
Double Debiased Machine Learning for Mediation Analysis with Continuous Treatments

Houssam Zenati
Inria MIND¹

Judith Abécassis
Inria SODA²

Julie Josse
Inria PreMedical³

Bertrand Thirion
Inria MIND¹

Abstract

Uncovering causal mediation effects is of significant value to practitioners seeking to isolate the direct treatment effect from the potential mediated effect. We propose a double machine learning (DML) algorithm for mediation analysis that supports continuous treatments. To estimate the target mediated response curve, our method uses a kernel-based doubly robust moment function for which we prove asymptotic Neyman orthogonality. This allows us to obtain asymptotic normality with nonparametric convergence rate while allowing for nonparametric or parametric estimation of the nuisance parameters. We then derive an optimal bandwidth strategy along with a procedure for estimating asymptotic confidence intervals. Finally, to illustrate the benefits of our method, we provide a numerical evaluation of our approach on a simulation along with an application to real-world medical data to analyze the effect of glycemic control on cognitive functions.

However, in many evaluation settings, understanding the causal mechanisms driving these total effects is equally important. For example, in studying the relationship between socio-demographic factors and cognitive function, one might hypothesize that the effect of these factors is mediated by the brain structure captured in brain imaging. This mediation insight comes from studies such as (Dadi et al., 2021) and (Cox et al., 2019), that have demonstrated that combining IDPs with socio-demographic variables often provides little to no improvement in predictive performance over socio-demographic data alone. However, the causal impact of socio-demographic variables on cognitive outcomes is expected to occur through their influence on brain structure captured in brain imaging. Understanding these mediation pathways can help researchers and health policymakers design more targeted interventions that focus e.g. on altering socio-demographic determinants before they alter irreversibly the brain tissues.

Mediation analysis aims to assess the specific contributions of the indirect and direct effects of a given treatment. The indirect effect quantifies the proportion of the total effect that is due to a mediator, while the direct effect measures the effect of the treatment without additional mediators (Robins and Greenland, 1992; Pearl, 2001). Several studies have used flexible (often nonparametric) models (Petersen et al., 2006; Flores and Flores-Lagunes, 2009; Imai et al., 2010; Hong et al., 2010; Imai and Yamamoto, 2013): this includes regression-based (Robins, 1986; Zheng and van der Laan, 2012) using the Pearl formula (Pearl, 2001) or models that use inverse probability weighting (IPW) (Horvitz and Thompson, 1952; Huber, 2014), hence requiring the computation of propensity scores, either conditional on both the mediator and covariates or on the covariates alone. However, these methods become inconsistent if the models for the conditional mean outcome or the treatment and mediator densities are misspecified. To improve on this practice, more sophisticated causal mediation analysis schemes have been proposed, based on efficient score functions (Tchetgen and Shpitser, 2012), with double machine learning (DML) as outlined in Chernozhukov et al. (2018), which

1 INTRODUCTION

In causal inference (Imbens and Rubin, 2015), traditional treatment evaluations often focus on assessing the total causal effect of a treatment on an outcome variable, such as the average treatment effect (ATE).

¹MIND, Inria-Saclay, Palaiseau, Université Paris Saclay, CEA Saclay, Paris, France

²SODA, Inria-Saclay, Palaiseau, Université Paris Saclay, Paris, France

³PreMedical, Inria-Sophia-Antipolis, Montpellier, France

relies on Neyman orthogonality and sample splitting. Typically, when combining efficient score-based estimation with sample splitting, $n^{-1/2}$ convergence for treatment effect estimation can be achieved, even with slower plug-in estimate convergence rates of $n^{-1/4}$.

While most nonparametric mediation studies have focused on binary treatments, many real-world problems involve continuous treatments, such as the dose of a medical treatment (Hirano and Imbens, 2005; Imai and Van Dyk, 2004; Bia and Mattei, 2012; Kluve et al., 2012; Galvao and Wang, 2015; Lee, 2018; Cox et al., 2019). This paper addresses nonparametric estimation of natural direct and indirect effects (Pearl, 2001) for continuous treatments. We propose a multiply robust estimator based on weighting by the inverse of conditional treatment densities (Hirano and Imbens, 2005; Imai and Van Dyk, 2004) and the estimation of conditional mean outcomes (Pearl, 2001; Singh et al., 2023). Our method is asymptotically normal and converges at the rate of one-dimensional nonparametric regression under specific regularity conditions.

Our first contribution is to show the asymptotic Neyman (1959) orthogonality of a kernelized moment function, similar to the efficient score functions of Tchetgen and Shpitser (2012); Farbmacher et al. (2022). This property makes the estimation of direct and indirect effects rather insensitive to (local) estimation errors in the plug-in estimates. Second, we propose a DML estimator with a Bayes transformation that avoids the conditional mediator density estimation and directly estimates a nested conditional mean outcome (Farbmacher et al., 2022). This appears particularly useful when the mediator is a vector of variables and/or continuous. Third, we analyze this estimator and prove that it is asymptotically normal with a nonparametric convergence rate under mild regularity conditions and mild convergence requirements for nuisance parameters. This allows us to derive an asymptotic mean squared error optimal bandwidth and an asymptotic confidence interval. Moreover, our analysis proves the multiple robustness of our estimator with regards to i) the consistency of the nuisance parameters and ii) achieving a faster non-parametric convergence rate with slow nuisance convergence rates. Eventually, we provide numerical experiments illustrating the practical performance of DML compared to other mediation analysis estimators and an application of continuous treatment mediation for cognitive function on real-world data.

1.1 Related Work

In the causal mediation literature, most studies assume sequential conditional independence, where (i) potential outcomes and treatment assignment are independent given covariates, and (ii) potential outcomes

and mediator are independent given treatment and covariates (Imai et al., 2010; Pearl, 2001; Flores and Flores-Lagunes, 2009). Under such assumptions, a common approach uses linear equations (Judd and Kenny, 1981; Baron and Kenny, 1986; VanderWeele, 2015) modeling of the mediator as a function of treatment and covariates, and the outcome as a function of the mediator, treatment, and covariates. However, such methods rely on the correct specification of linear relationships, which may not be appropriate for binary mediators or outcomes. In contrast, G-computation (Robins, 1986; Zheng and van der Laan, 2012) using Pearl formula (Pearl, 2001) allows for non-linear models. Moreover, inverse probability weighting (IPW) (Horvitz and Thompson, 1952; Huber, 2014) methods (in the binary treatment case) leverage propensity scores (either conditional on the mediator and covariates or on the covariates alone) to estimate mediation mechanisms. However, if either the conditional means or densities are misspecified, those methods become inconsistent.

Subsequently, multiply robust estimators—akin to doubly robust estimation in standard treatment effect models—have been proposed (Tchetgen and Shpitser, 2012; Zheng and van der Laan, 2012) and shown to be robust to model misspecification (Huber et al., 2015). Unlike Tchetgen and Shpitser (2012) which requires the mediator density estimation, Farbmacher et al. (2022) analyzed a doubly robust estimation method that relies on estimates of the conditional mean outcome, cross conditional mean outcome, and conditional treatment probabilities. Analogously to doubly robust estimation of average treatment effects (Robins et al., 1994; Robins and Rotnitzky, 1995), all multiply robust resulting estimators are semiparametrically efficient if all models underlying the plug-in estimates are correctly specified and remain consistent even if one model is misspecified. Our work builds on the results for semiparametric models in Ichimura and Newey (2022); Chernozhukov et al. (2022a) and follows the line of Farbmacher et al. (2022): a growing literature employs the double machine learning (DML) approach for non-regular nonparametric infinite-dimensional objects for estimating causal quantities (Chernozhukov et al., 2022b; Semenova and Chernozhukov, 2020; Fan et al., 2022a; Zimmert and Lechner, 2019; Bonvini and Kennedy, 2022).

However, the majority of research in mediation analysis has focused on binary treatments. Continuous treatments specifically pose a challenge in IPW-based methods, where propensity scores need to be adapted (Imbens, 2000; Hirano and Imbens, 2005; Flores, 2007; Galvao and Wang, 2015; Zenati et al., 2025), even in the mediation analysis task Hsu et al. (2020) which uses a kernel smoothing strategy. Sani et al. (2024)

proposes a multiply robust estimator for the mediation task with continuous treatment, but requires i) the mediator conditional density estimation and ii) consistency of the nuisance parameters. Instead, our work does not require the cumbersome mediator conditional density estimation, as in Farbmacher et al. (2022) and we prove the asymptotic Neyman orthogonality (Neyman, 1959) of our kernel moment function. Moreover, our estimator is multiply robust in the sense that our inference theory is valid even when one nuisance function is inconsistent, as in Kennedy et al. (2017) and Takatsu and Westling (2024). This is a stronger result than the usual multiply robustness on the consistency of the estimator used in Farbmacher et al. (2022) and Sani et al. (2024) which analysis does not consider inconsistent nuisance parameters. Eventually, unlike Sani et al. (2024) our method discussed and posed the problem of bandwidth selection and uncertainty quantification through confidence intervals.

2 BACKGROUND

In this section we introduce the necessary definitions and notations for the general mediation analysis task.

2.1 Natural direct and indirect effects

We label the random variable for the treatment as T , the outcome as Y , the mediator(s) as M , and the covariate(s) as X for each individual. We will also write $Z = (Y, T, X, M)$. The interrelations among these variables are depicted in Figure 1.

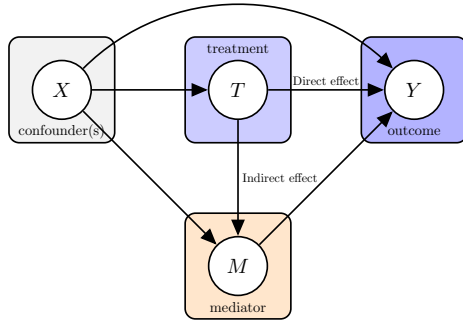


Figure 1: Causal graph for mediation analysis.

Given treatment values $t, t' \in \mathcal{T}$ we denote $M(t)$ and $Y(t, M(t'))$ the potential mediator state and potential outcome, expanding upon the potential outcomes framework (Imbens and Rubin, 2015). We make the assumption that the observed outcome and mediator correspond to the potential outcome and mediator under the actually assigned treatment.

We then define the total average treatment effect for a general treatment space \mathcal{T} . For the next definitions, let $t, t' \in \mathcal{T}$.

Definition 2.1 (Total average treatment effect).

$$\tau(t, t') = \mathbb{E}[Y(t', M(t'))] - \mathbb{E}[Y(t, M(t))]. \quad (1)$$

Next, we define the natural direct effect, denoted by $\theta(t, t')$, as the difference in expected potential outcomes when switching the treatment while keeping the potential mediator fixed, which blocks the causal mechanism via M :

Definition 2.2 (Direct effect).

$$\theta(t, t') = \mathbb{E}[Y(t', M(t)) - \mathbb{E}[Y(t, M(t))]]. \quad (2)$$

Eventually, the natural indirect effect, $\delta(t, t')$, equals the difference in expected potential outcomes when switching the potential mediator values while keeping the treatment fixed to block the direct effect.

Definition 2.3 (Indirect effect).

$$\delta(t, t') = \mathbb{E}[Y(t', M(t'))] - \mathbb{E}[Y(t', M(t))]. \quad (3)$$

Let $\mathcal{Z} = \mathcal{Y} \times \mathcal{T} \times \mathcal{X} \times \mathcal{M}$ represent the support of $Z = (Y, T, X, M)$, with an associated cumulative distribution function (CDF) $F_Z(Z)$.

2.2 Mediated response

When it comes to isolating direct and effects, it is more common to introduce the notion of mediated response for $t, t' \in \mathcal{T}$.

Definition 2.4 (Mediated response).

$$\eta_{t,t'} = \mathbb{E}[Y(t, M(t'))] \quad (4)$$

It is then possible to introduce simple and convenient expressions that establish relations between previously defined quantities:

$$\tau(t, t') = \theta(t, t') + \delta(t, t') \quad (5)$$

$$\theta(t, t') = \eta(t', t) - \eta(t, t) \quad (6)$$

$$\delta(t, t') = \eta(t', t') - \eta(t', t) \quad (7)$$

Therefore, one can rely the counterfactual potential outcome $\mathbb{E}[Y(t, M(t'))]$, that is the mediated response, to derive expressions of the mediated causal quantities of interest. To this end, we present in the next subsection classical assumptions necessary for its identification.

2.3 Identification assumptions

Since the mediated response curve is defined in terms of potential outcomes that are not directly observed,

we must consider assumptions under which it can be expressed in terms of observed data. A complete treatment of identification in the presence of continuous random variables has been given by Gill and Robins (2001); we refer the reader there for details. The assumptions most commonly employed (Imai et al., 2010; Pearl, 2001; Flores and Flores-Lagunes, 2009) for identification are as follows.

Assumption 1 (Conditional independence of the treatment). *For all $t', t \in \mathcal{T}, m \in \mathcal{M}$*

$$\{Y(t', m), M(t)\} \perp\!\!\!\perp T | X.$$

Assumption 1 requires the absence of unobserved confounders influencing the treatment or the outcome, or affecting their relationship.

Assumption 2 (Conditional independence of the mediator). *For all $t', t \in \mathcal{T}, m \in \mathcal{M}, x \in \mathcal{X}$*

$$Y(t', m) \perp\!\!\!\perp M | T = t, X = x.$$

Assumption 2 requires the absence of unobserved confounders influencing both the mediator and the outcome.

Assumption 3 (Positivity assumption). *For some positive constant c , the essential infimums $\inf_{t \in \mathcal{T}} \text{ess}_{\inf_{x \in \mathcal{X}}} f_{T|X}(t | x) \geq c$ and $\inf_{t \in \mathcal{T}} \text{ess}_{\inf_{x \in \mathcal{X}, m \in \mathcal{M}}} f_{T|X,M}(t | x, m) \geq c$.*

Note that the mild condition on the essential infimum is slightly stronger than the common positivity assumption; as in Colangelo and Lee (2020) this will prove useful for the continuous task with kernel localization in Section 3. Assumption 3 ensures that the treatment is not deterministic in the covariate X and in the pair of covariate and mediator X, M . Given the identifications conditions, we introduce the mediation formula at the core of the mediation estimators we propose.

2.4 Mediation formula

Let us note $f_{T|X}(t | X)$ (respectively $f_{T|X,M}(t | X, M)$) the conditional density of T given X (respectively given X and M) and $f_{M|T,X}(M | T, X)$ the conditional density of M given T and X (if M is discrete, this is a conditional probability). We now define fundamental quantities, starting with the conditional mean being defined as the expectation of the outcome given determined values of the treatment $t \in \mathcal{T}$, mediator $m \in \mathcal{M}$ and context $x \in \mathcal{X}$.

Definition 2.5 (Conditional mean outcome).

$$\mu_Y(t, m, x) = \mathbb{E}[Y | T = t, M = m, X = x] \quad (8)$$

Based on the previous expression, we can define the cross conditional mean outcome as follows:

Definition 2.6 (Cross conditional mean outcome).

$$\begin{aligned} \omega_Y(t, t', x) &= \int_{\mathcal{M}} \mu_Y(t, m, x) f_{M|T,X}(m | t', x) dm \\ &= \mathbb{E}[\mu_Y(t, m, X) | T = t', X = x] \end{aligned} \quad (9)$$

Using the previous definitions, Pearl's mediation formula (Pearl, 2001) is as follows:

Lemma 2.1 (Pearl's mediation formula).

$$\begin{aligned} \eta_{t,t'} &= \int_{\mathcal{M}, \mathcal{X}} \mu_Y(t, m, x) f_{M|T,X}(m | t', x) f_X(x) dm dx \\ &= \int_{\mathcal{X}} \omega_Y(t, t', x) f_X(x) dx \end{aligned} \quad (10)$$

where f_X is the density over the covariate space \mathcal{X} . This formula is the basis for several estimates that are defined in the next section.

3 CONTINUOUS TREATMENT MEDIATION

In this section, we introduce the tools used to analyze mediation with continuous treatments. For the sake of clarity, we consider a unidimensional treatment space \mathcal{T} , the generalization to multidimensional \mathcal{T} is given in Appendix 8 and 9. Let us then consider a kernel k that satisfies Assumption 4, which is standard in nonparametric kernel estimation and holds for commonly used kernel functions, such as Epanechnikov and Gaussian.

Assumption 4. *The second-order symmetric kernel function $k(\cdot)$ is bounded differentiable, i.e. $\int_{-\infty}^{\infty} k(u) du = 1$, $\int_{-\infty}^{\infty} uk(u) du = 0$, and $0 < \kappa < \infty$. For some finite positive constants C, \bar{U} , and for some $\nu > 1$, $|dk(u)/du| \leq C|u|^{-\nu}$ for $|u| > \bar{U}$.*

Let us consider a bandwidth h . We now define the kernel smoothing operator as

$$K_h(T - t) = k((T - t)/h)/h \quad (11)$$

With Assumptions 1, 2, 3 and the same reasoning for the binary treatment, it is possible to show the identification (Hsu et al., 2020) (see Appendix 8):

$$\begin{aligned} \eta_{t,t'} &= \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{K_h(T - t)Y}{f_{T|X}(t | X)} \frac{f_{M|T,X}(M | t', X)}{f_{M|T,X}(M | t, X)} \right] \\ &= \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{K_h(T - t)Y}{f_{T|X}(t' | X)} \frac{f_{T|X,M}(t' | X, M)}{f_{T|X,M}(t | X, M)} \right] \end{aligned} \quad (12)$$

for $t, t' \in \mathcal{T}$ and where the last equality follows a Bayes transformation on the mediator density as in Farbmacher et al. (2022).

The expression in Eq. (10) motivates the class of G-computation (or regression-based) estimators, while Eq. (12) motivates the class of inverse probability weighting estimators; see Appendix 8 for further discussion. The DML (Chernozhukov et al., 2018) estimator leverages a moment function that combines both classes of methods, as in influence functions for semiparametric estimators (Newey, 1994; Ichimura and Newey, 2022). Similar to Tchetgen and Shpitser (2012) and Farbmacher et al. (2022) for the mediation analysis task, we introduce a kernel smoothed moment function with provable asymptotic convergence.

Lemma 3.1. *Let $t, t' \in \mathcal{T}$ and let the moment function*

$$\begin{aligned} \psi_{t,t'}^h = & \frac{K_h(T-t)f_{T|X,M}(t' | X, M)}{f_{T|X}(t' | X)f_{T|X,M}(t | X, M)} [Y - \mu_Y(t, M, X)] \\ & + \frac{K_h(T-t')}{f_{T|X}(t' | X)} [\mu_Y(t, M, X) - \omega_Y(t, t', X)] \\ & + \omega_Y(t, t', X) - \eta_{t,t'}. \end{aligned} \quad (13)$$

then, the moment function is asymptotically Neyman-Orthogonal.

A proof of Lemma 3.1 is given in Appendix 8. Note that in the moment function we propose, we do not make use of the mediator density function and instead propose to estimate the nuisance parameters $\widehat{f_{T|X}}$, $\widehat{\mu_Y}$, $\widehat{f_{T|X,M}}$ and $\widehat{\omega_Y}$. Indeed, as explained in Farbmacher et al. (2022), the original multiply robust estimator for mediation (Tchetgen and Shpitser, 2012) proposed to only estimate the nuisances $\widehat{f_{T|X}}$, $\widehat{\mu_Y}$, and $\widehat{f_{M|T,X}}$ to build their efficient influence function but would estimate the cross conditional expectation ω_Y as $\int_M \widehat{\mu_Y}(t, M, X) \widehat{f_{M|T,X}}(M | t', X) dM$ as done in Sani et al. (2024). However, we argue that such a formulation hurts the estimation of ω_Y especially when the mediator is high dimensional.

Subsequently, aside from the the Neyman Orthogonality condition on the estimator which motivated the use of an efficient score function, the DML approach (Chernozhukov et al., 2018) also uses sample splitting to estimate separately the nuisance parameters and the causal quantity of interest. Let now $\{Y_i, T_i, M_i, X_i\}_{i=1}^n$ be i.i.d realizations of the random variables Y, T, M, X . An L -fold cross-fitting splits the sample into subsamples I_ℓ for $\ell \in \{1, \dots, L\}$. For a given split I_ℓ and $i \in I_\ell$, the nuisance function estimators for $\widehat{f_{T|X_{i\ell}}}(t') = f_{T|X}(t' | X_i)$, $\widehat{f_{T|X,M_{i\ell}}}(t) = f_{T|X,M}(t | X_i, M_i)$, $\widehat{\mu_{Y_{i\ell}}} = \mu_Y(t, M_i, X_i)$ and $\widehat{\omega_{Y_{i\ell}}} = \omega_Y(t, t', X_i)$ use observations in the other $(L-1)$ subsamples that do not contain observation i . The DML estimator then averages over the whole subsamples as stated in Algorithm 1.

Note that a key distinction of our method compared

Algorithm 1: Continous Treatment Double Machine Learning

Data: Observational $\{Y_i, T_i, M_i, X_i\}_{i=1}^n$

- Split the data in L subsamples. For each subsample ℓ , let n_ℓ denote its size, Z^ℓ the set of observations in the sample and $Z^{\ell,C}$ the complement set of all observations not in Z^ℓ .
- For each ℓ , use $Z^{\ell,C}$ to estimate $f_{T|X}$, $f_{T|X,M}$, μ_Y and ω_Y .
- Predict the nuisance parameters for each observation i in \mathcal{D}_ℓ , i.e. $\widehat{f_{T|X_{i\ell}}}$, $\widehat{f_{T|X,M_{i\ell}}}$, $\widehat{\mu_{Y_{i\ell}}}$ and $\widehat{\omega_{Y_{i\ell}}}$.
- For each ℓ , obtain an estimate of the kernel score function for each observation i in Z^ℓ , denoted by $\hat{\psi}_{t,t',i\ell}^h$:

$$\begin{aligned} \hat{\psi}_{t,t',i\ell}^h = & \frac{K_h(T_i-t)\widehat{f_{T|X,M_{i\ell}}}(t')}{\widehat{f_{T|X,M_{i\ell}}}(t)\widehat{f_{T|X_{i\ell}}}} \cdot [Y - \widehat{\mu_{Y_{i\ell}}}] \\ & + \frac{K_h(T_i-t')}{\widehat{f_{T|X_{i\ell}}}} \cdot [\widehat{\mu_{Y_{i\ell}}} - \widehat{\omega_{Y_{i\ell}}}] \\ & + \widehat{\omega_{Y_{i\ell}}} \end{aligned} \quad (14)$$

- Average the estimated scores $\hat{\psi}_{t,t',i\ell}^h$ over all observations across all L subsamples to obtain an estimate of $\eta_{t,t'}$ in the total sample, denoted by

$$\hat{\eta}_{t,t'} = \frac{1}{n} \sum_{\ell=1}^L \sum_{i=1}^{n_\ell} \hat{\psi}_{t,t',i\ell}^h \quad (15)$$

Result: Estimated $\hat{\eta}_{t,t'}$

to Sani et al. (2024) is that they integrate the conditional mean outcome against the mediator density to estimate the cross conditional mean outcome. While this works only for discrete mediators, we avoid integration over mediator densities, as in Farbmacher et al. (2022) and directly estimate the cross-conditional mean outcome with an implicit integration as in Singh et al. (2023) which recently addressed the challenge related to continuous cross-counterfactual treatments (t, t') . This makes our approach suitable for continuous or high-dimensional mediators.

4 ASYMPTOTIC ANALYSIS

In this section we analyze the asymptotic behavior of the DML estimator we proposed.

We define $\|\cdot\|_{tMX}^2$ (respectively $\|\cdot\|_X^2$ and $\|\cdot\|_{tX}^2$) the partial $L_2(tMX)$ norm (respectively $L_2(X)$ and $L_2(tX)$)

for any $t \in \mathcal{T}$ as

$$\|\cdot\|_{tMX}^2 = \int_{\mathcal{M}, \mathcal{X}} (\cdot)^2 f_{TMX}(t, m, x) dm dx$$

where the integral is taken with regards in $\mathcal{M} \times \mathcal{X}$ (respectively \mathcal{X}) with the joint density f_{TMX} (respectively f_X and f_{TX}). We will also write $\delta_{\mu_Y}^\ell = \widehat{\mu_{Y\ell}} - \overline{\mu_Y}$, $\delta_{\omega_Y}^\ell = \widehat{\omega_{Y\ell}} - \overline{\omega_Y}$, $\delta_{T|X}^\ell = \widehat{f_{T|X\ell}} - \overline{f_{T|X}}$ and $\delta_{T|X,M}^\ell = \widehat{f_{T|X,M\ell}} - \overline{f_{T|X,M}}$. We can now state conditions on the convergence of the nuisance parameters.

