) only asymptotically. Condition C.2 is an identification of the parameter. Condition C.3 is a smooth assumption of \( Z \) in \( h \) only at \( h_0 \). Condition C.4 requires that the nuisance parameter is consistently estimated. Condition C.5 is a high level assumption and can be stated in more primitive conditions for specific cases. Further, condition C.5 is implied by the following uniform convergence condition of \( Z \) to \( Z \).

C5S For any sequences \( \delta_n \downarrow 0 \),

\[ \sup_{\theta \in \Theta : ||h - h_0||_H \leq \delta_n} ||Z(\theta, h) - Z(\theta, h)||_L = o_p^*(1). \]

This set of conditions are similar to conditions of Theorem 1 of Chen, Linton, and Van Keilegom (2003).

The following lemma summarizes the consistency of a generic Z-estimator.

Lemma 1. Suppose that \( \theta_0 \in \Theta \) satisfies \( Z(\theta_0, h_0) = 0 \) with \( h_0 \in \mathcal{H} \) and that conditions C.1–C.5 hold. Then \( ||\hat{\theta} - \theta_0||_\Theta = o_p^*(1). \)

Proof. See Appendix B. \( \Box \)
7.1.2 Weak Convergence

Now we provide a general result for the Z-estimator. For the proof of weak convergence of the Z-estimator, consistency is assumed without loss of generality. Therefore the parameter space is replaced by $\Theta_\delta \times H_\delta$ where $\Theta_\delta := \{ \theta \in \Theta : ||\theta - \theta_0||_\Theta < \delta \}$ as in Chen, Linton, and Van Keilegom (2003) and $H_\delta := \{ h \in H : ||h - h_0||_H < \delta \}$.

Because the parameter spaces are a Banach and a normed space, we need a notion of derivatives for maps from a Banach or a normed space to a Banach space. Let $\Theta$ and $L$ denote Banach spaces, and $H$ a normed space. Fréchet differentiability of a map $\phi : \Theta \mapsto L$ at $\theta \in \Theta$ means that there exists a continuous, linear map $\phi'_\theta : \Theta \mapsto L$ with

$$\frac{||\phi(\theta + h_n) - \phi(\theta) - \phi'_\theta(h_n)||}{||h_n||} \rightarrow 0$$

for all sequences $\{h_n\} \subset \Theta$ with $||h_n|| \rightarrow 0$ and $\theta + h_n \in \Theta$ for all $n \geq 1$; see, e.g., p. 26 of Kosorok (2008). Pathwise derivative of a map $\varphi : H \mapsto L$ at $h \in H$ in the direction $[\bar{h} - h]$ is

$$\varphi'_h[\bar{h} - h] = \lim_{\rho \rightarrow 0} \frac{\varphi(h + \rho(\bar{h} - h)) - \varphi(h)}{\rho}$$

with $\{h + \rho(\bar{h} - h) : \rho \in [0, 1]\} \subset H$, provided that the limit exists. To obtain the weak limit, we impose the following sufficient conditions.

G.1 $||Z(\hat{\theta}, \hat{h})||_L = o_p^*(n^{-1/2})$.

G.2 The map $\theta \mapsto Z(\theta, h_0)$ is Fréchet differentiable at $\theta_0$ with a continuously invertible derivative $Z_1(\theta_0, h_0)$.

G.3 For all $\theta \in \Theta_\delta$ the pathwise derivative $Z_2(\theta, h_0)[h - h_0]$ of $Z(\theta, h_0)$ exists in all directions $[h - h_0] \in H$. Moreover, for all $(\theta, h) \in \Theta_\delta \times H_\delta$ with a positive sequence $\delta_n = o(1)$:

- G31 $||Z(\theta, h_0) - Z(\theta, h) - Z_2(\theta, h_0)[h - h_0]||_L = 0$ uniformly in $\theta$.
- G32 $||Z_2(\theta, h_0)[h - h_0] - Z_2(\theta_0, h_0)[h - h_0]||_L \leq o(1)\delta_n$.

G.4 The estimator $\hat{h} \in H$ with probability tending to one; and $||\hat{h} - h_0||_H = o_p^*(n^{-1/4})$.

G.5 For any $\delta_n \downarrow 0$,

$$\sup_{||\theta - \theta_0|| \leq \delta_n, ||h - h_0||_H \leq \delta_n} \frac{||\sqrt{n}(\bar{Z}(\theta, h) - \bar{Z}(\theta_0, h_0))||_L}{1 + \sqrt{n}||Z(\theta, h)||_L + \sqrt{n}||Z(\theta, h)||_L} = o_p^*(1).$$
\[ \sqrt{n}(Z_2(\theta_0, h_0)[\hat{h} - h_0] + (Z - Z)(\theta_0, h_0)) \] converges weakly to a tight random element \( G \) in \( L \).

Condition \( G.1 \) requires \( \hat{\theta} \) to solve the estimating equation only asymptotically. Conditions \( G.2 \) and \( G.3 \) are smoothness conditions for \( Z \). Condition \( G.4 \) is the same as condition (2.4) of Chen, Linton, and Van Keilegom (2003). Conditions \( G.5 \) and \( G.6 \) are high-level assumptions, and more primitive conditions are provided for more specific cases. Moreover, condition \( G.5 \) is implied by

\[ \text{G.5'} \quad \text{For any } \delta_n \downarrow 0, \]

\[ \sup_{||\theta - \theta_0|| \leq \delta_n, ||h - h_0|| \leq \delta_n} ||\sqrt{n}(Z - Z)(\theta, h) - \sqrt{n}(Z - Z)(\theta_0, h_0)||_L = o_p(1). \]

Now we provide a general result for \( Z \)-estimators.

**Lemma 2.** Suppose that \( \theta_0 \in \Theta_0 \) satisfies \( Z(\theta_0, h_0) = 0 \), that \( \hat{\theta} = \theta_0 + o_p(1) \), and that conditions \( G.1 \)–\( G.6 \) hold. Then \( \sqrt{n}(\hat{\theta} - \theta_0) \leadsto Z^{-1}_1(\theta_0, h_0)G \).

