A RIESZ REPRESENTER PERSPECTIVE ON TARGETED LEARNING

Salvador V. Balkus
Department of Biostatistics,
Harvard Chan School of Public Health,
sbalkus@g.harvard.edu

Christian Testa
Department of Biostatistics,
Harvard Chan School of Public Health,
ctesta@g.harvard.edu

Nima S. Hejazi
Department of Biostatistics,
Harvard Chan School of Public Health,
nhejazi@hsph.harvard.edu

April 13, 2026

ABSTRACT

As research in causal inference has sought to address more complex scientific questions, the number of specialized estimands in the field has proliferated. Recognition that many of these estimands share a common linear form has generated interest in simplifying estimation procedures using Riesz representers. In this work, we construct a targeted minimum loss-based estimation procedure for nested linear functionals, leveraging Riesz representers of a general recursive form. The proposed method unifies asymptotically efficient estimation for a variety of statistical estimands that originate in causal inference, including the effects of time-varying treatments under treatment-confounder feedback and direct and indirect effects from causal mediation analysis. We demonstrate how our proposal reduces the need for laborious and technically challenging mathematical derivations when constructing estimators of common statistical estimands under complex forms of censoring and sampling. We investigate and validate the properties of the proposed procedures in numerical experiments, discuss open-source software facilitating their implementation, and illustrate their application in a re-analysis of data from an HIV vaccine efficacy trial.

1 Introduction

Contemporaneous advances at the intersection of causal inference and machine learning in recent decades have led to a rich and thriving research interface between the two fields. Significant effort in causal inference has been devoted to developing interpretable, scientifically relevant estimands and deriving identification assumptions re-expressing counterfactual queries in terms of quantities that are estimable from observed data (Pearl, 2009; Hernán & Robins, 2026). Meanwhile, the library of machine learning algorithms for flexible, data-adaptive regression has continued to grow (Friedman et al., 2001).

Despite these dual advances, estimation and uncertainty quantification remained siloed, for a time, between causal inference and machine learning. The use of semi-parametric theory to bridge these areas (Bickel et al., 1993; van der Laan & Robins, 2003)—in particular the development of theory and methods to correct an asymptotic bias that arises from the use of flexible regression procedures for nuisance function estimation—helped give rise to formal frameworks for de-biased estimation. Among these frameworks, *targeted learning* (van der Laan & Rubin, 2006; van der Laan & Rose, 2011) has outlined a general, structured template for constructing asymptotically efficient substitution estimators robust to model misspecification, and, as such, has helped to drive an explosion of interest in this interface, recently termed *causal machine learning*.

Causal machine learning seeks to unify core principles of causal inference with advances in machine learning, using the former to derive target statistical estimands and the latter to facilitate flexible estimation of components of the data-generating process (i.e., nuisance functions) relevant to estimation of the target estimand. This use of flexible learning algorithms for nuisance estimation can help to curb an often severe bias—from model misspecification—in the final estimator, yet it comes at a cost: the resultant estimator incurs an asymptotic bias stemming from a mismatch between the typically slower convergence rates of machine learning algorithms and the comparatively faster rate attained by more commonly used parametric regression strategies.

To fix ideas, consider, for example, the statistical functional $\Psi(\mathsf{P}_0) = \mathsf{E}_{\mathsf{P}_0}(\mathsf{E}_{\mathsf{P}_0}(Y \mid A = a, L))$, which, under well-studied identification assumptions (Hernán & Robins, 2026), coincides with the counterfactual mean $\mathsf{E}[Y^a]$ of an outcome Y where all study units receive a specific level a of treatment A. Here, Y^a is the potential outcome of Y under treatment level $a \in \mathcal{A}$, and we use P_0 to denote the data-generating law of the data unit O, where we assume that $\mathsf{P}_0 \in \mathcal{M}$, for $\mathcal{M}$ an unrestricted statistical model (that is, a collection of candidate data-generating laws P); the statistical parameter then is a mapping $\Psi : \mathcal{M} \to \mathbb{R}^d$. We let $\overline{Q}_{\mathsf{P}_0}(A, L) := \mathsf{E}_{\mathsf{P}_0}(Y \mid A, L)$ denote the mean of the outcome Y conditional on treatment A and putative confounders L, and we will, going forward, abbreviate the nuisance function evaluated at the true distribution $\overline{Q}_{\mathsf{P}_0}$ as $\overline{Q}_0$, and the nuisance at an arbitrary distribution $\overline{Q}_{\mathsf{P}}$ as $\overline{Q}$. We note that $\Psi(\mathsf{P}_0) = \Psi(\overline{Q}_{\mathsf{P}_0}) = \mathsf{E}_{\mathsf{P}_0}(\overline{Q}_{\mathsf{P}_0}(a, L))$, clarifying that the parameter mapping $\Psi(\cdot)$ depends on only a specific component $\overline{Q}_{\mathsf{P}_0}$ of the data-generating law P_0.

Construction of an estimator ψ_n of $\psi_0 = \Psi(\mathsf{P}_0)$ reduces to construction of an estimator $\overline{Q}_n = \hat{\mathsf{E}}_{\mathsf{P}}(Y \mid A, L)$ of $\overline{Q}_0$, the nuisance function appearing in the parameter mapping. Learning $\overline{Q}$ can be accomplished by a variety of means, among them parametric regression (McCullagh & Nelder, 1989); semi-parametric regression, e.g., generalized additive models (Hastie & Tibshirani, 1987, 1990); and machine learning, e.g., decision trees and forests (Breiman et al., 1984; Breiman, 2001), gradient boosting machines (Friedman, 2001), neural networks (Ripley, 1996), and stacked regression or ensemble models (Wolpert, 1992; Breiman, 1996; van der Laan et al., 2007a). How the estimator $\overline{Q}_n$ of $\overline{Q}_0$ is constructed impacts the downstream properties of the estimator $\psi_n = n^{-1} \sum_{i=1}^n \overline{Q}_n(a, L)$; moreover, approaches for constructing the nuisance estimator $\overline{Q}_n$—which strive to obtain optimal estimators of the nuisance function $\overline{Q}_0$, e.g., by appealing to the asymptotic optimality of cross-validated loss-based estimation (van der Laan et al., 2004; Dudoit & van der Laan, 2005; van der Vaart et al., 2006)—are not generally optimally suited for use in constructing the estimator ψ_n. In particular, the use of data-adaptive regression strategies necessitates the negotiation of a misspecification–convergence tradeoff: the estimator $\overline{Q}_n$ is less likely to be misspecified, but it will attain a slower convergence rate, leaving the estimator ψ_n with a bias that does not vanish asymptotically.

To enable the use of machine learning in such statistical estimation tasks, a great deal of research in recent years has centered on developing techniques to remove such asymptotic bias. Targeted maximum likelihood estimation (TMLE), or targeted minimum loss-based estimation, introduced by van der Laan & Rubin (2006), is a general template for de-biasing estimators through an updating procedure generically described as "targeting," which recognizes that the bias incurred is proportional to how well (or poorly) an estimating equation based on the efficient influence function (EIF), a key object in semi-parametric efficiency theory (Bickel et al., 1993), is solved. By fluctuating the nuisance estimator $\overline{Q}_n$ in a manner that solves the EIF estimating equation, a targeting procedure produces an updated estimator $\overline{Q}_n^\star$ that results in a consistent and, under conditions, asymptotically efficient estimator $\psi_n^\star = n^{-1} \sum_{i=1}^n \overline{Q}_n^\star(a, L)$ of ψ_0; the resultant substitution estimator is also, under certain regularity conditions (van der Laan & Rose, 2011), asymptotically efficient, achieving the lowest possible variance among the class of regular asymptotically linear (RAL) estimators of the target estimand ψ_0. This latter property is tied to the EIF, which characterizes the semi-parametric efficiency bound in a non-parametric statistical model; the interested reader is invited to consult recent reviews of the relevant semi-parametric theory (Kennedy, 2016; Hines et al., 2022; Kennedy, 2024).

Broadly, derivation of a TMLE algorithm consists of three steps: (1) obtain the unique EIF in the non-parametric model $\mathcal{M}$ for RAL estimators of the target estimand; (2) set up a maximum likelihood (or loss-based minimization) problem with parameter(s) ε, such that the derivative of the loss with respect to ε equals the EIF; and (3) update the initial estimator $\overline{Q}_n$ of the nuisance function $\overline{Q}_0$ by solving the loss minimization problem for $\hat{\varepsilon}$. By nature of its construction, this final step fluctuates the initial estimator in a direction (in model space) that results in an updated estimator that better solves the EIF estimating equation than its counterpart obtained without reference to the EIF.

The generality of the TMLE framework leaves many details to the investigator. For any given target estimand, one must determine the unique EIF of estimators in the RAL class. In some instances, such calculations can be labor-intensive and require specialized expertise in semi-parametric theory, though accessible guidance on the nature of such calculations has become available in more recent years (Hines et al., 2022; Kennedy, 2024). Although TMLE procedures for target estimands of common interest are widely available, including for the statistical functionals that identify the counterfactual mean under treatment (Gruber & van der Laan, 2010) (and related contrasts such as the average treatment effect) or the effects of time-varying treatments (Stitelman et al., 2011; van der Laan & Gruber, 2012; Díaz et al., 2021), it remains challenging for investigators without expertise in semi-parametric theory to derive TMLE algorithms for novel estimands or for well-studied estimands in settings subject to complex forms of censoring.

Concurrent developments in causal machine learning have formulated alternative representations for the EIFs of common classes of target estimands. Consider statistical estimands of the form $\Psi(\mathsf{P}_0) \equiv \Psi(\eta_0) = \mathsf{E}_{\mathsf{P}_0}[h(O_i;\eta_0)]$, where $\eta_0 := \eta_{\mathsf{P}_0}$ are nuisance functions that depend on P_0 (e.g., $\overline{Q}$) and h is a linear functional of η; here, we stress our choice of notation $\Psi(\mathsf{P}_0) \equiv \Psi(\eta_{\mathsf{P}_0})$ to indicate that the target parameter depends on P_0 only through the nuisance functions η_0. Among recent developments, Hirshberg & Wager (2021) and Chernozhukov et al. (2022b), for example, recognized that when the target estimand admits the structure above, the EIF ϕ then takes the form

$$\phi(\mathsf{P})(O) = h(X;\eta) + \alpha(X)(Y - \eta(X)) - \Psi(\eta)\ , \tag{1}$$

where X denotes all variables in O upon which the nuisance function η depends and $h(\cdot)$ is a non-stochastic transformation of the nuisance function η. Here, α is called a *Riesz representer* and is analogous to inverse probability weights (Horvitz & Thompson, 1952) and balancing weights (Zubizarreta, 2015) commonly used for estimation in causal inference and missing data problems. Such a representation is motivated primarily by the convenience of estimating α using specialized machine learning algorithms (Chernozhukov et al., 2022a) or de-biasing well-known regression algorithms (Bruns-Smith et al., 2025). However, as Williams et al. (2025) argue based on the work of Newey (1994), this representation also considerably simplifies the process of deriving an EIF, especially in settings where the estimand is complex, such as the iterated conditional expectation functionals that arise in longitudinal data problems subject to time-varying treatment and confounding (Hernán & Robins, 2026).

This general class of expressions now forms the basis of the *Riesz regression* framework (Williams et al., 2025). While many of the above works focus on direct estimation of the EIF, none, to our knowledge, construct a TMLE. We argue that the TMLE framework can be adapted to yield simpler algorithms for estimator construction when the target estimand admits a representation compatible with Riesz regression.

Our contributions. Our work outlines how the Riesz representation theorem can be used to construct a TMLE procedure for an extremely broad class of statistical estimands. Generalizing previous developments in the Riesz learning framework such as Equation 1, as well as findings in longitudinal causal inference, we derive a common EIF for a sequence of nested linear functionals that depend on arbitrary nuisance parameters. Using this EIF, we provide a TMLE that fluctuates each nuisance parameter by employing its corresponding Riesz representer as a "clever covariate". We implement this TMLE as a software package, {`RieszCML`}, unifying estimation across a multitude of causal inference areas such as longitudinal data, mediation, quantile effects, and multi-phase sampling schemes.