Assumption 5 (Nuisance convergence). *There exist functions $\overline{\mu_Y}$, $\overline{\omega_Y}$, $\overline{f_{T|X,M}}$ and $\overline{f_{T|X}}$ that are three-times differentiable with $f_{T|X}(t | x)$, $f_{T|X,M}(t | x, m) \geq c$ for some positive constant c , and satisfy the following: For each $\ell = 1, \dots, L$,*

1. $\|\delta_{\mu_Y}^\ell\|_{tMX} = o_p(1)$, $\|\delta_{\omega_Y}^\ell\|_X = o_p(1)$, $\|\delta_{T|X}^\ell\|_{tX} = o_p(1)$ and $\|\delta_{T|X,M}^\ell\|_{tMX} = o_p(1)$
2. *Either $\overline{\mu_Y} = \mu_Y$ and $\overline{\omega_Y} = \omega_Y$, or $\overline{\mu_Y} = \mu_Y$ and $\overline{f_{T|X}} = f_{T|X}$, or $\overline{\omega_Y} = \omega_Y$ and $\overline{f_{T|X,M}} = f_{T|X,M}$, or $\overline{f_{T|X}} = f_{T|X}$ and $\overline{f_{T|X,M}} = f_{T|X,M}$.*

Assumption 6 (Nuisance rates). *There exist functions $\overline{\mu_Y}$, $\overline{\omega_Y}$, $\overline{f_{T|X,M}}$ and $\overline{f_{T|X}}$ that satisfy the following: For each $\ell = 1, \dots, L$,*

1. $\sqrt{nh} \|\delta_{T|X}^\ell\|_{tX} \|\delta_{\mu_Y}^\ell\|_{tMX} = o_p(1)$
2. $\sqrt{nh} \|\delta_{T|X,M}^\ell\|_{tMX} \|\delta_{\mu_Y}^\ell\|_{tMX} = o_p(1)$
3. $\sqrt{nh} \|\delta_{T|X}^\ell\|_{tX} \|\delta_{\omega_Y}^\ell\|_X = o_p(1)$

The nuisance function estimators $\widehat{\mu_{Y\ell}}$, $\widehat{\omega_{Y\ell}}$, $\widehat{f_{T|X\ell}}$ and $\widehat{f_{T|X,M\ell}}$ converge to some fixed functions $\overline{\mu_Y}$, $\overline{\omega_Y}$, $\overline{f_{T|X}}$ and $\overline{f_{T|X,M}}$ respectively in the sense of Assumption 5.1. Assumption 5.2 allows some of the nuisance functions to be misspecified. Assumption 6 and 5.2 imply that if one nuisance function is misspecified, then the other needs to be estimated consistently at a convergence rate faster than \sqrt{nh} . This is the cost of the multiply robust inference. For example, if $\overline{\mu_Y} \neq \mu_Y$, then $\|\widehat{\mu_{Y\ell}} - \mu_Y\|_{tMX} = O_p(1)$. So Assumption 6 requires $\sqrt{nh} \|\widehat{f_{T|X\ell}} - \overline{f_{T|X}}\|_{tX} = o_p(1)$.

Moreover, we will require additional regularity assumptions on the density functions f_Z and the nuisance parameters.

Assumption 7 (Regularity). *We list the following:*

1. $f_Z(y, t, m, x)$, μ_Y and ω_Y are three-times differentiable with respect to t with all three derivatives being bounded uniformly over $(y, t, x, m) \in \mathcal{Z}$
2. $\text{var}(Y | T = t, X = x, M = m)$ and its derivatives with respect to t are bounded uniformly over $(t', x) \in \mathcal{T} \times \mathcal{X}$.

We can now state the main result of the asymptotic normality of the DML estimator in mediation.

Theorem 8 (Asymptotic normality). *Let Assumptions 1-7 hold. Let $h \rightarrow 0$, $nh \rightarrow \infty$, and $nh^5 \rightarrow C \in [0, \infty)$. Then for any $t \in \mathcal{T}$,*

$$\begin{aligned} & \sqrt{nh} (\hat{\eta}_{t,t'} - \eta_{t,t'}) \\ &= \sqrt{\frac{h}{n}} \sum_{i=1}^n \left[\frac{K_h(T_i - t) \cdot \overline{f_{T|X,Mt',i}}}{\overline{f_{T|X,i}} \cdot \overline{f_{T|X,Mt,i}}} \cdot [Y - \overline{\mu_{Yi}}] \right. \\ & \quad \left. + \frac{K_h(T_i - t')}{\overline{f_{T|X,i}}} \cdot [\overline{\mu_{Yi}} - \overline{\omega_{Yi}}] \right. \\ & \quad \left. + \overline{\omega_{Yi}} \right] + o_p(1). \end{aligned}$$

where we abbreviated $\overline{f_{T|X,i}} = \overline{f_{T|X}}(t' | X_i)$, $\overline{\mu_{Yi}} = \overline{\mu_Y}(t, M_i, X_i)$, $\overline{\omega_{Yi}} = \overline{\omega_Y}(t, t', X_i)$, $\overline{f_{T|X,Mt',i}} = \overline{f_{T|X,M}}(t' | X_i, M_i)$ and $\overline{f_{T|X,Mt,i}} = \overline{f_{T|X,M}}(t | X_i, M_i)$.

Let $\mathbb{E} [|Y - \overline{\mu_Y}(T, M, X)|^3 | T = t, M, X]$, $\mathbb{E} [(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^3 | T = t', X]$ and their derivatives with respect to t be bounded uniformly over $(t, m, x) \in \mathcal{T} \times \mathcal{M} \times \mathcal{X}$. Let $\int_{-\infty}^{\infty} k(u)^3 du < \infty$. Then

$$\sqrt{nh} (\hat{\eta}_{t,t'} - \eta_{t,t'} - h^2 B_{t,t'}) \xrightarrow{d} \mathcal{N}(0, V_{t,t'})$$

where

$$\begin{aligned} V_{t,t'} &\equiv \mathbb{E} \left[\bar{V}_Y \frac{f_{T|X,M}(t | X, M)}{\overline{f_{T|X}}(t' | X)^2} \frac{\overline{f_{T|X,M}}(t' | X, M)^2}{\overline{f_{T|X,M}}(t | X, M)^2} \right. \\ & \quad \left. + \frac{f_{T|X}(t' | X)}{\overline{f_{T|X}}(t' | X)^2} \bar{V}_{\mu_Y} \right] R_k \\ B_{t,t'} &\equiv \mathbb{E} \left[\frac{\overline{f_{T|X,M}}(t' | X, M)}{\overline{f_{T|X,M}}(t | X, M)} \left(\partial_t \mu_Y(t, M, X) \cdot \frac{\partial_t f_{T|X,M}(t | X, M)}{\overline{f_{T|X}}(t' | X)} + \frac{\partial_t^2 \mu_Y(t, M, X)}{\overline{f_{T|X}}(t' | X)} \right. \right. \\ & \quad \left. \left. (f_{T|X,M}(t | X, M) + \mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)) \right) \right. \\ & \quad \left. + \left((\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)) \cdot \frac{\partial_t^2 f_{T|X,M}(t' | X, M)}{\overline{f_{T|X}}(t' | X)} \right) / 2 \right] \kappa. \end{aligned}$$

where $\bar{V}_Y = \mathbb{E} [Y - \bar{\mu}_Y(T, M, X)]^2 \mid T = t, M, X]$, $\bar{V}_{\mu_Y} = \mathbb{E} [(\bar{\mu}_Y(t, M, X) - \bar{\omega}_Y(t, t', X))^2 \mid T = t', X]$, and where we denote the roughness of k as $R_k \equiv \int_{-\infty}^{\infty} k(u)^2 du$ and $\kappa \equiv \int_{-\infty}^{\infty} u^2 k(u) du$.

A proof of Theorem 8 is given in Appendix 9. Note that the last part in the influence function in (13) $n^{-1} \sum_{i=1}^n \bar{\omega}_Y(t, t', X_i) - \eta_{t,t'} = O_p(1/\sqrt{n}) = o_p(1/\sqrt{nh})$ and hence does not contribute to the first-order asymptotic variance $V_{t,t'}$.

Remark 9. In Theorem 8, the scaling is by the square root of the local sample size \sqrt{nh} rather than by the usual parametric rate \sqrt{n} . This slower rate of convergence is a typical bias-variance trade-off as in Colangelo and Lee (2020) and Kennedy et al. (2017). Similar to them, the estimator is consistent but not quite centered at $\eta_{t,t'}$; there is a bias term of order $O(h^2)$. Moreover, balancing the depending terms on h requires $h \sim n^{-1/5}$, giving a scaling $1/\sqrt{nh} \sim h^2 \sim n^{-2/5}$ as in Kennedy et al. (2017).

Comparison with Sani et al. (2024) After we prove our moment function is asymptotically Neyman Orthogonal, our analysis establishes stronger multiply robustness than Sani et al. (2024) by requiring milder assumptions on nuisance parameter consistency. Unlike Sani et al. (2024), which assumes all nuisance parameters converge to their true values (see their Assumption 4), we allow for potential inconsistency and prove consistency of the DML estimator under weaker conditions (see our Assumption 5, 6), similar to Colangelo and Lee (2020) and Takatsu and Westling (2024).

Our DML estimator uses a kernel for which the bandwidth h needs to be selected. Theorem 8 is key for inference, such as constructing confidence intervals and estimating the bandwidth h that minimizes the asymptotic mean squared error in a plug-in fashion (Park and Marron, 1990; Sheather and Jones, 1991). We propose an estimator for the leading bias $B_{t,t'}$, inspired by the idea in Powell and Stoker (1996); Colangelo and Lee (2020). Let the notation $\hat{\eta}_{t,t'} = \hat{\eta}_{t,t',b_n}$ be explicit on the bandwidth b_n and

$$\hat{B}_{t,t'} \equiv \frac{\hat{\eta}_{t,t',b_n} - \hat{\eta}_{t,t',\epsilon b_n}}{b_n^2 (1 - \epsilon^2)} \quad (16)$$

with a pre-specified fixed scaling parameter $\epsilon \in (0, 1)$. Corollary 11 below shows the consistency of $\hat{B}_{t,t'}$ under Assumption 10.

We can estimate the asymptotic variance $V_{t,t'}$ by the sample variance of the estimated influence function,

$$\hat{V}_{t,t'} = b_n n^{-1} \sum_{\ell=1}^L \sum_{i \in I_t} (\hat{\psi}_{t,t',i\ell}^{b_n})^2, \quad (17)$$

where $\hat{\psi}_{t,t',i\ell}^{b_n}$ is defined in Eq. (14). Then we propose a data-driven bandwidth

$$\hat{h}_{t,t'} = \left(\hat{V}_{t,t'} / \left(4 \hat{B}_{t,t'}^2 \right) \right)^{1/5} n^{-1/5} \quad (18)$$

to consistently estimate the optimal bandwidth that minimizes the asymptotic mean squared error (AMSE) given in Theorem 8. We first state additional assumptions for it.

Assumption 10. For each $\ell = 1, \dots, L$ and for any $t \in \mathcal{T}$,

1. $\left\| \delta_{\mu_Y}^\ell \delta_{T|X}^\ell \right\|_{tX} = o_p(1)$, $\left\| \delta_{\mu_Y}^\ell \delta_{\omega_Y}^\ell \right\|_{tX} = o_p(1)$, $\left\| \delta_{\omega_Y}^\ell \delta_{T|X,M}^\ell \right\|_{tMX} = o_p(1)$, $\left\| \delta_{T|X}^\ell \delta_{T|X,M}^\ell \right\|_{tMX} = o_p(1)$,
2. $\left\| \delta_{\mu_Y}^\ell \delta_{T|X}^\ell \right\|_{tX}^2 = O_p(1)$, $\left\| \delta_{\mu_Y}^\ell \delta_{\omega_Y}^\ell \right\|_{tX}^2 = O_p(1)$, $\left\| \delta_{\omega_Y}^\ell \delta_{T|X,M}^\ell \right\|_{tMX}^2 = O_p(1)$, $\left\| \delta_{T|X}^\ell \delta_{T|X,M}^\ell \right\|_{tMX}^2 = O_p(1)$
3. $\left\| \delta_{\mu_Y}^\ell \right\|_{tX}^2 = O_p(1)$, $\left\| \delta_{\omega_Y}^\ell \right\|_{tX}^2 = O_p(1)$, $\left\| \delta_{T|X,M}^\ell \right\|_{tMX}^2 = O_p(1)$, $\left\| \delta_{T|X}^\ell \right\|_{tMX}^2 = O_p(1)$
4. $\mathbb{E} [(Y - \bar{\mu}_Y(t, M, X))^4 \mid T = t, M, X]$, $\mathbb{E} [(\bar{\mu}_Y(t, M, X) - \bar{\omega}_Y(t, t', X))^4 \mid T = t', M, X]$ and their derivatives with respect to t are bounded uniformly over $(t', x) \in \mathcal{T} \times \mathcal{X}$ and $\int_{-\infty}^{\infty} k(u)^4 du < \infty$
5. The bandwidth $b_n \rightarrow 0$ and $nb_n^5 \rightarrow \infty$. $\int_{-\infty}^{\infty} k(u)k(u/\epsilon)du < \infty$ for $\epsilon \in (0, 1)$.

We are now in position to state another result related to our DML approach.

Corollary 11 (AMSE optimal bandwidth for $\hat{\eta}_{t,t'}$). *Let the conditions in Theorem 8 hold. For $t, t' \in \mathcal{T}$, if $B_{t,t'}$ is non-zero, then the bandwidth that minimizes the asymptotic mean squared error is $h_{t,t'}^* = (V_{t,t'} / (4 B_{t,t'}^2))^{1/5} n^{-1/5}$. Further let Assumption 10 hold. Then $\hat{V}_{t,t'} - V_{t,t'} = o_p(1)$, $\hat{B}_{t,t'} - B_{t,t'} = o_p(1)$, and $\hat{h}_{t,t'} / h_{t,t'}^* - 1 = o_p(1)$.*

Assumption 10 are for the consistency of $\hat{V}_{t,t'}$. The first one is stricter than Assumption 5, and the second is mild boundedness conditions that are implied when

$\widehat{\mu_{Y_\ell}}, \widehat{\omega_{Y_\ell}}, \widehat{f_{T|X_\ell}}$ and $\widehat{f_{T|X,M_\ell}}$ are bounded uniformly, which is often verified in practice. We now state the last corollary of the asymptotic normality of the DML estimator.

Corollary 12 (Asymptotic confidence interval for $\widehat{\eta}_{t,t'}$). *Let the conditions in Corollary 11 hold. For $t, t' \in \mathcal{T}$, we can construct the usual $(1-\alpha) \times 100\%$ point-wise confidence interval $\left[\widehat{\eta}_{t,t'} \pm \Phi^{-1}(1-\alpha/2) \sqrt{\widehat{V}_{t,t'}/(nh)} \right]$, where Φ is the CDF of $\mathcal{N}(0,1)$.*

5 NUMERICAL EVALUATION

This section provides a simulation study to investigate the finite sample behaviour of our methods. The code to reproduce our experiments can be found at <https://github.com/houssamzenati/double-debiased-machine-learning-mediation-continuous-treatments>.

In our experiments, for the kernel smoothing, we considered second order Gaussian kernels with kernel bandwidth either with the optimal AMSE bandwidth or a Scott (2015)-type rule of thumb (see Appendix 11.1) verifying the regularity conditions in Theorem 8. Furthermore, for the estimation of the nuisances parameters we used nonparametric estimators for the conditional and cross conditional mean outcomes μ_Y and ω_Y as detailed in Appendix 10, following Singh et al. (2023) which uses kernel mean embeddings. We performed parametric estimation of the generalized propensity scores, that is to say the conditional densities of the treatment. To this end, as in Hsu et al. (2020), we assume T to be normally distributed given X (or given (X, M) , respectively) and learned linear predictors \hat{t} of T conditional on X (or given (X, M)). More specifically, $\widehat{f_{T|X}} = \mathcal{N}(\hat{t}, \hat{\sigma})$ (respectively for $\widehat{f_{T|X,M}}$) where the noise $\hat{\sigma}$ is the empirical standard deviation of treatments.

In addition, we also estimate the direct and indirect effects with baseline methods. First, we implement the coefficient product method (Baron and Kenny, 1986; VanderWeele, 2015), that is a linear OLS regression of the outcome on a the treatment, the mediator, and the covariate and respectively the mediator on a the treatment and covariate. Second, we consider the kernel mean embedding (KME) method (Singh et al., 2023) which is a G-computation method. We also implement the generalized importance weighting (GIPW) estimator from Hsu et al. (2020). Eventually, we also considered the multiply robust estimator of Sani et al. (2024), with their explicit integration of the conditional mean outcome for estimating the cross conditional mean outcome. We refer the reader to the Appendices 10 and 11

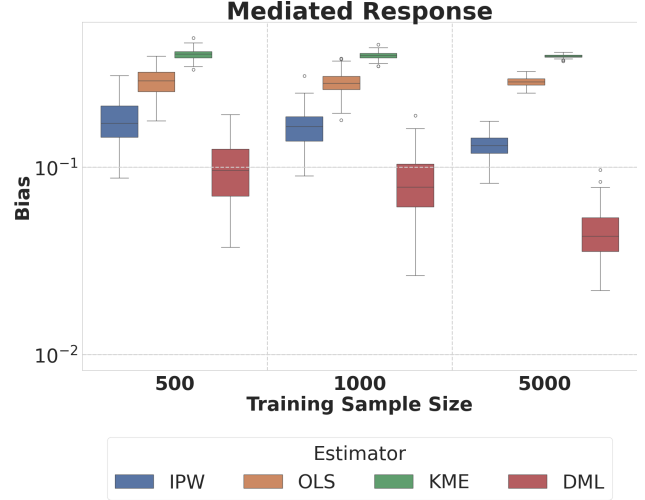


Figure 2: Bias of mediated response estimation on simulations with different sample sizes. DML significantly outperforms OLS and KME and also improves upon IPW.

for further details on the implementations.

5.1 Simulation

In this part we evaluate our approach following the simulations in Hsu et al. (2020), and Singh et al. (2023). In this simulation, the oracle for the mediated response $\eta_{t,t'}$, the direct effect $\theta(t, t')$ and the indirect effect by $\delta(t, t')$ are accessible (see all details in Appendix 11.2) and we can therefore evaluate the error of all estimators. We consider 100 simulations and three sample sizes $n = 500, 1000, 5000$ as in Singh et al. (2023) to investigate the performance of our approach.

Comparison to other methods In Figure 2 we compare the DML estimator to other approaches. We see that the DML significantly outperforms the OLS and KME approaches while improving upon the IPW method as well. We therefore find as in binary treatment studies (Tchetgen and Shpitser, 2012; Farbmacher et al., 2022) that the DML estimation of the mediated response practically outperforms alternative approaches.

Comparison to DML variants We compare different DML formulations for continuous treatments. We implemented the estimator from Sani et al. (2024), which estimates the cross-conditional mean outcome ω_Y using Explicit Integration (EI) of μ_Y and a ratio of mediator densities $f_{M|T,X}$ (MD). (EI) estimates $\widehat{\omega}_Y(t, t', x) = \sum_{\mathcal{M}} \widehat{\mu}_Y(t, m, x) \widehat{f}_{M|T,X}(m|t', x)$ (using Eq. (9)), while Implicit Integration (II) estimates $\widehat{\omega}_Y$ via two regressions (Farbmacher et al., 2022; Singh

et al., 2023). Our approach instead uses (II) of μ_Y and a ratio of treatment propensities $f_{T|X}, f_{T|X,M}$. Notably, (EI) is unsuitable for continuous mediators, and the mediator densities (MD) approach struggles with high-dimensional mediators. We also compare our DML estimator to (II) combined with (MD) to highlight its advantages.

Estimator	n	d_M	Bias	Std	RMSE
EI-MD	500	1	0.2429	0.1672	0.2979
II-MD	500	1	0.1011	0.0389	0.1208
(Ours)	500	1	0.1006	0.0375	0.1211
EI-MD	1000	1	0.3265	0.2745	0.3890
II-MD	1000	1	0.0834	0.0329	0.0995
(Ours)	1000	1	0.0828	0.0319	0.0985
EI-MD	5000	1	0.7319	0.6298	0.8475
II-MD	5000	1	0.0460	0.0146	0.0557
(Ours)	5000	1	0.0449	0.0145	0.0544
II-MD	500	5	0.7271	0.1742	0.9002
(Ours)	500	5	0.1709	0.0462	0.2178
II-MD	1000	5	0.5835	0.0924	0.7278
(Ours)	1000	5	0.1623	0.0369	0.2122
II-MD	5000	5	0.3393	0.0182	0.4209
(Ours)	5000	5	0.1319	0.0173	0.1843

Table 1: Comparison to different DML variants on our setting with mediator of dimension $d_M = 1$ and $d_M = 5$: (EI-MD) explicit integration and mediator densities (Sani et al., 2024), (II-MD) implicit integration and mediator densities, (Ours) with (II-TP) implicit integration and treatment propensities

Table 1 compares these variants for a one-dimensional mediator and of dimension 5. For a one-dimensional mediator our approach outperforms Sani et al. (2024) but performs similarly to the II-MD variant, as mediator density estimation is straightforward in this setting. However, for a higher dimension our approach outperforms the II-MD variant, where mediator density estimation is more challenging. In this multidimensional setting, the EI procedure is not suitable. To extend the comparison, we replicated the numerical simulation of Sani et al. (2024) in Appendix 11.3 using binary mediators, which favor (EI) of the conditional mean outcome. Using their metric—the average absolute bias—we compare our DML estimator to their variant, as well as IPS, OLS, and KME, and still find that our approach significantly outperforms theirs.

Bandwidth selection Moreover, in Appendix 11.2, we provide additional results where we compare the empirical performance of our DML approach under two bandwidth selection methods, one following the AMSE

strategy from Corollary 11 and the other following the Scott (2015) rule of thumb. Overall, the AMSE strategy provides less variance but shows slightly more bias than the heuristic. Moreover, we conclude that both approaches are suited for practical applications.

Additional experiments Eventually, we also provide in Appendix 11.2.2 an experiment to validate the coverage of the asymptotic confidence interval and an additional experiment to compare parametric and non-parametric estimators of the nuisances functions.

5.2 Application to cognitive function

We consider UK Biobank (Sudlow et al., 2015) imaging data, with around 40,000 brain scans, to study brain health. Prior research shows that while imaging-derived phenotypes (IDPs) have some predictive power for traits like fluid intelligence and neuroticism, adding IDPs to socio-demographic variables yields little to no improvement in predictive performance (Dadi et al., 2021; Cox et al., 2019). In this application, we assess whether the effect of poor glycemic control on cognitive function is mediated by brain structure, using glycated hemoglobin (as a proxy for the glycemic control (Jha et al., 2022)). In Appendix 11.4, along with additional details, we provide an experiment to measure the effect mediated by brain structural information. The DML results are on par with existing methods and show no total nor indirect effect of brain-derived features, which corroborates Dadi et al. (2021); Cox et al. (2019). However, we observe greater stability than IPW in regions where there is limited overlap in the data and for which identifiability is therefore not ensured.

6 DISCUSSION

In this work, we propose a DML estimator for mediation analysis with continuous treatments, using kernel smoothing (Hsu et al., 2020; Colangelo and Lee, 2020). Our estimator is root-n consistent under mild regularity conditions and, as a targeted DML estimator (Kennedy, 2022), exhibits asymptotic normality, allowing for asymptotic confidence intervals and mean squared error optimal bandwidth. We provide numerical evaluations in non-asymptotic regimes and apply the method to cognitive function analysis. Future work could explore targeted learning (van der Laan and Rubin, 2006; Zheng and van der Laan, 2012) with kernel smoothing, as it has proven effective without requiring cross-fitting, which can reduce effective sample size and does not improve finite-sample performance in simpler problems (Williamson et al., 2023; Fan et al., 2022b).

Acknowledgments

HZ and BT were supported by the KARAIB AI chair (ANR-20-CHIA-0025-01) and the H2020 Research Infrastructures Grant EBRAIN-Health 101058516. This research has been conducted using the UK Biobank Resource under Application Number 49314.

References

- F. Alfaro-Almagro, M. Jenkinson, N. K. Bangerter, J. L. Andersson, L. Griffanti, G. Douaud, S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, E. Vallee, et al. Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank. *Neuroimage*, 166:400–424, 2018.
- F. Alfaro-Almagro, P. McCarthy, S. Afyouni, J. L. Andersson, M. Bastiani, K. L. Miller, T. E. Nichols, and S. M. Smith. Confound modelling in uk biobank brain imaging. *NeuroImage*, 224:117002, 2021.
- R. Baron and D. Kenny. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51: 1173–1182, 01 1986.
- M. Bia and A. Mattei. Assessing the effect of the amount of financial aids to piedmont firms using the generalized propensity score. *Statistical Methods & Applications*, 21:485–516, 2012.
- M. Bonvini and E. H. Kennedy. Fast convergence rates for dose-response estimation, 2022. URL <https://arxiv.org/abs/2207.11825>.
- M. Carone, A. Luedtke, and M. Laan. Toward computerized efficient estimation in infinite-dimensional models. *Journal of the American Statistical Association*, 114, 08 2016.
- V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1), 2018.
- V. Chernozhukov, J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins. Locally Robust Semiparametric Estimation. *Econometrica*, 90(4):1501–1535, 2022a.
- V. Chernozhukov, W. Newey, and R. Singh. Debiased machine learning of global and local parameters using regularized riesz representer. *The Econometrics Journal*, 25, 2022b.
- K. Colangelo and Y.-Y. Lee. Double Debiased Machine Learning Nonparametric Inference with Continuous Treatments. Technical report, 2020. URL <https://arxiv.org/pdf/2004.03036>.
- S. R. Cox, S. J. Ritchie, C. Fawns-Ritchie, E. M. Tucker-Drob, and I. J. Deary. Structural brain imaging correlates of general intelligence in uk biobank. *Intelligence*, 76:101376, 2019.
- K. Dadi, G. Varoquaux, J. Houenou, D. Bzdok, B. Thirion, and D. Engemann. Population modeling with machine learning can enhance measures of mental health. *GigaScience*, 10, 10 2021.
- C. R. Doss, G. Weng, L. Wang, I. Moscovice, and T. Chantarat. A nonparametric doubly robust test for a continuous treatment effect, 2023. URL <https://arxiv.org/abs/2202.03369>.
- Q. Fan, Y.-C. Hsu, R. P. Lieli, and Y. Zhang. Estimation of Conditional Average Treatment Effects With High-Dimensional Data. *Journal of Business & Economic Statistics*, 40(1):313–327, 2022a.
- Q. Fan, Y.-C. Hsu, R. P. Lieli, and Y. Zhang. Estimation of Conditional Average Treatment Effects With High-Dimensional Data. *Journal of Business & Economic Statistics*, 40(1):313–327, January 2022b.
- H. Farbmacher, M. Huber, L. Laffers, H. Langen, and M. Spindler. Causal mediation analysis with double machine learning. *The Econometrics Journal*, 25, 01 2022.
- C. Fawns-Ritchie and I. J. Deary. Reliability and validity of the uk biobank cognitive tests. *PloS one*, 15 (4):e0231627, 2020.
- C. Flores. Estimation of dose-response functions and optimal doses with a continuous treatment. Working Papers 0707, University of Miami, Department of Economics, 2007.
- C. A. Flores and A. Flores-Lagunes. Identification and estimation of causal mechanisms and net effects of a treatment under unconfoundedness. IZA Discussion Papers 4237, Institute of Labor Economics (IZA), 2009.
- A. Galvao and L. Wang. Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment. *Journal of the American Statistical Association*, 110(512):1528–1542, 2015.
- R. D. Gill and J. M. Robins. Causal inference for complex longitudinal data: The continuous case. *The Annals of Statistics*, 29(6):1785–1811, 2001.
- K. Hirano and G. Imbens. *The Propensity Score with Continuous Treatments*, pages 73 – 84. 07 2005.
- G. Hong et al. Ratio of mediator probability weighting for estimating natural direct and indirect effects. In *Proceedings of the American Statistical Association, biometrics section*, pages 2401–2415, 2010.
- D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe.