**Proof.** See Appendix B. \qed

### 7.1.3 The Validity of the Bootstrap

There are two potential difficulties when constructing the confidence bands for the DRF. First, closed-form expressions of the covariance kernel are hard to calculate. This mainly is due to the estimation of the nuisance parameters. Second, even if closed-form expressions of the covariance kernel are available, they are useful only when the set \( \mathcal{T} \) is finite. Thus, we use the ordinary nonparametric bootstrap method to determine the rejection regions of the tests. We show that the bootstrap estimator of the asymptotic distribution of \( \sqrt{n}(\hat{\beta}(t) - \beta_0(t)) \) is consistent. It is without loss of generality to study the validity of bootstrap for \( \sqrt{n}(\hat{\beta}(t) - \beta(t)) \) by the continuous mapping theorem. Let \( \{(X_i^*, Y_i^*, T_i^*)\}_{i=1}^n \) be randomly drawn with replacement from \( \{(X_i, Y_i, T_i)\}_{i=1}^n \). Let \( \hat{\pi}^* \) be the estimator of \( \pi_0 \) using \( \{(X_i^*, Y_i^*, T_i^*)\}_{i=1}^n \). Let \( Z^*(\beta, \pi) \) denote the resampled average. The bootstrap estimator \( \hat{\beta}^* \) satisfies

\[ ||Z^*(\hat{\beta}^*, \hat{\pi}^*)|| = o_p(n^{-1/2}). \]

Following Chen, Linton, and Van Keilegom (2003), an asterisk denotes a probability or moment computed under the bootstrap distribution conditional on the original data set.
Consider the following conditions:

**G4** With $P^*$-probability tending to one, $\hat{\pi}^* \in \Pi$ and $||\hat{\pi}^* - \pi||_\Pi = o_p(n^{-1/4})$.

**G5** For any $\delta_n \downarrow 0$,

$$\sup_{||\beta - \beta_0|| \leq \delta_n, ||\pi - \pi_0||_\Pi \leq \delta_n} ||\sqrt{n}(Z^* - Z)(\beta, \pi) - \sqrt{n}(Z^* - Z)(\beta_0, \pi_0)||_L = o_p(1).$$

**G6** $\sqrt{n}(Z_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}] + (Z^* - Z)(\hat{\beta}, \hat{\pi}))$ converges weakly to a tight random element $G$ in $L$ in $P^*$-probability.

**Lemma 3.** Replace "$\theta$" and "$h$" by "$\beta$" and "$\pi$", respectively, in conditions **G.1**–**G.6** and **G.I**–**G.VI**. Suppose $\beta_0 \in \text{int}(B)$ and $\hat{\beta} \overset{a.s.}{\to} \beta_0$. Assume that conditions **G.1**, **G.4**, **G.V**, and **G.VI** are satisfied with “in probability” replaced by “almost surely”. Let conditions **G.2** and **G.3** hold with $\pi_0$ replaced by $\pi \in \Pi_{\delta_n}$. Also, assume that $Z_1(\beta; \pi)$ is continuous in $\pi$ at $\beta = \beta_0$ and $\pi = \pi_0$. Then $\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \Rightarrow Z_1^{-1}(\beta_0, \pi_0)G$ in $P^*$-probability.

**Proof.** See Appendix B.

### 7.2 Appendix B

This appendix collects the proofs for the asymptotic properties of a generic $Z$-estimator described in Lemmas 1-3 above.

**Proof of Lemma 1.** By condition **C.2**, it suffices to show that $||Z(\hat{\theta}, h_0)||_L = o_p(1)$. Using triangle inequality,

$$||Z(\hat{\theta}, h_0)||_L \leq ||Z(\hat{\theta}, h_0) - Z(\hat{\theta}, \hat{h})||_L + ||Z(\hat{\theta}, \hat{h}) - Z(\hat{\theta}, \hat{h})||_L + ||Z(\hat{\theta}, \hat{h})||_L.$$  \hspace{1cm} (7)

By conditions **C.3** and **C.4**, $||Z(\hat{\theta}, h_0) - Z(\hat{\theta}, \hat{h})||_L = o_p(1)$. By condition **C.1**, $||Z(\hat{\theta}, \hat{h})||_L = o_p(1)$. Also,$$

||Z(\hat{\theta}, \hat{h}) - Z(\hat{\theta}, \hat{h})||_L = o_p(1) + o_p(||Z(\hat{\theta}, \hat{h})||_L) + o_p(||Z(\hat{\theta}, \hat{h})||_L)$$

$$= o_p(1) + o_p(1) + o_p(||Z(\hat{\theta}, h_0)||_L) + o_p(1)$$

where the first equality is by condition **C.5** and the second equality is by conditions **C.1** and **C.3**. Therefore, inequality (7) implies $||Z(\hat{\theta}, h_0)||_L \leq o_p(1)$ and hence the result.  \hspace{1cm} $\square$
Proof of Lemma 2. Step 1: \( \sqrt{n} \)-consistency

We start the proof by showing that \( \hat{\theta} \) is \( \sqrt{n} \)-consistent for \( \theta_0 \) in \( \Theta \). By definition, the Fréchet differentiability of \( Z(\theta, h_0) \) implies the existence of a continuous linear map \( Z_1(\theta_0, f_0) \) such that

\[
\frac{||Z(\theta, f_0) - Z(\theta_0, f_0) - Z_1(\theta_0, f_0)(\theta - \theta_0)||_\mathcal{L}}{||\theta - \theta_0||_\Theta} = o(1).
\]

By triangle inequality, it follows

\[
||Z_1(\theta_0, h_0)(\theta - \theta_0)||_\mathcal{L} \leq ||Z(\theta, h_0) - Z(\theta_0, h_0)||_\mathcal{L} + o(||\theta - \theta_0||_\Theta).
\]

Since the derivative \( Z_1(\theta_0, h_0) \) is continuously invertible by condition G.2, there exists a positive constant \( c \) such that

\[
||Z_1(\theta_0, h_0)(\theta_1 - \theta_2)||_\mathcal{L} \geq c||\theta_1 - \theta_2||_\Theta \text{ for every } \theta_1 \text{ and } \theta_2 \in \Theta_{\delta}.
\]

Therefore, it follows

\[
(c - o(1))||\theta - \theta_0||_\Theta \leq ||Z(\theta, h_0) - Z(\theta_0, h_0)||_\mathcal{L},
\]

and

\[
(c - o_p(1))||\hat{\theta} - \theta_0||_\Theta \leq ||Z(\hat{\theta}, h_0) - Z(\theta_0, h_0)||_\mathcal{L} = ||Z(\hat{\theta}, h_0)||_\mathcal{L}
\]

with probability tending to one. By triangle inequality and conditions G.1 and G.6, the right hand side of the previous inequality is bounded by

\[
||Z(\hat{\theta}, h_0) - Z(\hat{\theta}, \hat{h})||_\mathcal{L} + ||Z(\hat{\theta}, \hat{h}) - Z(\hat{\theta}, h_0) + Z(\theta_0) - Z(\theta_0, h_0)||_\mathcal{L} + O_p(n^{-1/2}).
\]