Outline. The remainder of the manuscript is organized as follows. Section 2 describes the de-biasing properties of the Riesz representer α, when it exists. Next, Section 3 provides a recursively applicable "Riesz EIF", generalizing previous works in which the parameter mapping $\Psi(\mathsf{P}_0)$ depends on only a single conditional expectation as nuisance function. Section 4 describes a TMLE algorithm for this nested setting, and Section 5 demonstrates how it can be applied to develop estimators of a few commonly studied estimands. Finally, Section 6 presents simulation experiments demonstrating the use of this TMLE with Riesz regression for estimation in the longitudinal regime, and Section 7 illustrates use of the proposed methodology by its application to an immune correlates analysis of data from the HVTN 505 vaccine efficacy trial (NCT00865566). We conclude in Section 8 with a discussion of productive avenues for future investigation.

2 Mathematical preliminaries and notation

2.1 Observed data and statistical model

Throughout, we consider data on n study units, $O_1,\ldots,O_n$, where the data on a given study unit is represented by the random variable O. We allow details of the hypothetical data-generating study to vary across examples, and, accordingly, allow constituent random variables contained within O to vary across study contexts. Following standard conventions in causal inference, O is composed of a sequence of temporally ordered observations of several random variables. In a study with exposure occurring at a single time point, we have $O = (L, A, Y)$, with baseline confounders L, treatment A, and outcome Y. Meanwhile, in a study with time-varying confounders and exposure, we have $O = (L_0, L_1, A_1, \ldots, L_T, A_T, Y)$, where, for $t = 1,\ldots,T$ time points, L_t are time-varying confounders and A_t are time-varying treatments, with L_0 additionally denoting time-fixed confounders (e.g., sex-at-birth) and $Y := L_{T+1}$ denoting a realization of the outcome variable at the time point following (i.e., $T + 1$) the final treatment A_T. We introduce additional data structures, in which O contains alternative constituent random variables, in the sequel as needed.

We assume that $O \sim \mathsf{P}_0$, where $\mathsf{P}_0 \in \mathcal{M}$ is the true and unknown data-generating law, assumed only to be contained in a non-parametric statistical model $\mathcal{M}$. Further, we assume that the n copies of O are obtained by i.i.d. sampling from P_0, and denote by P_n the empirical distribution of the study sample $O_1, \ldots, O_n$. At times, when the context is clear, we will for convenience use P_n also to denote the empirical mean—that is, $\mathsf{P}_n f := \frac{1}{n}\sum_{i=1}^{n} f(O)$ for some arbitrary function $f(\cdot)$ of the study data O. We let the mapping $\Psi(\cdot)$ denote a statistical target parameter—often, a functional that, under assumptions, identifies a causal estimand. The (statistical) target parameter is a mapping from the model $\mathcal{M}$ containing the data-generating law P_0 to an appropriate outcome space (e.g., $\mathbb{R}^d$), that is $\Psi : \mathcal{M} \to \mathbb{R}^d$. When context is clear, we will also suppress dependence of $\Psi(\cdot)$ on the law P_0, writing $\psi_0 := \Psi(\mathsf{P}_0)$ and, generically, $\psi := \Psi(\mathsf{P})$ for arbitrary $\mathsf{P} \in \mathcal{M}$.

As noted in the preceding section, very often, the parameter Ψ does not depend on the entire data-generating law P_0, but rather, may instead be expressed in terms of a nuisance function η, which itself depends on $\mathsf{P} \in \mathcal{M}$; this allows us to write $\Psi(\mathsf{P}) = \Psi(\eta)$ and, adopting this convention, we will, going forward, use η to denote nuisance functions that depend on P and appear in Ψ.

2.2 Riesz representer preliminaries

Consider a function $f \in \mathcal{H}$, where $\mathcal{H}$ is a Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$; for simplicity, we limit our discussion to Hilbert spaces with inner products mapping to $\mathbb{R}$. Call the functional $\Psi(\cdot)$ *linear* if $\Psi(f) + \Psi(g) = \Psi(f + g)$ and $\Psi(a \cdot f) = a \cdot \Psi(f)$ for all $f, g \in \mathcal{H}$ and scalars $a \in \mathbb{R}$, and call $\Psi(\cdot)$ *bounded* if, for all $f \in \mathcal{H}$, $|\psi(f)| < M\|f\|_{\mathcal{H}}$ for some positive scalar $M \in \mathbb{R}^+$ that is bounded (i.e., finite) $M < \infty$. For a bounded linear functional $\Psi(\cdot)$, the Riesz representation theorem guarantees that there exists a unique function α such that

$$\Psi(f) = \langle \alpha, f \rangle \; . \tag{2}$$

In Equation (2), α is often referred to as a *Riesz representer*.

When $\mathcal{H} \equiv L_2(\mathsf{P})$, the Hilbert space of square-integrable functions, the inner product with which the space is equipped denotes integration with respect to the probability measure $d\mathsf{P}$, that is,

$$\langle f, g \rangle = \int fg d\mathsf{P} \; .$$

This representation is often useful when studying a functional $\Psi(\eta)$ that integrates a nuisance function η over some unknown probability distribution $d\mathsf{P}^\star$. Consider a setting where integration with respect to the probability measure $d\mathsf{P}$ is possible (e.g., sampling from P is possible) but that $d\mathsf{P}^\star$ is not readily accessible (i.e., we do not observe any samples from the corresponding $\mathsf{P}^\star$). When the target parameter depends on $\mathsf{P}^\star$, we may use the Riesz representation theorem to re-express it in terms of P as follows:

$$\Psi(\eta) = \int \eta d\mathsf{P}^\star = \int \eta \frac{d\mathsf{P}^\star}{d\mathsf{P}} d\mathsf{P} = \int \alpha \eta d\mathsf{P} = \langle \alpha, \eta \rangle \; , \tag{3}$$

where $\alpha = dP^\star / dP$, the Riesz representer, is the Radon-Nikodym derivative facilitating a "change of measure." Recognizing this form can allow one to easily determine the Riesz representer of a functional of interest.

2.3 Riesz representers in semi-parametric efficiency theory

Often, we take as our goal the statistical estimation problem of developing estimators of a statistical estimand that is a functional of the true but unknown data-generating law P_0 only through a nuisance function η_0, so that we can express the target parameter $\Psi(\mathsf{P}_0)$ as $\Psi(\eta_0)$. As and example, consider the problem of estimating the statistical functional that identifies the counterfactual mean of the outcome Y under treatment contrast $A = a$ from a collection of study units, sampled i.i.d. from P_0, with the individual-level data structure $O = (L, A, Y)$. Application of the g-formula yields the plug-in estimand (Hernán & Robins, 2026) $\Psi(\mathsf{P}_0) = \Psi(\eta_0) = \mathsf{E}[\mathsf{E}[Y \mid A = a, L]]$, where the nuisance function is $\eta_0(A, L) = \mathsf{E}_0[Y \mid A, L]$. We will return to this example to build intuition.

Two decades ago, van der Laan & Rubin (2006) proposed a targeted maximum likelihood estimator (TMLE) that considered functionals of the entire data distribution P_0. More often, though, we are interested in functionals, as in the example above, for which the nuisance function η is itself a regression function, such as a conditional expectation function. In such instances, we might proceed by constructing an initial estimator η_n of η_0 via a data-adaptive or non-parametric regression procedure that minimizes an appropriately selected loss function (van der Laan et al., 2004; Dudoit & van der Laan, 2005; van der Vaart et al., 2006). For problems following such a structure, we can construct consistent and asymptotically normal (CAN) estimators of the target parameter by using the unique Riesz representer, which we may combine with the targeting procedures developed in the TMLE literature (van der Laan & Rose, 2011).

Debiasing property. Suppose our goal is to estimate the functional $\Psi(\eta_0)$ and that we obtain an initial esitmator η_n via some data-adaptive or non-parametric regression procedure. Then, our estimator $\psi_n := \Psi(\eta_n)$ will typically exhibit an asymptotically non-negligible bias, unless the nuisance estimator η_n converges to the corresponding nuisance function η_0 at $o_{\mathsf{P}}(n^{-1/2})$ (e.g., as commonly attained in parametric regression). The bias may be expressed in terms of the corresponding Riesz representer α and the difference between the nuisance estimator η_n and the nuisance function η_0:

$$\Psi(\eta_n) - \Psi(\eta_0) = \Psi(\eta_n - \eta_0) = \langle \alpha, \eta_0 - \eta_n \rangle \ .$$

Hence, if α is known, it can be used to construct an unbiased estimator $\Psi(\eta_n) + \langle \alpha, \eta_0 - \eta_n \rangle$, in which the Riesz representer is used to remove the bias term. Note that the unbiased estimator we describe is analogous to the well-studied one-step de-biased estimator (Bickel et al., 1993).

Double-robust property. When α is unknown, as may be more commonly the case, we can construct an estimator α_n of α_0 and use it to de-bias the initial estimator $\Psi(\eta_n)$. In this case, the resulting "second-order" bias will be:

$$\Psi(\eta_n) + \langle \alpha_n, \eta_0 - \eta_n \rangle - \Psi(\eta) = \langle \alpha_0 - \alpha_n, \eta_0 - \eta_n \rangle \ ,$$

where the bias will vanish if the product of the errors between (α_n, α_0) and (η_n, η_0) is $o_{\mathsf{P}}(n^{-1/2})$. As it arises in statistical estimation problems rooted in causal inference repeatedly, this structure has been termed a *doubly-robust* remainder. Applying this to our running example of the treatment-specific mean $\Psi(\eta_0) = E[E[Y \mid A = a, L]]$, the nuisance is $\eta_0(A, L) = E[Y \mid A, L]$ and Riesz representer is the IPW-weight $\alpha(A, L) = \mathbb{1}(A = a)/g(A = a \mid L)$ where g denotes the density of A; hence, the second-order bias can be observed to take the common form $\int(\alpha_0 - \alpha_n)(\eta_0 - \eta_n)dP$. This remainder will converge to zero if either the outcome model η_n or the inverse probability weights α_n are correctly specified (model double-robustness) or if their rate *product* decays faster than $1/\sqrt{n}$ (rate double-robustness).

Efficiency property. Again suppose α is known so that $\Psi(\eta_n) + \langle \alpha, \eta_0 - \eta_n \rangle$ is an unbiased estimator. The existence of a Riesz regressor implies pathwise differentiability of the target parameter (Newey, 1994; Hirshberg & Wager, 2021; Chernozhukov et al., 2022c). In order for the Riesz representer to exist, the functional must be bounded by the norm of η; that is, $\Psi(\eta) \leq M\|\eta\|^2_{L_2(\mathsf{P}_0)}$ for some finite positive constant M. Consequently, it is always possible to construct an estimator that reaches the semi-parametric efficiency bound, which is related to the unique efficient influence function in the non-parametric model $\mathcal{M}$. In the next section, we derive the efficient influence function used to construct such an efficient estimator.

3 Riesz representers for efficient estimation

Consider the setting of having observed data from a cohort study comprised of n study units $O_1, \ldots, O_n$, where the study units are sampled i.i.d. from P_0, that is $O \sim \mathsf{P}_0$ and $O = (L, A, Y)$. To construct an efficient estimator of ψ_0, we begin with the class of regular asymptotically linear (RAL) estimators of the estimand ψ_0; RAL estimators ψ_n take the form

$$\psi_n - \psi_0 = \frac{1}{n}\sum_{i=1}^{n} \phi(\mathsf{P}_0)(O_i) + o_{\mathsf{P}}(n^{-1/2}) \ ,$$

where ϕ is a mean-zero function called the *influence function* of the estimator ψ_n; in the non-parametric model $\mathcal{M}$, there is a unique influence function for the class of RAL estimators with variance matching the semi-parametric efficiency bound, called the *efficient influence function* (van der Laan & Rose, 2011). The double-bracket notation $\phi(P_0)(O)$ is taken to mean that ϕ is a function of the true data distribution P_0 (whose components will appear as nuisances to be estimated) and the observed data O.