- Journal of the American statistical Association*, 47 (260):663–685, 1952.
- Y.-C. Hsu, M. Huber, Y.-Y. Lee, and L. Lettry. Direct and indirect effects of continuous treatments based on generalized propensity score weighting. *Journal of Applied Econometrics*, 35, 04 2020.
- M. Huber. Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics*, 29(6):920–943, 2014.
- M. Huber, M. Lechner, and G. Mellace. The finite sample performance of estimators for mediation analysis under sequential conditional independence. *Journal of Business and Economic Statistics*, 02 2015.
- H. Ichimura and W. K. Newey. The influence function of semiparametric estimators. *Quantitative Economics*, 13(1):29–61, January 2022.
- K. Imai and D. A. Van Dyk. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866, 2004.
- K. Imai and T. Yamamoto. Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*, 21(2):141–171, 2013.
- K. Imai, L. Keele, and T. Yamamoto. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*, 25(1):51 – 71, 2010.
- G. W. Imbens. The Role of the Propensity Score in Estimating Dose-Response Functions. *Biometrika*, 87(3):706–710, 2000.
- G. W. Imbens and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.
- M. K. Jha, C. R. C. Fatt, A. Minhajuddin, T. L. Mayes, J. D. Berry, and M. H. Trivedi. Accelerated brain aging in individuals with diabetes: association with poor glycemic control and increased all-cause mortality. *Psychoneuroendocrinology*, 145:105921, 2022.
- J. Josse and F. Husson. missmda: a package for handling missing values in multivariate data analysis. *Journal of statistical software*, 70:1–31, 2016.
- C. M. Judd and D. A. Kenny. Process analysis: Estimating mediation in treatment evaluations. *Evaluation review*, 5(5):602–619, 1981.
- B. Keith, D. Eleanor, H. Maggie, L. Greg, O. Paul, O. Miruna, and V. Syrgkanis. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. <https://github.com/pywhy/EconML>, 2019. Version 0.x.
- E. H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*, 2022.
- E. H. Kennedy, Z. Ma, M. D. McHugh, and D. S. Small. Non-parametric Methods for Doubly Robust Estimation of Continuous Treatment Effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1229–1245, 09 2017.
- J. Kluve, H. Schneider, A. Uhlenborff, and Z. Zhao. Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 175 (2):587–617, 2012.
- S. Lê, J. Josse, and F. Husson. Factominer: an r package for multivariate analysis. *Journal of statistical software*, 25:1–18, 2008.
- Y.-Y. Lee. Partial mean processes with generated regressors: Continuous treatment effects and non-separable models. *arXiv preprint arXiv:1811.00157*, 2018.
- K. L. Miller, F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. Andersson, et al. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature neuroscience*, 19(11):1523–1536, 2016.
- D. Newby and V. Garfield. Understanding the interrelationships of type 2 diabetes and hypertension with brain and cognitive health: A uk biobank study. *Diabetes, Obesity and Metabolism*, 24(5):938–947, 2022.
- W. K. Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):1349–1382, 1994.
- J. Neyman. Optimal asymptotic tests of composite statistical hypotheses. *Probability and Statistics*, 1959.
- B. U. Park and J. S. Marron. Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409):66–72, 1990.
- J. Pearl. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, page 411–420, 2001.
- M. L. Petersen, S. E. Sinisi, and M. J. van der Laan. Estimation of direct causal effects. *Epidemiology*, 17 (3):276–284, 2006.
- J. L. Powell and T. M. Stoker. Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics*, 75(2):291–316, 1996.
- J. Robins. A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect. *Mathematical Modeling*, 7:1393–1512, 1986.

- J. Robins. *Semantics of causal DAG models and the identification of direct and indirect effects*, pages 70–82. 05 2003.
- J. Robins, A. Rotnitzky, and L. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of The American Statistical Association - J AMER STATIST ASSN*, 89:846–866, 09 1994.
- J. M. Robins and S. Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3:143–155, 1992.
- J. M. Robins and A. Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429):122–129, 1995.
- N. Sani, Y. Xu, A. Ghassami, and I. Shpitser. Multiply robust causal mediation analysis with continuous treatments, 2024. URL <https://arxiv.org/abs/2105.09254>.
- M. Schurz, L. Q. Uddin, P. Kanske, C. Lamm, J. Sallet, B. C. Bernhardt, R. B. Mars, and D. Bzdok. Variability in brain structure and function reflects lack of peer support. *Cerebral Cortex*, 31(10):4612–4627, 2021.
- D. W. Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.
- V. Semenova and V. Chernozhukov. Debiased machine learning of conditional average treatment effects and other causal functions, 2020. URL <https://arxiv.org/abs/1702.06240>.
- S. J. Sheather and M. C. Jones. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):683–690, 1991.
- R. Singh, L. Xu, and A. Gretton. Sequential kernel embedding for mediated and time-varying dose response curves, 2023.
- C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3), 2015.
- K. Takatsu and T. Westling. Debiased inference for a covariate-adjusted regression function. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 05 2024.
- E. Tchetgen and I. Shpitser. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *The Annals of Statistics*, 40, 10 2012.
- A. Topiwala, K. P. Ebmeier, T. Maullin-Sapey, and T. E. Nichols. Alcohol consumption and mri markers of brain structure and function: cohort study of 25,378 uk biobank participants. *NeuroImage: Clinical*, 35:103066, 2022.
- M. J. van der Laan and D. Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1), 2006.
- T. J. VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015.
- B. D. Williamson, P. B. Gilbert, N. R. Simon, and M. Carone. A general framework for inference on algorithm-agnostic variable importance. *Journal of the American Statistical Association*, 118(543):1645–1658, 2023.
- H. Zenati, A. Bietti, M. Martin, E. Diemert, P. Gaillard, and J. Mairal. Counterfactual learning of stochastic policies with continuous actions. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856.
- W. Zheng and M. J. van der Laan. Targeted maximum likelihood estimation of natural direct effects. *The International Journal of Biostatistics*, 8(1), 2012.
- M. Zimmert and M. Lechner. Nonparametric estimation of causal heterogeneity under high-dimensional confounding, 2019. URL <https://arxiv.org/abs/1908.08779>.

Checklist

1. For all models and algorithms presented, check if you include:
 - (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model.
Yes The setting along with assumptions, algorithm are detailed in Section 2 and 3.
 - (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm.
Yes The analysis of the algorithm presented in Section 3 is shown in Section 4.
 - (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries.
Yes The anonymized source code is provided in the supplementary materials and contains a Read.me file to run the experiments.
2. For any theoretical claim, check if you include:
 - (a) Statements of the full set of assumptions of all theoretical results.
Yes The analysis of the algorithm presented in Section 3 is shown in Section 4.
 - (b) Complete proofs of all theoretical results.
Yes The analysis of the algorithm presented in Section 3 is shown in Section 4.
 - (c) Clear explanations of any assumptions.
Yes A commentary and explanation of all assumptions in Section 3 and 4 are provided.
3. For all figures and tables that present empirical results, check if you include:
 - (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL).
Yes The instructions to reproduce the tables, plots are in Read.me file of the code in the supplementary file.
 - (b) All the training details (e.g., data splits, hyperparameters, how they were chosen).
Yes The training details are provided in Appendix 11.1 for hyperparameter selection and Appendix 11.2 and 11.4 for the remaining.
 - (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times).
Yes The definition of the metrics and the way to obtain error bars are described in Section 5 and in Appendix 11.2 and 11.4.
 - (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider).
- Yes** The computational infrastructure is described in Appendix 11.5.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
 - (a) Citations of the creator If your work uses existing assets.
Yes We cite the original paper of the UKBB project (Sudlow et al., 2015).
 - (b) The license information of the assets, if applicable.
No We address the reader to the paper explaining the UKBB project (Sudlow et al., 2015) for this matter.
 - (c) New assets either in the supplemental material or as a URL, if applicable.
Not applicable.
 - (d) Information about consent from data providers/curators.
No We address the reader to the paper explaining the UKBB project (Sudlow et al., 2015) for this matter.
 - (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. **No.**
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
 - (a) The full text of instructions given to participants and screenshots. **Not applicable.**
 - (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. **Not applicable.**
 - (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. **Not applicable.**

APPENDIX

This appendix is organized as follows:

- Appendix 7: summary of the notations used in the analysis.
- Appendix 8: discussion on kernel smoothing and proof for the construction of the moment function in Section 3.
- Appendix 9: proof for the asymptotic analysis in Section 4.
- Appendix 10: contains the implementation information of the nuisance parameters used in Section 5.
- Appendix 11: details on the implementation of the algorithms and additional experiment details, discussions and results.

All the code to reproduce our experiment can be found at <https://github.com/houssamzenati/double-debiased-machine-learning-mediation-continuous-treatments>.

7 NOTATIONS

In this appendix, we recall useful notations that are used throughout the paper.

Below are generic notations related to the mediation task:

- T is the treatment random variable, Y the outcome random variable, M the mediator(s) and the covariate(s) is X . We respectively write t, t' for treatment values, y, m, x for the outcome mediator and covariate.
- $\mathcal{Y}, \mathcal{T}, \mathcal{X}$ and \mathcal{M} are respectively the support of Y, T, X and M .
- $M(t)$ and $Y(t, M(t'))$ are respectively the potential mediator state and the potential outcome.
- $\mathcal{Z} = \mathcal{Y} \times \mathcal{T} \times \mathcal{X} \times \mathcal{M}$ and $F_Z(Z)$ is the cumulative density function over such a space.
- $f_{T|X}(t | X)$ is the conditional density of T given X , $f_{T|X,M}(t | X, M)$ is the conditional density of T given X and M , $f_{M|T,X}(M | T, X)$ the conditional density of M given T and X (if M is discrete, this is a conditional probability).
- $\tau(t, t'), \theta(t, t')$ and $\delta(t, t')$ are respectively the total, direct and indirect effects for the treatment values $t, t' \in \mathcal{T}$.
- $\eta_{t,t'} = \mathbb{E}[Y(t, M(t'))]$ is the mediated response.
- $\mu_Y(t, m, x) = \mathbb{E}[Y | T = t, M = m, X = x]$ is the conditional mean outcome, and $\omega_Y(t, t', x) = \mathbb{E}[\mu_Y(t, m, X) | T = t', X = x]$ is the cross conditional mean outcome.

Below are notations related to the continuous mediation task with the kernel and finite samples.

- $k : \mathcal{T} \rightarrow \mathbb{R}$ is a bounded positive definite kernel
- $R_k \equiv \int_{-\infty}^{\infty} k(u)^2 du$ is the roughness of the kernel.
- $\kappa \equiv \int_{-\infty}^{\infty} u^2 k(u) du$.
- h is the bandwidth.
- $K_h(T - t) \equiv \prod_{j=1}^{d_T} k((T_j - t_j) / h) / h^{d_T}$ is the kernel product for smoothing.
- $\psi_{t,t'}^h$ is the kernel moment function associated to the bandwidth h and the treatment values $t, t' \in \mathcal{T}$.
- n is the total sample size, L is the number of splits and i refers to an index, typically belonging to a split I_ℓ for $\ell \in 1, \dots, L$. n_ℓ is the size of the ℓ -th split.
- Y_i, T_i, X_i, M_i are realizations of the random variables Y, T, X, M for $i \in I_\ell$ for $\ell \in 1, \dots, L$.
- d_T is the dimension of the treatment space.

Below are notations related to the asymptotic analysis:

- $o_p(1)$ and $O_p(1)$ respectively refer to convergence and boundedness in probability.
- $\|\cdot\|_{tMX}^2 = \int_{\mathcal{M}, \mathcal{X}} (\cdot)^2 f_{TMX}(t, m, x) dm dx$ is a $L_2(tMX)$ partial norm for a given $t \in \mathcal{T}$.
- $\|\cdot\|_X^2 = \int_{\mathcal{X}} (\cdot)^2 f_X(x) dx$ is a $L_2(X)$ partial norm.
- $\|\cdot\|_{tX}^2 = \int_{\mathcal{X}} (\cdot)^2 f_{TX}(t, x) dx$ is a $L_2(tX)$ partial norm for a given $t \in \mathcal{T}$.
- $\widehat{\mu}_{Y_\ell}, \widehat{\omega}_{Y_\ell}, \widehat{f_{T|X}_\ell}$ and $\widehat{f_{T|X,M}_\ell}$ are estimators for the nuisance functions $\mu_Y, \omega_Y, f_{T|X}$ and $f_{T|X,M}$ associated to the split $\ell \in 1 \dots L$.
- $\overline{\mu_Y}, \overline{\omega_Y}, \overline{f_{T|X}}$ and $\overline{f_{T|X,M}}$ are their limits.
- $B_{t,t'}$ and $V_{t,t'}$ are respectively the asymptotic bias and variance in the asymptotic normality.

- $\hat{B}_{t,t'}$ and $\hat{V}_{t,t'}$ are estimators of the latter quantities.
- h^* is the optimal bandwidth, b_n is another bandwidth.

Eventually, we highlight the most important notations related to the proofs used in the Appendix 8 and 9.

- $F^{\epsilon h}$ defines through ϵ a trajectory CDF between the true CDF F^0 and the CDF F^h induced by the bandwidth h . $f^{\epsilon h}, f^0, f^h$ are the respective PDF.
- $\lambda(t', X) = 1/f_{T|X}(t' | X)$ is the inverse of the propensity score. $r(t, t', M, X) = f_{T|X,M}(t', M, X)/f_{T|X,M}(t | X, M)$.
- Z_ℓ^c denote the observations $()$ for $i \in I_\ell$
- $\widehat{\mu_{Y_{i\ell}}} = \widehat{\mu_{Y_\ell}}(t, X_i, M_i)$, $\widehat{\omega_{Y_{i\ell}}} = \widehat{\omega_{Y_\ell}}(t, t', X_i)$, $\hat{\lambda}_{i\ell} = \widehat{f_{T|X_\ell}}(t' | X_i)^{-1}$ and $\hat{r}_{i\ell} = \widehat{f_{T|X,M_\ell}}(t' | X_i, M_i) / \widehat{f_{T|X,M_\ell}}(t | X_i, M_i)$ for $i \in I_\ell$.
- $\overline{\mu_{Y_i}} = \overline{\mu_Y}(t, X_i, M_i)$, $\overline{\omega_{Y_i}} = \overline{\omega_Y}(t, t', X_i)$, $\bar{\lambda}_i = \overline{f_{T|X_\ell}}(t' | X_i)^{-1}$ and $\bar{r}_i = \overline{f_{T|X,M}}(t' | X_i, M_i) / \overline{f_{T|X,M}}(t | X_i, M_i)$.

8 MOMENT FUNCTION AND KERNEL SMOOTHING

In this section we provide a discussion on the moment function and kernel smoothing techniques and start with a Lemma that is derived from the analysis of Colangelo and Lee (2020).

Lemma 8.1. *Let $W \in \mathcal{W}$ be a random variable and f_{TW} be the joint probability on $\mathcal{T} \times \mathcal{W}$, for any $t \in \mathcal{T}$*

$$\int_{\mathcal{T}} K_h(s - t) f_{TW}(s, x) ds = f_{TW}(t, w) + O(h^2) \quad (19)$$

uniformly in $x \in \mathcal{X}$.

Proof. We use change of variables $u = (u_1, u_2, \dots, u_{d_T}) = (T - t)/h$, a Taylor expansion, the mean value theorem where \bar{t} is between t and $t + uh$, $f_{TW}(t, w)$ being bounded away from zero, and the second derivatives of $f_{TW}(t, w)$ being bounded uniformly over $(t', x) \in \mathcal{T} \times \mathcal{W}$ to show that

$$\int_{\mathcal{T}} K_h(s - t) f_{TW}(s, w) ds \quad (20)$$

$$= \int_{\mathcal{R}^{d_T}} \Pi_{j=1}^{d_T} k(u_j) f_{TW}(t + uh, w) du \quad (21)$$

$$= \int_{\mathcal{R}^{d_T}} \Pi_{j=1}^{d_T} k(u_j) \left(f_{TW}(t, w) + h \sum_{j=1}^{d_T} u_j \frac{\partial f_{TW}(t, w)}{\partial t_j} \right. \quad (22)$$

$$\left. + \frac{h^2}{2} \sum_{j=1}^{d_T} u_j^2 \frac{\partial^2 f_{TW}(t, w)}{\partial t_j^2} \right|_{t=\bar{t}} + \frac{h^2}{2} \sum_{j=1}^{d_T} \sum_{l=1, l \neq j}^{d_T} u_j u_l \frac{\partial^2 f_{TW}(t, w)}{\partial t_j \partial t_l} \Big|_{t=\bar{t}} \Big) du_1 \cdots du_{d_T} \quad (23)$$

$$= \int_{\mathcal{R}^{d_T}} \Pi_{j=1}^{d_T} k(u_j) f_{TW}(t, w) \left\{ 1 + h \sum_{j=1}^{d_T} u_j \frac{\partial f_{TW}(t, w)}{\partial t_j} \frac{1}{f_{TW}(t, w)} \right. \quad (24)$$

$$\left. + \frac{h^2}{2} \sum_{j=1}^{d_T} \left(u_j^2 \frac{\partial^2 f_{TW}(t, w)}{\partial t_j^2} \right) \Big|_{t=\bar{t}} + \sum_{l=1, l \neq j}^{d_T} u_j u_l \frac{\partial^2 f_{TW}(t, x)}{\partial t_j \partial t_l} \Big|_{t=\bar{t}} \right) \frac{1}{f_{TW}(t, x)} \Big\} du_1 \cdots du_{d_T} \quad (25)$$

$$= f_{TW}(t, w) (1 + h^2 C) \quad (26)$$

$$= f_{TW}(t, w) + O(h^2) \quad (27)$$

for some positive constant C , for any $t \in \mathcal{T}$, uniformly over $w \in \mathcal{W}$ and where the term linear in h equals zero because $\int uk(u) = 0$ by Assumption 4. \square

The latter result will be at the core of the analysis that will follow in Appendix 9. We now discuss the construction of the doubly robust moment function with the Gateaux derivatives. Importantly the expression in (30) will be the building block to construct estimators for the mediated response $\eta_{t,t'}$ and its linear functionals. We will prove Lemma 3.1 that we restate below.

Lemma 3.1. *Let $t, t' \in \mathcal{T}$ and let the moment function*

$$\begin{aligned} \psi_{t,t',h} = & \frac{K_h(T-t)f_{T|X,M}(t' | X, M)}{f_{T|X}(t' | X)f_{T|X,M}(t | X, M)} [Y - \mu_Y(t, M, X)] \\ & + \frac{K_h(T-t')}{f_{T|X}(t' | X)} [\mu_Y(t, M, X) - \omega_Y(t, t', X)] \\ & + \omega_Y(t, t', X) - \eta_{t,t'}. \end{aligned}$$

then, the asymptotic moment function is Neyman-Orthogonal.

Proof. One way to obtain the influence function is to calculate the limit of the Gateaux derivative with respect to a smooth deviation from the true distribution, as the deviation approaches a point mass, following Carone et al. (2016) and Ichimura and Newey (2022).

Let \mathcal{F} be the set of cumulative density functions (CDF) on the observations $Z \in \mathcal{Z}$. Let F^0 be the true CDF of Z (respectively f^0 the PDF) and the CDF F_Z^h (respectively f_Z^h the PDF) approach a point mass at Z as the bandwidth $h \rightarrow 0$. Consider

$$F^{\epsilon h} = (1 - \epsilon)F^0 + \epsilon F_Z^h, \quad (28)$$

for $\epsilon \in [0, 1]$ such that for all small enough ϵ , $F^{\epsilon h} \in \mathcal{F}$ and the corresponding PDF

$$f^{\epsilon h} = f^0 + \epsilon (f_Z^h - f^0). \quad (29)$$

For any $t \in \mathcal{T}$, let $\eta_{t,t'}(\cdot) : \mathcal{F} \rightarrow \mathcal{R}$. We note that:

$$\begin{aligned} \eta_{t,t'}(F) &= \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} \mathbb{E}[Y | T = t, M = m, X = x] f_{M|T,X}(m|t', x) f_X(x) dx dm dy \\ &= \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} y \frac{f_Z(y, t, m, x) f_X(x) f_{TMX}(t', m, x)}{f_{TMX}(t, m, x) f_{TX}(t', x)} dx dm dy \end{aligned}$$

(Ichimura and Newey, 2022) show that if a semiparametric estimator is asymptotically linear and locally regular, then the influence function is $\lim_{h \rightarrow 0} d\eta_{t,t'}(F^{\epsilon h}) / d\epsilon|_{\epsilon=0}$. We calculate the Gateaux derivative of the functional $\eta_{t,t'}(F^{\epsilon h})$ with respect to a deviation $F_Z^h - F^0$ from the true distribution F^0 . Then we show that our estimator is asymptotically equivalent to a sample average of the moment function.

$$\begin{aligned} \left. \frac{d}{d\epsilon} \eta_{t,t'}(F^{\epsilon h}) \right|_{\epsilon=0} &= \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} y \frac{d}{d\epsilon} \left(\frac{f_Z(y, t, m, x) f_X(x) f_{TMX}(t', m, x)}{f_{TMX}(t, m, x) f_{TX}(t', x)} \right) dy dm dx \Big|_{\epsilon=0} \\ &= \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} \frac{y}{f_{TMX}(t, m, x) f_{TX}(t', x)} [(f_Z^h(y, t, m, x) - f_Z^0(y, t, m, x)) f_{TMX}(t', m, x) f_X(x) \\ &\quad + f_Z(y, t, m, x) (f_{TMX}^h(t', m, x) - f_{TMX}^0(t', m, x)) f_X(x) \\ &\quad + f_Z(y, t, m, x) (f_X^h(x) - f_X^0(x)) f_{TMX}(t', m, x)] dy dm dx \\ &\quad - \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} y \frac{f_Z(y, t, m, x) f_X(x) f_{TMX}(t', m, x)}{[f_{TMX}(t, m, x) f_{TX}(t', x)]^2} (f_{TMX}^h(t, m, x) - f_{TMX}^0(t, m, x)) f_{TX}(t', x) dy dm dx \\ &\quad - \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} y \frac{f_Z(y, t, m, x) f_X(x) f_{TMX}(t', m, x)}{[f_{TMX}(t, m, x) f_{TX}(t', x)]^2} (f_{TX}^h(t', x) - f_{TX}^0(t', x)) f_{TMX}(t, m, x) dy dm dx \end{aligned}$$

Specifically the term associated with f_{TMX}^h in the previous term contributes in the total term by calculating

$$\begin{aligned}
 & \lim_{h \rightarrow 0} \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} y \frac{f_Z(y, t, m, x) f_X(x) f_{TMX}(t', m, x)}{f_{TMX}(t, m, x)^2 f_{TX}(t', x)} f_{TMX}^h(t, m, x) dy dm dx \\
 &= \lim_{h \rightarrow 0} \int_{\mathcal{X}, \mathcal{M}} \mu_Y(t, m, x) \frac{f_X(x) f_{TMX}(t', m, x)}{f_{TMX}(t, m, x) f_{TX}(t', x)} f_{TMX}^h(t, m, x) dm dx \\
 &= \lim_{h \rightarrow 0} \int_{\mathcal{X}, \mathcal{M}} \frac{\mu_Y(t, m, x)}{f_{T|X}(t' | x)} \frac{f_{T|X, M}(t' | m, x)}{f_{T|X, M}(t | m, x)} f_{TMX}^h(t, m, x) dm dx \\
 &= \frac{\mu_Y(t, M, X)}{f_{T|X}(t' | X)} \frac{f_{T|X, M}(t' | M, X)}{f_{T|X, M}(t | M, X)} \lim_{h \rightarrow 0} f_T^h(t)
 \end{aligned}$$

Moreover, the term associated with f_{TX}^h in the previous term contributes in the total term by calculating

$$\begin{aligned}
 & \lim_{h \rightarrow 0} \int_{\mathcal{X}} \int_{\mathcal{M}} \int_{\mathcal{Y}} y \frac{f_Z(y, t, m, x) f_X(x) f_{TMX}(t', m, x)}{f_{TMX}(t, m, x) f_{TX}^2(t', x)} f_{TX}^h(t', x) dy dm dx \\
 &= \lim_{h \rightarrow 0} \int_{\mathcal{X}} \int_{\mathcal{M}} \mu_Y(t, m, x) \frac{f_{TMX}(t', m, x)}{f_{TX}(t', x)} \frac{f_X(x)}{f_{TX}(t', x)} f_{TX}^h(t', x) dm dx \\
 &= \lim_{h \rightarrow 0} \int_{\mathcal{X}} \frac{\omega_Y(t, t', x)}{f_{T|X}(t' | x)} f_{TX}^h(t', x) dx \\
 &= \frac{\omega_Y(t, t', X)}{f_{T|X}(t' | X)} \lim_{h \rightarrow 0} f_T^h(t')
 \end{aligned}$$

With same arguments for the other terms, we obtain that the Gateaux derivative for the direction $f_Z^h - f^0$ is

$$\begin{aligned}
 \lim_{h \rightarrow 0} \frac{d}{d\epsilon} \eta_{t, t'}(F^{\epsilon h}) \Big|_{\epsilon=0} &= \omega_Y(t, t', X) - \eta_{t, t'} + \lim_{h \rightarrow 0} \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} \frac{y - \mu_Y(t, m, x)}{f_{T|X}(t' | x)} \frac{f_{T|X, M}(t' | m, x)}{f_{T|X, M}(t | m, x)} f_Z^h(y, t, x) dy dx \\
 &+ \lim_{h \rightarrow 0} \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} \frac{\mu_Y(t, m, x) - \omega_Y(t, t', x)}{f_{T|X}(t' | x)} f_Z^h(y, t', x) dy dx \\
 &= \omega_Y(t, t', X) - \eta_{t, t'} + \frac{Y - \mu_Y(t, M, X)}{f_{T|X}(t' | X)} \frac{f_{T|X, M}(t' | M, X)}{f_{T|X, M}(t | M, X)} \lim_{h \rightarrow 0} f_T^h(t) \\
 &+ \frac{\mu_Y(t, M, X) - \omega_Y(t, t', X)}{f_{T|X}(t' | X)} \lim_{h \rightarrow 0} f_T^h(t')
 \end{aligned} \tag{30}$$

Note that the last term in (30) is a partial mean that is a marginal integration over $\mathcal{Y} \times \mathcal{M} \times \mathcal{X}$, fixing the value of T at t . Thus the Gateaux derivative depends on the choice of f_T^h .