For the first term, we have

\[
||Z(\hat{\theta}, h_0) - Z(\hat{\theta}, \hat{h})||_\mathcal{L} \leq ||Z(\hat{\theta}, h_0) - Z(\hat{\theta}, \hat{h}) - Z_2(\hat{\theta}, h_0)\hat{h} - h_0||_\mathcal{L}
\]

\[
+ ||Z_2(\hat{\theta}, h_0)\hat{h} - h_0||_\mathcal{L} - Z_2(\theta_0, h_0)\hat{h} - h_0||_\mathcal{L}
\]

\[
\leq o_p(1) ||\theta - \theta_0||_\Theta + O_p(1) + O_p(n^{-1/2})
\]

\[
\leq ||Z(\hat{\theta}, h_0)||_\mathcal{L} \times o_p(1) + O_p(n^{-1/2}),
\]

where the first inequality is by triangle inequality, the second one by conditions G.3 and G.6, and the third by inequality (8).
As for the second term in (10), by condition \textbf{G.5},

$$||Z(\hat{\theta}, h) - Z(\hat{\theta}, \hat{h}) + Z(\theta_0, h_0) - Z(\theta_0, h_0)||_\mathcal{L} = o_p^*(1/\sqrt{n}) + ||Z(\hat{\theta}, \hat{h})||_\mathcal{L} + ||Z(\hat{\theta}, h)||_\mathcal{L}$$

$$= o_p^*(1/\sqrt{n}) + o_p^*(||Z(\hat{\theta}, \hat{h})||_\mathcal{L})$$

The second equality is by condition \textbf{G.1}, \(||Z(\hat{\theta}, \hat{h})||_\mathcal{L} = o_p^*(1/\sqrt{n})\). By triangle inequality,

$$||Z(\hat{\theta}, h)||_\mathcal{L} \leq ||Z(\hat{\theta}, \hat{h}) - Z(\hat{\theta}, h) + Z(\theta_0, h_0) - Z(\theta_0, h_0)||_\mathcal{L} + O_p^*(1/\sqrt{n}).$$

It then follows

$$(1 - o_p^*(1))||Z(\hat{\theta}, h) - Z(\hat{\theta}, \hat{h}) + Z(\theta_0, h_0) - Z(\theta_0, h_0)||_\mathcal{L} \leq o_p^*(1/\sqrt{n})$$

Thus, formula (10) is bounded by

$$||Z(\hat{\theta}, h_0)||_\mathcal{L} \times o_p^*(1) + O_p^*(n^{-1/2})$$

and the right side of the equality in (9) satisfies

$$(1 - o_p^*(1))||Z(\hat{\theta}, h_0)||_\mathcal{L} \leq O_p^*(n^{-1/2}). \quad (11)$$

Therefore, \((c - o_p(1))\sqrt{n}||\hat{\theta} - \theta_0||_\Theta \leq O_p^*(1)\) and \(\hat{\theta}\) is \(\sqrt{n}\)-consistent for \(\theta_0\) in \(\Theta\).

\textbf{Step 2: Weak Convergence}

Now we show the weak convergence. By conditions \textbf{G.2} and \textbf{G.3},

$$||Z(\hat{\theta}, h) - Z(\theta_0, h_0) - Z_1(\theta_0, h_0)(\hat{\theta} - \theta_0) - Z_2(\theta_0, h_0)\hat{h} - h_0||_\mathcal{L}$$

$$= ||Z(\hat{\theta}, h) - Z(\hat{\theta}, h_0) - Z_2(\hat{\theta}, h_0)(\hat{h} - h_0) + Z(\hat{\theta}, h_0) - Z(\theta_0, h_0) - Z_1(\theta_0, h_0)(\hat{\theta} - \theta_0) + Z_2(\hat{\theta}, h_0)(\hat{h} - h_0)||_\mathcal{L}$$

$$\leq ||Z(\hat{\theta}, h) - Z(\hat{\theta}, h_0) - Z_2(\hat{\theta}, h_0)(\hat{h} - h_0)||_\mathcal{L} + ||Z(\hat{\theta}, h_0) - Z(\theta_0, h_0) - Z_1(\theta_0, h_0)(\hat{\theta} - \theta_0)||_\mathcal{L}$$

$$+ ||Z_2(\hat{\theta}, h_0)(\hat{h} - h_0) - Z_2(\theta_0, h_0)(\hat{h} - h_0)||_\mathcal{L}$$

$$= o_p^*(n^{-1/2}) + o_p^*(n^{-1/2}) = o_p^*(n^{-1/2}).$$
Therefore, it follows

\[ Z_1(\theta_0, h_0) \sqrt{n} (\hat{\theta} - \theta_0) + \sqrt{n} Z_2(\theta_0, h_0) (\hat{h} - h_0) = \sqrt{n} (Z(\theta_0, h_0) - Z(\theta_0, h_0)) + o_p(1) \]

and

\[ Z_1(\theta_0, h_0) \sqrt{n} (\hat{\theta} - \theta_0) = -\sqrt{n} (Z_2(\theta_0, h_0) (\hat{h} - h_0) + (Z - Z(\theta_0, h_0)) + o_p(1) \approx \mathbb{G} \]

by condition G.6.

Now by condition G.2 and the continuous mapping theorem, we have \( \sqrt{n}(\hat{\theta} - \theta_0) \approx Z_1^{-1}(\theta_0, h_0)\mathbb{G} \).

\[ \square \]

**Proof of Lemma 3.** The assertion that \( ||\hat{\beta}^* - \hat{\beta}|| = O_p(n^{-1/2}) \) a.s. [P] can be shown in a similar way as the proof of the \( \sqrt{n} \)-consistency of \( \hat{\beta} \). Therefore we omit the proof and only show the weak convergence in probability of the bootstrap estimator.