Let us now consider the class of bounded linear functionals, so that the target parameter can be expressed $\Psi(\eta_0) = \mathsf{E}(h(O_i; \eta_0))$, where h denotes a transformation of the nuisance function η_0 consistent with the parameter definition. For example, returning to our running example, the estimand $\psi_0 = \mathsf{E}_{\mathsf{P}_0}[\mathsf{E}_{\mathsf{P}_0}[Y \mid A = a, L]]$ is a bounded linear functional $\psi_0 \equiv \Psi(\eta_0) = \mathsf{E}(h(O_i; \eta_0))$, where, as noted previously, $\eta_0 = \mathsf{E}_{\mathsf{P}_0}[Y \mid A, L]$ and, here, $h(\cdot)$ is the point-evaluation transformation yielding $h(O; \eta_0) = \mathsf{E}_{\mathsf{P}_0}[Y \mid A = a, L]$. To construct a RAL estimator of such a functional, we can use the representation above as a guide: an efficient RAL estimator will act as a solution to the EIF estimating equation, that is, $\mathsf{P}_0\phi(\mathsf{P})(O; \psi_n) = 0$ and have variance matching that of the EIF $\mathsf{P}_0[\{\phi(\mathsf{P})(O; \psi_0)]^2\}$. Several causal machine learning frameworks have been proposed for the construction of asymptotically efficient RAL estimators, including one-step de-biased estimation and targeted minimum loss-based estimation, all of which allow for an initial estimator of η_n to be obtained from data-adaptive or non-parametric regression methods, including machine learning procedures. All such frameworks require knowledge of the EIF for RAL estimators of the target parameter of interest. Riesz representation can be used to simplify the derivation of the EIF. We describe this next.

3.1 The single-step Riesz EIF

Let us first consider the case where the plug-in nuisance η may be estimated directly from the data.

Theorem 1 (Riesz EIF). *Let $\Psi(\eta) = E[h(O_i;\eta)]$ be a bounded linear functional of η with Riesz representer α. Suppose $\eta = \eta_{\mathsf{P}}$, a function of the data distribution P_0. The efficient influence function of $\Psi(\eta_0)$ takes the form*

$$\phi(\mathsf{P})(O) = h(O;\eta_{\mathsf{P}}) - \Psi(\eta_{\mathsf{P}}) + \int \alpha(O)\phi_\eta(\mathsf{P})(O)d\mathsf{P} \ ,$$

where $\phi_\eta(\mathsf{P})(O)$ denotes the efficient influence function of the nuisance parameter η.

Proof. We will use a pointwise contamination strategy (Hines et al., 2022), coupled with the Riesz representation identity. With the pointwise contamination approach, we construct a pathwise $\mathsf{P}_\varepsilon = \varepsilon\delta_O + (1-\varepsilon)\mathsf{P}$ through the data-generating law $\mathsf{P}_0 = \mathsf{P}_{\epsilon=0}$ and compute its pathwise derivative:

$$\begin{aligned}
\nabla_\varepsilon \psi(\eta_{\mathsf{P}_\varepsilon};\mathsf{P}_\varepsilon)\Big|_{\varepsilon=0} &= \int \nabla_\varepsilon h(O;\eta_{\mathsf{P}_\varepsilon}) d\mathsf{P}_\varepsilon\Big|_{\varepsilon=0} \\
&= \int \Big(\nabla_\varepsilon h(O;\eta_{\mathsf{P}_\varepsilon})\Big|_{\varepsilon=0}\Big) d\mathsf{P}_0 + \int h(O;f_{\mathsf{P}_0})\nabla_\varepsilon \mathsf{P}_\varepsilon\Big|_{\varepsilon=0} && \text{(Product rule)} \\
&= \int h(O;\nabla_\varepsilon \eta_{\mathsf{P}_\varepsilon})\Big|_{\varepsilon=0}\Big) d\mathsf{P}_0 + \int h(O_i;\eta_{\mathsf{P}_0})(\delta_O + d\mathsf{P}_0) \\
& && \text{(Linearity of } h \text{ and } \nabla_\varepsilon\text{; evaluate } \nabla_\varepsilon P_\varepsilon \text{ at } \varepsilon=0) \\
&= \int \alpha(X) \underbrace{\Big(\nabla_\varepsilon \eta_{\mathsf{P}_\varepsilon}(O)\Big|_{\varepsilon=0}\Big)}_{\substack{\text{pathwise derivative} \\ \text{of nuisance}}} dP + \underbrace{h(O;\eta_{P_0}) - \Psi(\eta_{P_0})}_{\substack{\text{pathwise derivative} \\ \text{of mean}}} \\
& && \text{(Apply Riesz representation and simplify)}
\end{aligned}$$

The proof concludes by noting that the pathwise derivative $\nabla_\varepsilon \eta_{\mathsf{P}_\varepsilon}(O)\Big|_{\varepsilon=0}$ equals the efficient influence function of the nuisance η_{P}. □

To summarize, we compute the EIF by calculating the pathwise derivative of our desired parameter using the product rule, and then applying the Riesz representation theorem. Our strategy was to delay use of the Riesz identity until we had obtained a form involving integration with respect to the true data-generating law $d\mathsf{P}_0$. Other than Riesz representation, the technique used to derive this result follows existing strategies from semi-parametric theory; chiefly, computation of a pathwise derivative (Kennedy, 2016, 2024; van der Laan & Rose, 2011). Its form mirrors familiar influence function results such as those for counterfactual means and average treatment effects.

The result generalizes Riesz-based EIF representations presented by previous authors such as Newey (1994); Hirshberg & Wager (2021) and Chernozhukov et al. (2022b) in the sense that any nuisance may be considered, not just conditional expectations. When η_{P} is a regression function, its EIF is often simple and known, or otherwise easily derivable. When the nuisance function η is a conditional expectation function $\overline{Q}(A,L) = \mathsf{E}_{\mathsf{P}}(Y \mid A, L)$, its EIF is known to be

$$\frac{\delta_{L,A}}{d\mathsf{P}(A,L)}[Y - \overline{Q}(A,L)] \ ,$$

where δ denotes a Dirac delta function that places all mass at the observed (L,A) and zero elsewhere. This was derived by Hines et al. (2022). Noting that integration over a Dirac delta function performs point evaluation, we have the following corollary:

Corollary 1 (Riesz EIF with conditional mean). *Let $\Psi(\overline{Q}) = \mathsf{E}_{\mathsf{P}}[\overline{Q}(A,L)]$, a bounded linear functional of the conditional mean function $\overline{Q}$. Then, the EIF of $\Psi(\overline{Q})$ is*

$$\phi(\mathsf{P})(O) = h(O;\overline{Q}) - \psi + \alpha(A,L)(Y - \overline{Q}(A,L)) \ .$$

This Riesz EIF recovers the form discussed at length by previous authors including Newey (1994); Hirshberg & Wager (2021); Chernozhukov et al. (2022b) and Williams et al. (2025). It mirrors familiar presentations of the EIF for parameters such as the treatment-specific mean, whose EIF admits the form

$$\phi(\mathsf{P})(O) = \overline{Q}(a,L) - \psi + \frac{1(A=a)}{g(A,L)}\Big(Y - \overline{Q}(A,L)\Big)$$

Though convenient when applied directly, Theorem 1 can become even more powerful when applied *recursively*. To assist in this effort, we introduce a preliminary lemma.

Lemma 1 (Conditional Riesz EIF). *Let $V \subset O$ denote some set of baseline covariates. Consider the "conditional" estimand $\Psi(\eta) = E[h(O;\eta) \mid V = v]$. The EIF of this estimand is*

$$\mathsf{E}[\alpha(O)\phi_\eta(\mathsf{P})(O) \mid V = v] + \frac{\delta_v}{d\mathsf{P}_V(v)}\Big(h(O;\eta_\mathsf{P}) - \mathsf{E}[h(O;\eta_\mathsf{P}) \mid V = v]\Big)$$

where $\phi_\eta(\mathsf{P})(O)$ denotes the EIF of the nuisance η, δ_v a Dirac delta function equal to 1 at v and 0 elsewhere, and $d\mathsf{P}_V$ the density of V.

Proof. Using the point mass contamination strategy, this results from applying the product rule to the EIF for a conditional mean functional, following Example 6 of Hines et al. (2022) to note

$$\begin{aligned}\nabla_\varepsilon \Phi(\eta_{\mathsf{P}_\varepsilon};\mathsf{P}_\varepsilon) = \mathsf{E}\Big[h\Big(O;\nabla_\varepsilon \eta_{\mathsf{P}_\varepsilon}\Big|_{\varepsilon=0}\Big) \mid V = v\Big] \\ + \frac{\delta_v}{d\mathsf{P}_{0,V}(v)}\Big(h(O;\eta_{\mathsf{P}_0}) - \mathsf{E}[h(O;\eta_{\mathsf{P}_0}) \mid V = v]\Big)\end{aligned}$$

and then applying the Riesz representation theorem to the first term, noting the pathwise derivative of $\eta_{\mathsf{P}_\varepsilon}$ at $\varepsilon = 0$ is its EIF. □

Unless V is discrete, this parameter is not pathwise differentiable on its own due to the presence of the Dirac delta (hence, we only employ Riesz representation for the first component under expectation). However, it will be a useful intermediate result for applying Theorem 1 recursively. We apply this lemma in the next two sections to derive Riesz EIFs for more complex parameters of interest.

3.2 Sequential representation of a Riesz EIF

Suppose now, rather than single time-point data, one instead observes a sequence of time-ordered variables, where the data unit may now be represented $O = (L_0, L_1, A_1, L_2, A_2, \ldots, L_T, A_T, Y)$. The prototypical example is a longitudinal study with time-varying treatment and time-varying treatment-confounder feedback (Robins, 1986; Hernán & Robins, 2026), but this also encompasses settings such as mediation. As before, we assume that we have an i.i.d. sample of n study units, $O_1, \ldots, O_n$, where each $O \sim \mathsf{P}_0$.

In such a setting, we can consider an estimand represented by a sequence of regression functions. For instance, updating our running example, a common goal in causal inference is to evaluate the counterfactual mean under a fixed treatment sequence, for example, $(A_1 = a, A_2 = a, \ldots, A_T = a)$, which indicates that treatment is set to $a \in \mathcal{A}$ at all time points $t = 1, \ldots, T$. The effect of such a treatment regime may be identified via the g-formula (Robins, 1986; Hernán & Robins, 2026) and leads to an estimand of the form $\Psi(P_0) =$

$$\mathsf{E}[\mathsf{E}[\cdots \mathsf{E}[\mathsf{E}[Y \mid \bar{A}_T = \bar{a}_T, \bar{L}_T] \mid \bar{A}_{T-1} = \bar{a}_{T-1}, \bar{L}_{T-1}] \cdots \mid A_1 = a, L_1, L_0]] \,,$$

where we use $\bar{Z}_t$ to denote the history of random variable Z up until time t, that is $\bar{Z}_t = (Z_1, \ldots, Z_t)$, and each expectation is taken over P_0. Because a conditional expectation is *itself* a linear functional, the corresponding EIF of such an estimand may be expressed in terms of sequential expressions involving Riesz representers, as elaborated upon next.

