A Riesz representer perspective on targeted learning

Salvador Balkus
sbalkus@g.harvard.edu
Harvard Biostatistics

Christian Testa
ctesta@hsph.harvard.edu
Harvard Biostatistics

Nima Hejazi
nhejazi@hsph.harvard.edu
Harvard Biostatistics

April 15, 2026

Causal Inference
Modus Operandi

1. Choose an estimand ψ and gather data $O_1, \ldots, O_n \sim \mathsf{P}$

2. Construct an estimator of ψ

1. Choose an estimand ψ and gather data $O_1, \ldots, O_n \sim \mathsf{P}$

2. Construct an estimator of ψ

1. Choose an estimand ψ and gather data $O_1, \ldots, O_n \sim \mathsf{P}$

2. Construct an estimator of ψ

1. Choose an estimand ψ and gather data $O_1, \ldots, O_n \sim \mathsf{P}$

2. Construct an estimator of ψ

Targeted Minimum Loss-Based Estimation

Even if P totally unknown (nonparametric), can construct "good" plug-in estimator of ψ by...

1. Deriving an "efficient influence function" $\phi(\mathsf{P})(O_i;\psi)$

2. **TMLE**: Choosing and optimizing loss $L(O,\varepsilon)$ satisfying, for some v,

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

→ choose L to respect problem constraints (i.e. bounded outcomes)

Targeted Minimum Loss-Based Estimation

Even if P totally unknown (nonparametric), can construct "good" plug-in estimator of ψ by...

1. Deriving an "efficient influence function" $\phi(\mathsf{P})(O_i;\psi)$

2. **TMLE**: Choosing and optimizing loss $L(O,\varepsilon)$ satisfying, for some v,

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

→ choose L to respect problem constraints (i.e. bounded outcomes)

Targeted Minimum Loss-Based Estimation

Even if P totally unknown (nonparametric), can construct "good" plug-in estimator of ψ by...

1. Deriving an "efficient influence function" $\phi(\mathsf{P})(O_i;\psi)$
2. **TMLE**: Choosing and optimizing loss $L(O,\varepsilon)$ satisfying, for some v,

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

→ choose L to respect problem constraints (i.e. bounded outcomes)

Targeted Minimum Loss-Based Estimation

Even if P totally unknown (nonparametric), can construct "good" plug-in estimator of ψ by...

1. Deriving an "efficient influence function" $\phi(\mathsf{P})(O_i;\psi)$
2. **TMLE**: Choosing and optimizing loss $L(O,\varepsilon)$ satisfying, for some v,

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

→ choose L to respect problem constraints (i.e. bounded outcomes)

> *The derivation of the efficient influence function is often regarded as somewhat of a "dark art."*
>
> Hines *et al.* (2022)

> *The derivation of the efficient influence function is often regarded as somewhat of a "dark art."*
>
> Hines *et al.* (2022)

TMLE.jl

Search docs (Ctrl + /)

R/`txshift`

Efficient Estimation of the Causal

R/`tmle3mediate`

Targeted Learning for Causal Mediation Analysis

Journal of Statistical Software

Information on Mission | Information for Authors | Style Guide | Volumes | About

Home / Archives / Vol. 81 (2017) / Issue 1

ltmle: An R Package Implementing Targeted Minimum Loss-Based Estimation for Longitudinal Data

Samuel D. Lendle, Joshua Schwab, Maya L. Petersen, Mark J. van der Laan

Authors: Nima Hejazi, James Duncan, David McCoy, and Mark van der Laan

Q: Can we derive a TMLE for a *general* class of estimands?