We then choose $f_Z^h(z) = K_h(Z - z)1\{f^0(z) > h\}$, following Ichimura and Newey (2022), so $\lim_{h \rightarrow 0} f_T^h(t) = \lim_{h \rightarrow 0} K_h(T - t)$. Indeed, we specify F_Z^h following (Ichimura and Newey, 2022). Let $K_h(Z) = \Pi_{j=1}^{d_z} k(Z_j/h)/h$, where $Z = (Z_1, \dots, Z_{d_z})'$ and k satisfies Assumption 4 and is continuously differentiable of all orders with bounded derivatives. Let $F^{\epsilon h} = (1 - \epsilon)F^0 + \epsilon F_Z^h$ with pdf with respect to a product measure given by $f^{\epsilon h}(z) = (1 - \epsilon)f^0(z) + \epsilon f^0(z)\delta_Z^h(z)$, where $\delta_Z^h(z) = K_h(Z - z)/f^0(z)$, a ratio of a sharply peaked pdf to the true density. Thus

$$f_Z^h(y, t, m, x) = K_h(Y - y)K_h(T - t)K_h(M - m)K_h(X - x)$$

It follows that $\lim_{h \rightarrow 0} f_T^h(t) = \lim_{h \rightarrow 0} K_h(T - t)$ and

$$\begin{aligned}
 & \lim_{h \rightarrow 0} \int_{\mathcal{X}, \mathcal{M}, \mathcal{Y}} \frac{y - \mu_Y(t, m, x)}{f_{T|X}(t' | x)} \frac{f_{T|X, M}(t' | m, x)}{f_{T|X, M}(t | m, x)} f_Z^h(y, t, m, x) dy dm dx \\
 &= \frac{Y - \mu_Y(t, M, X)}{f_{T|X}(t' | X)} \frac{f_{T|X, M}(t' | M, X)}{f_{T|X, M}(t | M, X)} \lim_{h \rightarrow 0} K_h(T - t) \\
 & \mathbb{E} \left[\frac{d}{d\epsilon} \eta_{t, t'}(F^{\epsilon h}) \Big|_{\epsilon=0} \right] = \mathbb{E} \left[\omega_Y(t, t', X) - \eta_{t, t'} + \frac{\mu_Y(t, M, X) - \omega_Y(t, t', X)}{f_{T|X}(t' | X)} K_h(T - t') \right. \\
 & \quad \left. + \frac{Y - \mu_Y(t, M, X)}{f_{T|X}(t' | X)} \frac{f_{T|X, M}(t' | M, X)}{f_{T|X, M}(t | M, X)} K_h(T - t) \right]
 \end{aligned}$$

As in Lemma 8.1, we can show that

$$\begin{aligned}
 & \mathbb{E} \left[\frac{Y - \mu_Y(t, M, X)}{f_{T|X}(t' | X)} \frac{f_{T|X, M}(t' | M, X)}{f_{T|X, M}(t | M, X)} K_h(T - t) \right] \\
 &= \mathbb{E} \left[\mathbb{E} [Y - \mu_Y(t, M, X) K_h(T - t) | M, X] \frac{f_{T|X, M}(t' | M, X)}{f_{T|X, M}(t | M, X) f_{T|X}(t' | X)} \right] \\
 &= \mathbb{E} \left[\mathbb{E} [Y - \mu_Y(t, M, X) | M, X] \frac{f_{T|X, M}(t' | M, X) f_{T|X, M}(t | M, X)}{f_{T|X, M}(t | M, X) f_{T|X}(t' | X)} \right] + O(h^2) \\
 &= O(h^2)
 \end{aligned}$$

and similarly by definition of μ_Y and ω_Y :

$$\begin{aligned}
 \mathbb{E} \left[\frac{\mu_Y(t, M, X) - \omega_Y(t, t', X)}{f_{T|X}(t' | X)} K_h(T - t') \right] &= \mathbb{E} \left[\frac{(\mu_Y(t, M, X) - \omega_Y(t, t', X)) f_{T|X, M}(t' | X, M)}{f_{T|X}(t' | X)} \right] + O(h^2) \\
 &= O(h^2)
 \end{aligned}$$

Therefore,

$$\mathbb{E} \left[\frac{d}{d\epsilon} \eta_{t, t'}(F^{\epsilon h}) \Big|_{\epsilon=0} \right] = O(h^2)$$

So Neyman orthogonality holds when $h \rightarrow 0$.

□

Note also that another motivation for this moment function lies in the common combination of two alternative estimators for $\eta_{t, t'}$, the regression estimator $\hat{\eta}_{t, t'}^{REG} = n^{-1} \sum_{i=1}^n \hat{\eta}(t, M_i, X_i)$ that is based on the identification in Eq.(10) in the main text, and the inverse probability weighting (IPW) estimator $\hat{\eta}_{t, t'}^{IPW} = n^{-1} \sum_{i=1}^n K_h(T_i - t) Y_i \frac{\widehat{f_{T|X, M}(t', X_i, M_i)}}{\widehat{f_{T|X, M}(t, X_i, M_i)} \widehat{f_{T|X}(t' | X_i)}}$ that is based on the identification in 12 in the main text. Adding the influence function that accounts for the one-step estimation partials out the first order effect of the one-step estimation on the final estimator, as discussed in (van der Laan and Rubin, 2006; Chernozhukov et al., 2022a). For $\hat{\eta}_{t, t'}^{IPW}$, when $\widehat{f_{T|X}}$ and $\widehat{f_{T|X, M}}$ are standard kernel density estimator with bandwidth h , Hsu et al. (2020) derive the asymptotically linear representation of $\hat{\eta}_{t, t'}^{IPW}$ that is first-order equivalent to our DML estimator. The moment function here is constructed by adding the influence function adjustments for estimating the nuisance functions $f_{T|X}, f_{T|X, M}$ to the original moment function $K_h(T - t) Y \frac{\widehat{f_{T|X, M}(t', X, M)}}{\widehat{f_{T|X, M}(t, X, M)} \widehat{f_{T|X}(t' | X)}}$.

Moreover, note that in binary treatment, under Assumptions 1, 2, and 3, the counterfactual mediated response $\mathbb{E}[Y(t, M(t'))]$ is identified by the following efficient moment function:

$$\begin{aligned} \mathbb{E}[Y(t, M(t'))] &= \mathbb{E}[\psi_{t,t'}], \\ \text{with } \psi_{t,t'} &= \frac{\mathbf{1}\{T=t\} \cdot f_{M|T,X}(M|t', X)}{\mathbb{P}(T=t|X) \cdot f_{M|T,X}(M|t, X)} \cdot [Y - \mu_Y(t, M, X)] \\ &\quad + \frac{\mathbf{1}\{T=t'\}}{\mathbb{P}(T=t'|X)} \cdot [\mu_Y(t, M, X) \\ &\quad - \int_{m \in \mathcal{M}} \mu_Y(t, m, X) \cdot f_{M|T,X}(m|t', X) dm] \\ &\quad + \int_{m \in \mathcal{M}} \mu_Y(t, m, X) \cdot f_{M|T,X}(m|t', X) dm \end{aligned} \quad (31)$$

Moreover, the efficient score function can alternatively be written as:

$$\begin{aligned} \eta(t, t') &= \mathbb{E}[\psi_{t,t'}^*], \\ \text{with } \psi_{t,t'}^* &= \frac{\mathbf{1}\{T=t\} \cdot (\mathbb{P}(T=t'|M, X))}{\mathbb{P}(T=t|M, X) \cdot (\mathbb{P}(T=t'|X))} \cdot [Y - \mu_Y(t, M, X)] \\ &\quad + \frac{\mathbf{1}\{T=t'\}}{\mathbb{P}(T=t'|X)} \cdot [\mu_Y(t, M, X) - \mathbb{E}[\mu_Y(t, M, X) | T=t', X]] \\ &\quad + \mathbb{E}[\mu_Y(t, M, X) | T=t', X]. \end{aligned} \quad (32)$$

This comes from:

$$\mathbb{E}[Y(t, M(t'))] = \mathbb{E}\left[Y \frac{\mathbf{1}\{T=t\}}{\mathbb{P}(T=t|M, X)} \frac{\mathbb{P}(T=t'|M, X)}{\mathbb{P}(T=t'|X)}\right] \quad (33)$$

and Bayes' law:

$$\frac{\mathbb{P}(T=t'|X, M)}{\mathbb{P}(T=t|X, M)\mathbb{P}(T=t'|X)} = \frac{1}{\mathbb{P}(T=t|X)} \frac{\mathbb{P}(T=t'|X, M)f_M(M|X)}{\mathbb{P}(T=t'|X)} \frac{\mathbb{P}(T=t|X)}{\mathbb{P}(T=t|X, M)f_M(M|X)} \quad (34)$$

$$= \frac{f(M|T=t', X)}{\mathbb{P}(T=t|X)f(M|T=t, X)} \quad (35)$$

$$(36)$$

9 ASYMPTOTIC ANALYSIS

In this section we will prove the main analysis results of this paper. We start by restating Assumption 6 with multidimensional d_T .

Assumption 13 (Nuisance rates). *There exist functions $\overline{\mu_Y}$, $\overline{\omega_Y}$, $\overline{f_{T|X,M}}$ and $\overline{f_{T|X}}$ that satisfy the following: For each $\ell = 1, \dots, L$,*

1. $\sqrt{nh^{d_T}} \left\| \delta_{T|X}^\ell \right\|_{t_X} \left\| \delta_{\mu_Y}^\ell \right\|_{t_{MX}} = o_p(1)$
2. $\sqrt{nh^{d_T}} \left\| \delta_{T|X,M}^\ell \right\|_{t_{MX}} \left\| \delta_{\mu_Y}^\ell \right\|_{t_{MX}} = o_p(1)$
3. $\sqrt{nh^{d_T}} \left\| \delta_{T|X}^\ell \right\|_{t_X} \left\| \delta_{\omega_Y}^\ell \right\|_X = o_p(1)$

We will now state again Theorem 8 with general dimension d_T of the treatment space \mathcal{T} and prove it.

Theorem 8. Let Assumptions 1-6 hold. Let $h \rightarrow 0$, $nh^{d_T} \rightarrow \infty$, and $nh^{d_T+4} \rightarrow C \in [0, \infty)$. Then for any $t \in \mathcal{T}$,

$$\sqrt{nh^{d_T}} (\hat{\eta}_{t,t'} - \eta_{t,t'}) = \sqrt{\frac{h^{d_T}}{n}} \sum_{i=1}^n \left[\frac{K_h(T_i - t) \cdot \overline{f_{T|X,M} t', i}}{\hat{f}_{T|X} i \cdot \overline{f_{T|X,M} t, i}} \cdot [Y - \overline{\mu_Y} i] + \frac{K_h(T_i - t')}{\hat{f}_{T|X} i} \cdot [\overline{\mu_Y} i - \overline{\omega_Y} i] + \overline{\omega_Y} i \right] + o_p(1).$$

Further $\mathbb{E} [|Y - \overline{\mu_Y}(T, M, X)|^3 | T = t, M, X]$, $\mathbb{E} [(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^3 | T = t', X]$ and their derivatives with respect to t be bounded uniformly over $(t, m, x) \in \mathcal{T} \times \mathcal{M} \times \mathcal{X}$. Let $\int_{-\infty}^{\infty} k(u)^3 du < \infty$. Then

$$\sqrt{nh^{d_T}} (\hat{\eta}_{t,t'} - \eta_{t,t'} - h^2 B_{t,t'}) \xrightarrow{d} \mathcal{N}(0, V_{t,t'})$$

where

$$\begin{aligned} V_{t,t'} &\equiv \mathbb{E} \left[\mathbb{E} [(Y - \overline{\mu_Y}(t, M, X))^2 | T = t, M, X] \frac{f_{T|X,M}(t, X, M)}{\hat{f}_{T|X}(t', X)^2} \frac{\overline{f_{T|X,M}}(t', X, M)^2}{\overline{f_{T|X,M}}(t, X, M)^2} \right. \\ &\quad \left. + \frac{f_{T|X}(t' | X)}{\hat{f}_{T|X}(t' | X)^2} \mathbb{E} [(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^2 | T = t', X] \right] R_k^{d_T} \\ B_{t,t'} &\equiv \sum_{j=1}^{d_T} \mathbb{E} \left[\frac{\overline{f_{T|X,M}}(t', X, M)}{\overline{f_{T|X,M}}(t, X, M)} \left(\partial_{t_j} \mu_Y(t, M, X) \frac{\partial_{t_j} f_{T|X,M}(t | X, M)}{\hat{f}_{T|X}(t' | X)} \right. \right. \\ &\quad \left. \left. + \partial_{t_j}^2 \mu_Y(t, M, X) \frac{(f_{T|X,M}(t | X, M) + \mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X))}{\hat{f}_{T|X}(t' | X)} \right. \right. \\ &\quad \left. \left. + (\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)) \partial_{t_j}^2 \frac{f_{T|X,M}(t', X, M)}{\hat{f}_{T|X}(t' | X)} \right) / 2 \right] \kappa \end{aligned}$$

Proof. The proof is twofolds:

- Asymptotical linear representation of the estimated mediated response $\hat{\eta}_{t,t'}$
- Asymptotic normality

9.0.1 Asymptotical linear representation of the estimated mediated response $\hat{\eta}$

We give an outline of deriving the asymptotically linear representation in Theorem 8. Let the moment function for identification

$$m(Z_i, \eta_{t,t'}, \omega_Y) \equiv \omega_Y(t, t', X_i) - \eta_{t,t'} \quad (37)$$

by (10), $\mathbb{E}[m(Z_i, \eta_{t,t'}, \omega_Y(t, t', X_i))] = 0$ uniquely defines β_t . Let the adjustment terms

$$\phi_t(Z_i, \eta_{t,t'}, \mu_Y, f_{T|X}, \overline{f_{T|X,M}}) \equiv \frac{K_h(T_i - t) f_{T|X,M}(t' | x, m)}{\hat{f}_{T|X}(t' | x) \overline{f_{T|X,M}}(t | x, m)} (Y_i - \mu_Y(t, X_i, \psi_i)) \quad (38)$$

$$\varphi_{t'}(Z_i, \eta_{t,t'}, \mu_Y, \omega_Y, f_{T|X}) \equiv \frac{K_h(T_i - t')}{\hat{f}_{T|X}(t' | x)} (\mu_Y(t, X_i, \psi_i) - \omega_Y(t, t', X_i)) \quad (39)$$

The doubly robust moment function then writes as

$$\psi(Z_i, \eta_{t,t'}, \mu_Y, \omega_Y, f_{T|X}, \overline{f_{T|X,M}}) \equiv m(Z_i, \eta_{t,t'}, \omega_Y) + \phi_t(Z_i, \eta_{t,t'}, \mu_Y, f_{T|X}, \overline{f_{T|X,M}}) + \varphi_{t'}(Z_i, \eta_{t,t'}, \mu_Y, \omega_Y, f_{T|X}).$$

Let $r_i \equiv \frac{f_{T|X,M}(t' | M_i, X_i)}{\overline{f_{T|X,M}}(t | M_i, X_i)}$, Z^ℓ denote the observations z_i for $i \in I_\ell$. Let $\widehat{\mu_Y}_{i\ell} \equiv \widehat{\mu_Y}_\ell(t, X_i, M_i)$, $\widehat{\omega_Y}_{i\ell} \equiv \widehat{\omega_Y}_\ell(t, t', X_i)$ and $\widehat{f_{T|X}}_{i\ell} \equiv \widehat{f_{T|X}}_\ell(t' | X_i)$, $\hat{\lambda}_{i\ell} \equiv \widehat{f_{T|X}}_\ell(t' | X_i)^{-1}$, $\hat{r}_{i\ell} \equiv \frac{\widehat{f_{T|X,M}}_\ell(t' | X_i, M_i)}{\widehat{f_{T|X,M}}_\ell(t | X_i, M_i)}$ using Z^ℓ for $i \in I_\ell$. Let $\overline{\mu_Y}_i \equiv$

$\overline{\mu_Y}(t, X_i, M_i)$, $\overline{\omega_Y}_i \equiv \overline{\omega_Y}(t, t', X_i)$ and $\overline{f_{T|X}i} \equiv \overline{f_{T|X}}(t', X_i)$, $\bar{\lambda}_i \equiv \overline{f_{T|X}i}(t' | X_i)^{-1}$, $\bar{r}_i \equiv \frac{\overline{f_{T|X,M}}(t' | X_i, M_i)}{\overline{f_{T|X,M}}(t | X_i, M_i)}$. We can write, with $n_\ell = n/L$. Then

$$\begin{aligned}\widehat{\eta}_{t,t'} &= L^{-1} \sum_{\ell=1}^L \widehat{\eta}_{t,t',\ell} \\ \widehat{\eta}_{t,t',\ell} &= n_\ell^{-1} \sum_{i \in I_\ell} \psi \left(Z_i, \eta_{t,t'}, \widehat{\mu_{Yi}}, \widehat{\omega_{Yi}}, \widehat{f_{T|X}i}, \widehat{f_{T|X,M}i} \right) + \eta_{t,t'} \\ \sqrt{nh^{d_T}} (\widehat{\eta}_{t,t'} - \eta_{t,t'}) &= \sqrt{nh^{d_T}} L^{-1} \sum_{\ell=1}^L (\widehat{\eta}_{t,t',\ell} - \eta_{t,t'}) = L^{-1/2} \sum_{\ell=1}^L \sqrt{n_\ell h^{d_T}} (\widehat{\eta}_{t,t',\ell} - \eta_{t,t'})\end{aligned}\quad (40)$$

We show below $\sqrt{n_\ell h^{d_T}} (\widehat{\eta}_{t,t',\ell} - \eta_{t,t'}) = \sqrt{h^{d_T}/n_\ell} \sum_{i \in I_\ell} \psi(Z_i, \eta_{t,t'}, \overline{\mu_Y}_i, \overline{\omega_Y}_i, \overline{f_{T|X}i}, \overline{f_{T|X,M}i}) + o_p(1)$ for each $\ell \in \{1, \dots, L\}$.

Since L is fixed and $\{I_\ell\}_{\ell=1, \dots, L}$ are randomly partitioned distinct subgroups, the result follows from

$$\begin{aligned}\sqrt{nh^{d_T}} (\widehat{\eta}_{t,t'} - \eta_{t,t'}) &= L^{-1/2} \sum_{\ell=1}^L \sqrt{n_\ell h^{d_T}} (\widehat{\eta}_{t,t',\ell} - \eta_{t,t'}) \\ &= L^{-1/2} \sum_{\ell=1}^L \sqrt{h^{d_T}/n_\ell} \sum_{i \in I_\ell} \psi(Z_i, \eta_{t,t'}, \overline{\mu_Y}_i, \overline{\omega_Y}_i, \overline{f_{T|X}i}, \overline{f_{T|X,M}i}) + o_p(1) \\ &= \sqrt{h^{d_T}/n} \sum_{\ell=1}^L \sum_{i \in I_\ell} \psi(Z_i, \eta_{t,t'}, \overline{\mu_Y}_i, \overline{\omega_Y}_i, \overline{f_{T|X}i}, \overline{f_{T|X,M}i}) + o_p(1)\end{aligned}$$

We decompose the remainder term for each $\ell \in \{1, \dots, L\}$,

$$\begin{aligned}&\sqrt{n_\ell h^{d_T}} \frac{1}{n_\ell} \sum_{i \in I_\ell} \left\{ \psi \left(Z_i, \eta_{t,t'}, \widehat{\mu_{Yi}}, \widehat{\omega_{Yi}}, \widehat{f_{T|X}i}, \widehat{f_{T|X,M}i} \right) - \psi \left(Z_i, \eta_{t,t'}, \overline{\mu_Y}_i, \overline{\omega_Y}_i, \overline{f_{T|X}i}, \overline{f_{T|X,M}i} \right) \right\} \\ &= K_h(T_i - t) \left\{ \widehat{r}_{i\ell} \widehat{\lambda}_i \{Y_i - \widehat{\mu_{Yi}}\} - \bar{r}_i \bar{\lambda}_i \{Y_i - \overline{\mu_Y}_i\} \right\} + K_h(T_i - t') \left\{ \widehat{\lambda}_{i\ell} \{\widehat{\mu_{Yi}} - \widehat{\omega_{Yi}}\} - \bar{\lambda}_i \{\overline{\mu_Y}_i - \overline{\omega_Y}_i\} \right\} \\ &\quad + \widehat{\omega_{Yi\ell}} - \overline{\omega_Y}_i\end{aligned}$$

The method for the analysis below lies on the decomposition of the outcome residual:

$$\begin{aligned}\widehat{r}_{i\ell} \widehat{\lambda}_i \{Y_i - \widehat{\mu_{Yi}}\} - \bar{r}_i \bar{\lambda}_i \{Y_i - \overline{\mu_Y}_i\} &= (\widehat{r}_{i\ell} - \bar{r}_i) (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) (Y_i - \overline{\mu_Y}_i) - (\widehat{r}_{i\ell} - \bar{r}_i) (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu_{Yi}} - \overline{\mu_Y}_i) \\ &\quad - (\widehat{r}_{i\ell} - \bar{r}_i) (\widehat{\mu_{Yi}} - \overline{\mu_Y}_i) \bar{\lambda}_i - (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu_{Yi}} - \overline{\mu_Y}_i) \bar{r}_i \\ &\quad + (\widehat{r}_{i\ell} - \bar{r}_i) (Y_i - \overline{\mu_Y}_i) \bar{\lambda}_i + (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) (Y_i - \overline{\mu_Y}_i) \bar{r}_i - (\widehat{\mu_{Yi}} - \overline{\mu_Y}_i) \bar{r}_i \bar{\lambda}_i,\end{aligned}\quad (41)$$

and the decomposition of the conditional mean outcome residual:

$$\begin{aligned}\widehat{\lambda}_i \{\widehat{\mu_{Yi}} - \widehat{\omega_{Yi}}\} - \bar{\lambda}_{i\ell} \{Y_i - \overline{\mu_Y}_i\} &= (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu_{Yi}} - \overline{\mu_Y}_i) - (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\omega_{Yi}} - \overline{\omega_Y}_i) + (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) \overline{\mu_Y}_i \\ &\quad + (\widehat{\mu_{Yi}} - \overline{\mu_Y}_i) \bar{\lambda}_i - (\widehat{\omega_{Yi}} - \overline{\omega_Y}_i) \bar{\lambda}_i - (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) \overline{\omega_Y}_i.\end{aligned}\quad (42)$$

This intuition is to use such decompositions to leverage the assumptions on convergence and multiply robust inference.

We then expand the remainder into the terms:

$$\begin{aligned} & \sqrt{n_\ell h^{d_T}} \frac{1}{n_\ell} \sum_{i \in I_\ell} \left[K_h(T_i - t) \left\{ \hat{r}_{i\ell} \hat{\lambda}_i \{Y_i - \widehat{\mu}_{Y_{i\ell}}\} - \bar{r}_i \bar{\lambda}_i \{Y_i - \overline{\mu}_{Y_i}\} \right\} \right. \\ & \left. + K_h(T_i - t') \left\{ \hat{\lambda}_{i\ell} \{\widehat{\mu}_{Y_{i\ell}} - \widehat{\omega}_{Y_{i\ell}}\} - \bar{\lambda}_i \{\overline{\mu}_{Y_i} - \overline{\omega}_{Y_i}\} \right\} + \widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i} \right] \end{aligned}$$

$$= \sqrt{\frac{h^{d_T}}{n_\ell}} \sum_{i \in I_\ell} \left[\widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i} - \mathbb{E}[\widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i} \mid Z^\ell] \right. \quad (\text{R1-1})$$

$$- K_h(T_i - t') (\widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i}) \bar{\lambda}_i - \mathbb{E}[K_h(T_i - t') (\widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i}) \bar{\lambda}_i \mid Z_{i\ell}^c] \quad (\text{R1-2})$$

$$+ \mathbb{E}[(\widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i}) (1 - K_h(T_i - t') \bar{\lambda}_i) \mid Z_{i\ell}^c] \quad (\text{MR-1})$$

$$- K_h(T_i - t) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{\lambda}_i \bar{r}_i - \mathbb{E}[K_h(T_i - t) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{\lambda}_i \bar{r}_i \mid Z_{i\ell}^c] \quad (\text{R1-3})$$

$$+ K_h(T_i - t') (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{\lambda}_i - \mathbb{E}[K_h(T_i - t') (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{\lambda}_i \mid Z_{i\ell}^c] \quad (\text{R1-4})$$

$$+ \mathbb{E}[(\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \{K_h(T_i - t') \bar{\lambda}_i - K_h(T_i - t) \bar{\lambda}_i \bar{r}_i\} \mid Z_{i\ell}^c] \quad (\text{MR-2})$$

$$+ K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\overline{\mu}_{Y_i} - \overline{\omega}_{Y_i}) - \mathbb{E}[K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\overline{\mu}_{Y_i} - \overline{\omega}_{Y_i}) \mid Z_{i\ell}^c] \quad (\text{R1-5})$$

$$+ \mathbb{E}[K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) \{\overline{\mu}_{Y_i} - \overline{\omega}_{Y_i}\} \mid Z_{i\ell}^c] \quad (\text{MR-3})$$

$$+ K_h(T_i - t) (\hat{r}_{i\ell} - \bar{r}_i) \bar{\lambda}_i (Y_i - \overline{\mu}_{Y_i}) - \mathbb{E}[K_h(T_i - t) (\hat{r}_{i\ell} - \bar{r}_i) \bar{\lambda}_i (Y_i - \overline{\mu}_{Y_i}) \mid Z_{i\ell}^c] \quad (\text{R1-6})$$

$$+ \mathbb{E}[K_h(T_i - t) (\hat{r}_{i\ell} - \bar{r}_i) \bar{\lambda}_i (Y_i - \overline{\mu}_{Y_i}) \mid Z_{i\ell}^c] \quad (\text{MR-4})$$

$$+ K_h(T_i - t) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) \bar{r}_i (Y_i - \overline{\mu}_{Y_i}) - \mathbb{E}[K_h(T_i - t) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) \bar{r}_i (Y_i - \overline{\mu}_{Y_i}) \mid Z_{i\ell}^c] \quad (\text{R1-7})$$

$$+ \mathbb{E}[K_h(T_i - t) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) \bar{r}_i (Y_i - \overline{\mu}_{Y_i}) \mid Z_{i\ell}^c] \quad (\text{MR-5})$$

$$- K_h(T_i - t) (\hat{r}_{i\ell} - \bar{r}_i) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \quad (\text{R2-1})$$

$$- K_h(T_i - t) (\hat{r}_{i\ell} - \bar{r}_i) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{\lambda}_i \quad (\text{R2-2})$$

$$- K_h(T_i - t) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{r}_i \quad (\text{R2-3})$$

$$+ K_h(T_i - t) (\hat{r}_{i\ell} - \bar{r}_i) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (Y_i - \overline{\mu}_{Y_i}) \quad (\text{R2-4})$$

$$+ K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \quad (\text{R2-5})$$

$$- K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i}) \quad (\text{R2-6})$$

We will now bound (R1-1)-(R1-7), (R2-1)-(R2-6) and (MR-1)-(MR-5). The statements in the following hold for $i \in I_\ell, \ell \in \{1, \dots, L\}$, and for all t, t' .