Theorem 2 (Sequential Riesz EIF). *Consider the bounded linear functional*

$$\Psi(\mathsf{P}) = \mathsf{E}_\mathsf{P}[h_1(\bar{A}_1, \bar{L}_1; \overline{Q}_1)]$$

where $\overline{Q}_t$ is a bounded linear functional defined sequentially such that, for $t = 1, \ldots, T$, we have

$$\overline{Q}_t(\bar{A}_t, \bar{L}_t) = \mathsf{E}_\mathsf{P}[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \overline{Q}_{t+1}) \mid \bar{A}_t, \bar{L}_t] \,,$$

with $\overline{Q}_T(\bar{A}_T, \bar{L}_T) := \mathsf{E}_\mathsf{P}(Y \mid \bar{A}_T, \bar{L}_T)$. Let α_t denote the Riesz representer for $\overline{Q}_{t+1}$ in the functional $\mathsf{E}_\mathsf{P}[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \overline{Q}_{t+1}) \mid \bar{A}_t, \bar{L}_t]$. Then, letting with $h_T(\bar{A}_{T+1}, \bar{L}_{T+1}; \overline{Q}_{T+1}) := Y$, the EIF of $\Psi(\mathsf{P})$ is $\phi(\mathsf{P})(O) =$

$$h_1(\bar{A}_1, \bar{L}_1; \overline{Q}_1) - \psi + \sum_{t=1}^{T} \prod_{k=1}^{t} \alpha_k(\bar{A}_k, \bar{L}_k)[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \overline{Q}_{t+1}) - \overline{Q}_t(\bar{A}_t, \bar{L}_t)] \,.$$

Proof. The idea will be to apply Lemma 1 recursively over a changing sequence of conditional probability distributions. Note that, by the sequential nature of the data structure, the probability measure of the data can be decomposed, letting $\bar{A}_0 = \emptyset$, as

$$d\mathsf{P} = d\mathsf{P}(Y \mid \bar{A}_T, \bar{L}_T) \prod_{t=0}^{T} d\mathsf{P}(A_t, L_t \mid \bar{A}_{t-1}, \bar{L}_{t-1})$$

To ease notational burden, this proof will employ the shorthand $d\mathsf{P}^t = d\mathsf{P}(A_t, L_t \mid \bar{A}_{t-1}, \bar{L}_{t-1})$. We will also denote the EIF of $\overline{Q}_t(\bar{A}_t, \bar{L}_t)$ as $\phi_{\overline{Q}_t}(\mathsf{P}_0)(O)$. The proof will consist of three derivations: (1) the EIF of the first step $t = 1$, (2) the EIF of the last step $t = T$, and (3) the EIF of intermediate steps.

(1) In the first base case at $t = 1$, Theorem 1 implies

$$\begin{aligned}\phi(\mathsf{P})(O) &= h_1(A_1, L_1, L_0; \overline{Q}_1) - \psi_0 \\ &\quad + \int \alpha_1(A_1, L_1, L_0)\phi_{\overline{Q}_1}(A_1, L_1, L_0) d\mathsf{P}_0(A_1, L_1, L_0)\end{aligned}$$

(2) In the second base case at $t = T$, the EIF of the final regression $\overline{Q}_T(\bar{A}_T, \bar{L}_T) := \mathsf{E}(Y \mid \bar{A}_T, \bar{L}_T)$ follows from that of a conditional mean:

$$\phi_{\overline{Q}_T}(\mathsf{P}_0)(O) = \frac{\delta_{\bar{A}_T, \bar{L}_T}}{d\mathsf{P}(\bar{A}_T, \bar{L}_T)}[Y - \overline{Q}_T(\bar{A}_T, \bar{L}_T)]$$

(3) In each intermediate step, applying Lemma 1, we have that the EIF of an arbitrary element $\overline{Q}_t$ of the sequence is

$$\begin{aligned}\phi_{\overline{Q}_t}(\mathsf{P})(O) = &\int \alpha_t(A_{t+1}, L_{t+1})\phi_\eta(\mathsf{P})(A_{t+1}, L_{t+1}) d\mathsf{P}^{t+1} + \\ &\frac{\delta_{\bar{A}_t, \bar{L}_t}}{\prod_{k=1}^{t} d\mathsf{P}^t}\Big(h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \eta_\mathsf{P}) - \mathsf{E}[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \eta_\mathsf{P}) \mid \bar{A}_t, \bar{L}_t]\Big).\end{aligned}$$

Finally, plug in the intermediate EIF (3) into the EIF (1) repeatedly until $t = T$, and plug in EIF (2) for the final step. The weights $\delta_{\bar{A}_t, \bar{L}_t} / \prod_{k=1}^{t} dP_t$ eliminate each integral operator by turning the integration into a point-evaluation at A_t, L_t at each step and canceling out the accumulating measures $\prod_{k=1}^{t} P_k$ □

Note that this EIF differs slightly from previous Riesz representer-based EIF expressions such as those presented by Williams et al. (2025) in that the final form involves a *product* of Riesz representers corresponding to each nested nuisance. Interestingly, each product is *itself* a Riesz representer for its corresponding nuisance function with respect to the entire functional. Viewing the products this way recovers the form presented by Williams et al. (2025). In the context of estimation for large T, in some cases it may be more numerically stable to estimate $\omega_t = \prod_{k}^{t} \alpha_k$ directly instead of estimating each individual α_k and multiplying them.

3.3 Riesz representation for more complex functionals

Beyond their use in obtaining efficient influence functions directly, the Riesz representation strategies of Theorems 1 and 2 can also prove useful for deriving *components* of an EIF for a functional that may not be linear. The simplest example of this can be seen for functionals such as a ratio of counterfactual means

$$\Psi(\mathsf{P}) = \frac{\mathsf{E}[\mathsf{E}(Y \mid A = a_1, L)]}{\mathsf{E}[\mathsf{E}(Y \mid A = a_2, L)]}$$

whose EIF and asymptotic distribution can be derived in terms of Riesz representers by first applying Theorem 1 to the numerator and denominator and then using the delta method. In general, this strategy can be easily applied to any functional $\psi = b(\psi_1, \ldots, \psi_k)$ where the components $\psi_1, \ldots, \psi_k$ are bounded linear functionals and b is differentiable and does not depend on the data distribution P.

A more advanced use case arises in the context of alternative sampling schemes or incomplete data applications. Suppose one would like to perform inference on "complete-case" data $X \sim \mathsf{P}_0^X$, but they only observe data $O = (V, \Delta, \Delta X) \sim \mathsf{P}_0$; this indicates X is only observed according to the sampling indicator $\Delta \in \{0, 1\}$ which may depend on covariates $V \subset X$. In this case, one can combine Theorems 1 and 2 to derive an EIF analogous to those discussed by Rose & van der Laan (2011).

Theorem 3. *Let $\Psi(\mathsf{P})$ be a bounded linear functional whose EIF under a "complete" set of data $X \sim \mathsf{P}^X$ is given as $\phi(\mathsf{P}^X)(O)$. Then, its EIF under the observed data $O \sim \mathsf{P}$ takes the form*

$$\phi(\mathsf{P})(O) = \mathsf{E}[\phi^{uc}(\mathsf{P}^X)(X) \mid \Delta = 1, V] - \psi$$
$$+ \alpha(\Delta, V)(\phi^{uc}(\mathsf{P}^X)(X) - \mathsf{E}[\phi^{uc}(\mathsf{P}^X)(X) \mid \Delta = 1, V])$$

where $\phi^{uc}(\mathsf{P}^X)(X)$ denotes the uncentered complete-data EIF $\phi(\mathsf{P}^X)(X) + \psi$.

Proof. This theorem follows by noting that

$$\psi = \mathsf{E}[\phi^{\text{uc}}(\mathsf{P})(X)] = \mathsf{E}(\mathsf{E}[\phi^{\text{uc}}(\mathsf{P}^X)(X) \mid \Delta = 1, V])$$

and applying Theorem 1 with $\eta(\Delta, V) = \mathsf{E}[\phi^{\text{uc}}(\mathsf{P}^X)(X) \mid \Delta, V]$ as the nuisance function to obtain

$$\phi(\mathsf{P})(O) = \mathsf{E}[\phi^{\text{uc}}(\mathsf{P}^X)(X) \mid \Delta = 1, V] - \psi_0$$
$$+ \int \alpha(\Delta, V)\phi_\eta(\mathsf{P})(\Delta, V) d\mathsf{P}(\Delta, V) \tag{4}$$

By Lemma 1, for some Riesz representer $\tilde{\alpha}$ (distinct from α) the nuisance EIF ϕ_η is given

$$\phi_\eta(\mathsf{P})(\Delta, V) = \mathsf{E}[\tilde{\alpha}(X)\phi_{\phi^{uc}}(\mathsf{P})(X) \mid \Delta, V]$$
$$+ \frac{\delta_{\Delta,V}}{d\mathsf{P}(\Delta, V)}\Big(\phi^{uc}(\mathsf{P}^X)(X) - \mathsf{E}[\phi^{uc}(\mathsf{P}^X)(X) \mid \Delta, V]\Big)$$

Since an EIF is a mean-zero projection, the EIF $\phi_{\phi^{uc}}$ is simply the centered version $\phi(\mathsf{P}^X)(X)$. Plugging in the expression for $\phi_\eta(\Delta, V)$ into the second term of Equation (4), we observe that

$$\int \alpha(\Delta, V)\mathsf{E}[\tilde{\alpha}(X)\phi(\mathsf{P}^X)(X) \mid \Delta, V] dP(\Delta, V) = 0 \qquad \text{(EIF is mean-zero)}$$

$$\int \alpha(\Delta, V)\frac{\delta_{\Delta,V}}{d\mathsf{P}(\Delta, V)}\Big(\phi^{uc}(\mathsf{P}^X)(X) - \mathsf{E}[\phi^{uc}(\mathsf{P}^X)(X) \mid \Delta, V]\Big) dP(\Delta, V)$$
$$= \alpha(\Delta, V)\Big(\phi^{uc}(\mathsf{P}^X)(X) - \mathsf{E}[\phi^{uc}(\mathsf{P}^X)(X) \mid \Delta, V]\Big)$$

(Cancel probability measures and integrate over Dirac delta)

The above steps follow by algebraic cancellation. Plugging these results into Equation (4) concludes the proof. □

4 Riesz TMLE

When ψ_0 is a linear functional, a TMLE can be constructed to exploit the Riesz representer. The key idea used by the TMLE framework is that *any* estimator solving the EIF estimating equation will be asymptotically efficient and doubly-robust. TMLE uses this to construct a plug-in estimator.

The "classical" approach to efficient estimation is to solve the estimating equation $\mathsf{P}_n\phi(P)(O_i; \psi) = 0$ directly for $\psi(\eta_0)$; this is called "one-step" estimation. However, one can go further. Suppose we have an initial estimator ψ_n (such as the plug-in estimator). A more general approach would be to propose some loss function $L(X; \varepsilon)$ that depends on one or more parameters ε such that for some vector v of known constants,

$$v^\top \nabla_\varepsilon L(O; \varepsilon)\Big|_{\varepsilon=0} = \frac{1}{n}\sum_{i=1}^{n} \phi(\mathsf{P})(O_i) \,. \tag{5}$$

In this situation, optimizing L with respect to ε is equivalent to solving $\nabla_\varepsilon L(X; \varepsilon)\Big|_{\varepsilon=0} = 0$, and by extension, solving $\mathsf{P}_n\phi(\mathsf{P})(O_i) = 0$, which produces an efficient estimator. If the estimating equation is not solved in one step, it can be iterated until convergence (van der Laan & Rubin, 2006).

We follow similar logic to van der Laan & Rubin (2006) and Gruber & van der Laan (2010) to construct a TMLE for Riesz-representable functionals by "fluctuating" around the Riesz representer. As a warm-up, suppose $O = (L, A, Y)$ and our EIF takes the form

$$\phi_{\text{EIF}}(O) = \underbrace{h(A, L; \overline{Q}) - \psi}_{\phi_1 \text{ Sample mean EIF}} + \underbrace{\alpha(A, L)[Y - \overline{Q}(A, L)]}_{\phi_2 \text{ Riesz-represented sub-EIF}}$$

We can construct a loss $L = L_1 + L_2$ whose pathwise derivatives equal each part. For the first part, let $\mathsf{P}_{\varepsilon_1} = (1 + \varepsilon_1\phi_1)\mathsf{P}$, and the loss $L_1(\mathsf{P}) = -\log(\mathsf{P})$, so that $L_1(\mathsf{P}_{\varepsilon_1}) = -\log\Big((1 + \varepsilon_1\phi_1)\mathsf{P}\Big)$. Taking the derivative, $\nabla_{\varepsilon_1} L_1(\phi_1)\Big|_{\varepsilon=0} = \frac{\phi_1\mathsf{P}}{(1+(0)\phi_1)\mathsf{P}} = \phi_1$, so optimizing this loss solves the first part of the estimating equation. Also, $\varepsilon = 0$ is the MLE of $\mathsf{E}(L_1(P_{\varepsilon_1}))$, because $\mathsf{E}(\phi_1) = 0$ since it represents the EIF of a sample mean, which must be mean-zero by definition of a score. So, it may be excluded from the loss, because its solution is known.