Estimand Name	**Formula for ψ**
Counterfactual mean (Static)	$\mathsf{E}[\mathsf{E}(Y \mid A = a, L)]$
Counterfactual mean (Stochastic)	$\mathsf{E}[\mathsf{E}(Y \mid A = A^\star, L)]$
Average quantile effect	$\mathsf{E}[Q_\tau(Y \mid A = 1, L) - Q_\tau(Y \mid A = 0, L)]$
Average treatment effect on treated	$\mathsf{E}(\mathsf{E}(Y \mid A = 1, L) - \mathsf{E}(Y \mid A = 0, L) \mid A = 1)$
Mediation formula	$\mathsf{E}[\mathsf{E}(\mathsf{E}(Y \mid A = 1, M, L) \mid A = 0, L)]$
Longitudinal treatment-specific mean	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{a}_T, \bar{L}_T) \ldots \mid A_1 = a_1, L_1, L_0)]$
Longitudinal modified treatment effect	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{A}_t^d, \bar{L}_T) \ldots \mid A_1 = A_1^d, L_1, L_0)]$

Estimand Name	**Formula for ψ**
Counterfactual mean (Static)	$\mathsf{E}[\mathsf{E}(Y \mid A = a, L)]$
Counterfactual mean (Stochastic)	$\mathsf{E}[\mathsf{E}(Y \mid A = A^{\star}, L)]$
Average quantile effect	$\mathsf{E}[Q_\tau(Y \mid A = 1, L) - Q_\tau(Y \mid A = 0, L)]$
Average treatment effect on treated	$\mathsf{E}(\mathsf{E}(Y \mid A = 1, L) - \mathsf{E}(Y \mid A = 0, L) \mid A = 1)$
Mediation formula	$\mathsf{E}[\mathsf{E}(\mathsf{E}(Y \mid A = 1, M, L) \mid A = 0, L)]$
Longitudinal treatment-specific mean	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{a}_T, \bar{L}_T) \ldots \mid A_1 = a_1, L_1, L_0)]$
Longitudinal modified treatment effect	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{A}^d_t, \bar{L}_T) \ldots \mid A_1 = A^d_1, L_1, L_0)]$

Estimand Name	**Formula for ψ**
Counterfactual mean (Static)	$\mathsf{E}[\mathsf{E}(Y \mid A = a, L)]$
Counterfactual mean (Stochastic)	$\mathsf{E}[\mathsf{E}(Y \mid A = A^{\star}, L)]$
Average quantile effect	$\mathsf{E}[Q_{\tau}(Y \mid A = 1, L) - Q_{\tau}(Y \mid A = 0, L)]$
Average treatment effect on treated	$\mathsf{E}(\mathsf{E}(Y \mid A = 1, L) - \mathsf{E}(Y \mid A = 0, L) \mid A = 1)$
Mediation formula	$\mathsf{E}[\mathsf{E}(\mathsf{E}(Y \mid A = 1, M, L) \mid A = 0, L)]$
Longitudinal treatment-specific mean	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{a}_T, \bar{L}_T) \ldots \mid A_1 = a_1, L_1, L_0)]$
Longitudinal modified treatment effect	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{A}_t^d, \bar{L}_T) \ldots \mid A_1 = A_1^d, L_1, L_0)]$

Estimand Name	**Formula for ψ**
Counterfactual mean (Static)	$\mathsf{E}[\mathsf{E}(Y \mid A = a, L)]$
Counterfactual mean (Stochastic)	$\mathsf{E}[\mathsf{E}(Y \mid A = A^{\star}, L)]$
Average quantile effect	$\mathsf{E}[Q_{\tau}(Y \mid A = 1, L) - Q_{\tau}(Y \mid A = 0, L)]$
Average treatment effect on treated	$\mathsf{E}(\mathsf{E}(Y \mid A = 1, L) - \mathsf{E}(Y \mid A = 0, L) \mid A = 1)$
Mediation formula	$\mathsf{E}[\mathsf{E}(\mathsf{E}(Y \mid A = 1, M, L) \mid A = 0, L)]$
Longitudinal treatment-specific mean	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{a}_T, \bar{L}_T) \ldots \mid A_1 = a_1, L_1, L_0)]$
Longitudinal modified treatment effect	$\mathsf{E}[\mathsf{E}(\ldots \mathsf{E}(Y \mid \bar{A}_t = \bar{A}_t^d, \bar{L}_T) \ldots \mid A_1 = A_1^d, L_1, L_0)]$

Estimand Name	**Formula for ψ**
Counterfactual mean (Static)	
Counterfactual mean (Stochastic)	
Average quantile effect	
Average treatment effect on treated	$\mathsf{E}[h(O;\eta)]$
Mediation formula	
Longitudinal treatment-specific mean	
Longitudinal modified treatment effect	

$$\mathrm{E}\left[\underset{\substack{\text{function}\\\text{"evaluator"}}}{h}\left(\overset{\text{data}}{O};\underset{\substack{\text{nuisance}\\\text{function}}}{\eta}\right)\right]$$

Riesz Representation

Riesz Representation Theorem (statistics version).