Note that

\[
\begin{align*}
||Z^* (\hat{\beta}^*, \hat{\pi}^*) - Z^*(\hat{\beta}, \hat{\pi}) - Z_1 (\hat{\beta}^*, \hat{\pi}^*) (\hat{\beta}^* - \hat{\beta}) - Z_2 (\hat{\beta}, \hat{\pi}) (\hat{\pi}^* - \hat{\pi}) || \\
= ||Z(\hat{\beta}^*, \hat{\pi}^*) - Z(\hat{\beta}, \hat{\pi}) - Z_2 (\hat{\beta}, \hat{\pi}) (\hat{\pi}^* - \hat{\pi}) + Z(\hat{\beta}^*, \hat{\pi}^*) - Z(\hat{\beta}, \hat{\pi}) - Z_1 (\hat{\beta}, \hat{\pi}) (\hat{\beta}^* - \hat{\beta}) \\
+ ([Z^*(\hat{\beta}^*, \hat{\pi}^*) - Z(\hat{\beta}, \hat{\pi})) - (Z^*(\hat{\beta}, \hat{\pi}) - Z(\hat{\beta}, \hat{\pi})) ] \\
+ ([Z(\hat{\beta}^*, \hat{\pi}^*) - Z(\hat{\beta}, \hat{\pi}^*)) - (Z(\hat{\beta}, \hat{\pi}) - Z(\hat{\beta}, \hat{\pi})) ] \\
+ Z_2 (\hat{\beta}, \hat{\pi}) (\hat{\pi}^* - \hat{\pi}) - Z_2 (\hat{\beta}^*, \hat{\pi}) (\hat{\pi}^* - \hat{\pi}) || \\
\leq ||Z(\hat{\beta}^*, \hat{\pi}^*) - Z(\hat{\beta}, \hat{\pi}) - Z_2 (\hat{\beta}, \hat{\pi}) (\hat{\pi}^* - \hat{\pi}) || + ||Z(\hat{\beta}^*, \hat{\pi}) - Z(\hat{\beta}, \hat{\pi}) - Z_1 (\hat{\beta}, \hat{\pi}) (\hat{\beta}^* - \hat{\beta}) || \\
+ ||(Z^*(\hat{\beta}^*, \hat{\pi}^*) - Z(\hat{\beta}^*, \hat{\pi}^*)) - (Z^*(\hat{\beta}, \hat{\pi}) - Z(\hat{\beta}, \hat{\pi})) || \\
+ ||(Z(\hat{\beta}^*, \hat{\pi}^*) - Z(\hat{\beta}^*, \hat{\pi}^*)) - (Z(\hat{\beta}, \hat{\pi}) - Z(\hat{\beta}, \hat{\pi})) || \\
+ ||Z_2 (\hat{\beta}, \hat{\pi}) (\hat{\pi}^* - \hat{\pi}) - Z_2 (\hat{\beta}^*, \hat{\pi}) (\hat{\pi}^* - \hat{\pi}) || \\
= o_p(n^{-1/2}).
\]

The first term is \( o_p(n^{-1/2}) \) by condition G3 (version of Theorem 4) and G4B. The second term is \( o_p(n^{-1/2}) \) by condition G2 (version of Theorem 4) and \( \sqrt{n} \)-consistency of \( \hat{\beta}^* \). The third and fourth terms are \( o_p(n^{-1/2}) \) by the triangular inequality and conditions G5' (almost sure version) and G5B. And the fifth term is \( o_p(n^{-1/2}) \) by condition G3 (version of Theorem 4) and \( \sqrt{n} \)-consistency of \( \hat{\beta}^* \).
Therefore, it follows
\[
Z_1(\hat{\beta}, \hat{\pi})\sqrt{n}(\hat{\beta}^* - \hat{\beta}) + \sqrt{n}Z_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}] = \sqrt{n}(Z^*(\hat{\beta}^*, \hat{\pi}^*) - Z^*(\hat{\beta}, \hat{\pi})) + o_p(1)
\]
\[
= - \sqrt{n}(Z^*(\hat{\beta}, \hat{\pi}) - Z(\hat{\beta}, \hat{\pi})) + o_p(1)
\]
and
\[
Z_1(\hat{\beta}, \hat{\pi})\sqrt{n}(\beta_0 - \hat{\beta}) = - \sqrt{n}Z_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}] - \sqrt{n}(Z^*(\hat{\beta}, \hat{\pi}) - Z(\hat{\beta}, \hat{\pi})) + o_p(1) \Rightarrow \mathbb{G}
\]
in \mathcal{L} in \(P^\ast\)-probability by condition \(G.6\). We can replace \(Z_1(\hat{\beta}, \hat{\pi})\) by \(Z_1(\beta_0, \pi_0)\) with probability one. Now by condition \(G.2\) (version of \textit{Theorem 4}) and the continuous mapping theorem, we have \(\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow Z^{-1}_1(\beta_0, \pi_0)\mathbb{G} \)

\[\square\]

\section{Appendix C}

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

\textit{Proof of Theorem 1.} 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,

\[
E[m(Y(t_0); \beta(t_0))|X] = E[m(Y(t_0); \beta(t_0))|X, T = t_0] = E[m(Y; \beta(t_0))|X, T = t_0] = \lim_{\Delta t \downarrow 0} E[m(Y; \beta(t_0))|X, T \in [t_0, t_0 + \Delta t]],
\]

where the first equality is by condition \textit{I.III.1}, the second equation is by the fact that if \(T = t_0\), then \(Y = Y(t_0)\), and the third equality is by condition \textit{I.III.2}. Moreover, we have

\[
E[m(Y; \beta(t_0))|X, T \in [t_0, t_0 + \Delta t]] = E[1(T \in [t_0, t_0 + \Delta t])m(Y; \beta(t_0))|X, T \in [t_0, t_0 + \Delta t]]. \quad (12)
\]

By law of total expectation

\[
E[1\{T \in [t_0, t_0 + \Delta t]\}m(Y; \beta(t_0))|X]
\]
\[
= E[1\{T \in [t_0, t_0 + \Delta t]\}m(Y; \beta(t_0))|X, T \in [t_0, t_0 + \Delta t]]P(T \in [t_0, t_0 + \Delta t]|X)
\]
\[
+ E[1\{T \in [t_0, t_0 + \Delta t]\}m(Y; \beta(t_0))|X, T \notin [t_0, t_0 + \Delta t]]P(T \notin [t_0, t_0 + \Delta t]|X),
\]
the right hand side of equation (12) equals

\[ \frac{\mathbb{E}[\mathbf{1}\{T \in [t_0, t_0 + \Delta t]\}m(Y; \beta(t_0))|\mathbf{X}]}{\mathbb{P}(T \in [t_0, t_0 + \Delta t]|\mathbf{X})} = \frac{\mathbb{E}[\mathbf{1}\{T \leq t_0 + \Delta t\} - \mathbf{1}\{T < t_0\}]m(Y; \beta(t_0))|\mathbf{X}]}{F_{T|\mathbf{X}}(t_0 + \Delta t|\mathbf{X}) - F_{T|\mathbf{X}}(t_0|\mathbf{X})}, \]

where \( F_{T|\mathbf{X}} \) denotes the conditional distribution function of \( T \) given \( \mathbf{X} \). Noting that

\[
\mathbb{E}[\mathbf{1}\{T \leq t_0 + \Delta t\} - \mathbf{1}\{T \leq t_0\}]m(Y; \beta(t_0))|\mathbf{X} = \mathbf{x} = \int \int (1\{t \leq t_0 + \Delta t\} - 1\{t \leq t_0\})m(y; \beta(t_0))f_{T,Y|X}(t,y|\mathbf{x})\,dt\,dy
\]

it follows

\[
\lim_{\Delta t \downarrow 0} \mathbb{E}[m(Y; \beta(t_0))|\mathbf{X} = \mathbf{x}, T \in [t_0, t_0 + \Delta t]] = \lim_{\Delta t \downarrow 0} \int \int_{t_0}^{t_0+\Delta t} m(y; \beta(t_0))f_{T,Y|X}(t,y|\mathbf{x})\,dt\,dy
\]