For L_2, we can define $\overline{Q}_{\varepsilon_2} = \overline{Q} - \varepsilon_2\alpha$, and a loss that, at $\varepsilon_2 = 0$, equals ϕ_2. For example, the squared loss satisfies

$$\nabla_{\varepsilon_2} L_2(O; \varepsilon)\Big|_{\varepsilon=0} = \nabla_{\varepsilon_2}\frac{1}{2}[Y - \overline{Q}(A, L) + \varepsilon_2\alpha(A, L)]^2\Big|_{\varepsilon_2=0} = \alpha(L, A)[Y - \overline{Q}(A, L)]$$

which is equivalent to ϕ_2. Since we used the squared loss, ε_2 can be estimated by simply fitting a linear regression with offset $\overline{Q}(A, L)$ to estimate Y. If L_2 were some other loss, we could fit a similar regression using that loss (for example, Gruber & van der Laan (2010) use log-loss for bounded outcomes). In general we will consider regressions of the form $\text{link}(Y) = \text{link}(\overline{Q}(A, L)) + \varepsilon\alpha(A, L)$

Now, suppose instead our EIF was instead defined as in Theorem 2. Our goal should be to construct a loss function $L = \sum_{t=1}^{T} L_t$ where the loss of each over a parameter ε_j will be minimized at the jth component of the recursive EIF. How can such a loss be constructed?

There are a few different ways to do this. One strategy is to construct a sequential TMLE analogous to Stitelman et al. (2011) or van der Laan & Gruber (2012). The algorithm proceeds as described in Algorithm 1. This algorithm provides a TMLE for estimating the general sequential regression functional $\Psi(\mathsf{P}) = \mathsf{E}_{\mathsf{P}}[h_1(\bar{A}_1, \bar{L}_1; \overline{Q}_1)]$, where $\overline{Q}_t$ is recursively defined $\overline{Q}_t(\bar{A}_t, \bar{L}_t) := \mathsf{E}[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \overline{Q}_{t+1}) \mid \bar{A}_t, \bar{L}_t]$, and the final regression is $\overline{Q}_T(\bar{A}_T, \bar{L}_T) := \mathsf{E}(Y \mid \bar{A}_T, \bar{L}_T)$. To estimate this parameter, one first fits each sequential regression $\overline{Q}_t$ in reverse time order, and then similarly estimates the Riesz representers. Then, each sequential regression is updated by constructing a univariate fluctuation model (each of whose outputs depend on the previous regression through the offset). Finally, predictions are produced from each fluctuated sequential regression to construct the final plug-in estimator ψ_n.

Algorithm 1: Riesz TMLE

1. Fit sequential regressions $\overline{Q}_T, \dots, \overline{Q}_1$ and Riesz representers $\alpha_1, \dots, \alpha_T$.
2. For $t = 1, \dots, T$, compute the weights $\omega_t(\bar{A}_t, \bar{L}_t) = \prod_{k=1}^{t} \alpha_k(\bar{A}_k, \bar{L}_k)$.
3. Set $h_T(\bar{A}_T, \bar{L}_T; \overline{Q}_{T,\hat{\varepsilon}_T}) = Y$
4. For $t = T - 1, \dots, 1$, fit 1-D parametric model $\overline{Q}_{t,\hat{\varepsilon}_t}$ that regresses $\text{link}[h_{t+1}(\bar{A}_{(t+1)}, \bar{L}_{(t+1)}, \overline{Q}_{(t+1),\hat{\varepsilon}_{(t+1)}})] = \text{link}[\overline{Q}_t(\bar{A}_t, \bar{L}_t)] + \varepsilon_t\omega_t$ and set $\overline{Q}_t = \overline{Q}_{t,\hat{\varepsilon}_t}$
5. The final TMLE update is the substitution estimator constructed using the plug-in $h_1(\bar{A}_1, \bar{L}_1; \overline{Q}_1^\star) = \text{link}[h_1(\bar{A}_1, \bar{L}_1; \overline{Q}_{1,\hat{\varepsilon}_1})]$, which takes the form
$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} h_1(\bar{A}_1, \bar{L}_1; \overline{Q}_1^\star)$$

In this TMLE, at each step, the score equation L_t is solved, meaning it solves the entire recursive estimating equation. It is doubly-robust in the sense that consistency follows from the consistent estimation of *either* all of the nuisance estimates of $\bar{Q}_1, \dots, \bar{Q}_T$ *or* all of the Riesz representers $\alpha_1, \dots, \alpha_T$, similar to van der Laan & Gruber (2012) or Díaz et al. (2021). Alternatively, one could also use a common fluctuation coefficient ε, as proposed in van der Laan & Gruber (2012), and iteratively update it, but this requires more computational resources. Since $\prod_t \alpha_t$ is itself a Riesz representer, one can also estimate ω_t directly to improve stability, rather than multiplying many potentially large numbers together.

An alternative strategy can achieve the slightly stronger condition of *sequential* double-robustness, which only requires either $\overline{Q}_t$ *or* α_t to be modeled correctly at a given step. This condition is explored by Luedtke et al. (2017) and Díaz et al. (2021). It can be achieved by fitting a fluctuation *function* $\hat{\varepsilon}(\bar{A}_t, \bar{L}_t)$ at each step instead of a single parameter in a 1-D parametric model. Following a nearly identical construction as Algorithm 4 of Luedtke et al. (2017), we can define a TMLE that is sequentially doubly robust as in Algorithm 2. This follows similarly to Algorithm 1, but instead of a 1-D parameter, ε_t is now a *function* estimated nonparametrically via an arbitrary learning algorithm, and every time a fluctuation is performed, one must update the ε_t for all of the prior time steps as well.

Algorithm 2: Sequentially Doubly-Robust Riesz TMLE

1. Fit Riesz representers $\alpha_1, \ldots, \alpha_T$.
2. Set $h_T(\bar{A}_T, \bar{L}_T; \overline{Q}_T) = Y$
3. For $t = T-1, \ldots, 1$:
   (a) Initialize $\overline{Q}_s^{s+1}(\bar{A}_t, \bar{L}_t) = 1/2$
   (b) For $s = t, \ldots, 1$, fit a nonparametric regression model $\overline{Q}_{t,\hat{\varepsilon}_t}^{s}$ that regresses link$[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \overline{Q}_{t+1})] =$ link$[\overline{Q}_{t+1}^{s+1}(\bar{A}_t, \bar{L}_t)] + \hat{\varepsilon}_t^s(\bar{A}_t, \bar{L}_t)$ with weights $\prod_{u=s+1}^{t} \alpha_u$ and then set $\overline{Q}_t^{s} = \overline{Q}_{t,\hat{\varepsilon}_t}^{s+1}$
   (c) Set $\overline{Q}_{t,\hat{\varepsilon}_t} = \overline{Q}_{t,\hat{\varepsilon}_t}^{1}$.
4. The final TMLE update is the substitution estimator constructed using the plug-in $h_1(\bar{A}_1, \bar{L}_1; \overline{Q}_1^\star) =$ link$[h(\bar{A}_1, \bar{L}_1; \overline{Q}_{1,\hat{\varepsilon}_1})]$, which takes the form
$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} h_1(\bar{A}_1, \bar{L}_{i1}; \overline{Q}_1^\star)$$

Finally, we can consider a TMLE for parameters satisfying Theorem 3, which we define formally in Algorithm 3. In this algorithm, the complete-data EIF is first computed on the second-phase sample, and is then learned as the outcome of a regression to predict on the first-phase data. This regression is fluctuated using the Riesz representer of the sampling structure as the clever covariate. This method differs from prior literature (Hejazi et al., 2020b; Rose & van der Laan, 2011); instead of targeting a particular nuisance, such as the conditional probability of second-phase sample inclusion, one instead tilts the conditional mean model to ensure agnosticism to the form of the Riesz representer. Some authors might consider this a "quasi-TMLE" because only the outer regression is targeted, but similar procedures can be constructed depending on the desired plug-in constraints (for example, see section 4.2 of Qiu et al. (2026) for an alternative construction that also targets the inner regression).

As TMLE estimators that solve the efficient score equation, the outputs of Algorithms 1, 2, and 3 will be asymptotically normal and achieve the semi-parametric efficiency bound for ψ, permitting the construct of Wald-style confidence intervals.

Algorithm 3: Two-phase sampling Riesz TMLE

1. Estimate the complete-data EIF $\phi^{\text{uc}}(\mathsf{P}_n^X)(X)$ and fit both the regression $\overline{Q}_{\text{obs}}(X, \Delta, V) = \mathsf{E}[\phi^{\text{uc}}(\mathsf{P}_n^X)(X) \mid \Delta, V]$ and the Riesz representer $\alpha(\Delta, V)$.
2. Fit a 1-D parametric model $\overline{Q}_{\text{obs},\hat{\varepsilon}}$ that regresses link$[\phi^{\text{uc}}(\mathsf{P}_n^X)(X)] =$ link$[\overline{Q}_{\text{obs}}(X, \Delta, V)] + \varepsilon\alpha(\Delta, V)$
3. Construct a substitution estimator using the plug-in $\overline{Q}_{\text{obs}}^\star(X, \Delta, V) =$ link$[\overline{Q}_{\text{obs},\hat{\varepsilon}}(X, \Delta = 1, V)]$, which takes the form
$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} \overline{Q}_{\text{obs}}^\star(X, \Delta, V)$$

5 Examples

The EIF given by Riesz representation in Theorem 2 encompasses a wide range of common estimands, especially in causal inference, to which our TMLE procedures may be applied. In this section, we show how this EIF appears in a variety of common statistical estimation problems.

5.1 Treatment-specific mean and quantile

Consider observing $i = 1, \ldots, n$ units of data (L, A, Y), where Y represents an outcome, A a treatment of interest, and L a set of confounders. One of the most basic parameters in causal inference is the treatment-specific mean:

$$\psi_0 = \mathsf{E}[\mathsf{E}(Y \mid A = a, L)]$$

Its EIF takes the form

$$\phi(\mathsf{P})(O) = \frac{\mathbb{1}(A = a)}{\mathsf{P}(A = a \mid L)}\Big(Y - \mathsf{E}(Y \mid A, L)\Big) + \mathsf{E}(Y \mid A = a, L) - \psi\ ,$$

which can easily be seen to follow the pattern of Corollary 1 with nuisances

$$\eta(X) = \mathsf{E}(Y \mid A, L); \qquad\qquad \alpha(X) = \frac{\mathbb{1}(A = a)}{\mathsf{P}(A = a \mid L)}$$

But, we can also use Theorem 1 to obtain results that extend beyond those obtained by previous literature; for example, when η is not a conditional expectation function. Now instead of being interested in a treatment-specific mean, suppose one wanted to study a treatment-specific *quantile*, such as a median (Firpo, 2007; Díaz, 2017; Sun et al., 2021). The average quantile effect can be expressed

$$E[Q^\tau(A = a, L)]$$

In this case, η is the conditional τ-quantile function $\eta(O) = Q^\tau(A, L)$. Following Appendix B of Hines et al. (2022), the EIF of the conditional quantile function is

$$\frac{\delta_{A,L}}{d\mathsf{P}(A, L)}\frac{\tau - \mathbb{1}(Y > Q^\tau(A, L))}{d\mathsf{P}(Q^\tau(A, L) \mid A, L)}\ ,$$

so that the EIF of an average quantile effect $\mathsf{E}[h(O; Q^\tau)]$ is

$$h(O; Q^\tau) - \psi + \alpha(A, L)\left(\frac{\tau - \mathbb{1}(Y > Q^\tau(A, L))}{d\mathsf{P}(Q^\tau(A, L) \mid A, L)}\right)\ .$$

The above has a quite similar form to the EIF of the treatment-specific mean, except it multiplies the Riesz representer not by the residuals of a conditional mean regression, but by a scaled derivative of the "pinball loss" typically optimized to fit a quantile regression.