Suppose $\eta \in \mathcal{L}_2(\mathsf{P})$, and $\psi := \Psi(\eta) = \mathsf{E}[h(O;\eta)]$ is a bounded linear functional. Then, there exists a **Riesz representer** $\alpha \in \mathcal{L}_2(\mathsf{P})$ such that

$$\Psi(\eta) = \mathsf{E}[\alpha(O)\eta(O)]$$

Think of α like a "balancing weight"!

Theorem (Riesz EIF). The efficient influence function of $\mathsf{E}[h(O;\eta)]$ is

$$\phi(\mathsf{P})(O) = \underbrace{h(O;\eta) - \Psi(\eta)}_{\text{expected value EIF}} + \underbrace{\int \alpha(O)\phi_\eta(\mathsf{P})(O)d\mathsf{P}}_{\text{"reweighted nuisance bias"}}$$

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

*Generalizes previous work, like Hirshberg and Wager (**2021**), Chernozhukov et al. (**2022**), or Williams et al. (**2025**)*

Example 1: Counterfactual mean $\Psi(\eta) = \mathsf{E}[\mathsf{E}(Y \mid A = a, L)]$ where $\eta(A, L) = \mathsf{E}(Y \mid A, L)$. Its EIF is

$$\underbrace{\mathsf{E}(Y \mid A = a, L)}_{\substack{\text{evaluator} \\ h(A,L;\eta)}} - \Psi(\eta) + \underbrace{\frac{\mathbb{1}(A = a)}{d\mathsf{P}(A = a \mid L)}}_{\substack{\text{Riesz representer} \\ \alpha(A,L)}} \underbrace{(Y - \mathsf{E}(Y \mid A, L))}_{\substack{\text{derivative of} \\ \text{squared loss}}}$$

Integral cancels out because $\phi_\eta(\mathsf{P})(O) = \frac{\delta_{A,L}}{d\mathsf{P}(A,L)}(Y - \mathsf{E}(Y \mid A, L))$

Example 2: Counterfactual mean $\Psi(\eta) = \mathsf{E}[\mathsf{E}(Y \mid A = A + \delta, L)]$ of a policy setting $A = A + \delta$ where $\eta(A, L) = \mathsf{E}(Y \mid A, L)$. Its EIF is

$$\underbrace{\mathsf{E}(Y \mid A = A + \delta, L)}_{\substack{\text{evaluator} \\ h(A,L;\eta)}} - \Psi(\eta) + \underbrace{\frac{d\mathsf{P}(A - \delta \mid L)}{d\mathsf{P}(A \mid L)}}_{\substack{\text{Riesz representer} \\ \alpha(A,L)}} \underbrace{(Y - \mathsf{E}(Y \mid A, L))}_{\substack{\text{derivative of} \\ \text{squared loss}}}$$

Integral cancels out because $\phi_\eta(\mathsf{P})(O) = \frac{\delta_{A,L}}{d\mathsf{P}(A,L)}(Y - \mathsf{E}(Y \mid A, L))$

Example 3: Mean τ-th quantile $\Psi(\eta) = \mathsf{E}[Q^\tau(Y \mid A = a, L)]$ under treatment where $\eta(A, L) = Q^\tau(Y \mid A, L)$. Its EIF is

$$\underbrace{Q^\tau(Y \mid A = a, L)}_{\substack{\text{evaluator} \\ h(A,L;\eta)}} - \Psi(\eta) + \underbrace{\frac{\mathbb{1}(A = a)}{d\mathsf{P}(A = a \mid L)}}_{\substack{\text{Riesz representer} \\ \alpha(A,L)}} \underbrace{\left(\frac{\tau - \mathbb{1}(Y > Q^\tau(A, L))}{d\mathsf{P}(Q^\tau(A, L) \mid A, L)}\right)}_{\substack{\text{reweighted derivative of} \\ \text{"pinball loss"}}}$$