Bounding the residuals (R1-1)-(R1-7) The remainder terms R1-1 to R1-7 are stochastic equicontinuous terms that are controlled to be $o_p(1)$ by the mean-square consistency conditions in Assumption 5 and cross-fitting.

For (R1-1), define $\Delta_{i\ell} = \widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i} - \mathbb{E}[\widehat{\omega}_{Y_{i\ell}} - \overline{\omega}_{Y_i} \mid Z^\ell]$. By construction and independence of Z^ℓ and Z_i for $i \in I_\ell, \mathbb{E}[\Delta_{i\ell} \mid Z^\ell] = 0$ and $\mathbb{E}[\Delta_{i\ell} \Delta_{j\ell} \mid Z^\ell] = 0$ for $i, j \in I_\ell$. By Assumptions 3 and 5,

$$h^{d_T} \mathbb{E}[\Delta_{i\ell}^2 \mid Z^\ell] = O_p \left(h^{d_T} \int_{\mathcal{X}} (\widehat{\omega}_{Y_\ell}(t, t', x) - \overline{\omega}_{Y_\ell}(t, t', x))^2 f_X(x) dx \right) = o_p(1)$$

$$\begin{aligned} \mathbb{E} \left[\left(\sqrt{\frac{h^{d_T}}{n}} \sum_{i \in I_\ell} \Delta_{i\ell} \right)^2 \mid Z^\ell \right] &= \frac{h^{d_T}}{n} \sum_{i \in I_\ell} \mathbb{E} [\Delta_{i\ell}^2 \mid Z^\ell] = O_p \left(h^{d_T} \int_{\mathcal{X}} (\widehat{\omega}_{Y_\ell}(t, t', x) - \overline{\omega_Y}(t, t', x))^2 f_X(x) dx \right) \\ &= o_p(1) \end{aligned}$$

The conditional Markov's inequality implies that $\sqrt{h^{d_T}/n} \sum_{i \in I_\ell} \Delta_{i\ell} = o_p(1)$.

The analogous results hold for (R1-2), (R1-3), (R1-4) due to convergence Assumption 5 and the boundedness Assumption 7. For (R1-5), (R1-6) and (R1-7), a standard algebra using change of variables, a Taylor expansion, the mean value theorem, and Assumption 7 yields for example for $\Delta_{i\ell} = K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\overline{\mu_{Y_i}} - \overline{\omega_{Y_i}}) - \mathbb{E} [K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\overline{\mu_{Y_i}} - \overline{\omega_{Y_i}}) \mid Z_{i\ell}^c]$ in (R1-5):

$$\begin{aligned} &h^{d_T} \mathbb{E} [\Delta_{i\ell}^2 \mid Z^\ell] \\ &\leq h^{d_T} \mathbb{E} \left[K_h(T_i - t')^2 (\hat{\lambda}_{i\ell} - \bar{\lambda}_i)^2 (\overline{\mu_{Y_i}} - \overline{\omega_{Y_i}})^2 \mid Z^\ell \right] \\ &= h^{d_T} \int_{\mathcal{X}, \mathcal{M}, \mathcal{T}} K_h(s - t')^2 (\hat{\lambda}_\ell(t', x) - \bar{\lambda}(t', x))^2 \mathbb{E} [(\overline{\mu_Y}(t, x, m) - \overline{\omega_Y}(t, t', x))^2 \mid T = s, X = x, M = m] \\ &\quad \times f_{TXM}(s, x, m) ds dx dm \\ &= \int_{\mathcal{X}, \mathcal{M}} \int_{\mathcal{R}^{d_T}} \Pi_{j=1}^{d_T} k(u_j)^2 (\hat{\lambda}_\ell(t', x) - \bar{\lambda}(t', x))^2 \mathbb{E} [(\overline{\mu_Y}(t, x, m) - \overline{\omega_Y}(t, t', x))^2 \mid T = t' + uh, X = x, M = m] \\ &\quad \times f_{TXM}(t' + uh, x, m) du dx dm \\ &= \int_{\mathcal{X}, \mathcal{M}} \int_{\mathcal{R}^{d_T}} (\mathbb{E} [(\overline{\mu_Y}(t, x, m) - \overline{\omega_Y}(t, t', x))^2 \mid T = t', X = x] \\ &\quad + \sum_{j=1}^{d_T} u_j h \frac{\partial}{\partial t_j} \mathbb{E} [(\overline{\mu_Y}(t, x, m) - \overline{\omega_Y}(t, t', x))^2 \mid T = t', X = x] \Big|_{t=\bar{t}}) \\ &\quad \times \left(f_{TXM}(t', x, m) + \sum_{j=1}^{d_T} u_j h \frac{\partial}{\partial t_j} f_{TXM}(t', x, m) \Big|_{t=\bar{t}} \right) \Pi_{j=1}^{d_T} k(u_j)^2 du (\hat{\lambda}_\ell(t', x) - \bar{\lambda}(t', x))^2 dx dm \\ &= O_p \left(\int_{\mathcal{X}, \mathcal{M}} (\hat{\lambda}_\ell(t', x) - \bar{\lambda}(t', x))^2 f_{TXM}(t', x, m) dx dm \right), \tag{KD} \end{aligned}$$

where \bar{t} and \bar{t} are between t and $t + uh$. So $h^{d_T} \mathbb{E} [\Delta_{i\ell}^2 \mid Z^\ell] = o_p(1)$ by Assumption 5. The conditional Markov's inequality implies that $\sqrt{h^{d_T}/n} \sum_{i \in I_\ell} \Delta_{i\ell} = o_p(1)$.

We have then proven that all residuals (R1-1) = $o_p(1)$ up until (R1-7) = $o_p(1)$.

Bounding the error differences (R2-1)-(R2-6) The second-order remainder terms (R2-1)-(R2-6) are controlled by Assumption 5.2. All of these terms involve the product of two or more errors and can be addressed in a similar manner. We present a detailed proof for (R2-6), with the same approach applicable to the remaining terms.

For (R2-6),

$$\begin{aligned}
 & \mathbb{E} \left[\left| \sqrt{h^{d_T}/n_\ell} \sum_{i \in I_\ell} K_h(T_i - t') \left(\hat{\lambda}_{i\ell} - \bar{\lambda}_i \right) (\bar{\omega}_{Yi} - \widehat{\omega}_{Yi\ell}) \right| \mid Z^\ell \right] \\
 & \leq \sqrt{n_\ell h^{d_T}} \int_{\mathcal{X}, \mathcal{T}} \left| \left(\hat{\lambda}_\ell(t', x) - \bar{\lambda}(t', x) \right) (\bar{\omega}_Y(t, t', x) - \widehat{\omega}_{Y\ell}(t, t', x)) \right| K_h(s - t') f_{TX}(s, x) ds dx \\
 & \leq \sqrt{n_\ell h^{d_T}} \left(\int_{\mathcal{X}, \mathcal{T}} \left(\hat{\lambda}_\ell(t', x) - \bar{\lambda}(t', x) \right)^2 K_h(s - t') f_{TX}(s, x) ds dx \right)^{1/2} \\
 & \quad \times \left(\int_{\mathcal{X}, \mathcal{T}} (\bar{\omega}_Y(t, t', x) - \widehat{\omega}_{Y\ell}(t, t', x))^2 K_h(s - t') f_{TX}(s, x) ds dx \right)^{1/2} \\
 & = \sqrt{n_\ell h^{d_T}} \left(\int_{\mathcal{X}} \left(\hat{\lambda}_\ell(t', x) - \bar{\lambda}(t', x) \right)^2 f_{TX}(t, x) dx \right)^{1/2} \left(\int_{\mathcal{X}} (\widehat{\omega}_{Y\ell}(t, t', x) - \bar{\omega}_Y(t, t', x))^2 f_{TX}(t, x) dx \right)^{1/2} \\
 & \quad + o_p(h^2) \\
 & = o_p(1)
 \end{aligned}$$

by Cauchy-Schwartz inequality, Assumption 5.2, and an application of Lemma (8.1). So (R2-6) = $o_p(1)$ follows by the conditional Markov's and triangle inequalities.

Bounding (MR-1)-(MR-5) The remainder terms MR-1-MR-5 are the key to multiply robust inference. Note that in the binary treatment case when $K_h(T_i - t)$ is replaced by $1\{T_i = t\}$, the sum of those terms is zero because ψ is the Neyman-orthogonal influence function, under correct specification $\bar{\mu}_Y = \mu_Y$, $\bar{\omega}_Y = \omega_Y$, $\bar{\lambda} = \lambda$ and $\bar{r} = r$. In our continuous treatment case, the Neyman orthogonality holds as $h \rightarrow 0$.

For (MR-1), we write

$$\begin{aligned}
 & \sqrt{nh^{d_T}} \mathbb{E} [(\widehat{\omega}_{Yi\ell} - \bar{\omega}_{Yi}) (1 - K_h(T_i - t') \bar{\lambda}_i) \mid Z_{i\ell}^c] \\
 & = \sqrt{nh^{d_T}} \int_{\mathcal{X}, \mathcal{T}} (\widehat{\omega}_Y(t, t', x) - \bar{\omega}_Y(t, t', x)) (1 - K_h(s - t') \bar{\lambda}(t', x)) f_{TX}(s, x) ds dx \\
 & = \sqrt{nh^{d_T}} \int_{\mathcal{X}} (\widehat{\omega}_Y(t, t', x) - \bar{\omega}_Y(t, t', x)) \left(1 - \left\{ \int_{\mathcal{T}} K_h(s - t') f_{T|X}(s \mid x) ds \right\} \bar{\lambda}(t', x) \right) f_X(x) dx \\
 & = \sqrt{nh^{d_T}} \int_{\mathcal{X}} (\widehat{\omega}_Y(t, t', x) - \bar{\omega}_Y(t, t', x)) (1 - f_{T|X}(t' \mid x) \bar{\lambda}(t', x)) f_X(x) dx \\
 & \quad + \sqrt{nh^{d_T}} \left(\int_{\mathcal{X}} (\widehat{\omega}_Y(t, t', x) - \bar{\omega}_Y(t, t', x)) \bar{\lambda}(t', x) f_X(x) dx \right) \times O(h^2) \\
 & = \sqrt{nh^{d_T}} \mathbb{E} [(\widehat{\omega}_{Yi\ell} - \bar{\omega}_{Yi}) (1 - f_{T|X_{i\ell}} \bar{\lambda}_i) \mid Z_{i\ell}^c] + o_p(1)
 \end{aligned}$$

where the second last equality follows from Lemma 8.1, and the last equality follows from the definition of $\bar{\lambda}_i$, $nh^{d_T+4} \rightarrow C_h$, Assumption 5, and Assumption 7 along with an application of the Cauchy-Schwartz inequality.

Now, we use the multiply robust assumptions to see that the remaining expectation term is 0 if $\lambda = \bar{\lambda}$, and otherwise if $\lambda \neq \bar{\lambda}$, then by Assumption 6, $\sqrt{nh^{d_T}} \|\widehat{\omega}_{Y\ell} - \bar{\omega}_Y\|_X = o_p(1)$. Therefore (MR-1) = $o_p(1)$.

For (MR-2), we write it as

$$\sqrt{nh^{d_T}} \mathbb{E} [(\widehat{\mu}_{Yi\ell} - \bar{\mu}_{Yi}) \{K_h(T_i - t') \bar{r}_i\}] - \sqrt{nh^{d_T}} \mathbb{E} [(\widehat{\mu}_{Yi\ell} - \bar{\mu}_{Yi}) \{K_h(T_i - t) \bar{\lambda}_i \bar{r}_i\}]$$

The left term with t' can be written as

$$\begin{aligned}
 & \sqrt{nh^{d_T}} \mathbb{E} [(\widehat{\mu}_{Yi\ell} - \bar{\mu}_{Yi}) \{K_h(T_i - t') \bar{\lambda}_i\}] \\
 & = \sqrt{nh^{d_T}} \int_{\mathcal{M} \times \mathcal{X}} (\widehat{\mu}_Y(t, m, x) - \bar{\mu}_Y(t, m, x)) \bar{\lambda}(t', x) \left\{ \int_{\mathcal{T}} K_h(s - t') f_{T|X, M}(s \mid x, m) ds \right\} f_{MX}(m, x) dm dx
 \end{aligned}$$

An application of Lemma 8.1 once again provides

$$\begin{aligned} &= \sqrt{nh^{d_T}} \int_{\mathcal{M} \times \mathcal{X}} (\widehat{\mu_Y}(t, m, x) - \overline{\mu_Y}(t, m, x)) \bar{\lambda}(t', x) f_{T|X, M}(t' | x, m) f_{MX}(m, x) dm dx \\ &+ \sqrt{nh^{d_T}} \left(\int_{\mathcal{M} \times \mathcal{X}} (\widehat{\mu_Y}(t, m, x) - \overline{\mu_Y}(t, m, x)) \bar{\lambda}(t', x) f_{MX}(m, x) dm dx \right) \times O(h^2) \end{aligned}$$

Similarly the right term of (MR-2) writes as

$$\begin{aligned} &\sqrt{nh^{d_T}} \int_{\mathcal{M} \times \mathcal{X}} (\widehat{\mu_Y}(t, m, x) - \overline{\mu_Y}(t, m, x)) \bar{\lambda}(t', x) \bar{r}(t, t', m, x) f_{T|X, M}(t | x, m) f_{MX}(m, x) dm dx \\ &+ \sqrt{nh^{d_T}} \left(\int_{\mathcal{M} \times \mathcal{X}} (\widehat{\mu_Y}(t, m, x) - \overline{\mu_Y}(t, m, x)) \bar{\lambda}(t', x) \bar{r}(t, t', m, x) f_{MX}(m, x) dm dx \right) \times O(h^2) \end{aligned}$$

Therefore, due to boundedness Assumption 7 and the convergence Assumption 5, we obtain:

$$(MR-2) = \mathbb{E} [(\widehat{\mu_Y}_{i\ell} - \overline{\mu_Y}_i) \bar{\lambda}_i \{f_{T|X, M}(t' | x_i, \psi_i) - \bar{r}_i f_{T|X, M}(t | x_i, \psi_i)\} | Z_{i\ell}^c] + o_p(1)$$

We use the multiply robust assumptions to see that the remaining expectation term is 0 when $f_{T|X, M} = \overline{f_{T|X, M}}$, and otherwise by Assumption 6, $\sqrt{nh^{d_T}} \|\widehat{\mu_Y}_\ell - \overline{\mu_Y}\|_{t_{XM}} = o_p(1)$. Therefore $(MR-2) = o_p(1)$.

For (MR-3), we note that

$$\begin{aligned} &\sqrt{nh^{d_T}} \mathbb{E} \left[K_h(T_i - t') (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) \{\overline{\mu_Y}_i - \overline{\omega_Y}_i\} | Z_{i\ell}^c \right] \\ &= \sqrt{nh^{d_T}} \int K_h(s - t') (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \{\overline{\mu_Y}(t, m, x) - \overline{\omega_Y}(t, t', x)\} f_{TMX}(s, m, x) ds dm dx \\ &= \sqrt{nh^{d_T}} \int \left\{ \int K_h(s - t') f_{T|X, M}(s | x, m) ds \right\} (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \\ &\quad \times \{\overline{\mu_Y}(t, m, x) - \overline{\omega_Y}(t, t', x)\} f_{MX}(m, x) dm dx \\ &= \sqrt{nh^{d_T}} \int (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \{\overline{\mu_Y}(t, m, x) - \overline{\omega_Y}(t, t', x)\} f_{TMX}(t', m, x) dm dx \\ &+ \sqrt{nh^{d_T}} \left(\int (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \{\overline{\mu_Y}(t, m, x) - \overline{\omega_Y}(t, t', x)\} f_{MX}(m, x) dm dx \right) O(h^2). \end{aligned}$$

by application of Lemma 8.1. The term

$$\sqrt{nh^{d_T}} \left(\int (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \{\overline{\mu_Y}(t, m, x) - \overline{\omega_Y}(t, t', x)\} f_{MX}(m, x) dm dx \right) O(h^2) = o_p(1)$$

by an application of Cauchy-Schwartz and the use of Assumption 5. Next note that:

$$\overline{\mu_Y} - \overline{\omega_Y} = \overline{\mu_Y} - \mu_Y - (\overline{\omega_Y} - \omega_Y) + \mu_Y - \omega_Y \quad (43)$$

Therefore with this decomposition and by definition of ω_Y ,

$$\begin{aligned} &\sqrt{nh^{d_T}} \int (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \{\overline{\mu_Y}(t, m, x) - \overline{\omega_Y}(t, t', x)\} f_{TMX}(t', m, x) dm dx \\ &= \sqrt{nh^{d_T}} \int (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \{\overline{\mu_Y}(t, m, x) - \mu_Y(t, m, x)\} f_{TMX}(t', m, x) dm dx \\ &+ \sqrt{nh^{d_T}} \int (\hat{\lambda}(t', x) - \bar{\lambda}(t', x)) \{\omega_Y(t, t', x) - \overline{\omega_Y}(t, t', x)\} f_{TX}(t', x) dx \end{aligned}$$

Therefore,

$$\begin{aligned} (\text{MR-3}) &= \mathbb{E} \left[\left(\hat{\lambda}_{i\ell} - \bar{\lambda}_i \right) \{ \overline{\mu_{Yi}} - \mu_{Yi\ell} \} \mid Z_{i\ell}^c \right] \\ &\quad + \mathbb{E} \left[\left(\hat{\lambda}_{i\ell} - \bar{\lambda}_i \right) \{ \omega_{Yi\ell} - \overline{\omega_{Yi}} \} \mid Z_{i\ell}^c \right] + o_p(1) \end{aligned}$$

Using the multiply robust assumptions, the first remaining expectation term is 0 when $\mu_Y = \overline{\mu_Y}$, and otherwise by Assumption 6, $\sqrt{nh^{d_T}} \left\| \hat{\lambda}_\ell - \bar{\lambda} \right\|_{tX} = o_p(1)$. The second remaining expectation is 0 when $\omega_Y = \overline{\omega_Y}$, and otherwise by Assumption 6, $\sqrt{nh^{d_T}} \left\| \hat{\lambda}_\ell - \bar{\lambda} \right\|_{tX} = o_p(1)$. Therefore $(\text{MR-3}) = o_p(1)$.

With similar expansions on the kernel smoothing, we can show that:

$$(\text{MR-4}) = \mathbb{E} \left[(\hat{r}_{i\ell} - \bar{r}_i) f_{T|X, M_{i\ell}} \bar{\lambda}_i (\mu_{Yi\ell} - \overline{\mu_{Yi}}) \mid Z_{i\ell}^c \right] + o_p(1)$$

the remaining expectation term is 0 when $\mu_Y = \overline{\mu_Y}$, and otherwise by Assumption 6, $\sqrt{nh^{d_T}} \left\| \widehat{f_{T|X, M_\ell}} - \widehat{f_{T|X, M}} \right\|_{tXM} = o_p(1)$. Thus $(\text{MR-4}) = o_p(1)$.

and

$$(\text{MR-5}) = \mathbb{E} \left[\left(\hat{\lambda}_{i\ell} - \bar{\lambda}_i \right) f_{T|X, M_{i\ell}} \bar{r}_i (\mu_{Yi\ell} - \overline{\mu_{Yi}}) \mid Z_{i\ell}^c \right] + o_p(1)$$

the remaining expectation term is 0 when $\mu_Y = \overline{\mu_Y}$, and otherwise by Assumption 6, $\sqrt{nh^{d_T}} \left\| \hat{\lambda}_\ell - \bar{\lambda} \right\|_{tXM} = o_p(1)$. Thus $(\text{MR-5}) = o_p(1)$.

Assembling all terms By the triangle inequality, we obtain the asymptotically linear representation

$$\sqrt{nh^{d_T}} n^{-1} \sum_{i=1}^n \left(\hat{\psi} \left(Z_i, \eta_{t,t'}, \widehat{\mu_{Yi\ell}}, \widehat{\omega_{Yi\ell}}, \widehat{f_{T|X, i\ell}}, \widehat{f_{T|X, M_{i\ell}}} \right) - \psi \left(Z_i, \eta_{t,t'}, \overline{\mu_{Yi}}, \overline{\omega_{Yi}}, \overline{f_{T|X, i}}, \overline{f_{T|X, M_{i\ell}}} \right) \right) = o_p(1). \quad (44)$$

9.0.2 Asymptotic normality

The proof for asymptotic normality follows from an application of the Lyapunov Central Limit theorem to the terms $\sqrt{nh^{d_T}} n^{-1} \psi \left(Z_i, \overline{f_{T|X}}, \overline{f_{T|X, M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'} \right)$. We will compute the expectation and variance of:

$$\sqrt{nh^{d_T}} n^{-1} \psi \left(Z_i; f_{T|X}, f_{T|X, M}, \mu_Y, \omega_Y, \eta_{t,t'} \right)$$

Calculation for $B_{t,t'}$ and μ_i Given

$$\psi \left(Z_i; \overline{f_{T|X}}, \overline{f_{T|X, M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'} \right) = \frac{K_h(T_i - t) \overline{f_{T|X, M}}(t' \mid M_i, X_i)}{f_{T|X, M}(t \mid M_i, X_i) \overline{f_{T|X}}(t' \mid X_i)} \{Y_i - \overline{\mu_Y}(X_i, \psi_i, t)\} \quad (\text{R2-1})$$

$$+ \frac{K_h(T_i - t')}{f_{T|X}(t' \mid X_i)} \{ \overline{\mu_Y}(X_i, \psi_i, t) - \overline{\omega_{Yi}} \} \quad (\text{R2-2})$$

$$+ \overline{\omega_{Yi}} - \eta_{t,t'} \quad (\Delta-3)$$

We start by focusing on $\frac{K_h(T-t) \overline{f_{T|X, M}}(t' \mid M, X)}{f_{T|X, M}(t \mid M, X) \overline{f_{T|X}}(t' \mid X)} \{Y - \overline{\mu_Y}(t, M, X)\} + \frac{K_h(T-t')}{f_{T|X}(t' \mid X)} \{ \overline{\mu_Y}(X, M, t) - \overline{\omega_Y}(t, t', X) \}$. We start by computing the expectation of each the individual terms one at a time.