5.2 Longitudinal Treatment Regimes

Efficient estimation for the longitudinal causal inference problem with a sequence of discrete treatments has been described by van der Laan & Gruber (2012) and Tran et al. (2019), and an approach for continuous treatments by Díaz et al. (2021). In the discrete treatment regime, we can consider as a parameter of interest the counterfactual mean under treatment regime $\bar{a}$. This estimand is a recursively defined sequence of conditional means identified from the g-formula:

$$\psi_0 = \mathsf{E}[\mathsf{E}[\ldots \mathsf{E}[\mathsf{E}[Y \mid \bar{A}_t, \bar{L}_t] \mid \bar{A}_t, \bar{L}_t] \ldots \mid A_1, L_1, L_0]]$$

Let $\overline{Q}_{T+1}(\bar{a}_{T+1}, \bar{L}_{T+1}) = Y$ and a given step in the sequence of nuisance regressions be $\overline{Q}_t(\bar{A}_t, \bar{L}_t) = \mathsf{E}[\overline{Q}_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}) \mid \bar{A}_t, \bar{L}_t]$. The EIF is given $\phi(\mathsf{P})(O) = \bar{Q}(a_1, L_1, L_0) - \psi_0 + \sum_{t=1}^T D_t$ where

$$D_t = \frac{\mathbb{1}(\bar{A}_t = \bar{a}_t)}{g_t(\bar{a}_t \mid \bar{L}_t)}\Big(\overline{Q}_{t+1}(\bar{a}_{t+1}, \bar{L}_{t+1}) - \overline{Q}_t(\bar{A}_t, \bar{L}_t)\Big) \tag{6}$$

with $g_t(\bar{a}_t \mid \bar{L}_t) = P(\bar{A} = \bar{a}_t \mid \bar{L}_t)$. From visual inspection, it is clear to see that this form satisfies the form of Theorem 2 with

$$\alpha_t(\bar{A}_t, \bar{L}_t) = \frac{\mathbb{1}(\bar{A}_{t-1} = \bar{a}_{t-1})}{g_t(\bar{a}_{t-1} \mid \bar{L}_{t-1})}$$

Here, each $\alpha_t(\cdot)$ represents the Radon-Nikodym derivative of the counterfactual density with respect to the observed density.

One can also consider a *modified treatment policy*. In this setting, rather than a binary contrast between $A = 1$ to $A = 0$, we consider an intervention that changes the natural value of A_t to some "intervened" value A_t^d. Díaz et al. (2021) describe this setting, which is very similar; the EIF is $\phi(\mathsf{P})(O) = \bar{Q}(A_1^d, L_1, L_0) - \psi_0 + \sum_{t=1}^{T} D_t$ but this time,

$$D_t = \frac{g_t^d(\bar{A}_t \mid \bar{L}_t)}{g_t(\bar{A}_t \mid \bar{L}_t)}\Big(\overline{Q}_{t+1}(\bar{A}_{t+1}^d, \bar{L}_{t+1}) - \overline{Q}(\bar{A}_t, \bar{L}_t)\Big) \tag{7}$$

where g_t^d represents the density of $\bar{A}_t^d$. From visual inspection, this is the same as the discrete case, but with a different Riesz representer

$$\alpha_t(X) = \frac{g^d(\bar{a}_{t-1} \mid \bar{L}_{t-1})}{g(\bar{a}_{t-1} \mid \bar{L}_{t-1})} .$$

Hence, despite being composed of complex sequences of nested parameters, the EIFs of relevant estimands in the longitudinal setting are easily obtained.

5.3 Mediation

In mediation analysis, one observes data units $O = (L, A, M, Y)$; the goal is to analyze the effect of the treatment A on the outcome Y through the effect of the mediating variable M. A common target estimand is the "natural direct effect" (NDE) (Robins & Greenland, 1992; Pearl, 2001), which represents the contrast

$$\psi = \underbrace{\mathsf{E}(\mathsf{E}[\mathsf{E}(Y \mid A = 1, M, L) \mid A = 0, L])}_{\theta} - \underbrace{\mathsf{E}(\mathsf{E}(Y \mid A = 0, M, L))}_{\text{Counterfactual mean}}$$

For brevity we will consider the M-functional θ, as the additional contribution to the EIF from the counterfactual mean is already known. The EIF of θ is given by Tchetgen Tchetgen & Shpitser (2012) as the expression

$$\begin{aligned}\phi_{\text{EIF}}(O) =& \frac{\mathbb{1}(A=1)}{P(A=1 \mid L)}\frac{P(M \mid A=0, L)}{P(M|A=1, L)}\Big(Y - \mathsf{E}(Y \mid A=1, M, L)\Big)\\ &+ \frac{\mathbb{1}(A=0)}{P(A=0|L)}(\mathsf{E}(Y \mid A=1, M, L) - \eta(1,0,L))\\ &+ \eta(1,0,L) - \theta\end{aligned}$$

which uses the nuisance parameter shorthand

$$\eta(1, 0, L) = \mathsf{E}[\mathsf{E}(Y \mid A = 1, M, L) \mid A = 0, L]$$

Seeing how this EIF fits the form of Theorem 2 requires a slight re-expression using two key identities. First, for any function $z(M)$, $\mathbb{1}(A = 0)/P(A = 0 \mid L)$ can be re-expressed under expectation like so:

$$\begin{aligned}\mathsf{E}\Big(\frac{\mathbb{1}(A=0)}{\mathsf{P}(A=0|L)}z(M)\Big) &= \int z(M)\frac{d\mathsf{P}(M=m, A=0, L=l)}{\mathsf{P}(A=0 \mid L=l)}\\ &= \int z(M)d\mathsf{P}(M=m \mid A=0, L)d\mathsf{P}(L=l)\\ &= \int \mathsf{E}(z(M) \mid A, L)\frac{d\mathsf{P}(M=m \mid A=0, L)}{d\mathsf{P}(M=m \mid A=a, L)}d\mathsf{P}(L=l)\\ &= E\Big(\frac{d\mathsf{P}(M=m \mid A=0, L)}{d\mathsf{P}(M=m \mid A, L)}\mathsf{E}(z(M) \mid A, L)\Big)\end{aligned}$$

Second, note that $\mathbb{1}(A = 1)z(1) = \mathbb{1}(A = 1)z(A)$. Using these two identities, the EIF can be re-expressed

$$\begin{aligned}\phi_{\text{EIF}}(O) =& \frac{\mathbb{1}(A=1)}{\mathsf{P}(A=1 \mid L)}\frac{\mathsf{P}(M \mid A=0, L)}{\mathsf{P}(M \mid A, L)}\Big(Y - \mathsf{E}(Y \mid A, M, L)\Big)\\ &+ \frac{\mathsf{P}(M \mid A=0, L)}{\mathsf{P}(M \mid A, L)}(\mathsf{E}(Y \mid A=1, M, L) - \eta(1, A, L))\\ &+ \eta(1,0,L) - \theta\end{aligned}$$

Since θ is a "two-layer" conditional expectation, its EIF has two Riesz representers and two "sub-EIFs". From visual inspection, the Riesz representers, following from change-of-measure, are

$$\alpha_1(O) = \frac{P(M \mid A=0, L)}{P(M \mid A=a, L)} \qquad \alpha_2(O) = \frac{\mathbb{1}(A=1)}{P(A=1|L)}$$

and the learnable nuisances are

$$\eta_1(O_1) = \eta(1, A, L) \qquad \eta_2(O_2) = \mathsf{E}(Y \mid A, M, L)$$

Hence, Theorem 2 holds for mediation with the nuisances above, and can be estimated with Algorithm 1. A similar strategy can be employed when mediation is expanded to other settings. For example, Theorem 1 of Zheng & van der Laan (2017) provides the natural direct effect's EIF under a longitudinal data structure, which can be cast in the form of this manuscript's Theorem 2 using the same logic as above. However, one must be careful, as some mediational functionals, such as those under stochastic intervention (Hejazi et al., 2022) may not be directly representable as sequences of conditional expectations, necessitating a more careful application of Theorem 1 instead.

6 Numerical Studies

We implemented our proposed Riesz TMLE algorithms along with one-step estimators in the R package {`RieszCML`}, available at `https://github.com/nshlab/RieszCML`. The package supports one-step and TMLE procedures, along with flexible nuisance function estimation via the Super Learner algorithm. Our sequential Riesz TMLE algorithm and Theorem 2 naturally support our package's ability for estimation in both the cross-sectional and longitudinal setting, as well as supporting estimation of novel "composed" parameters. A central design feature of the {`RieszCML`} package is the separation of (i) the definition of the target parameter through a Riesz representer and (ii) the estimation of nuisance components. In the following sections, we present some basic simulation studies demonstrating the comparable performance of this package to existing software.

6.1 Data Generating Process

To set up the simulations, we consider a single time point setting with a binary treatment A, covariates $W = (L_1, L_2, L_3, L_4, L_5)$, and a binary outcome Y. The target parameter is the ATE, $\psi_0 = \mathsf{E}[\overline{Q}(1, L) - \overline{Q}(0, L)]$, the risk difference between treatment and control. Covariates are generated as

$$L_1, L_2, L_3 \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad L_4 \overset{i.i.d.}{\sim} \text{Bernoulli}(0.5), \quad L_5 \overset{i.i.d.}{\sim} \text{Uniform}(-1,1) \,.$$

Treatment assignment follows a logistic model depending on main terms and interactions:

$$\mathsf{P}(A=1 \mid W) = \text{expit}(-0.4 + 0.6L_2 - 0.5L_3 + 0.5L_4L_5 - 0.4L_1L_2),$$

and the outcome regression is given by

$$\mathsf{P}(Y=1 \mid A=a, W) = \text{expit}(-0.8 + 0.9a + 0.6L_2 + 0.8L_4L_5 + 0.7aL_1 - 0.6aL_2).$$

This data generating process deliberately includes interactions and omits some covariates so that flexible machine learning methods such as the Super Learner algorithm are beneficial. The true ψ_0 and semi-parametric efficiency bound were approximated via a large sample Monte Carlo simulation with 2×10^6 draws from the DGP.

6.2 Comparison of Methods

We compared the Riesz TMLE to the standard TMLE method as implemented by the {`tmle3`} package. Both were coded to apply a logistic fluctuation to the initial plugin estimator so as to respect the binary nature of the Y outcomes.

We considered sample sizes ranging from $n = 30$ to 3000. For each sample, we performed 2,000 Monte Carlo replications. Both methods were implemented using super learning for the nuisance functions. For each estimator, we calculated estimated bias, coverage, mean squared error (MSE), and $\sqrt{n}$ times the MSE, summarized in Figure 1.

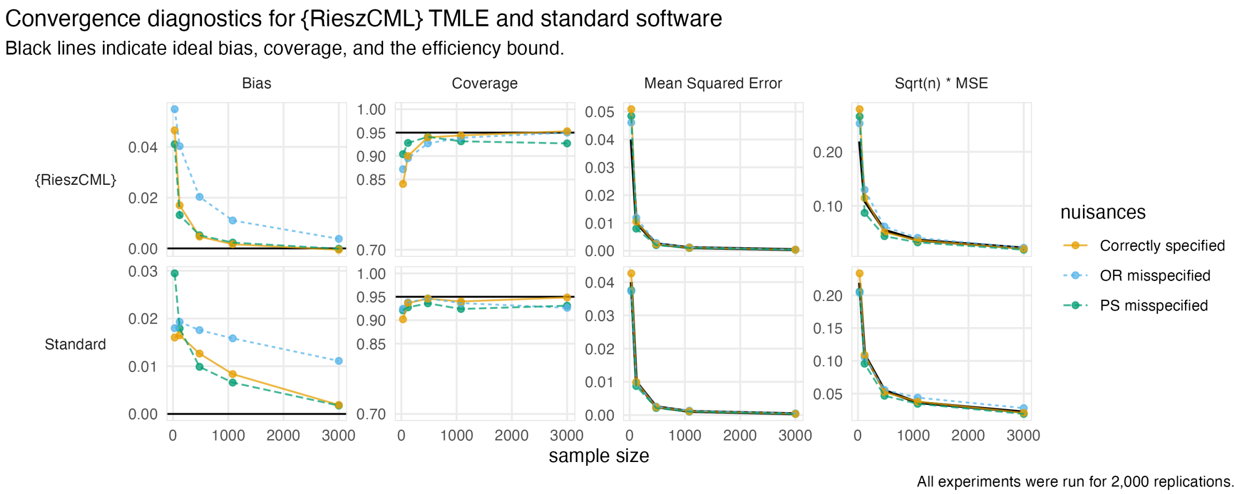

Figure 1: Convergence diagnostics for ATE estimators under nonlinear binary-treatment data generating process. Panels display bias, coverage, mean squared error (MSE) and $\sqrt{n} \times \text{MSE}$ as functions of sample size. Results are shown separately for the `RieszCML` and standard (`tmle3`) frameworks, and for three nuisance configurations: both correctly specified, outcome regression misspecified, and propensity score misspecified.