Integral cancels out because $\phi_\eta(\mathsf{P})(O) = \frac{\delta_{A,L}}{d\mathsf{P}(A,L)} \left(\frac{\tau - \mathbb{1}(Y > Q^\tau(A,L))}{d\mathsf{P}(Q^\tau(A,L) \mid A,L)}\right)$

Many more ways to use this

Consider a general time-ordered data structure

$$O = (L_1, A_1, \dots, L_T, A_T, Y)$$

Denote the histories at time t as $\bar{A}_t$ and $\bar{L}_t$. For example:

- Longitudinal data
- Mediation

Theorem (Sequential Riesz EIF)

Consider the estimand $\Psi(\eta_1) = \mathsf{E}_{\mathsf{P}}[h_1(A_1, L_1; \eta_1)]$, where η_t is a bounded linear functional defined sequentially such that, for $t = 1, \ldots, T$, we have

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

with $h_{T+1}(\bar{A}_{T+1}, \bar{L}_{T+1}; \eta_{T+1}) := Y$. Let α_t denote the Riesz representer for η_t in the functional $\mathsf{E}[h_t(\bar{A}_t, \bar{L}_t; \eta_t) \mid \bar{A}_{t-1}, \bar{L}_{t-1}]$. Then, the EIF of the estimand $\Psi(\eta_1)$ is

$$\underbrace{h_1(A_1, L_1; \eta_1) - \Psi(\eta_1)}_{\text{Expected value EIF}} + \sum_{t=1}^{T} \underbrace{\prod_{k=1}^{t} \alpha_t(\bar{A}_t, \bar{L}_t)}_{\substack{\text{Riesz}\\\text{representer}\\\text{reweighting}}} \underbrace{[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \eta_{t+1}) - \eta_t(\bar{A}_t, \bar{L}_t)]}_{\text{Residuals of sequential regressions}}$$

Theorem (Sequential Riesz EIF)

Consider the estimand $\Psi(\eta_1) = \mathsf{E}_\mathsf{P}[h_1(A_1, L_1; \eta_1)]$, where η_t is a bounded linear functional defined sequentially such that, for $t = 1, \dots, T$, we have

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

with $h_{T+1}(\bar{A}_{T+1}, \bar{L}_{T+1}; \eta_{T+1}) := Y$. Let α_t denote the Riesz representer for η_t in the functional $\mathsf{E}[h_t(\bar{A}_t, \bar{L}_t; \eta_t) \mid \bar{A}_{t-1}, \bar{L}_{t-1}]$. Then, the EIF of the estimand $\Psi(\eta_1)$ is

$$\underbrace{h_1(A_1, L_1; \eta_1) - \Psi(\eta_1)}_{\text{Expected value EIF}} + \sum_{t=1}^{T} \underbrace{\prod_{k=1}^{t} \alpha_t(\bar{A}_t, \bar{L}_t)}_{\substack{\text{Riesz}\\\text{representer}\\\text{reweighting}}} \underbrace{\left[h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}; \eta_{t+1}) - \eta_t(\bar{A}_t, \bar{L}_t)\right]}_{\text{Residuals of sequential regressions}}$$

Sequential TMLE

1. Fit sequential regressions $\eta_1, \ldots, \eta_T$ and Riesz representers $\alpha_1, \ldots, \alpha_T$.

Sequential TMLE

1. Fit sequential regressions $\eta_1, \dots, \eta_T$ and Riesz representers $\alpha_1, \dots, \alpha_T$.
2. For $t = 1, \dots T$, compute the weights $\omega_t(\bar{A}_t, \bar{L}_t) = \prod_{k=1}^{t} \alpha_k(\bar{A}_k, \bar{L}_k)$