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 I.III.1 and dominated convergence theorem. Hence \( \mathbb{E}[m(Y(t_0); \beta(t_0))] = \mathbb{E} \left[ m(Y; \beta(t_0)) f_{T,Y|X(t_0),Y|X} \right]. \)

**Proof of Theorem 2.** The general result in the previous lemma for consistency of the Z-estimator can be applied to our continuous treatment model as stated in the following theorem with \( \theta_0 = \beta_0(t) \), \( h_0 = \pi(\cdot; t) \), \( \mathbb{Z}(\theta, h)(t) = \mathbb{E}\psi_{\theta, \pi, t} \), and \( Z(\theta, h)(t) = \mathbb{E}\psi_{\theta, \pi, t} \), where \( \psi_{\theta, \pi, t} = m(y; \beta(t))\pi(u; t) \). In this case, \( \Theta = \mathcal{L} = \ell^\infty(\mathcal{T}) \) and \( ||\cdot||_\Theta = ||\cdot||_\mathcal{L} = ||\cdot||_\infty \), while \( \mathcal{H} = \Pi, \) a function class with domain \( \mathcal{U} \times \mathcal{T} \), and \( ||\cdot||_\mathcal{H} = ||\cdot||_\Pi = \sup_{t \in \mathcal{T}} \sup_{u \in \mathcal{U}} ||\cdot|| = ||\cdot||_\infty. \) For any \( \delta > 0 \), \( \Pi_\delta = \{ \pi \in \Pi : ||\pi - \pi_0||_\infty < \delta \}. \)

39
First we show that condition C.3, the continuity of \( E m(Y; \beta(t)) \pi(U; t) \) at \( \pi_0 \) uniformly over \( \beta(t) \in \ell^\infty(T) \), is satisfied. For any \( ||\pi - \pi_0||_\infty \leq \delta \), which is equivalent to \( \sup_{t \in T} \sup_u |\pi(u; t) - \pi_0(u; t)| \leq \delta \), we have

\[
\begin{align*}
|E[m(Y; \beta(t))\pi(U; t)] - E[m(Y; \beta(t))\pi_0(U; t)]|_\infty \\
= |E[m(Y; \beta(t))(\pi(U; t) - \pi_0(U; t))]|_\infty \leq |E[m(Y; \beta(t))]|_\infty \delta.
\end{align*}
\]

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


\[ \square \]

**Proof of Corollary 1, Consistency of \( \hat{\mu}(t) \).** To show consistency, we verify the conditions of **Theorem 2**. Note that in this case we have a closed form solution \( \mu_0(t) = \frac{E Y \pi_0(U; t)}{E \pi_0(U; t)} \).

Condition C.I is verified by the fact that \( \hat{\mu} = \frac{E Y \pi_0(U; t)}{E \pi_0(U; t)} \) is the exact zero of the estimating equation.

For condition C.II, if there is a sequence \( \mu_n(t) \) such that \( E[(Y - \mu_n(t))\pi_0(U; t)] \to 0 \) uniformly, then \( \mu_n(t) \to \frac{E Y \pi_0(U; t)}{E \pi_0(U; t)} = \mu_0(t) \) uniformly. To see this, we note that \( E[(Y - \mu_n(t))\pi_0(U; t)] = o(1) \) implies \( \mu_n(t) = \frac{E Y \pi_0(U; t)}{E \pi_0(U; t)} + o(1) = \frac{E Y \pi_0(U; t)}{E \pi_0(U; t)} + o(1) \).

Condition C.III is verified by the direct calculation

\[
|E[m(Y; \mu(t))]|_\infty = |E[Y - \mu(t)]|_\infty = E|Y| + |\mu(t)|_\infty,
\]

both of which are finite by condition AC.I.

As for condition C.V', noting that \( \psi_{1\mu,t} = y - \mu(t) \), automatically \( \{\psi_{1\beta,t} : \beta \in \ell^\infty(T), t \in T\} \) is Glivenko-Cantelli. To see this, note

\[
\sup_{\mu(t) \in \ell^\infty(T)} |E[Y - \mu(t)] - E[Y - \mu(t)]|_\infty = |EY - EY| \overset{p}{\to} 0
\]

by Khintchine's weak law of large numbers. It has envelope \( F_1(y) = y + |\mu(t)|_\infty \) by condition AC.I. Condition AC.II completes the verification of condition C.V'.

Hence, all the conditions of **Theorem 2** are satisfied. \[ \square \]
Proof of Corollary 2. Consistency of \( \hat{q}_r(t) \). 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.4 and QC.II

\[
\left| \mathbb{E}[(\tau - 1\{Y < \hat{q}_r(t)\})\hat{\pi}(U_t)] \right| \leq \text{const} \cdot \sup_{i \leq n} \hat{\pi}(U_i; t) \leq \text{const} \cdot \frac{\|\hat{\pi}(U; t)\|_\Pi + o_p(1)}{n} = O_p(1/n) .
\]

Condition C.II holds by condition QC.I. Condition C.III is satisfied because \( \tau - 1\{y < q_r(t)\} \) is a bounded function. Condition CV' is implied by the fact that the function class \( \{\psi_{1q,t} : q \in \ell^\infty(T), t \in T\} \) is Glivenko-Cantelli because it is a Vapnik-Červonenkis class and by condition QC.III. \( \square \)

Proof of Theorem 3. 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 \( \tilde{\pi} \) such that \( \{\pi_0 + \alpha(\tilde{\pi} - \pi_0) : \alpha \in [0, 1]\} \subset \Pi \),

\[
\frac{\big|E[m(Y; \beta) - E[m(Y; \beta)(\pi_0 + \alpha(\tilde{\pi} - \pi_0))]| - E[m(Y; \beta)\pi_0] \big|}{\alpha} = E[m(Y; \beta)(\tilde{\pi} - \pi_0)]
\]

and has the limit \( E[m(Y; \beta)(\tilde{\pi} - \pi_0)] \) as \( \alpha \to 0 \). Therefore \( Z_2(\beta, \pi_0)[\pi - \pi_0] = E[m(Y; \beta)(\pi - \pi_0)] \) in all directions \([\pi - \pi_0] \in \Pi \).