For the Riesz-based estimators, nuisance functions are estimated via `nadir::super_learner()`, while for `tmle3` we use `sl3::Lrnr_sl`. In both cases, the Super Learner library consisted of the `mean` learner (a learner that predicts a simple scalar mean/average every time), GLM, MARS, and glmnet, combined using a non-negative least squares metalearner. for nadir (citation to CRAN) and for sl3 (cite the website or TLVerse handbook)

To assess double robustness, we considered three nuisance specifications.

- Correctly specified: both $\overline{Q}_0(a, W)$ and $g_0(W)$ are estimated using a flexible super learner with a library of learners that can represent the true data generating process.
- Outcome regression misspecified: $\overline{Q}_0(a, W)$ is replaced with a constant mean estimator, while $g_0(W)$ is estimated flexibly as in the correctly specified case.
- Propensity score misspecified: $g_0(W)$ is replaced with a constant mean estimator, while $\overline{Q}_0(a, W)$ is estimated flexibly as in the correctly specified case.

These results allow us to verify that the Riesz TMLE (like the standard TMLE) retains consistency whenever either nuisance component is correctly specified. The Riesz TMLE and standard TMLE exhibit nearly identical performance across all settings. In particular, all estimators show small bias that decreases with sample size. Coverage is close to the nominal 95% level across all sample sizes and nuisance specifications. The MSE of both TMLE estimators closely tracks the semi-parametric efficiency bound, and $\sqrt{n} \times \text{MSE}$ stabilizes as expected.

Across all experiments, the Riesz TMLE performs comparably to the standard TMLE, demonstrating that the Riesz representation based construction yields a valid and efficient estimator in practice. Overall, the simulation study supports that the Riesz TMLE attains the same semi-parametric efficiency bound as the standard TMLE.

7 Application to two-phase sampling in the HVTN 505 Trial

In this illustrative data analysis, we re-analyze data from the HIV Vaccine Trials Network's (HVTN) 505 trial (Hammer et al., 2013) using the Riesz representation TMLE presented in Algorithm 3. This trial randomized 2504 HIV-negative participants 1-to-1 to receive an active vaccine or placebo. One goal of this trial was to examine how vaccine-induced immune responses—specifically, CD4+ and CD8+ polyfunctionality scores—induce protection against human immunodeficiency virus (HIV). These polyfunctionality scores were collected according to a two-phase case-control sampling procedure (Janes et al., 2017): blood was drawn at week 26 and immune responses measured from week 28 to month 24 for all HIV-1 cases, along with five controls matched based on BMI and race/ethnicity.

To understand the causal effects of such continuous-valued exposures, one can consider studying the counterfactual outcome that would have occurred under a *modified treatment policy* (MTP) (Haneuse & Rotnitzky, 2013); that is, a

δ-unit shift in the exposure variable for each unit. For instance, in the context of the HTVN 505 trial, one might be interested in estimating the counterfactual one-year risk of HIV-1 infection had the polyfunctionality scores of treated participants been increased or decreased by δ units. Such knowledge can help inform the role that such immune responses play in the efficacy of HIV vaccines.

Identification and semi-parametric efficiency theory for causal effects of modified treatment policies was developed under complete data by Muñoz & Van Der Laan (2012) and Haneuse & Rotnitzky (2013). Hejazi et al. (2020b) constructed TMLE procedures for MTPs under two-phase sampling and applied them to the HVTN polyfunctionality data using the `txshift` package in `R` (Hejazi & Benkeser, 2020). Constructing this TMLE and similar estimators requires considerably involved theory and an understanding of how to target and estimate particular conditional probabilities in the two-phase sampling procedure, and can suffer from instability. The application of Riesz representation theory, however, simplifies the procedure, thus motivating our re-analysis of the data.

To do this, we apply Theorem 3. The observed data follow $O = (L, \Delta, \Delta S, Y)$; the sampling indicator Δ satisfies

$$P(\Delta = 1 \mid Y, L) = \begin{cases} 1, & Y = 1 \\ \pi(L), & \text{otherwise} \end{cases}$$

The goal is to estimate the population intervention effect of a modified treatment policy $\mathsf{E}[\mathsf{E}(Y \mid S + \delta, L)]$ under the two-phase sampling. Applying Theorem 3 with the known EIF for an MTP effect, this EIF takes the form

$$\begin{aligned}\phi(\mathsf{P})(O) =& \mathsf{E}\Big(E[Y \mid S + \delta, L] + \alpha_2(S, L)[Y - \mathsf{E}(Y \mid S, L)] \mid \Delta = 1, Y, L\Big) - \psi_0 \\ &+ \alpha_1(\Delta, Y, L)\mathsf{E}[Y \mid S + \delta, L] - \alpha_1(\Delta, Y, L)\mathsf{E}\Big(\mathsf{E}[Y \mid S + \delta, L] \mid \Delta, Y, L\Big) \\ &- \alpha_1(\Delta, L, Y)\mathsf{E}\Big(\alpha_2(S, L)[Y - \mathsf{E}(Y \mid S, L)] \mid \Delta, Y, L\Big) \\ &+ \alpha_1(\Delta, Y, L)\alpha_2(S, L)[Y - \mathsf{E}(Y \mid S, L)] \; .\end{aligned}$$

where the Riesz representers are given by

$$\alpha_2(S, L) = \frac{dP(S - \delta \mid L)}{dP(S \mid L)} \qquad \qquad \alpha_1(\Delta, Y, L) = \frac{\mathbb{1}(\Delta = 1)}{P(\Delta = 1 \mid Y, L)}$$

Visually, this is evidently a quite complicated EIF, but by representing its form in terms of Riesz representers and regression functions, practical estimation procedures can be simplified by using Algorithm 3.

In our re-analysis, we estimated mean counterfactual risks of HIV-1 infection across a range of possible shifts in standardized polyfunctionality scores on the standard deviation scale within the strata of vaccinated participants. Estimates were produced using the TMLE procedure outlined in Algorithm 3 (see Figure 2). Except for the EIF regression $\overline{Q}_{\text{obs}}$, we reuse nuisance estimates computed by Hejazi et al. (2020b), including second-stage sampling probabilities and conditional densities of polyfunctionality scores used to manually construct first- and second-stage Riesz representers. To fit the regression $\overline{Q}_{\text{obs}}$, we used super learning (van der Laan et al., 2007b); our weighted ensemble of statistical learning models included a generalized additive model, random forest, gradient boosted trees, penalized and unpenalized generalized linear models with all 1st-order interaction terms, and both 0th and 1st-order highly adaptive lasso models (Benkeser & van der Laan, 2016; van der Laan, 2017; Hejazi et al., 2020a).

Our analysis reaches the same conclusion as Janes et al. (2017) and Hejazi et al. (2020b): that reductions in polyfunctionality scores increase the risk of HIV infection. For negative shifts, we obtain very similar point estimates as those shown in Figure 2 of Hejazi et al. (2020b), matching the downward trend. For positive shifts, our analysis shows this downward trend continuing, albeit at a much slower rate. It does deviate slightly from the weak upward trend in HIV risk at higher polyfunctionality scores observed by Hejazi et al. (2020b), but we posit that our method, by targeting a conditional expectation instead of a conditional sampling probability, actually serves to more effectively correct possible instabilities in the estimates of this previous work.

8 Discussion and Conclusions

In this work, we demonstrated how the Riesz representation theorem can be used as a shortcut for constructing a semi-parametric-efficient targeted minimum loss-based estimator (TMLE) for a broad class of statistical functionals. Estimands that can be expressed as a sequence of one or more linear functionals of nuisance functions share a common efficient influence function (EIF) consisting of a series of reweighted nuisance residuals. This representation permits construction of a TMLE based on sequential regressions, generalizing results from the longitudinal causal inference

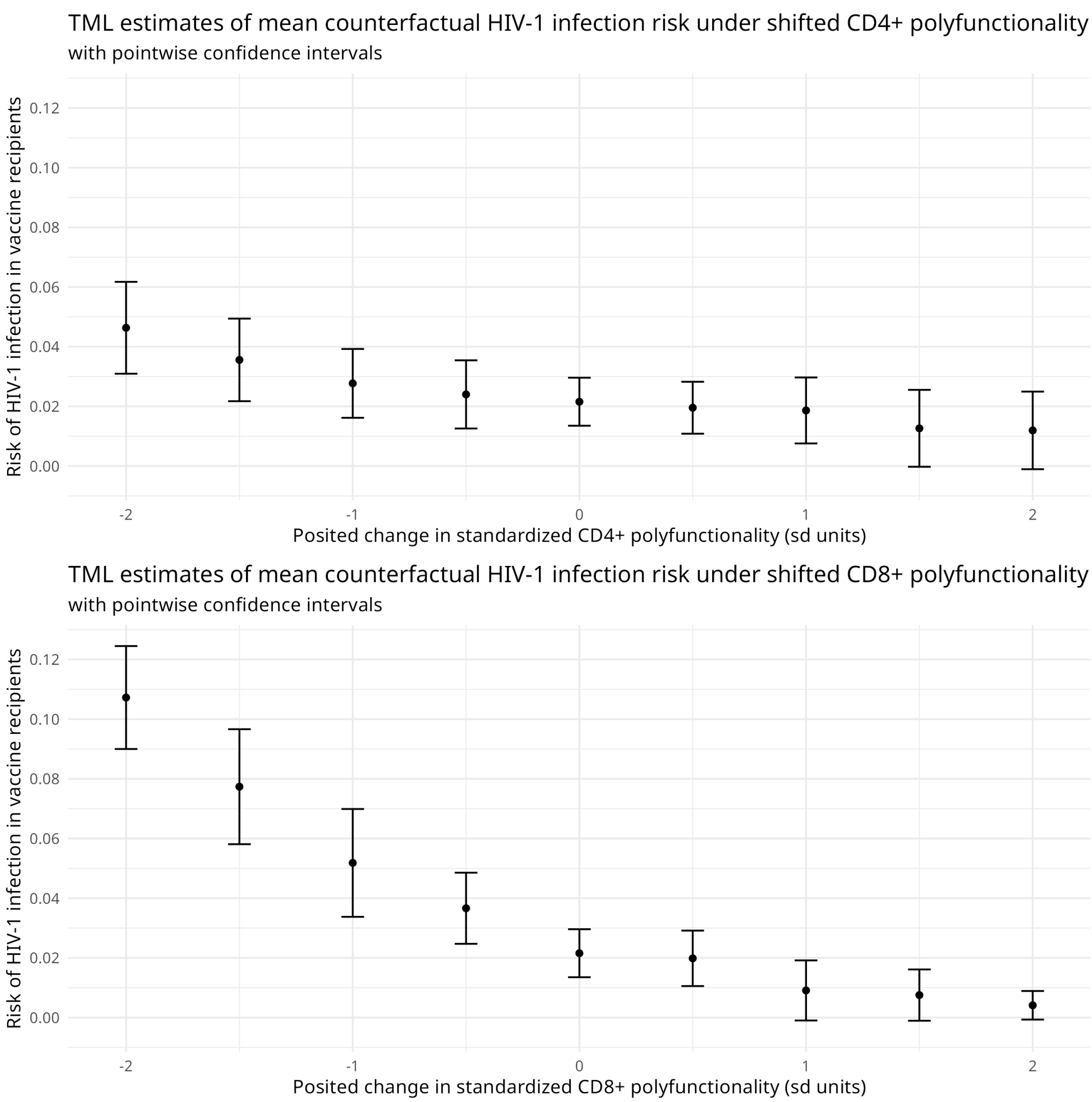

Figure 2: TML estimates of the counterfactual mean of HIV-1 infection in vaccine recipients under additive shifts in CD4+ (top) and CD8+ (bottom) standardized polyfunctionality scores, including pointwise Wald-style confidence intervals. These estimates compare favorably with the analogous estimates of Figure 2 in Hejazi et al. (2020b).

setting (van der Laan & Gruber, 2012; Luedtke et al., 2017; Díaz et al., 2021). Consequently, our work unifies efficient estimation across settings such as longitudinal data, mediation, subgroup analysis, quantile effects, and two-phase sampling.