Sequential TMLE

1. Fit sequential regressions $\eta_1, \dots, \eta_T$ 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. For $t = T-1, \dots, 1$, fit 1-D parametric model $\eta_{t,\hat{\varepsilon}_t}$ that regresses $\text{link}[\underbrace{h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}, \eta_{t+1,\hat{\varepsilon}_{t+1}})}_{\text{outcome } \eta_t \text{ was fitted on}}] = \text{link}[\underbrace{\eta_t(\bar{A}_t, \bar{L}_t)}_{\substack{\text{offset: original} \\ \text{regression}}}] + \underbrace{\varepsilon_t \omega_t(\bar{A}_t, \bar{L}_t)}_{\substack{\text{clever} \\ \text{covariate}}}$ and set $\eta_t = \eta_{t,\hat{\varepsilon}_t}$

Sequential TMLE

1. Fit sequential regressions $\eta_1, \dots, \eta_T$ and Riesz representers $\alpha_1, \dots, \alpha_T$.

2. For $t = 1, \dots T$, compute the weights $\omega_t(\bar{A}_t, \bar{L}_t) = \prod_{k=1}^{t} \alpha_k(\bar{A}_k, \bar{L}_k)$

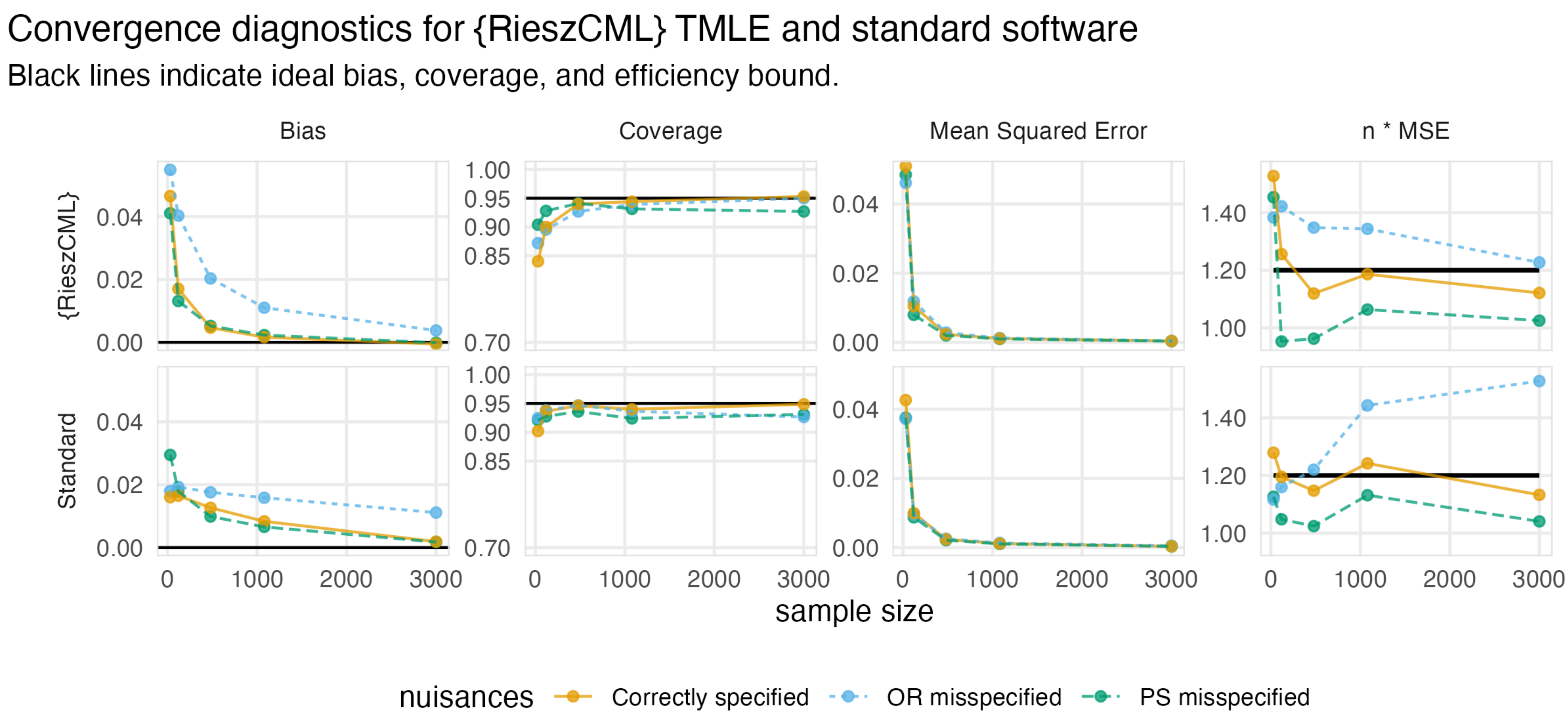
3. For $t = T-1, \dots, 1$, fit 1-D parametric model $\eta_{t,\hat{\varepsilon}_t}$ that regresses $\mathrm{link}[\underbrace{h_{t+1}(\bar{A}_{t+1}, \bar{L}_{t+1}, \eta_{t+1,\hat{\varepsilon}_{t+1}})}_{\text{outcome } \eta_t \text{ was fitted on}}] = \mathrm{link}[\underbrace{\eta_t(\bar{A}_t, \bar{L}_t)}_{\text{offset: original regression}}] + \underbrace{\varepsilon_t \omega_t(\bar{A}_t, \bar{L}_t)}_{\text{clever covariate}}$ and set $\eta_t = \eta_{t,\hat{\varepsilon}_t}$