Expanding the outcome residual (R2-1)

$$\begin{aligned} & \mathbb{E} [K_h(T-t)\bar{\lambda}(t', X)\bar{r}(t, t', M, X)\overline{f_{T|X,M}}(t' | M, X) \{Y - \overline{\mu_Y}(t, M, X)\}] \\ &= \mathbb{E} \{ \bar{\lambda}(t', X)\bar{r}(t, t', M, X) \mathbb{E} [K_h(T-t)(\mu_Y(T, M, X) - \overline{\mu_Y}(t, M, X)) | X, M] \} \end{aligned}$$

The inner product further expands as follows,

$$\begin{aligned} & \mathbb{E} [K_h(T-t)(\mu_Y(T, M, X) - \overline{\mu_Y}(t, M, X)) | X, M] \\ &= \int_{\mathcal{T}} K_h(s-t)(\mu_Y(s, M, X) - \overline{\mu_Y}(t, M, X))f_{T|X,M}(s | X, M)ds \\ &= \int_{\mathcal{R}^{d_T}} k(u)(\mu_Y(t+uh, M, X) - \overline{\mu_Y}(t, M, X))f_{T|X,M}(t+uh | X, M)du \\ &= \int_{\mathcal{R}^{d_T}} \left(\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X) + \sum_{j=1}^{d_T} hu_j \partial_{t_j} \mu_Y(t, M, X) + \frac{h^2}{2} u_j^2 \partial_{t_j}^2 \mu_Y(t, M, X) \right. \\ & \quad \left. + \frac{h^2}{2} \sum_{l=1, l \neq j}^{d_T} u_j u_l \partial_{t_j} \partial_{t_l} \mu_Y(t, M, X) \right) \left(f_{T|X,M}(t | X, M) + \sum_{j=1}^{d_T} hu_j \partial_{t_j} f_{T|X,M}(t | X, M) + \frac{h^2}{2} u_j^2 \partial_{t_j}^2 f_{T|X,M}(t | X, M) \right. \\ & \quad \left. + \frac{h^2}{2} \sum_{l=1, l \neq j}^{d_T} u_j u_l \partial_{t_j} \partial_{t_l} f_{T|X,M}(t | X, M) \right) k(u_1) \cdots k(u_{d_T}) du_1 \cdots du_{d_T} + O(h^3) \\ &= (\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X))f_{T|X,M}(t | X, M) + h^2 \sum_{j=1}^{d_T} (\partial_{t_j} \mu_Y(t, M, X) \partial_{t_j} f_{T|X,M}(t | X, M) \\ & \quad + \frac{1}{2} \partial_{t_j}^2 \mu_Y(t, M, X) f_{T|X,M}(t | X, M) + (\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)) \frac{1}{2} \partial_{t_j}^2 \mu_Y(t, M, X)) \int_{-\infty}^{\infty} u^2 k(u) du + O(h^3) \end{aligned}$$

for all X, M in respective range. Inserting this back into the original expectation we get,

$$\begin{aligned} & \mathbb{E} [\bar{\lambda}(t', X)\bar{r}(t, t', M, X) \mathbb{E} [K_h(T-t)(\mu_Y(t, M, X) - \overline{\mu_Y}(X, M, t)) | X]] \\ &= \mathbb{E} [(\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X))\bar{\lambda}(t', X)\bar{r}(t, t', M, X)f_{T|X,M}(t | X, M)] \\ & \quad + h^2 \sum_{j=1}^{d_T} \mathbb{E} [\partial_{t_j} \mu_Y(t, M, X) \partial_{t_j} f_{T|X,M}(t | X, M) \bar{\lambda}(t', X)\bar{r}(t, t', M, X) \\ & \quad + \partial_{t_j}^2 \mu_Y(t, M, X) \frac{\bar{\lambda}(t', X)\bar{r}(t, t', M, X)f_{T|X,M}(t | X, M)}{2} \\ & \quad + (\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)) \partial_{t_j}^2 \mu_Y(t, M, X) \frac{\bar{\lambda}(t', X)\bar{r}(t, t', M, X)}{2}] \int_{-\infty}^{\infty} u^2 k(u) du + O(h^3) \\ &= (\Delta-1) + h^2 B_{t,t'}^1 + O(h^3) \end{aligned} \tag{45}$$

where

$$\Delta-1 = \mathbb{E} [(\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X))\bar{\lambda}(t', X)\bar{r}(t, t', M, X)f_{T|X,M}(t | X, M)] \tag{\Delta-1}$$

$$\begin{aligned}
 B_{t,t'}^1 &= \sum_{j=1}^{d_T} \mathbb{E} \left[\partial_{t_j} \mu_Y(t, M, X) \partial_{t_j} f_{T|X,M}(t | X, M) \bar{\lambda}(t', X) \bar{r}(t, t', M, X) \right. \\
 &\quad + \partial_{t_j}^2 \mu_Y(t, M, X) \frac{\bar{\lambda}(t', X) \bar{r}(t, t', M, X) f_{T|X,M}(t | X, M)}{2} \\
 &\quad \left. + (\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)) \partial_{t_j}^2 \mu_Y(t, M, X) \frac{\bar{\lambda}(t', X) \bar{r}(t, t', M, X)}{2} \right] \int_{-\infty}^{\infty} u^2 k(u) du
 \end{aligned}$$

Expanding the conditional mean outcome residual (R2-2)

$$\begin{aligned}
 &\mathbb{E} [K_h(T - t') \bar{\lambda}(t', X) \{\overline{\mu_Y}(t, M, X) - \overline{\mu_Y}(t, t', X)\}] \\
 &= \mathbb{E} [(\overline{\mu_Y}(t, M, X) - \overline{\mu_Y}(t, t', X)) \bar{\lambda}(t', X) \mathbb{E} [K_h(T - t') | X, M]]
 \end{aligned}$$

As in Lemma 8.1, the inner expectation can be written as

$$\mathbb{E} [K_h(T - t') | X, M] = f_{T|X,M}(t' | X, M) + \frac{1}{2} h^2 \int u^2 k(u) du \sum_{j=1}^{d_T} \partial_{t_j}^2 f_{T|X,M}(t' | X, M) + O(h^3) \quad (46)$$

Plugging this back into the above expectation

$$\begin{aligned}
 &\mathbb{E} \left[(\overline{\mu_Y}(t, M, X) - \overline{\mu_Y}(t, t', X)) \bar{\lambda}(t', X) \left(f_{T|X,M}(t' | X, M) + \frac{1}{2} h^2 \int u^2 k(u) du \sum_{j=1}^{d_T} \partial_{t_j}^2 f_{T|X,M}(t' | X, M) \right) \right] \\
 &\quad + O(h^3) \\
 &= (\Delta-2) + h^2 B_{t,t'}^2 + O(h^3).
 \end{aligned}$$

where

$$B_{t,t'}^2 = h^2 \left[\int_{\mathcal{R}^{d_T}} u^2 k(u) du \right] \mathbb{E} \left[\{\overline{\mu_Y}(t, M, X) - \overline{\mu_Y}(t, t', X)\} \frac{1}{2} \sum_{j=1}^{d_T} \partial_{t_j}^2 f_{T|X,M}(t' | X, M) \bar{\lambda}(t', X) \right] \quad (47)$$

and

$$\Delta-2 = \mathbb{E} [(\overline{\mu_Y}(t, M, X) - \overline{\mu_Y}(t, t', X)) \bar{\lambda}(t', X) f_{T|X,M}(t' | X, M)] \quad (\Delta-2)$$

.

We use the same reasoning as (43) to write:

$$(\Delta-2) = \mathbb{E} \left[\{\overline{\mu_Y}(t, M, X) - \mu_Y(t, M, X)\} \frac{f_{T|X,M}(t' | X, M)}{f_{T|X}(t' | X)} \right] + \mathbb{E} \left[\{\omega_Y(t, t', X) - \overline{\omega_Y}(t, t', X)\} \frac{f_{T|X}(t' | X)}{f_{T|X}(t' | X)} \right]$$

Dealing with $(\Delta-1)$, $(\Delta-2)$ and $(\Delta-3)$ with the MR property Let

$$\Delta = (\Delta-1) + (\Delta-2) + (\Delta-3) \quad (\Delta)$$

Expanding the term (Δ) ,

$$\begin{aligned}
 \Delta = \mathbb{E} & \left[(\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)) \frac{\overline{f_{T|X,M}}(t' | X, M) f_{T|X,M}(t | X, M)}{\overline{f_{T|X,M}}(t | X, M) \overline{f_{T|X}}(t' | X)} \right. \\
 & \left. \{ \overline{\mu_Y}(t, M, X) - \mu_Y(t, M, X) \} \frac{f_{T|X,M}(t' | X, M)}{\overline{f_{T|X}}(t' | X)} \right. \\
 & \left. + \{ \omega_Y(t, t', X) - \overline{\omega_Y}(t, t', X) \} \frac{f_{T|X}(t' | X)}{\overline{f_{T|X}}(t' | X)} \right. \\
 & \left. + \overline{\omega_Y}(t, t', X) - \eta_{t,t'} \right]
 \end{aligned}$$

Due to the multiply robust property, (Δ) is zero as stipulated in Assumption 5.2 when one of the conditions is met:

- $\overline{\mu_Y} = \mu_Y$ and $\overline{\omega_Y} = \omega_Y$
- $\overline{\mu_Y} = \mu_Y$ and $\overline{f_{T|X}} = f_{T|X}$
- $\overline{\omega_Y} = \omega_Y$ and $\overline{f_{T|X,M}} = f_{T|X,M}$
- $\overline{f_{T|X}} = f_{T|X}$ and $\overline{f_{T|X,M}} = f_{T|X,M}$.

Conclusion Hence, letting

$$B_{t,t'} = B_{t,t'}^1 + B_{t,t'}^2 \quad (48)$$

we have $\mathbb{E} [\psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'})] = h^2 B_{t,t'} + O(h^3)$.

Next, we prove the properties of variance.

Calculation for $V_{t,t'}$ and s_n^2 From the definition of s_n^2 , we have

$$\begin{aligned}
 s_n^2 &= \sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n \text{var} \left(\sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right) \\
 &= h^{d_T} \text{var} \left(\psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right)
 \end{aligned}$$

To compute the variance, we need to compute the squared expectation and the second moment. The squared expectation can be obtained using the previous calculation of the bias:

$$\begin{aligned}
 & h^{d_T} \mathbb{E} [\psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'})]^2 \\
 &= h^{d_T} \mathbb{E} \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right. \\
 & \quad \left. + \frac{K_h(T-t')}{\overline{f_{T|X}}(t' | X)} \{ \overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X) \} + \overline{\omega_Y}(t, t', X) \right]^2 \\
 &= O(h^{d_T+4})
 \end{aligned}$$

Let us then consider computations of the second moment of the following quantity:

$$\begin{aligned}
 & h^{d_T} \mathbb{E} \left\{ \psi \left(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'} \right)^2 \right\} \\
 &= h^{d_T} \mathbb{E} \left\{ \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right. \right. \\
 & \quad \left. \left. + \frac{K_h(T-t')}{\overline{f_{T|X}}(t' | X)} \{\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)\} + \overline{\omega_Y}(t, t', X) \right]^2 \right\}
 \end{aligned}$$

Examining each of the terms above one by one, the first term can be expanded as

$$= h^{d_T} \mathbb{E} \left\{ \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right]^2 \right\} \quad (\text{V1-1})$$

$$+ h^{d_T} \mathbb{E} \left\{ \left[\frac{K_h(T-t')}{\overline{f_{T|X}}(t' | X)} \{\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)\} \right]^2 \right\} \quad (\text{V1-2})$$

$$+ h^{d_T} \mathbb{E} \{\overline{\omega_Y}(t, t', X)\} \quad (\text{V1-3})$$

$$\begin{aligned}
 & + 2h^{d_T} \mathbb{E} \left\{ \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right] \right. \\
 & \quad \left. \times \left[\frac{K_h(T-t')}{\overline{f_{T|X}}(t' | X)} \{\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)\} \right] \right\} \quad (\text{V1-4})
 \end{aligned}$$

$$+ 2h^{d_T} \mathbb{E} \left\{ \overline{\omega_Y}(t, t', X) \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right] \right\} \quad (\text{V1-5})$$

$$+ 2h^{d_T} \mathbb{E} \left\{ \overline{\omega_Y}(t, t', X) \left[\frac{K_h(T-t')}{\overline{f_{T|X}}(t' | X)} \{\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)\} \right] \right\} \quad (\text{V1-6})$$

We analyze each of these terms part by part

Computing (V1-1)

$$\begin{aligned}
 & h^{d_T} \mathbb{E} \left\{ \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right]^2 \right\} \\
 &= h^{d_T} \mathbb{E} \left\{ \mathbb{E} \left\{ \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right]^2 \middle| X, M \right\} \right\} \\
 &= h^{d_T} \mathbb{E} \left\{ \frac{\overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \mathbb{E} \{K_h(T-t)^2 (Y - \overline{\mu_Y}(t, M, X))^2 | X, M\} \right\} \\
 &= h^{d_T} \mathbb{E} \left\{ \frac{\overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \times \mathbb{E} \{K_h(T-t)^2 \mathbb{E} \{(Y - \overline{\mu_Y}(t, M, X))^2 | X, M, T = t\} | X, M\} \right\}
 \end{aligned}$$

With a computation similar to (KD) the inner expectation can be written as

$$\begin{aligned}
 & h^{d_T} \mathbb{E} \{K_h(T-t)^2 \mathbb{E} \{(Y - \overline{\mu_Y}(t, M, X))^2 | X, M, T\} | X, M\} \\
 &= \left[\int_{\mathcal{R}^{d_T}} \Pi_{j=1}^{d_T} k(u_j)^2 du \right] \times \mathbb{E} \{(Y - \overline{\mu_Y}(t, M, X))^2 | X, M, T\} \overline{f_{T|X,M}}(t | X, M) + O(h^2)
 \end{aligned}$$

Hence, the part (V1-1) of the variance boils down to

$$\left[\int k(u)^2 du \right]^{d_T} \mathbb{E} \left\{ \frac{\overline{f_{T|X,M}}(t' | M, X)^2 f_{T|X,M}(t | X, M)}{\overline{f_{T|X,M}}(t | M, X)^2 \overline{f_{T|X}}(t' | X)^2} \mathbb{E} \{ (Y - \overline{\mu_Y}(t, M, X))^2 | X, M, T = t \} \right\} + O(h^2).$$

Computing (V1-2) Similarly, we can obtain:

$$\begin{aligned} & h^{d_T} \mathbb{E} \left\{ \left[\frac{K_h(T - t')}{\overline{f_{T|X}}(t' | X)^2} \{ \overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X) \} \right]^2 \right\} \\ &= h^{d_T} \mathbb{E} \left\{ \frac{1}{\overline{f_{T|X}}(t' | X)^2} \mathbb{E} \left[K_h(T - t')^2 (\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^2 | X \right] \right\} \\ &= h^{d_T} \mathbb{E} \left\{ \frac{1}{\overline{f_{T|X}}(t' | X)^2} \mathbb{E} \left[K_h(T - t')^2 \mathbb{E} \left[(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^2 | X, M, T = t' \right] | X \right] \right\} \end{aligned}$$

Again, similarly to (KD) with a Taylor expansion on the density, we obtain for the inner expectation

$$\begin{aligned} & h^{d_T} \mathbb{E} \left[K_h(T - t')^2 \mathbb{E} \left[(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^2 | X, M, T = t' \right] | X \right] \\ &= \left[\int k(u)^2 du \right]^{d_T} \overline{f_{T|X}}(t' | X) \mathbb{E} \left[(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^2 | X, M, T = t' \right] + O(h^2) \end{aligned}$$

Hence, the part (V1-2) of the variance accounts for

$$\left[\int k(u)^2 du \right]^{d_T} \times \mathbb{E} \left\{ \frac{\overline{f_{T|X}}(t' | X)}{\overline{f_{T|X}}(t' | X)^2} \mathbb{E} \left[(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^2 | X, M, T = t' \right] \right\} + O(h^2)$$

Computing (V1-3)

$$h^{d_T} \mathbb{E} [\overline{\omega_Y}^2(t, t', X)] = O(h^{d_T})$$

This holds because we assume $\overline{\omega_Y}$ is bounded.

Computing (V1-4)

$$\begin{aligned} & h^{d_T} \mathbb{E} \left\{ \left[\frac{K_h(T - t) \overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X) \overline{f_{T|X}}(t' | X)} \{ Y - \overline{\mu_Y}(t, M, X) \} \right] \right. \\ & \quad \times \left. \left[\frac{K_h(T - t')}{\overline{f_{T|X}}(t' | X)} \{ \overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X) \} \right] \right\} \\ &= h^{d_T} \mathbb{E} \left\{ \frac{K_h(T - t) K_h(T - t')}{\overline{f_{T|X}}(t' | X)^2} \frac{\overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X)} [Y - \overline{\mu_Y}(t, M, X)] [\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)] \right\} \\ &= h^{d_T} \mathbb{E} \left\{ \frac{1}{\overline{f_{T|X}}(t' | X)^2} \frac{\overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X)} [\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)] \right. \\ & \quad \times \mathbb{E} \{ K_h(T - t) K_h(T - t') [Y - \overline{\mu_Y}(t, M, X)] | X, M \} \} \\ &= h^{d_T} \mathbb{E} \left\{ \frac{1}{\overline{f_{T|X}}(t' | X)^2} \frac{\overline{f_{T|X,M}}(t' | M, X)}{\overline{f_{T|X,M}}(t | M, X)} [\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)] \right. \\ & \quad \times \mathbb{E} \{ K_h(T - t) K_h(T - t') [Y - \overline{\mu_Y}(t, M, X)] | X, M \} \} \end{aligned}$$

The inner expectation

$$\begin{aligned}
 & h^{d_T} \mathbb{E} \{ K_h(T-t) K_h(T-t') [\mu_Y(T, M, X) - \overline{\mu_Y}(t, M, X)] \mid X, M \} \\
 &= h^{d_T} \int \left[\prod_{j=1}^{d_T} \frac{1}{h^2} k\left(\frac{s-t}{h}\right) k\left(\frac{s-t'}{h}\right) \right] [\mu_Y(s, M, X) - \overline{\mu_Y}(t, M, X)] f_{T|X,M}(s \mid X, M) ds \\
 &= \int k(u_1) \cdots k(u_{d_T}) k\left(u_1 + \frac{t-t'}{h}\right) \cdots k\left(u_{d_T} + \frac{t-t'}{h}\right) [\mu_Y(uh+t, M, X) - \overline{\mu_Y}(t, M, X)] f_{T|X,M}(uh+t \mid X, M) du \\
 &= \int k(u_1) \cdots k(u_{d_T}) k\left(u_1 + \frac{t-t'}{h}\right) \cdots k\left(u_{d_T} + \frac{t-t'}{h}\right) \\
 &\quad \times \left[[\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)] + \sum_{j=1}^{d_T} u_j h \partial_{t_j} \mu_Y(t, M, X) + \frac{u_j^2 h^2}{2} \partial_{t_j}^2 \mu_Y(t, M, X) + \frac{u_j^3 h^3}{6} \partial_{t_j}^3 \mu_Y(t, M, X) \right] \\
 &\quad \times \left[f_{T|X,M}(t \mid X, M) + \sum_{j=1}^{d_T} u_j h \partial_{t_j} f_{T|X,M}(t \mid X, M) + \frac{u_j^2 h^2}{2} \partial_{t_j}^2 f_{T|X,M}(t \mid X, M) \right] du_1 \cdots du_{d_T} \\
 &= [\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)] f_{T|X,M}(t \mid X, M) \int k(u_1) \cdots k(u_{d_T}) k\left(u_1 + \frac{t-t'}{h}\right) \cdots k\left(u_{d_T} + \frac{t-t'}{h}\right) + O(h) \\
 &= O(h)
 \end{aligned}$$

Computing (V1-5)

$$\begin{aligned}
 & 2h^{d_T} \mathbb{E} \left\{ \overline{\omega_Y}(t, t', X) \left[\frac{K_h(T-t) \overline{f_{T|X,M}}(t' \mid M, X)}{\overline{f_{T|X,M}}(t \mid M, X) \overline{f_{T|X}}(t' \mid X)} \{Y - \overline{\mu_Y}(t, M, X)\} \right] \right\} \\
 &= 2h^{d_T} \mathbb{E} \left\{ \frac{\overline{\omega_Y}(t, t', X) \overline{f_{T|X,M}}(t' \mid M, X)}{\overline{f_{T|X,M}}(t \mid M, X) \overline{f_{T|X}}(t' \mid X)} \mathbb{E}[K_h(T-t) \{Y - \overline{\mu_Y}(t, M, X) \mid X, M\}] \right\}
 \end{aligned}$$

Applying the same expansion as in (R2-1), we can write the inner expectation as

$$= (\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)) f_{T|X,M}(t \mid X, M) + O(h^2)$$

Inserting this back into the full expectation, combined with the boundedness of $\mu_Y, \omega_Y, f_{T|X} f_{T|X,M}$ and their limits, we get

$$(V1-5) = O(h^{d_T})$$

Computing (V1-6)

$$\begin{aligned}
 & 2h^{d_T} \mathbb{E} \left\{ \overline{\omega_Y}(t, t', X) \left[\frac{K_h(T-t')}{\overline{f_{T|X}}(t' \mid X)} \{\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)\} \right] \right\} \\
 &= 2h^{d_T} \mathbb{E} \left\{ \frac{\overline{\omega_Y}(t, t', X) (\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))}{\overline{f_{T|X}}(t' \mid X)} \mathbb{E}[K_h(T-t') \mid X, M] \right\}
 \end{aligned}$$

Using (46) on $\mathbb{E}[K_h(T-t') \mid X, M]$ and plugging this back into the full expectation, we get

$$(V1-6) = O(h^{d_T})$$

Computing the total variance Finally, with all terms of the variance together, we have shown that

$$h^{d_T} \times \text{var}(\psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'})) = V_{t,t'} + O(h)$$

where the term converges to $V_{t,t'}$ as $h \rightarrow 0$ and

$$\begin{aligned} V_{t,t'} = & R_k^{d_T} \\ & \times \mathbb{E} \left\{ \frac{\overline{f_{T|X,M}}(t' | M, X)^2 \overline{f_{T|X,M}}(t | X, M)}{\overline{f_{T|X,M}}(t | M, X)^2 \overline{f_{T|X}}(t' | X)^2} \mathbb{E} \{ (Y - \overline{\mu_Y}(t, M, X))^2 | X, M, T = t \} \right. \\ & \left. + \frac{\overline{f_{T|X}}(t' | X)}{\overline{f_{T|X}}(t' | X)^2} \mathbb{E} \left[(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^2 | X, M, T = t' \right] \right\} \end{aligned}$$

Asymptotic normality follows from the Lyapunov central limit theorem with the third absolute moment. Specifically, having derived the bias and variance terms, we will now prove the Lyapunov condition for $\delta = 1$, i.e.

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^3} \sum_{i=1}^n \mathbb{E} \left[\left| \sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) - \mu_i \right|^3 \right] = 0$$

Where μ_i equals $\mathbb{E} \left[\sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \mu_Y, \omega_Y, \eta_{t,t'}) \right]$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$ where σ_i^2 is the variance of $\sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \mu_Y, \omega_Y, \eta_{t,t'})$.

Noticing that we can expand the previous term

$$\begin{aligned} \left| \sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) - \mu_i \right|^3 & \leq (h^{d_T} n^{-1})^{3/2} \left| \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right|^3 + |\mu_i|^3 \\ & \quad + 3(h^{d_T} n^{-1}) \left| \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right|^2 |\mu_i| \\ & \quad + 3(h^{d_T} n^{-1})^{1/2} \left| \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right| |\mu_i|^2 \end{aligned}$$

Then it suffices to note that

$$\begin{aligned} & \mathbb{E} \left[\left| \sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right|^3 \right] \\ & = O \left((n^{-1} h^{d_T})^{3/2} \mathbb{E} [K_h(T - t)^3 |Y - \overline{\mu_Y}(t, M, X)|^3 \bar{r}(t, t', M, X)^3 \bar{\lambda}(t', X)^3] \right) \\ & \quad + O \left((n^{-1} h^{d_T})^{3/2} \mathbb{E} [K_h(T - t')^3 |\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)|^3 \bar{\lambda}(t', X)^3] \right) \\ & = O \left((n^3 h^{d_T})^{-1/2} \right) \end{aligned}$$

by the same arguments as (KD) under the condition that $\mathbb{E} [|Y - \overline{\mu_Y}(t, M, X)|^3 | T = t, M, X]$, $\mathbb{E} [|\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X)|^3 | T = t', M, X]$ and their first derivative w.r.t. t are bounded uniformly in $x \in \mathcal{X}$.

Hence

$$\sum_{i=1}^n \mathbb{E} \left[\left| \sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right|^3 \right] = O \left((nh^{d_T})^{-1/2} \right) = o(1)$$

Moreover, recall that

$$s_n^2 \equiv \sum_{i=1}^n \text{var} \left(\sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right) = h^{d_T} \text{var}(\psi) = V_{t,t'} + o(1). \quad (49)$$

Thus the Lyapunov condition holds:

$$\sum_{i=1}^n \mathbb{E} \left[\left| \sqrt{nh^{d_T}} n^{-1} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) \right|^3 \right] / s_n^3 = O \left((nh^{d_T})^{-1/2} \right) = o(1) \quad (50)$$

and hence,

$$\frac{1}{s_n} \sum_{i=1}^n \left(\sqrt{\frac{h^{d_T}}{n}} \psi(Z_i; \overline{f_{T|X}}, \overline{f_{T|X,M}}, \overline{\mu_Y}, \overline{\omega_Y}, \eta_{t,t'}) - \mu_i \right) \xrightarrow{d} \mathcal{N}(0, 1)$$

To finish, an application of Slutsky's theorem provides the desired result that

$$\sqrt{nh^{d_T}} (\hat{\eta}_{t,t'} - \eta_{t,t'} - h^2 \mathbf{B}_{t,t'}) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}_{t,t'})$$

□

9.1 Optimal bandwidth

In this part we will prove the corollaries of the Theorem 8. We will start by stating Assumption 10 with the general dimension d_T .

Assumption 14. For each $\ell = 1, \dots, L$ and for any $t \in \mathcal{T}$,

1. $\|\delta_{\mu_Y}^\ell \delta_{T|X}^\ell\|_{tX} = o_p(1)$, $\|\delta_{\mu_Y}^\ell \delta_{\omega_Y}^\ell\|_{tX} = o_p(1)$, $\|\delta_{\omega_Y}^\ell \delta_{T|X,M}^\ell\|_{tMX} = o_p(1)$, $\|\delta_{T|X}^{\ell 2} \delta_{T|X,M}^\ell\|_{tMX} = o_p(1)$,
2. $\|\delta_{\mu_Y}^{\ell 2} \delta_{T|X}^{\ell 2}\|_{tX} = O_p(1)$, $\|\delta_{\mu_Y}^{\ell 2} \delta_{\omega_Y}^{\ell 2}\|_{tX} = O_p(1)$, $\|\delta_{\omega_Y}^{\ell 2} \delta_{T|X,M}^{\ell 2}\|_{tMX} = O_p(1)$, $\|\delta_{T|X}^{\ell 2} \delta_{T|X,M}^{\ell 2}\|_{tMX} = O_p(1)$,
3. $\|\delta_{\mu_Y}^{\ell 2}\|_{tX} = O_p(1)$, $\|\delta_{\omega_Y}^{\ell 2}\|_{tX} = O_p(1)$, $\|\delta_{T|X,M}^{\ell 2}\|_{tMX} = O_p(1)$, $\|\delta_{T|X}^{\ell 2}\|_{tMX} = O_p(1)$
4. $\mathbb{E}[(Y - \overline{\mu_Y}(t, M, X))^4 | T = t, M, X]$, $\mathbb{E}[(\overline{\mu_Y}(t, M, X) - \overline{\omega_Y}(t, t', X))^4 | T = t', M, X]$ and their derivatives with respect to t are bounded uniformly over $(t', x) \in \mathcal{T} \times \mathcal{X}$ and $\int_{-\infty}^{\infty} k(u)^4 du < \infty$
5. The bandwidth $b_n \rightarrow 0$ and $nb_n^{d_T+4} \rightarrow \infty$. $\int_{-\infty}^{\infty} k(u)k(u/\epsilon)du < \infty$ for $\epsilon \in (0, 1)$.