Condition G31 is satisfied by noting

\[
|E[m(Y; \beta(t))\pi_0(U_t)] - E[m(Y; \beta(t))\pi(U_t)] - E[m(Y; \beta(t))(\pi - \pi_0)(U_t)]|_\infty = 0 .
\]

And condition G32 is verified by

\[
|E[m(Y; \beta(t))(\pi - \pi_0)(U_t)] - E[m(Y; \beta_0(t))(\pi - \pi_0)(U_t)]|_\infty = |E[m(Y; \beta(t)) - m(Y; \beta_0(t))(\pi - \pi_0)(U_t)]|_\infty \leq |E[m(Y; \beta(t)) - E[m(Y; \beta_0(t))]|_\infty o(1) = \delta_n o(1) ,
\]

where the last equality is by condition G.III.

As for condition G.5, by Corollary 9.32 (iii) of Kosorok (2008), condition G.V implies that \( \{\psi_{\beta,\pi,t} : \beta \in \ell^\infty(\mathcal{T}), \pi \in \Pi_\delta, t \in \mathcal{T}\} \) is Donsker, which in turn implies G.5' by Lemma
3.3.5 of van der Vaart and Wellner (1996). Therefore, we obtain condition G.5 by condition G.1 and inequality (11).

Finally, G.VI is a representation of G.6.

\[ \text{Proof of Corollary 3, Weak Convergence of } \hat{\mu}(t). \] Condition G.1 is satisfied because the estimator is an exact Z-estimator.

The map \( \mu \mapsto E(Y - \mu)f_0(U) \) is Fréchet differentiable and is verified by the following calculation

\[
\frac{|E[(Y - \mu(t))\pi_0(U; t)] - E[(Y - \mu_0(t))\pi_0(U; t)] - E[\pi_0(U; t)(\mu(t) - \mu_0(t))]|_\infty}{|\mu(t) - \mu_0(t)|_\infty} = 0.
\]

Thus the Fréchet derivative is \( E\pi_0(U; t) \). For any \( \mu_1 \) and \( \mu_2 \),

\[
|E[\pi_0(U; t)\mu_1(t)] - E[\pi_0(U; t)\mu_2(t)]|_\infty \geq c|\mu_1(t) - \mu_2(t)|_\infty,
\]

and therefore is continuously invertible.

Condition G.III is verified by

\[
|E[Y - \mu(t)] - E[Y - \mu_0(t)]|_\infty = |\mu(t) - \mu_0(t)|_\infty.
\]

Condition G.V is implied by Conditions AG.I and AG.II and Corollary 9.32 (i) of Kosorok (2008) completes the verification.

Finally condition G.VI is implied by condition AG.III.

\[ \text{□} \]

\[ \text{Proof of Corollary 4, Weak Convergence of } \hat{\tau}_\tau(t). \] Condition G.I was verified in the proof of Corollary 2. For condition G.II, note that

\[
|E[(\tau - 1\{Y \leq q_{\tau_0}(t)\})\pi_0(U; t)] - E[(\tau - 1\{Y \leq q_{\tau_0}(t)\})\pi_0(U; t)] + E[\pi_0(U; t)f_Y(q_{\tau_0})(q_{\tau}(t) - q_{\tau_0}(t))]|_\infty \\
=|E[(1\{Y \leq q_{\tau_0}(t)\} - 1\{Y \leq q_{\tau}(t)\}) + f_Y(q_{\tau_0})(q_{\tau}(t) - q_{\tau_0}(t))]|_\infty M_\pi \\
\asymp|E[(1\{Y \leq q_{\tau_0}(t)\} - 1\{Y \leq q_{\tau}(t)\}) + f_Y(q_{\tau_0})(q_{\tau}(t) - q_{\tau_0}(t))]|_\infty M_\pi \\
=|F_Y(q_{\tau_0}(t)) - F_Y(q_{\tau}(t)) + f_Y(q_{\tau_0})(q_{\tau}(t) - q_{\tau_0}(t))|_\infty M_\pi = o(|q_{\tau}(t) - q_{\tau_0}(t)|_\infty).
\]
Condition G.III is satisfied because the distribution function of $Y$ is continuous. And condition G.V was verified in the proof of Corollary 2. Condition G.VI holds by condition QG.II. 

**Proof of Proposition 1.** A regular parametric submodel of the joint distribution of $(Y, T, X)$ with distribution function $F(y, t, x; \gamma)$ has the log-likelihood

$$\log f(y, t, x; \gamma) = \int_{\varsigma \in \mathcal{T}} \left[ \log f_{Y(\varsigma)}(y|\varsigma; \gamma) + \log f_{T|X}(\varsigma|\varsigma; \gamma) \right] d1\{\varsigma \geq t\} + \log f_X(x; \gamma)$$

with $F(y, t, x; \gamma_0) = F(y, t, x)$. The score of this model is

$$S(y, t, x; \gamma_0) = \frac{d}{d\gamma} \log f(y, t, x; \gamma_0)$$

$$= S_y(y, t, x) + S_T(t, x) + S_X(x),$$

where

$$S_y(y, t, x) = \frac{d}{d\gamma} \log f_{Y(t)}(y|x; \gamma_0)$$

$$S_T(t, x) = \frac{d}{d\gamma} \log f_{T|X}(t|x; \gamma_0)$$

$$S_X(x) = \frac{d}{d\gamma} \log f_X(x; \gamma_0),$$

Therefore, the tangent set of this model is the collection of the score functions of the form above and the tangent space $\mathcal{P}_F$ is the closed linear span of the tangent set.

Recall that $E[m(Y(t); \beta(t))] = 0$ if and only if $\beta(t) = \beta_0(t)$ for each $t \in \mathcal{T}$. By implicit function theorem,

$$\frac{\partial \beta_0(t)}{\partial \gamma}(\gamma) = -\Gamma_0^{-1}(t) \Upsilon(\gamma_0)$$
and
\[ \varUpsilon(\gamma_0) = \frac{\partial}{\partial \gamma} \int m(Y(t); \beta(t)) \, dF(y, t, x; \gamma_0) \]
\[ = E[m(Y(t); \beta_0(t)) S_y(Y(t), t, X)] + E[m(Y(t); \beta_0(t)) S_X(X)] \]

We need to find \( \Psi_{\beta}(y, t, x) \) such that
\[ \frac{\partial \beta_0(t)}{\partial \gamma} (\gamma) = E[\Psi_{\beta}(Y, T, X) S(Y, T, X; \gamma_0)] \]
for all regular parametric submodels.