4. The final TMLE is the updated plug-in estimator

$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} \mathrm{link}[h_1(\bar{A}_{i1}, \bar{L}_{i1}; \eta_{1,\hat{\varepsilon}_1})]$$

→ *Consistent, asymptotically normal, and semi-parametric efficient*

The `RieszCML` package: Simulations



All experiments were run for 2,000 replications.

Conclusions

- Unifies semi-parametric efficient estimation across…
  - → *Data*: Longitudinal, mediation, two-phase sampling, etc.
  - → *Interventions*: Binary, stochastic, derivatives, quantile effects, etc.
- Support theory and software re-use (*Lego bricks*)
- Agnostic about how α is learned; see, e.g. *Riesz regression* (Chernozhukov *et al.*, 2022)

Conclusions

- Unifies semi-parametric efficient estimation across…
  - → *Data*: Longitudinal, mediation, two-phase sampling, etc.
  - → *Interventions*: Binary, stochastic, derivatives, quantile effects, etc.

- Support theory and software re-use (*Lego bricks*)

- Agnostic about how α is learned; see, e.g. *Riesz regression* (Chernozhukov *et al.*, 2022)

Conclusions

- Unifies semi-parametric efficient estimation across...
  - → *Data*: Longitudinal, mediation, two-phase sampling, etc.
  - → *Interventions*: Binary, stochastic, derivatives, quantile effects, etc.
- Support theory and software re-use (*Lego bricks*)
- Agnostic about how α is learned; see, e.g. *Riesz regression* (Chernozhukov *et al.*, 2022)

Thank you! Questions?