We now restate Corollary 11 with the general treatment dimension as well.

Corollary 11. Let the conditions in Theorem 8 hold. For $t \in \mathcal{T}$, if $\mathbf{B}_{t,t'}$ is non-zero, then the bandwidth that minimizes the asymptotic mean squared error is $h_t^* = (d_T \mathbf{V}_t / (4 \mathbf{B}_t^2))^{1/(d_T+4)} n^{-1/(d_T+4)}$. Further let Assumption 10 hold. Then $\hat{\mathbf{V}}_t - \mathbf{V}_t = o_p(1)$, $\hat{\mathbf{B}}_t - \mathbf{B}_t = o_p(1)$, and $\hat{h}_t/h_t^* - 1 = o_p(1)$.

Proof. By Theorem 8, the asymptotic MSE is $h^4 \mathbf{B}_{t,t'}^2 + \mathbf{V}_{t,t'} / (nh^{d_T})$. Solving the first-order condition yields the optimal bandwidth h_t^* .

Note that once the consistency of $\hat{\mathbf{V}}_{t,t'}$ and $\hat{\mathbf{B}}_{t,t'}$ is proven, the continuous mapping theorem implies $\hat{h}_t/h_t^* - 1 = o_p(1)$. Therefore, we show below the consistency of $\hat{\mathbf{V}}_{t,t'}$ and $\hat{\mathbf{B}}_{t,t'}$.

Consistency of $\hat{\mathbf{V}}_{t,t'}$: Let $\hat{\mathbf{V}}_{t,t'} = L^{-1} \sum_{\ell=1}^L \hat{\mathbf{V}}_{t,t',\ell}$, where $\hat{\mathbf{V}}_{t,t',\ell} \equiv h^{d_T} n_\ell^{-1} \sum_{i \in I_\ell} \hat{\psi}_{i\ell}^2$. It suffices to show that $\hat{\mathbf{V}}_{t,t',\ell}$ is consistent for $\mathbf{V}_{t,t'}$ as $n_\ell \rightarrow \infty$, for $\ell = 1, \dots, L$. Toward that end, we show that:

1. $h^{d_T} n_\ell^{-1} \sum_{i \in I_\ell} \psi_i^2 - \mathbf{V}_{t,t'} = o_p(1)$, where $\psi_i \equiv K_h(T_i - t) \bar{r}_i \bar{\lambda}_i (Y_i - \overline{\mu_Y}_i) + K_h(T_i - t') \bar{\lambda}_i (\overline{\mu_Y}_i - \overline{\omega_Y}_i) + \overline{\omega_Y}_i - \eta_{t,t'}$,
2. $h^{d_T} n_\ell^{-1} \sum_{i \in I_\ell} \mathbb{E}[\hat{\psi}_{i\ell}^2 - \psi_i^2 | Z^\ell] = o_p(1)$, where $\hat{\psi}_{i\ell} \equiv K_h(T_i - t) \hat{r}_{i\ell} \hat{\lambda}_{i\ell} \cdot [Y_i - \widehat{\mu_{Y|t}}_{i\ell}] + K_h(T_i - t') \hat{\lambda}_{i\ell} \cdot [\widehat{\mu_{Y|t}}_{i\ell} - \widehat{\omega_{Y|t}}_{i\ell}] + \widehat{\omega_{Y|t}}_{i\ell} - \hat{\psi}_i$
3. $h^{d_T} n_\ell^{-1} \sum_{i \in I_\ell} \Delta_{i\ell} = o_p(1)$, where $\Delta_{i\ell} \equiv \hat{\psi}_{i\ell}^2 - \psi_i^2 - \mathbb{E}[\hat{\psi}_{i\ell}^2 - \psi_i^2 | Z^\ell]$

Showing $h^{dT} n_\ell^{-1} \sum_{i \in I_\ell} \psi_i^2 - V_{t,t'} = o_p(1)$. Let $\ell \in \{1, \dots, L\}$, and let $i \in I_\ell$. As computed in the proof of Theorem 8, $h^{dT} \mathbb{E} [\psi_i^2] = V_{t,t'} + o(1)$. We will now compute $h^{dT} \mathbb{E} [\psi_i^4]$. Recall the decomposition of ψ_i into m_i, ϕ_i, φ_i in Eq. (37), (39) and (38),

$$\begin{aligned}\phi_i &= K_h(T_i - t) \bar{r}_i \bar{\lambda}_i (Y_i - \bar{\mu}_{Yi}) \\ \varphi_i &= K_h(T_i - t') \bar{\lambda}_i (\bar{\mu}_{Yi} - \bar{\omega}_{Yi}) \\ m_i &= \bar{\omega}_{Yi} - \eta_{t,t'}\end{aligned}$$

Therefore, noticing that $\psi_i = \phi_i + \varphi_i + m_i$, we will consider different quantities that are in the expansion $\mathbb{E} [(\phi_i + \varphi_i + m_i)^4]$. By similar arguments as in (KD) and Assumption 10.3,

$$\begin{aligned}\mathbb{E} [\phi_i^4] &= h^{-3dT} \mathbb{E} [\mathbb{E} [(Y - \bar{\mu}_Y(t, X, M))^4 \mid T = t, X, M] f_{T|X,M}(t \mid M, X) \bar{\lambda}(t', X)^4 \bar{r}(M, X)^4] \left(\int_{-\infty}^{\infty} k(u)^4 du \right)^{d_T} \\ &\quad + o(h^{-3dT}) \\ &= O(h^{-3dT})\end{aligned}$$

and

$$\begin{aligned}\mathbb{E} [\varphi_i^4] &= h^{-3dT} \mathbb{E} [\mathbb{E} [(\bar{\mu}_Y(t, X, M) - \bar{\omega}_Y(t, t', X))^4 \mid T = t', X, M] f_{T|X,M}(t' \mid M, X) \bar{\lambda}(t', X)^4] \left(\int_{-\infty}^{\infty} k(u)^4 du \right)^{d_T} \\ &\quad + o(h^{-3dT}) \\ &= O(h^{-3dT})\end{aligned}$$

Moreover, let us consider integers p, q such that $p + q = 4$. $\mathbb{E} (\phi_i^p \varphi_i^q)$ can be written as follows

$$\begin{aligned}\mathbb{E} \left[\mathbb{E} \left(K_h(T - t)^p K_h(T - t')^q [Y - \mu_Y(t, M, X)]^p \lambda(t', X)^p r(t, t', M, X)^p \mid X, M \right) \right. \\ \left. \times (\mu_Y(t, M, X) - \omega_Y(t, t', X))^q \lambda(t', X)^q \right]\end{aligned}$$

Take the inner expectation:

$$\begin{aligned}
 &= \mathbb{E} \left\{ K_h(T-t)^p K_h(T-t')^q [[Y - \overline{\mu_Y}(t, M, X)] \lambda(t', X) r(t, t', M, X)]^p \mid X, M \right\} \\
 &= \int \left[\prod_{j=1}^{d_T} \frac{1}{h^{p+q}} k\left(\frac{s-t}{h}\right)^p k\left(\frac{s-t'}{h}\right)^q \right] [\mu_Y(s, M, X) - \overline{\mu_Y}(t, M, X)] f_{T|X, M}(s \mid X, M) ds \\
 &= \frac{1}{h^{(p+q-1)d_T}} \int \left[\prod_{j=1}^{d_T} k(u_j)^p k\left(u_j + \frac{t_j - t'_j}{h}\right)^q \right] [\mu_Y(uh + t, M, X) - \overline{\mu_Y}(t, M, X)] f_{T|X, M}(uh + t \mid X, M) du \\
 &= \frac{1}{h^{(p+q-1)d_T}} \int \left[\prod_{j=1}^{d_T} k(u_j)^p k\left(u_j + \frac{t_j - t'_j}{h}\right)^q \right] \\
 &\quad \times \left[[\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)] + \sum_{j=1}^{d_T} u_j h \partial_{t_j} \mu_Y(t, M, X) + \frac{u_j^2 h^2}{2} \partial_{t_j}^2 \mu_Y(\bar{t}, M, X) \right] \\
 &\quad \times \left[f_{T|X, M}(t \mid X, M) + \sum_{j=1}^{d_T} u_j h \partial_{t_j} f_{T|X, M}(t \mid X, M) + \frac{u_j^2 h^2}{2} \partial_{t_j}^2 f_{T|X, M}(\tilde{t} \mid X, M) \right] du_1 \cdots du_{d_T} \\
 &= \frac{1}{h^{(p+q-1)d_T}} [\mu_Y(t, M, X) - \overline{\mu_Y}(t, M, X)] f_{T|X, M}(t \mid X, M) \int \left[\prod_{j=1}^{d_T} k(u_j)^p k\left(u_j + \frac{t_j - t'_j}{h}\right)^q \right] \\
 &\quad + O\left(\frac{1}{h^{(p+q-1)d_T}}\right) \\
 &= O\left(\frac{1}{h^{(p+q-1)d_T}}\right) = O\left(\frac{1}{h^{3d_T}}\right)
 \end{aligned}$$

where \bar{t} and \tilde{t} are between t and $t + uh$.

For the calculation of $\mathbb{E}(\phi_i^p m_i^q)$, since we assume the boundedness of $\omega_Y(t, t', X) - \eta_{t, t'}$, and taking the inner expectation of $\mathbb{E}[(Y - \overline{\mu_Y}(t, X, M))^p \mid T = t, X, M] f_{T|X, M}(t \mid M, X) \bar{\lambda}(t', X)^p \bar{r}(M, X)^p (\omega_Y(t, t', X) - \eta_{t, t'})^q$ in a similar fashion as in $\mathbb{E}(\phi_i^4)$, we obtain that $\mathbb{E}(\phi_i^p m_i^q) = O\left(\frac{1}{h^{(c_1-1)d_T}}\right)$. With a reasoning similar to those two terms and with $\mathbb{E}(\varphi_i^p m_i^q) = O\left(\frac{1}{h^{(p-1)d_T}}\right)$.

Now considering integers p, q, m such that $p + q + m = 4$ similarly $\mathbb{E}(\phi_i^p \varphi_i^q m_i^m) = O\left(\frac{1}{h^{(p+q-1)d_T}}\right)$.

Combining all the terms in the expansion $\mathbb{E}[(\phi_i + \varphi_i + m_i)^4]$, we obtain $\mathbb{E}(\psi_i^4) = O(h^{-3d_T})$. Then by Markov

inequality, for any $\epsilon > 0$,

$$\begin{aligned}
 P\left(\left|h^{d_T} n^{-1} \sum_{i \in I_\ell} \psi_i^2 - V_{t,t'}\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2} \mathbb{E} \left\{ \left[h^{d_T} n^{-1} \sum_{i \in I_\ell} \psi_i^2 - V_{t,t'} \right]^2 \right\} \\
 &= \frac{1}{\epsilon^2} \mathbb{E} \left\{ \left[h^{d_T} n^{-1} \sum_{i \in I_\ell} \psi_i^2 - h^{d_T} \mathbb{E}[\psi_i^2] + o_p(1) \right]^2 \right\} \\
 &= \frac{h^{2d_T}}{n^2 \epsilon^2} \mathbb{E} \left\{ \left[\sum_{i \in I_\ell} \psi_i^2 - \mathbb{E} \left(\sum_{i \in I_\ell} \psi_i^2 \right) \right]^2 \right\} + o_p(1) \\
 &= \frac{h^{2d_T}}{n^2 \epsilon^2} \text{var} \left(\sum_{i \in I_\ell} \psi_i^2 \right) + o_p(1) \\
 &= \frac{h^{2d_T}}{n \epsilon^2} \text{var} (\psi_i^2) + o_p(1) \\
 &= O \left(\frac{1}{n h^{d_T}} \right) = o_p(1)
 \end{aligned}$$

where the equality in the last row comes from $\text{var}(\psi_i^2) = O(\mathbb{E}(\psi_i^4)) = O(h^{-3d_T})$.

Showing $h^{d_T} n_\ell^{-1} \sum_{i \in I_\ell} \mathbb{E}[\hat{\psi}_{i\ell}^2 - \psi_i^2 \mid Z^\ell] = o_p(1)$. Let us now introduce:

$$\hat{\phi}_{i\ell} = K_h(T_{i\ell} - t) \hat{r}_{i\ell} \hat{\lambda}_{i\ell} (Y_i - \widehat{\mu}_{Y_{i\ell}}) \quad (51)$$

$$\hat{\varphi}_{i\ell} = K_h(T_{i\ell} - t') \hat{\lambda}_{i\ell} (\widehat{\mu}_{Y_{i\ell}} - \widehat{\omega}_{Y_{i\ell}}) \quad (52)$$

$$\hat{m}_{i\ell} = \widehat{\omega}_{Y_{i\ell}} - \hat{\psi}_i \quad (53)$$

Then again $\hat{\psi}_{i\ell} = \hat{\phi}_{i\ell} + \hat{\varphi}_{i\ell} + \hat{m}_{i\ell}$ and in the expansion of $\hat{\psi}_{i\ell}^2 - \psi_i^2$, we first compute for $\hat{\phi}_{i\ell}^2$

$$h^{d_T} \mathbb{E}(\hat{\phi}_{i\ell}^2 \mid Z^\ell) = \mathbb{E} \left[\mathbb{E} \left[(Y_i - \widehat{\mu}_{Y_{i\ell}})^2 \hat{r}_{i\ell} \hat{\lambda}_{i\ell}^2 \mid T = t, X, M, Z^\ell \right] f_{T|X,M}(t \mid X_i, M_i) \hat{r}_{i\ell} \hat{\lambda}_{i\ell}^2 \mid Z^\ell \right] R_k^{d_T} + o_p(1)$$

By pairing $\hat{\phi}_{i\ell}^2$ to ϕ_i^2 we show that

$$\mathbb{E} \left[\mathbb{E} \left[(Y_i - \widehat{\mu}_{Y_{i\ell}})^2 \hat{r}_{i\ell} \hat{\lambda}_{i\ell}^2 - (Y_i - \overline{\mu}_{Y_i})^2 \bar{r}_i \bar{\lambda}_i^2 \mid T = t, X, Z^\ell \right] f_{T|X,M}(t \mid X_i, M_i) \mid Z^\ell \right] = o_p(1)$$

Using the decomposition of the outcome residual in Eq. (41), the previous writes

$$\begin{aligned}
 &\mathbb{E} \left[\left\{ (\hat{r}_{i\ell} - \bar{r}_i) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (Y_i - \overline{\mu}_{Y_i}) - (\hat{r}_{i\ell} - \bar{r}_i) (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \right. \right. \\
 &\quad \left. \left. - (\hat{r}_{i\ell} - \bar{r}_i) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{\lambda}_i - (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{r}_i \right. \right. \\
 &\quad \left. \left. + (\hat{r}_{i\ell} - \bar{r}_i) (Y_i - \overline{\mu}_{Y_i}) \bar{\lambda}_i + (\hat{\lambda}_{i\ell} - \bar{\lambda}_i) (Y_i - \overline{\mu}_{Y_i}) \bar{r}_i - (\widehat{\mu}_{Y_{i\ell}} - \overline{\mu}_{Y_i}) \bar{r}_i \bar{\lambda}_i \right\}^2 \mid Z^\ell \right].
 \end{aligned}$$

which is $o_p(1)$ by Assumption 7 and Assumption 10.1.

Next, for $\hat{\varphi}_{i\ell}$, we similarly write:

$$h^{d_T} \mathbb{E}(\hat{\varphi}_{i\ell}^2 \mid Z^\ell) = \mathbb{E} \left[\mathbb{E} \left[(\widehat{\mu}_{Y_{i\ell}} - \widehat{\omega}_{Y_{i\ell}})^2 \mid T = t, M, X, Z^\ell \right] f_{T|X,M}(t' \mid X_i, M_i) \hat{\lambda}_{i\ell}^2(t, X)^2 \mid Z^\ell \right] R_k^{d_T} + o_p(1)$$

By pairing $\widehat{\varphi}_{i\ell}^2$ to φ_i^2 we show that

$$\mathbb{E} \left[\mathbb{E} \left[(\widehat{\mu}_{Yi\ell} - \widehat{\omega}_{Yi\ell})^2 \widehat{\lambda}_{i\ell}^2 - (\overline{\mu}_{Yi} - \overline{\omega}_{Yi})^2 \bar{\lambda}_i^2 \middle| T = t, X, Z^\ell \right] f_{T|X,M}(t' | X_i, M_i) \middle| Z^\ell \right] = o_p(1)$$

This time using the decomposition of the conditional mean outcome residual in Eq. (42), the previous writes as

$$\begin{aligned} & \mathbb{E} \left[\left\{ (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) (\widehat{\mu}_{Yi\ell} - \overline{\mu}_{Yi}) - (\widehat{\lambda}_{i\ell} - \bar{\lambda}_{i\ell}) (\widehat{\omega}_{Yi\ell} - \overline{\omega}_{Yi}) + (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) \overline{\mu}_{Yi} \right. \right. \\ & \left. \left. + (\widehat{\mu}_{Yi\ell} - \overline{\mu}_{Yi}) \bar{\lambda}_i - (\widehat{\omega}_{Yi\ell} - \overline{\omega}_{Yi}) \bar{\lambda}_i - (\widehat{\lambda}_{i\ell} - \bar{\lambda}_i) \overline{\omega}_{Yi} \right\}^2 \middle| Z^\ell \right]. \end{aligned}$$

which is $o_p(1)$ by Assumption 7 and Assumption 10.1.

Next, if we pair $\widehat{m}_{i\ell}^2$ to m_i^2 through

$$h^{d_T} \mathbb{E} \left\{ \left[\widehat{\omega}_{Yi\ell} - \widehat{\psi}_i \right]^2 \middle| Z_{I_\ell}^c \right\} = o_p(1)$$

This holds because we assume the nuisance estimators are bounded, and following a similar calculation as Eq. (V1-3) it can be seen that $h^{d_T} \mathbb{E} [\widehat{\psi}_i | Z^\ell] = o_p(1)$. Combined with Jensen's inequality, this can be used to obtain the desired result. Eventually, $h^{d_T} \mathbb{E} (\widehat{\phi}_{i\ell} \widehat{\varphi}_{i\ell} - \phi_i \varphi_i | Z^\ell) = o_p(1)$ following a computation similar to (V1-5) and by Assumption 7 and Assumption 10.1. Similar reasoning applies to $\mathbb{E} (\widehat{\phi}_{i\ell} \widehat{m}_{i\ell} - \phi_i \widehat{\psi}_i | Z^\ell)$, $\mathbb{E} (\widehat{\varphi}_{i\ell} \widehat{m}_{i\ell} - \varphi_i \widehat{\psi}_i | Z^\ell)$.

Showing $h^{d_T} n_\ell^{-1} \sum_{i \in I_\ell} \Delta_{i\ell} = o_p(1)$. We write $\Delta_{i\ell} \equiv \widehat{\psi}_{i\ell}^2 - \psi_i^2 - \mathbb{E} [\widehat{\psi}_{i\ell}^2 - \psi_i^2 | Z^\ell]$ and will use the same decompositions using $\widehat{\phi}_{i\ell}, \widehat{\varphi}_{i\ell}, \widehat{m}_{i\ell}$.

Writing $\Delta_{i\ell, \phi} \equiv \widehat{\phi}_{i\ell}^2 - \phi_i^2 - \mathbb{E} [\widehat{\phi}_{i\ell}^2 - \phi_i^2 | Z^\ell]$, with the decomposition in Eq. (41) and Assumption 10.2-3., the similar arguments as for $\mathbb{E} [\phi_i^4]$ above yields

$$\mathbb{E} [\Delta_{i\ell, \phi}^2 | Z^\ell] = O_p \left(\mathbb{E} \left[K_h(T_i - t)^4 \left[(Y_i - \widehat{\mu}_{Yi\ell})^2 \widehat{r}_{i\ell} \widehat{\lambda}_{i\ell}^2 - (Y_i - \overline{\mu}_{Yi})^2 \bar{r}_i \bar{\lambda}_i^2 \right]^2 \middle| Z^\ell \right] \right) = O_p(h^{-3d_T}). \quad (54)$$

We can apply a similar reasoning for $\Delta_{i\ell, \varphi} \equiv \widehat{\varphi}_{i\ell}^2 - \varphi_i^2 - \mathbb{E} [\widehat{\varphi}_{i\ell}^2 - \varphi_i^2 | Z^\ell]$, and other cross terms.

Then $\text{var}(h^{d_T} n_\ell^{-1} \sum_{i \in I_\ell} \Delta_{i\ell} | Z^\ell) = O_p(h^{2d_T} n_\ell^{-1} h^{-3d_T}) = O_p(n_\ell^{-1} h^{-d_T}) = o_p(1)$. The result follows by the conditional Markov's inequality.

Consistency of $\widehat{B}_{t,t'}$: Theorem 8 provides $\mathbb{E}[\widehat{\eta}_{t,t'}] = b_n^2 \epsilon^2 B_{t,t'} + o(b_n^2)$ and therefore $\mathbb{E}[\widehat{B}_{t,t'}] = B_{t,t'} + o(1)$. Compute

$$\begin{aligned}
 & \mathbb{E}[\hat{\eta}_{t,t',b_n} \hat{\eta}_{t,t',\epsilon b_n}] \\
 &= O\left(\frac{1}{n} \mathbb{E} \left[\int_{\mathcal{T}} K_{b_n}(s-t) K_{\epsilon b_n}(s-t) \mathbb{E} \left[(Y - \hat{\mu}_{Y_\ell}(t, M, X))^2 \mid T = s, X, M, Z^\ell \right] \right. \right. \\
 & \quad \times \left. \left. f_{T|X,M}(s \mid M, X) ds \hat{r}_\ell(M, X)^2 \hat{\lambda}_\ell(t', X)^2 \mid Z^\ell \right] \right) \\
 & \quad + O\left(\frac{1}{n} \mathbb{E} \left[\int_{\mathcal{T}} K_{b_n}(s-t') K_{\epsilon b_n}(s-t') \mathbb{E} \left[(\hat{\mu}_{Y_\ell}(t, M, X) - \hat{\omega}_{Y_\ell}(t, t', X))^2 \mid T = s, X, Z^\ell \right] \right. \right. \\
 & \quad \times \left. \left. f_{T|X,M}(s \mid M, X) ds \hat{\lambda}_\ell(t', X)^2 \mid Z^\ell \right] \right) \\
 &= O\left(n^{-1} b_n^{-d_T} \epsilon^{-d_T} \int_{\mathcal{R}^{d_T}} \prod_{j=1}^{d_T} k(u_j) k\left(\frac{u_j}{\epsilon}\right) du \mathbb{E} \left[\mathbb{E} \left[(Y - \hat{\gamma}_\ell(t, X))^2 \mid T = t, X, Z^\ell \right] \right. \right. \\
 & \quad \times \left. \left. f_{T|X,M}(t \mid X, M) \hat{r}_\ell(M, X)^2 \hat{\lambda}_\ell(t', X)^2 \mid Z^\ell \right] \right) \\
 & \quad + O\left(n^{-1} b_n^{-d_T} \epsilon^{-d_T} \int_{\mathcal{R}^{d_T}} \prod_{j=1}^{d_T} k(u_j) k\left(\frac{u_j}{\epsilon}\right) du \mathbb{E} \left[\mathbb{E} \left[(\hat{\mu}_{Y_\ell}(t, M, X) - \hat{\omega}_{Y_\ell}(t, t', X))^2 \mid T = t', X, M, Z^\ell \right] \right. \right. \\
 & \quad \times \left. \left. f_{T|X,M}(t' \mid X, M) \hat{\lambda}_\ell(t', X)^2 \mid Z^\ell \right] \right) \\
 &= O\left((n b_n^{d_T})^{-1}\right)
 \end{aligned}$$

with the same arguments as in Lemma 8.1. So $\text{cov}(\hat{\eta}_{t,t',b_n}, \hat{\eta}_{t,t',\epsilon b_n}) = O(\text{var}(\hat{\eta}_{t,t',\epsilon b_n})) = O\left((n b_n^{d_T})^{-1}\right)$. It follows that

$$\text{var}(\hat{B}_{t,t'}) = b_n^{-4} (1 - \epsilon^2)^{-2} (\text{var}(\hat{\eta}_{t,t',b_n}) + \text{var}(\hat{\eta}_{t,t',\epsilon b_n}) - 2 \text{cov}(\hat{\eta}_{t,t',b_n}, \hat{\eta}_{t,t',\epsilon b_n})) = O\left(b_n^{-4} (n b_n^{d_T})^{-1}\right).$$

By the Markov's inequality and Assumption 10.4.,

$$P\left(\left|\hat{B}_{t,t'} - \mathbb{E}[\hat{B}_{t,t'}]\right| > \epsilon\right) \leq \text{var}(\hat{B}_{t,t'}) / \epsilon^2 = O\left((n b_n^{d_T+4})^{-1}\right) = o(1).$$

So $\hat{B}_{t,t'} - B_{t,t'} = o_p(1)$.

□

10 NON PARAMETRIC LEARNING OF NUISANCE PARAMETERS

We present the non parametric method to estimate the conditional and cross conditional mean outcomes following (Singh et al., 2023). Using kernel mean embeddings, the mediated response curve $\eta_{t,t'}$ defined in Eq. (4) is identified as a sequential integral of the form $\int \mu_Y(t', m, x) dQ$ for the distribution $Q = f_{M|T,X}(m|t', x) f(x)$ in the Pearl formula in Eq. (10).