It can be verified that
\[ \Psi_{\beta}(y, t, x) = -\Gamma_0^{-1}(t) \psi(y, t, x, \beta_0, \pi_0, e_0) \]
where \( \psi(y, t, x, \beta_0, \pi_0, e_0) = \pi(y, x, t) m(Y; \beta_0(t)) - (\pi(y, x, t) - 1)e_0(X, \beta_0(t)) \) with \( e_0(X, \beta(t)) = E[m(Y; \beta(t)) | X] \).

**Proof of Theorem 4.** We first verify that \( \sqrt{n} E[\psi(Y_i, X_i, t, \beta_0, \pi_0, e_0)] \) converges weakly in \( \ell^\infty(T) \). By condition GV, \( \psi_t = \psi(y, x, t, \beta_0, \pi_0, e_0) \) is Donsker, implying the weak convergence.

The uniform semiparametric efficiency follows from the weak convergence above and the pointwise semiparametric efficiency of Lemma 3 by Theorem 18.9 of Kosorok (2008).

Now we verify that the formula in condition GVI equals the left hand-side of condition E., which implies that the influence function of the two-step estimator is efficient. To this end, we begin with the formula in condition GVI,
\[ \sqrt{n}(E_m(Y; \beta_0(t))(\pi(U; t) - \pi_0(U; t)) | \pi = \hat{\pi} + E_m(Y; \beta_0(t))\pi_0(U; t)) \]
\[ = \sqrt{n}(E_m(Y; \beta_0(t))(\hat{\pi}(U; t) - \pi_0(U; t)) + E_m(Y; \beta_0(t))\pi_0(U; t)) = \sqrt{n}(E_m(Y; \beta_0(t))\hat{\pi}(U; t), \]
where the first equality is by condition G5’ which in turn is implied by G5’.

**Proof of Proposition 2.** By Mean Value Theorem, \( \pi(u; t; \hat{\vartheta}) - \pi(u; t; \vartheta_0) = \pi'(u; t; \vartheta^*)(\hat{\vartheta} - \vartheta_0) \)
\( \vartheta_0 \), where \( \vartheta^* \) is a convex combination of \( \hat{\vartheta} \) and \( \vartheta_0 \). Therefore,

\[
|\pi(\mathbf{u}; t; \hat{\vartheta}) - \pi(\mathbf{u}; t; \vartheta_0)|_{\infty} = |\pi'(\mathbf{u}; t; \vartheta^*)(\hat{\vartheta} - \vartheta_0)|_{\infty} \leq |\pi'(\mathbf{u}; t; \vartheta^*)|_{\infty}||\hat{\vartheta} - \vartheta_0|| = O_p(n^{-1/2})
\]
since \( |\pi'(\mathbf{u}; t; \vartheta^*)|_{\infty} \) is bounded and \( ||\hat{\vartheta} - \vartheta_0|| = O_p(n^{-1/2}) \). There, conditions C.IV and G.IV are verified.

For condition C.V, by Theorem 19.7 of van der Vaart (1998), this parametric Lipschitz continuous functional class is Donsker.

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 \( \phi : D_{\varphi} \mapsto E \) is Hadamard differentiable at \( \theta \in D \), tangentially to a set \( D_0 \), if there exists a continuous linear map \( \phi'_\theta : D \mapsto E \) such that

\[
\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h),
\]
as \( n \to \infty \), for all converging sequences \( t_n \to 0 \) and \( h_n \to h \in D_0 \), with \( h_n \in D \) and \( \theta + t_n h_n \in D_{\varphi} \) for all \( n \geq 1 \) sufficiently large; see p. 22 of Kosorok (2008).

We first verify that \( \eta \) is Hadamard differentiable at \( \hat{\vartheta} \) tangentially to \( \mathbb{R}^{d_{\varphi}} \). For any \( l_n \to 0 \) and \( h_n \to h \in \mathbb{R}^{d_{\varphi}}, \)

\[
\frac{\eta(\hat{\vartheta} + l_n h_n) - \eta(\hat{\vartheta})}{l_n} = \frac{\pi(\mathbf{u}; t; \hat{\vartheta} + l_n h_n) - \pi(\mathbf{u}; t; \hat{\vartheta})}{l_n} = \frac{\pi'(\mathbf{u}; t; \vartheta^*)l_n h_n}{l_n} \to \pi'(\mathbf{u}; t; \hat{\vartheta})h.
\]

Using the functional delta method, since \( \pi'(\mathbf{u}; t; \hat{\vartheta}) \) is uniformly bounded,

\[
\sqrt{n}(\pi(\mathbf{u}; t; \hat{\vartheta}) - \pi(\mathbf{u}; t; \hat{\vartheta})) \sim \pi'(\mathbf{u}; t; \hat{\vartheta})Z_\vartheta \text{ in } \ell^\infty(U \times T)
\]
where \( Z_\vartheta \sim N(0, \mathbb{I}_{\vartheta}^{-1}) \).
References


Figure 1: Distribution of the Months of First Prenatal Care Visit

Note: The number 10 means “did not have prenatal care”.
Figure 2: The Mother’s Weight Gain during Pregnancy and Level of Birthweight

The top and bottom horizontal lines represent the thresholds of high and low birthweight, respectively. The solid curve is the average of birthweights and the dashed curves are the 90%, 75%, 50%, 25%, and 10% quantiles of birthweights.
Figure 3: Mother’s Weight Gain during Pregnancy and Level of Birthweight with 90% Confidence Bands

(a) 10% Quantile
(b) 25% Quantile
(c) 50% Quantile
(d) 75% Quantile
(e) 90% Quantile
(f) Average

The top and bottom horizontal lines represent the thresholds of high and low birthweight, respectively.
Figure 4: The Mother’s Age and Level of Birthweight

The top and bottom horizontal lines represent the thresholds of high and low birthweight, respectively. The solid curve is the average of birthweights and the dashed curves are the 90%, 75%, 50%, 25%, and 10% quantiles of birthweights.
Figure 5: Mother’s Age and Level of Birthweight with 90% Confidence Bands

(a) 10% Quantile  (b) 25% Quantile  (c) 50% Quantile

(d) 75% Quantile  (e) 90% Quantile  (f) Average

The thresholds for low and high birthweights are 2.5 kilograms and 4 kilograms, respectively.