One advantage of this theory is that it supports the development of statistical software that can be generalized across a wide variety of problems. Historically, specific semi-parametric estimation procedures in causal inference have been implemented piecemeal across a wide array of different packages (van der Laan et al., 2019). We implemented the proposed TMLE algorithm for Riesz-representable parameters as an R package, {`RieszCML`}; simulation studies demonstrated comparable performance to existing packages. Hence, our TMLEs allow investigators to avoid the labor-intensive nature of writing new software packages for every new estimand.

Agnosticism to how the Riesz representer nuisance is estimated is another well-documented advantage of our framework. Rather than deriving its form and estimating its components manually, several authors propose estimating the Riesz representer "automatically" using knowledge of h, by optimizing a particular squared loss. This approach is called *Riesz regression* (Chernozhukov et al., 2022b), and it can avoid numerical instability when the Riesz representer takes the form of an inverse probability or density. Gradient boosting, random forest, and neural network algorithms for Riesz regression have been proposed (Lee & Schuler, 2025; Chernozhukov et al., 2022a), though software implementations remain in their infancy (Williams et al., 2025).

Regardless of the nuisance estimation strategy, this work intends to simplify semi-parametric estimation across potentially complex estimands common in fields such as, but not limited to, causal inference. We hope that simplifying EIF derivation and providing some examples of highly general TMLE constructions will spur the application of targeted learning to new scientific application areas.

Acknowledgements

SVB acknowledges support from the National Science Foundation (award no. DGE 2140743).

References

BENKESER, D. & VAN DER LAAN, M. (2016). The highly adaptive lasso estimator. In *2016 IEEE international conference on data science and advanced analytics (DSAA)*. IEEE.

BICKEL, P. J., KLAASSEN, C. A., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press Baltimore.

BREIMAN, L. (1996). Stacked regressions. *Machine Learning* **24**, 49–64.

BREIMAN, L. (2001). Random forests. *Machine Learning* **45**, 5–32.

BREIMAN, L., FRIEDMAN, J., OLSHEN, R. A. & STONE, C. J. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC.

BRUNS-SMITH, D., DUKES, O., FELLER, A. & OGBURN, E. L. (2025). Augmented balancing weights as linear regression. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* .

CHERNOZHUKOV, V., NEWEY, W., QUINTAS-MARTINEZ, V. M. & SYRGKANIS, V. (2022a). Riesznet and forestriesz: Automatic debiased machine learning with neural nets and random forests. In *International Conference on Machine Learning*. PMLR.

CHERNOZHUKOV, V., NEWEY, W. K. & SINGH, R. (2022b). Automatic debiased machine learning of causal and structural effects. *Econometrica* **90**, 967—1027.

CHERNOZHUKOV, V., NEWEY, W. K. & SINGH, R. (2022c). Debiased machine learning of global and local parameters using regularized Riesz representers. *The Econometrics Journal* **25**, 576—601.

DÍAZ, I. (2017). Efficient estimation of quantiles in missing data models. *Journal of Statistical Planning and Inference* **190**, 39–51.

DUDOIT, S. & VAN DER LAAN, M. J. (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology* **2**, 131–154.

DÍAZ, I., WILLIAMS, N., HOFFMAN, K. L. & SCHENCK, E. J. (2021). Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association* **118**, 846—857.

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

FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2001). *The Elements of Statistical Learning*. Springer.

FRIEDMAN, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* , 1189–1232.

GRUBER, S. & VAN DER LAAN, M. J. (2010). A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. *The International Journal of Biostatistics* **6**.

HAMMER, S. M., SOBIESZCZYK, M. E., JANES, H., KARUNA, S. T., MULLIGAN, M. J., GROVE, D., KOBLIN, B. A., BUCHBINDER, S. P., KEEFER, M. C., TOMARAS, G. D., FRAHM, N., HURAL, J., ANUDE, C., GRAHAM, B. S., ENAMA, M. E., ADAMS, E., DEJESUS, E., NOVAK, R. M., FRANK, I., BENTLEY, C., RAMIREZ, S., FU, R., KOUP, R. A., MASCOLA, J. R., NABEL, G. J., MONTEFIORI, D. C., KUBLIN, J., MCELRATH, M. J., COREY, L. & GILBERT, P. B. (2013). Efficacy trial of a dna/rad5 hiv-1 preventive vaccine. *New England Journal of Medicine* **369**, 2083–2092.

HANEUSE, S. & ROTNITZKY, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine* **32**, 5260–5277.

HASTIE, T. J. & TIBSHIRANI, R. J. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association* **82**, 371–386.

HASTIE, T. J. & TIBSHIRANI, R. J. (1990). *Generalized Additive Models*, vol. 43. CRC Press.

HEJAZI, N. S. & BENKESER, D. (2020). txshift: Efficient estimation of the causal effects of stochastic interventions inr. *Journal of Open Source Software* **5**, 2447.

HEJAZI, N. S., COYLE, J. R. & VAN DER LAAN, M. J. (2020a). hal9001: Scalable highly adaptive lasso regression inr. *Journal of Open Source Software* **5**, 2526.

HEJAZI, N. S., RUDOLPH, K. E., VAN DER LAAN, M. J. & DÍAZ, I. (2022). Nonparametric causal mediation analysis for stochastic interventional (in)direct effects. *Biostatistics* **24**, 686—707.

HEJAZI, N. S., VAN DER LAAN, M. J., JANES, H. E., GILBERT, P. B. & BENKESER, D. C. (2020b). Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics* **77**, 1241–1253.

HERNÁN, M. & ROBINS, J. M. (2026). *Causal Inference: What If.* Chapman & Hall/CRC.

HINES, O., DUKES, O., DIAZ-ORDAZ, K. & VANSTEELANDT, S. (2022). Demystifying statistical learning based on efficient influence functions. *The American Statistician* **76**, 292–304.

HIRSHBERG, D. A. & WAGER, S. (2021). Augmented minimax linear estimation. *The Annals of Statistics* **49**.

HORVITZ, D. G. & THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

JANES, H. E., COHEN, K. W., FRAHM, N., DE ROSA, S. C., SANCHEZ, B., HURAL, J., MAGARET, C. A., KARUNA, S., BENTLEY, C., GOTTARDO, R., FINAK, G., GROVE, D., SHEN, M., GRAHAM, B. S., KOUP, R. A., MULLIGAN, M. J., KOBLIN, B., BUCHBINDER, S. P., KEEFER, M. C., ADAMS, E., ANUDE, C., COREY, L., SOBIESZCZYK, M., HAMMER, S. M., GILBERT, P. B. & MCELRATH, M. J. (2017). Higher t-cell responses induced by dna/rad5 hiv-1 preventive vaccine are associated with lower hiv-1 infection risk in an efficacy trial. *The Journal of Infectious Diseases* **215**, 1376–1385.

KENNEDY, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In *Statistical Causal Inferences and Their Applications in Public Health Research*, H. He, P. Wu & D.-G. D. Chen, eds. Springer, pp. 141–167.

KENNEDY, E. H. (2024). Semiparametric doubly robust targeted double machine learning: A review. In *Handbook of Statistical Methods for Precision Medicine*, E. Laber, B. Chakraborty, E. E. M. Moodie, T. Cai & M. J. van der Laan, eds. Chapman and Hall/CRC, pp. 207–236.

LEE, K. J. & SCHULER, A. (2025). Rieszboost: Gradient boosting for riesz regression. *arXiv* .

LUEDTKE, A. R., SOFRYGIN, O., VAN DER LAAN, M. J. & CARONE, M. (2017). Sequential double robustness in right-censored longitudinal models. *arXiv* .

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*. Routledge.

MUÑOZ, I. D. & VAN DER LAAN, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics* **68**, 541–549.

NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society* , 1349–1382.

PEARL, J. (2001). Direct and indirect effects. In *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence*.

PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

QIU, S., GRUBER, S., SHAW, P. A., WILLIAMSON, B. D. & VAN DER LAAN, M. J. (2026). Efficient targeted maximum likelihood estimators for two-phase design problems .

RIPLEY, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

ROBINS, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393–1512.

ROBINS, J. M. & GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–155.

ROSE, S. & VAN DER LAAN, M. J. (2011). A targeted maximum likelihood estimator for two-stage designs. *The International Journal of Biostatistics* **7**.

STITELMAN, O. M., DE GRUTTOLA, V., & VAN DER LAAN, M. J. (2011). A general implementation of TMLE for longitudinal data applied to causal inference in survival analysis. *UC Berkeley Division of Biostatistics Working Paper Series* .

SUN, S., MOODIE, E. E. M. & NEŠLEHOVÁ, J. G. (2021). Causal inference for quantile treatment effects. *Environmetrics* **32**.

TCHETGEN TCHETGEN, E. J. & SHPITSER, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis. *The Annals of Statistics* **40**.

TRAN, L., YIANNOUTSOS, C., WOOLS-KALOUSTIAN, K., SIIKA, A., VAN DER LAAN, M. & PETERSEN, M. (2019). Double robust efficient estimators of longitudinal treatment effects: Comparative performance in simulations and a case study. *The International Journal of Biostatistics* **15**.

VAN DER LAAN, M. (2017). A generally efficient targeted minimum loss based estimator based on the highly adaptive lasso. *The international journal of biostatistics* **13**, 20150097.

VAN DER LAAN, M., PHILLIPS, R., COYLE, J., HUBBARD, A., HEJAZI, N. & MALENICA, I. (2019). Targeted (machine) learning for real-world data science and causal inference with the tlverse software ecosystem. *hist* **25**, 75.

VAN DER LAAN, M. J., DUDOIT, S. & KELES, S. (2004). Asymptotic optimality of likelihood-based cross-validation. *Statistical Applications in Genetics and Molecular Biology* **3**.

VAN DER LAAN, M. J. & GRUBER, S. (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The International Journal of Biostatistics* **8**.

VAN DER LAAN, M. J., POLLEY, E. C. & HUBBARD, A. E. (2007a). Super learner. *Statistical Applications in Genetics and Molecular Biology* **6**.

VAN DER LAAN, M. J., POLLEY, E. C. & HUBBARD, A. E. (2007b). Super learner .

VAN DER LAAN, M. J. & ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer Series in Statistics. Springer New York.

VAN DER LAAN, M. J. & ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer.

VAN DER LAAN, M. J. & RUBIN, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics* .

VAN DER VAART, A. W., DUDOIT, S. & VAN DER LAAN, M. J. (2006). Oracle inequalities for multi-fold cross validation. *Statistics & Decisions* **24**, 351–371.

WILLIAMS, N. T., HINES, O. J. & RUDOLPH, K. E. (2025). Riesz representers for the rest of us .

WOLPERT, D. H. (1992). Stacked generalization. *Neural Networks* **5**, 241–259.

ZHENG, W. & VAN DER LAAN, M. J. (2017). Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. *Journal of Causal Inference* **5**.

ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association* **110**, 910—922.