To begin, we construct an RKHS \mathcal{H} for μ_Y . We define a RKHS for the treatment \mathcal{H}_T , mediator \mathcal{H}_M and covariates \mathcal{H}_X , then assume that the conditional mean μ_Y is an element of the RKHS \mathcal{H} with the kernel $k((t, m, x), (t', m', x')) = k_T(t, t') k_M(m, m') k_X(x, x')$. Formally, this choice of kernel corresponds to the tensor product : $\mathcal{H} = \mathcal{H}_T \otimes \mathcal{H}_M \otimes \mathcal{H}_X$. As such,

$$\mu_Y(t, m, x) = \langle \mu_Y, \phi_T(t) \otimes \phi_M(m) \otimes \phi_X(x) \rangle, \tag{55}$$

where ϕ_T, ϕ_M and ϕ_X are the feature maps respectively associated with k_T, k_M and k_X .

Then, define:

$$\mu_m(t, x) = \int \phi_M(m) df_{M|T,X}(m|t, x) \quad (56)$$

$$\mu_{m,x}(t) = \int \{\mu_m(t, x) \otimes \phi_X(x)\} df(x) \quad (57)$$

Consider also the cross conditional mean outcome $\omega_Y(t', t, x) = \int \mu_Y(t', m, x) df_{M|T,X}(m|t, x)$ we previously defined in Eq. 9. We can then formulate it with the kernel mean embedding.

Then assuming regularity conditions on the RKHS, we have:

$$\omega_Y(t', t, x) = \langle \mu_Y, \phi_T(t') \otimes \mu_m(t, x) \otimes \phi_X(x) \rangle_{\mathcal{H}} \quad (58)$$

$$\eta_{t,t'} = \langle \mu_Y, \phi_T(t') \otimes \mu_{m,x}(t) \rangle_{\mathcal{H}} \quad (59)$$

Moreover, denote the kernel matrices by $K_{TT}, K_{MM}, K_{XX} \in \mathbb{R}^{n,n}$. Let \odot be the element-wise product. Mediated response curves have closed form solutions:

$$\begin{aligned} \hat{\omega}_Y(t', t, x) &= Y^\top (K_{TT} \odot K_{MM} \odot K_{XX} + n\lambda I)^{-1} \cdot \\ &\quad [K_{Tt'} \odot \{K_{MM}(K_{TT} \odot K_{XX} + n\lambda_1 I)^{-1}\}] \\ &\quad ((K_{Tt} \odot K_{Xx}) \odot K_{Xx}) \end{aligned} \quad (60)$$

$$\hat{\eta}_{t,t'} = \frac{1}{n} \sum_{i=1}^n \hat{\omega}_Y(t', t, x_i) \quad (61)$$

where (λ, λ_1) are ridge regression penalty parameters. Moreover, the conditional mean outcome written in Eq. 55 has the following closed form:

$$\begin{aligned} \hat{\mu}_Y(t, m, x) &= Y^\top (K_{TT} \odot K_{MM} \odot K_{XX} + n\lambda I)^{-1} \\ &\quad (K_{Tt} \odot K_{Mm} \odot K_{Xx}) \end{aligned} \quad (62)$$

11 EXPERIMENT DETAILS

All the code to reproduce our experiment can be found at <https://github.com/houssamzenati/double-debiased-machine-learning-mediation-continuous-treatments>.

11.1 Tuning of hyperparameters

In this part we discuss the selection strategy for hyperparameters.

Kernel smoothing bandwidth For the Scott (2015) heuristic of bandwidth selection, the bandwidth h is determined by multiplying the respective standard deviations of T with $Cn^{-1/5}$, where $C = 1.06$.

Kernel mean embedding ridge penalty We experiment grid search and generalized cross validation (GCV) as in (Singh et al., 2023), where for a variable $W \in \mathcal{W}$ we construct the matrices

$$H_\lambda = I - K_{WW} (K_{WW} + n\lambda I)^{-1} \in \mathbb{R}^{n \times n}$$

and set

$$\lambda^* = \operatorname{argmin}_{\lambda \in \Lambda} \frac{1}{n} \left\| \{\operatorname{tr}(H_\lambda)\}^{-1} \cdot H_\lambda Y \right\|_2^2, \quad \Lambda \subset \mathbb{R}$$

Kernel mean embedding bandwidth The exponentiated quadratic kernel is widely used among machine learning practitioners:

$$k(w, w') = \exp \left\{ -\frac{1}{2} \frac{(w - w')_{\mathcal{W}}^2}{\iota^2} \right\}$$

Importantly, it satisfies the required properties; it is continuous, bounded, and characteristic. Its hyperparameter is called the bandwidth ι . A convenient heuristic as in (Singh et al., 2023) is to set the lengthscale equal to the median interpoint distance of (W_i) , where the interpoint distance between the observations i and j is $\|W_i - W_j\|_{\mathcal{W}}$. When the input W is multidimensional, we use the kernel obtained as the product of scalar kernels for each input dimension.

11.2 Synthetic simulations

In this part we provide all additional details and results for the experiment of Hsu et al. (2020) and (Sani et al., 2024).

11.2.1 Generating process and causal experiment of Hsu et al. (2020)

The simulation is based on the following data generating process.

$$\begin{aligned} X &\sim \text{Uniform}(-1.5, 1.5), \\ U, V, W &\sim \text{Uniform}(-2, 2) \end{aligned}$$

and then

$$\begin{aligned} T &= 0.3X + U \\ M &= 0.3T + 0.3X + V \\ Y &= 0.3T + 0.3M + \alpha TM + 0.3X + \beta T^3 + W \end{aligned}$$

independently of each other. Outcome Y is a function of the observed variables T, M, X and an unobserved term W . α gauges the interaction effect between T and M . $\alpha = 0$ satisfies the assumption of no interaction as discussed in (Robins, 2003). In contrast, for $\alpha \neq 0$, direct and indirect effects are heterogeneous. The coefficient β determines whether the direct effect of T on Y is linear ($\beta = 0$) or nonlinear, namely cubic ($\beta \neq 0$).

In this simulation, the direct effect is given by $\theta(t, t') = 0.3(t - t') + 0.3\alpha t(t' - t) + \beta(t'^3 - t^3)$ and the indirect effect by $\delta(t, t') = 0.09(t' - t) + 0.3\alpha t'(t' - t)$. We will consider the setting of Hsu et al. (2020) where $\alpha = 0.25$ and $\beta = 0.5$ as Singh et al. (2023) did. Since the outcome function is smooth and has regular properties, we use a single fold for training the DML estimator.

For the causal experiment, the definition of the direct and indirect effects Hsu et al. (2020) sets $t' = 0$, for t , they consider a sequence of values defined by an equidistant grid between (and including) -1.5 and 1.5 with step size 0.1 (i.e. $t \in \{-1.5, -1.4, \dots, 1.4, 1.5\}$).

11.2.2 Additional results on the synthetic setting of Hsu et al. (2020)

Comparison of DML to other methods Table 2 reports the averages of the absolute bias (bias), standard deviation (std), and root mean squared error (RMSE) for each effect, where we consider the average and standard deviations of averaging the error over all treatment comparisons $(t - t')$ in the grid defined above. The DML approach provides significantly better estimation for the direct and total effect and also provides satisfactory performance for the indirect effect. Not surprisingly, the performances increase with larger sample sizes.

Strategy for bandwidth selection In this numerical experiment we compare two bandwidth selection procedures: we investigate the practical performance of the theoretical optimal bandwidth which minimizes the asymptotic mean squared error (AMSE) of Corollary 11 and the Scott (2015) rule of thumb which satisfies the assumptions of our theorems and corollaries. Figure 3 illustrates the average bias of both approaches under the 100 simulation, and shows how the AMSE strategy provides less variance than the heuristic but slightly more bias.

Table 2: Average absolute bias (bias), standard deviation (std), and root mean squared error (RMSE) for each effect across different sample sizes.

n	Method	direct			indirect			total		
		bias	std	rmse	bias	std	rmse	bias	std	rmse
500	OLS	0.3065	0.0411	0.3499	0.0133	0.0099	0.0151	0.3067	0.0413	0.3501
	IPW	0.1785	0.0687	0.2167	0.0698	0.0000	0.0794	0.1610	0.0732	0.1922
	KME	0.3748	0.0372	0.4898	0.4174	0.0588	0.5443	0.0842	0.0336	0.1090
	DML	0.0992	0.0361	0.1240	0.0339	0.0163	0.0427	0.0989	0.0391	0.1235
1000	OLS	0.3013	0.0332	0.3445	0.0079	0.0075	0.0090	0.3014	0.0340	0.3446
	IPW	0.1437	0.0443	0.1773	0.0698	0.0000	0.0794	0.1255	0.0450	0.1510
	KME	0.3726	0.0234	0.4860	0.4156	0.0440	0.5399	0.0681	0.0268	0.0903
	DML	0.0809	0.0320	0.0996	0.0233	0.0098	0.0299	0.0790	0.0323	0.0970
5000	OLS	0.3022	0.0139	0.3461	0.0040	0.0029	0.0045	0.3027	0.0135	0.3466
	IPW	0.0899	0.0183	0.1116	0.0698	0.0000	0.0794	0.0607	0.0207	0.0737
	KME	0.3759	0.0107	0.4889	0.4134	0.0191	0.5364	0.0444	0.0123	0.0606
	DML	0.0463	0.0135	0.0576	0.0154	0.0054	0.0196	0.0417	0.0148	0.0523

Validation of the asymptotic confidence interval We also perform a numerical experiment for the asymptotic confidence interval to test its empirical coverage. We run 100 simulations for the DML estimator where we report in Table 3 the number of times the true mediated response is inside the asymptotic confidence interval. We take a 95% level of confidence and therefore, the empirical coverage should ideally be close to that value. We can observe that this coverage has a strong value for a low sample size $n = 500$ but gets closer to this target and increases along with larger sample sizes which are closer to asymptotic regimes.

Table 3: Percentage of successful coverage of the asymptotic 95% confidence interval on the synthetic setting of Hsu et al. (2020) with regards to sample size.

Sample size	500	1000	5000
Coverage	0.915	0.894	0.898

Benefits of cross-fitting Regarding the finite sample performance of cross-fitting, we did not observe significant improvements in our numerical simulation. As expected, cross-fitting is less effective for simple problems involving Donsker class functions (Williamson et al., 2023). Practical studies, such as (Fan et al., 2022a) and (Williamson et al., 2023), have also noted the limited benefits of cross-fitting in such cases, while its advantages become more evident in more complex nonparametric problems (Kennedy, 2022). We consider simple class of estimators in our simulation following (Hsu et al., 2020) and (Singh et al., 2023) with a simple outcome function, which may explain the limited impact of cross-fitting.

Parametric versus nonparametric nuisance estimation We design an experiment similar in spirit to the misspecification experiments of (Kennedy et al., 2017) and (Doss et al., 2023), where we aim at simulating the evaluating variants of the treatment densities and the conditional mean outcomes (μ and ω_Y). We either use simple parametric models (Gaussian parametrization of the treatments, linear model for the outcome), or leverage non-parametric methods (conditional kernel estimation and Gaussian kernel mean embedding). We design four scenarios in which we compare our DML (with KME) against IPW and the G-computation (with KME) in our numerical simulation (Hsu et al., 2020; Singh et al., 2023), (i) treatment densities and conditional mean outcomes nonparametric ii) parametric well specified treatment densities and nonparametric conditional mean outcomes iii) non parametric treatment densities and parametric conditional mean outcome iv) parametric treatment densities and conditional mean outcomes.

In Table 4, we observe that for all the scenarios with all sample sizes DML consistently outperforms the two baselines, demonstrating the best performance even under all adverse conditions.

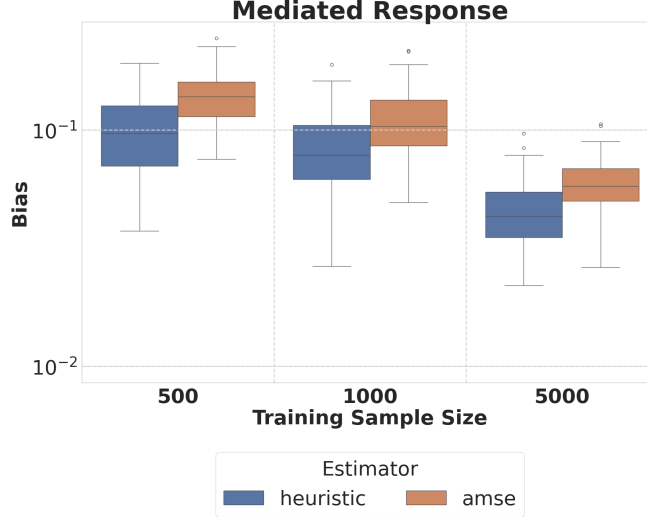


Figure 3: Bias of mediated response estimation with the DML estimator on simulations with different sample sizes and with two bandwidth selection strategies.

Table 4: Performance summary for all nuisance parameter estimation scenarios

Estimator	n_{samples}	Scenario i)		Scenario ii)		Scenario iii)		Scenario iv)	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
IPW	500	0.1222	0.1472	0.1366	0.1655	0.1173	0.1424	0.1344	0.1631
DML	500	0.0824	0.0968	0.0822	0.0966	0.0800	0.0944	0.0786	0.0928
KME	500	0.1783	0.2095	0.1802	0.2123	0.1671	0.1933	0.1665	0.1930
IPW	1000	0.0945	0.1144	0.1194	0.1426	0.0946	0.1146	0.1180	0.1422
DML	1000	0.0624	0.0742	0.0620	0.0738	0.0631	0.0747	0.0633	0.0749
KME	1000	0.1731	0.2015	0.1739	0.2021	0.1637	0.1885	0.1629	0.1885
IPW	5000	0.0515	0.0638	0.0818	0.0982	0.0506	0.0627	0.0820	0.0983
DML	5000	0.0325	0.0390	0.0330	0.0393	0.0323	0.0391	0.0310	0.0369
KME	5000	0.1694	0.1932	0.1697	0.1936	0.1631	0.1862	0.1634	0.1865

11.3 Comparison to (Sani et al., 2024) in their simulation setting

Here we compare our DML formulation to the multiply robust estimator of Sani et al. (2024) in the numerical experiment they did. We implemented their estimator and reproduced their numerical simulation, which is done with binary mediators. In Table 5, we report the same metric they use, the average absolute bias to compare our DML to their variant, and also IPS, OLS, KME. Once again we see that our formulation outperforms theirs even in their setting; this illustrates the benefits of using an implicit integration of the conditional mean outcome μ_Y to estimate the cross conditional mean outcome ω_Y and to use treatment propensities instead of the mediator density.

11.4 Application to cognitive function

In this part we provide details on our application of mediation analysis on the UKBB project (Sudlow et al., 2015). We analyse the effect of a continuous measure, the glycated hemoglobin (HbA1c) on cognitive functions, and its possible mediation by the brain structure.

11.4.1 Causal estimand identification and data preparation

The treatments in our study are well-defined, with a consistent definition among subjects and no interaction between subjects, as they are independent participants in the UKBB study, so the Stable Unit Treatment Value Assumption (SUTVA) is satisfied. Assumptions 1 and 2 focus on confounding variables between treatment,

Table 5: Average Absolute Bias for different estimators in the setting of (Sani et al., 2024)

Estimator	n_{samples}	Average Absolute Bias
OLS	2000	13.65
IPW	2000	0.5298
KME	2000	1.861
DML (Sani et al., 2024)	2000	0.4099
DML (ours)	2000	0.2507
OLS	5000	13.61
IPW	5000	0.2883
KME	5000	1.832
DML (Sani et al., 2024)	5000	0.2997
DML (ours)	5000	0.1445
OLS	8000	13.62
IPW	8000	0.2481
KME	8000	1.828
DML (Sani et al., 2024)	8000	0.3111
DML (ours)	8000	0.1280

mediators, and outcomes, ensuring identifiability of the total, direct, and indirect causal effects. We accounted for a broad range of confounders, including behavioral, physiological, and societal factors, as well as technical artifacts such as the evaluation center, head positioning in imaging, and known brain diseases (Alfaro-Almagro et al., 2021; Newby and Garfield, 2022; Schurz et al., 2021; Topiwala et al., 2022). The positivity assumption (Assumption 3) was assessed by defining a grid of evaluation over values on which there was coverage in Figure 4.

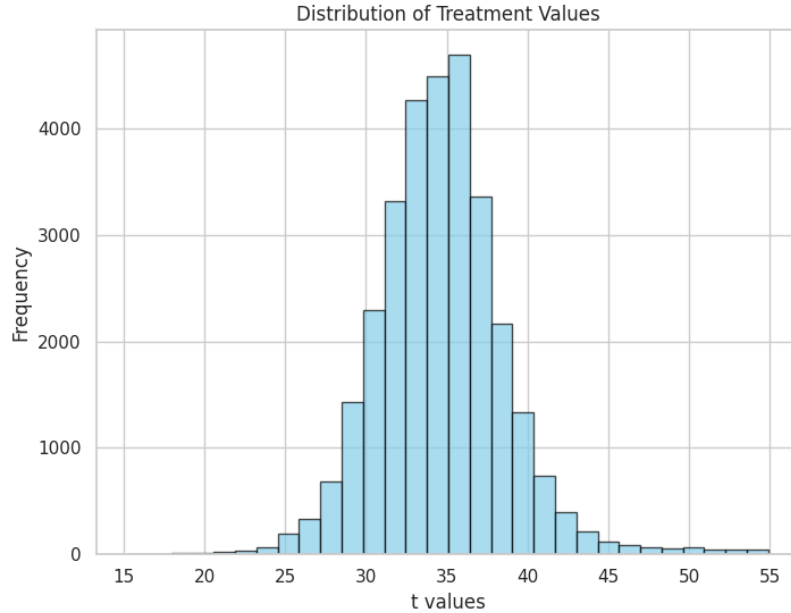


Figure 4: Empirical histogram of glycated hemoglobin levels (treatments) in the UKBB data.

A key aspect of establishing identifiability is preparing the variables for estimation models. For the outcome, we aggregated cognitive test results (e.g., working memory, fluid intelligence, vocabulary) into a single "G-factor" measuring general cognitive ability, obtained via Principal Component Analysis (PCA) as in (Fawns-Ritchie and Deary, 2020). This score was based on four cognitive tests administered at the imaging visit, with the G-factor correlating well with individual test results. For imaging data, we reduced the dimensionality of preprocessed imaging-derived phenotypes (IDPs) as in (Miller et al., 2016; Alfaro-Almagro et al., 2018) using PCA after

standard scaling, retaining six components from categories like dMRI skeleton, subcortical volumes, and regional gray matter volumes.

We also consolidated lifestyle variables to reduce dimensionality and address missing data due to the design of questionnaires. After transforming categorical variables via one-hot encoding and retaining those with more than 10,000 non-zero values (over the full dataset with 500,000 individuals), we further imputed missing data and reduced dimensionality using the R packages `missMDA` 485 and `FactoMineR` Josse and Husson (2016); Lê et al. (2008). The analysis was restricted to participants with brain MRI data, and we used two components for imputation.

After selecting and preparing the variables for causal mediation analysis, we limited the dataset to participants with brain imaging data and complete data for all relevant variables, resulting in a final sample of $n = 30,595$ independent participants.

11.4.2 Experiment details

To estimate the total, direct and indirect causal effects for each treatment, mediated by brain structure, measured by brain MRI, we applied the estimators able to handle multidimensional mediators: the OLS estimator, the KME G-computation, the IPW and our double machine learning estimator as explained in Section 5. We used cross-fitting with 2 folds. To assess uncertainty in the estimation, we applied the estimators to 100 bootstrapped samples for each exposure and compare the bootstrap procedure with the asymptotic confidence interval we derived for the mediated response.

11.4.3 Experiment results

Previous studies (Newby and Garfield, 2022) show that suffering from diabetes has deleterious consequences and results in a lower G-factor score. Diabetes tends to have a negative total causal effect on cognitive functions, so we have explored the role of glycemic control, through the measure of glycated hemoglobin (HbA1c) at the initial visit. HbA1c is a proxy of the three-month average blood sugar level, so high levels of HbA1c can be found in individuals with diabetes with difficulties balancing their diet, physical activity, and medication to control their glucose level or in individuals with an undiagnosed issue. Exploring the role of HbA1c rather than the diabetic status allows us to directly assess the role of blood glucose in the health damages associated with this condition.

As illustrated in Figure 5 (left panel), the total effect is very low, indicating a lack of evidence of an effect of HbA1c at the initial visit on the cognitive functions a few years later (imaging visit). This lack of effect has been confirmed with causal inference methods without mediation (Double Machine learning, as implemented in the Python package `econML` Keith et al. (2019)).

Regarding the indirect effect estimates for diabetes and alcohol (Newby and Garfield, 2022; Topiwala et al., 2022), they are typically of weaker amplitude than the total effect. We present the total and indirect effect results for the estimators for exposure to glycated hemoglobin in Figure 5. With DML, we found a slightly negative but not statistically significant effect of exposure to glycated hemoglobin on cognitive function.

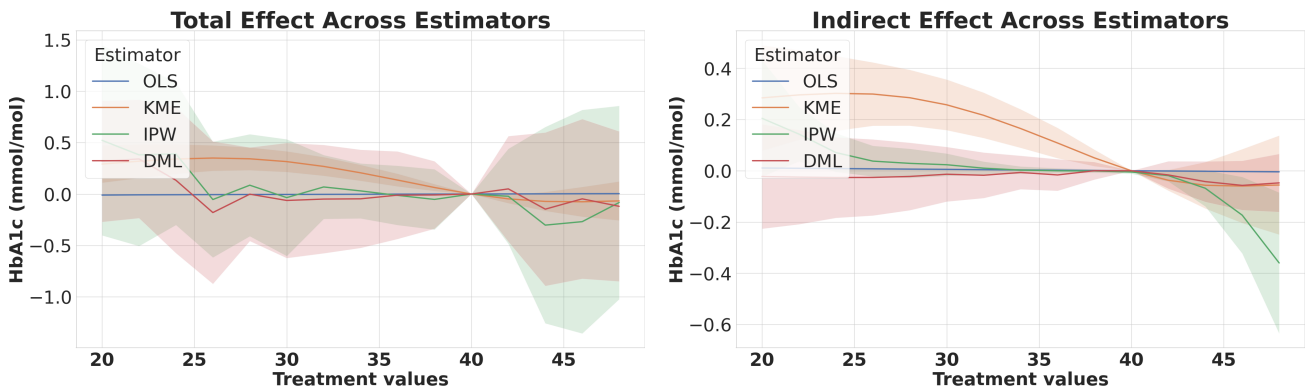


Figure 5: Effect estimation on the UKBB dataset for the total effect (left) and the indirect effect (right)

Overall, the results depended weakly on the choice of the estimation method. However, we notice an important

difference in the variance between the different estimators, with a very small variance for the OLS and KME, and larger ones for the DML and IPW estimators. Our results also illustrate how IPW suffers more from overlap issues at the boundaries of the treatment space. Eventually, as a validation, we provide in Figure 6 a comparison of our asymptotic confidence interval and the bootstrap procedure, showing a more regular estimation (and less computationally expensive).

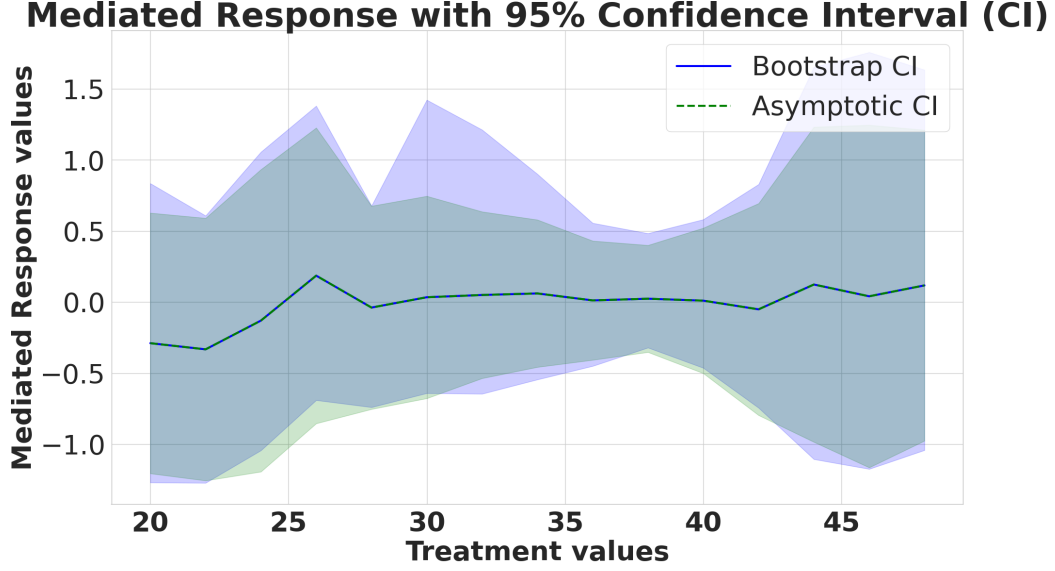


Figure 6: Comparison of the asymptotic confidence interval and the bootstrap procedure for uncertainty of the mediated response on the UKBB dataset.

11.5 Computation infrastructure

We ran our experiments on a CPU clusters with the following characteristics.

- **Dell C6320:**
 - 2 Intel(R) Xeon(R) CPU E5-2660 v2 @ 2.20GHz
 - 256G of RAM
 - Ethernet 1G
 - Infiniband Mellanox QDR
- **Dell R7525:**
 - 2 AMD EPYC 7742 64-Core Processor
 - 512G of RAM
 - Ethernet 10G
- **Dell R7525:**
 - 2 AMD EPYC 7702 64-Core Processor
 - 512G of RAM
 - Ethernet 10G
- **Dell R7525:**
 - 2 AMD EPYC 7742 64-Core Processor
 - 1024G of RAM
 - Ethernet 10G
- **Dell R640:**

- 2 Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz
- 256G of RAM
- Ethernet 1G (pending upgrade to 10G)
- **HPE Proliant DL385:**
 - 2 AMD EPYC 7763 64-Core Processor
 - 1024G of RAM
 - Ethernet 10G