55
### Table 1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>Birthweight</th>
<th>Age</th>
<th>WG</th>
<th>Cigarettes</th>
<th>No. Care</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0.26</td>
<td>14.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>3.06</td>
<td>20.00</td>
<td>25.00</td>
<td>0.00</td>
<td>10.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Median</td>
<td>3.37</td>
<td>24.00</td>
<td>33.00</td>
<td>0.00</td>
<td>12.00</td>
<td>13.00</td>
</tr>
<tr>
<td>Mean</td>
<td>3.35</td>
<td>24.88</td>
<td>34.20</td>
<td>1.10</td>
<td>11.80</td>
<td>13.02</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>3.69</td>
<td>29.00</td>
<td>42.00</td>
<td>0.00</td>
<td>14.00</td>
<td>15.00</td>
</tr>
<tr>
<td>Max</td>
<td>5.67</td>
<td>45.00</td>
<td>95.00</td>
<td>40.00</td>
<td>49.00</td>
<td>16.00</td>
</tr>
<tr>
<td>SD.</td>
<td>0.54</td>
<td>5.45</td>
<td>13.69</td>
<td>3.25</td>
<td>3.40</td>
<td>2.27</td>
</tr>
</tbody>
</table>

### Table 2: Treatment Effects of the Mother’s Weight Gain During Pregnancy

<table>
<thead>
<tr>
<th>WG change</th>
<th>10% Qt.</th>
<th>25% Qt.</th>
<th>50% Qt.</th>
<th>75% Qt.</th>
<th>90% Qt.</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–20</td>
<td>2.23</td>
<td>2.52</td>
<td>2.68</td>
<td>2.77</td>
<td>2.72</td>
<td>2.53</td>
</tr>
<tr>
<td>SD.</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>20–40</td>
<td>0.37</td>
<td>0.28</td>
<td>0.23</td>
<td>0.25</td>
<td>0.25</td>
<td>0.30</td>
</tr>
<tr>
<td>SD.</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>40–60</td>
<td>0.21</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>SD.</td>
<td>0.05</td>
<td>0.07</td>
<td>0.06</td>
<td>0.09</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>60–80</td>
<td>0.25</td>
<td>0.25</td>
<td>0.28</td>
<td>0.31</td>
<td>0.37</td>
<td>0.30</td>
</tr>
<tr>
<td>SD.</td>
<td>0.05</td>
<td>0.08</td>
<td>0.07</td>
<td>0.10</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>0–40</td>
<td>2.59</td>
<td>2.80</td>
<td>2.91</td>
<td>3.03</td>
<td>2.98</td>
<td>2.83</td>
</tr>
<tr>
<td>SD.</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>20–60</td>
<td>0.57</td>
<td>0.48</td>
<td>0.43</td>
<td>0.46</td>
<td>0.48</td>
<td>0.52</td>
</tr>
<tr>
<td>SD.</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>40–80</td>
<td>0.45</td>
<td>0.45</td>
<td>0.48</td>
<td>0.51</td>
<td>0.60</td>
<td>0.51</td>
</tr>
<tr>
<td>SD.</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>0–60</td>
<td>2.80</td>
<td>3.00</td>
<td>3.11</td>
<td>3.23</td>
<td>3.20</td>
<td>3.04</td>
</tr>
<tr>
<td>SD.</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>20–80</td>
<td>0.82</td>
<td>0.74</td>
<td>0.71</td>
<td>0.77</td>
<td>0.85</td>
<td>0.81</td>
</tr>
<tr>
<td>SD.</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>0–80</td>
<td>3.05</td>
<td>3.25</td>
<td>3.39</td>
<td>3.54</td>
<td>3.57</td>
<td>3.34</td>
</tr>
<tr>
<td>SD.</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>

56
Table 3: Treatment Effects of the Mother’s Age

<table>
<thead>
<tr>
<th>Age Change</th>
<th>10% Qt.</th>
<th>25% Qt.</th>
<th>50% Qt.</th>
<th>75% Qt.</th>
<th>90% Qt.</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>15–20</td>
<td>-0.08</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.03</td>
<td>-0.06</td>
</tr>
<tr>
<td>SD.</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>20–25</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>SD.</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>25–30</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td>SD.</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>30–35</td>
<td>-0.06</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>SD.</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>35–40</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.04</td>
</tr>
<tr>
<td>SD.</td>
<td>0.11</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>40–45</td>
<td>-0.17</td>
<td>-0.11</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.11</td>
</tr>
<tr>
<td>SD.</td>
<td>0.32</td>
<td>0.15</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
<td>0.16</td>
</tr>
<tr>
<td>15–25</td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.09</td>
</tr>
<tr>
<td>SD.</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>20–30</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.06</td>
</tr>
<tr>
<td>SD.</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>25–35</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.06</td>
</tr>
<tr>
<td>SD.</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>30–40</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.07</td>
</tr>
<tr>
<td>SD.</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>35–45</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.08</td>
<td>-0.10</td>
<td>-0.11</td>
<td>-0.15</td>
</tr>
<tr>
<td>SD.</td>
<td>0.30</td>
<td>0.13</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>15–30</td>
<td>-0.14</td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.11</td>
<td>-0.09</td>
<td>-0.11</td>
</tr>
<tr>
<td>SD.</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>20–35</td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.07</td>
<td>-0.08</td>
<td>-0.09</td>
</tr>
<tr>
<td>SD.</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>25–40</td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.09</td>
</tr>
<tr>
<td>SD.</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>30–45</td>
<td>-0.25</td>
<td>-0.17</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.13</td>
<td>-0.18</td>
</tr>
<tr>
<td>SD.</td>
<td>0.28</td>
<td>0.11</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>15–35</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.11</td>
<td>-0.12</td>
<td>-0.11</td>
<td>-0.14</td>
</tr>
<tr>
<td>SD.</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>20–40</td>
<td>-0.14</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.11</td>
<td>-0.13</td>
</tr>
<tr>
<td>SD.</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>25–45</td>
<td>-0.28</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.16</td>
<td>-0.20</td>
</tr>
<tr>
<td>SD.</td>
<td>0.27</td>
<td>0.10</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>15–40</td>
<td>-0.23</td>
<td>-0.17</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.18</td>
</tr>
<tr>
<td>SD.</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>20–45</td>
<td>-0.31</td>
<td>-0.23</td>
<td>-0.17</td>
<td>-0.17</td>
<td>-0.19</td>
<td>-0.23</td>
</tr>
<tr>
<td>SD.</td>
<td>0.25</td>
<td>0.09</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>15–45</td>
<td>-0.40</td>
<td>-0.28</td>
<td>-0.20</td>
<td>-0.23</td>
<td>-0.22</td>
<td>-0.29</td>
</tr>
<tr>
<td>SD.</td>
<td>0.22</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.07</td>
</tr>
</tbody>
</table>