Preprint

More about me

References

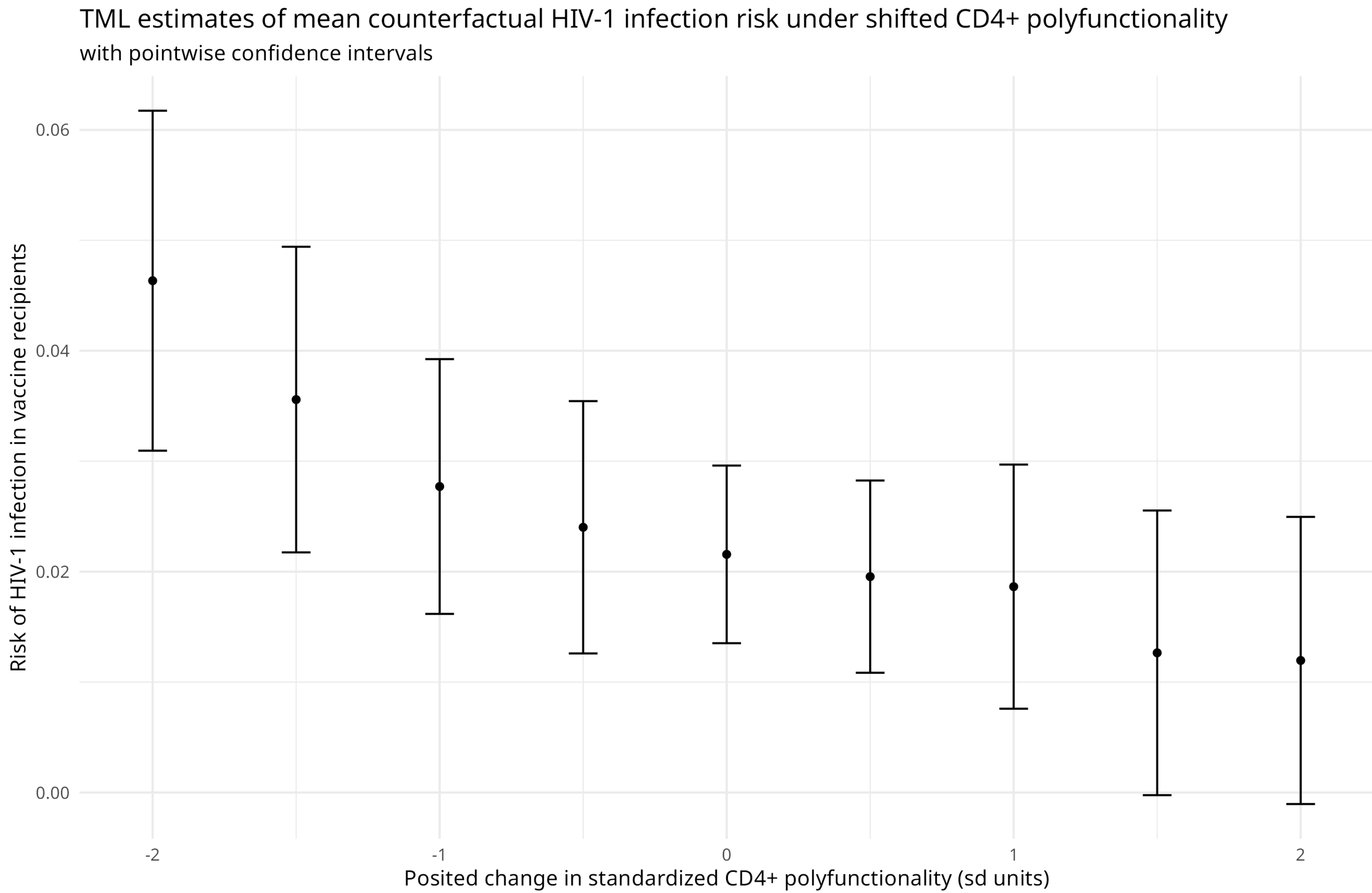
Chernozhukov, V., Newey, W. K. and Singh, R. (2022) Automatic debiased machine learning of causal and structural effects. *Econometrica*, **90**, 967--1027. The Econometric Society. DOI: 10.3982/ecta18515.

Hines, O., Dukes, O., Diaz-Ordaz, K., et al. (2022) Demystifying statistical learning based on efficient influence functions. *The American Statistician*, **76**, 292–304. Informa UK Limited. DOI: 10.1080/00031305.2021.2021984.

Hirshberg, D. A. and Wager, S. (2021) Augmented minimax linear estimation. *The Annals of Statistics*, **49**. Institute of Mathematical Statistics. DOI: 10.1214/21-aos2080.

Williams, N. T., Hines, O. J. and Rudolph, K. E. (2025) Riesz representers for the rest of us. arXiv. DOI: 10.48550/ARXIV.2507.19413.

Appendix A: Data Analysis



Appendix B: Functional analysis

Theorem (Riesz Representation, general).

Suppose $\eta \in \mathcal{H}$, a Hilbert space, and that $\psi(\eta) : \mathcal{H} \mapsto \mathbb{R}$ is a bounded linear functional.

Then, there exists $\alpha \in \mathcal{H}$ such that

$$\psi(\eta) = \langle \alpha, \eta \rangle$$