

Ensemble Learning Targeted Maximum Likelihood Estimation for Stata Users

Miguel Angel Luque-Fernandez, PhD

Assistant Professor of Epidemiology
Faculty of Epidemiology and Population Health
Department of Non-communicable Disease Epidemiology
Cancer Survival Group

<https://github.com/migariane/SUGML>

2017 London Stata Users Group meeting

September 7, 2017



Table of Contents

- 1 Causal Inference Background
- 2 ATE estimators
 - Estimators: Drawbacks
- 3 Targeted Maximum Likelihood Estimation
 - Why care about TMLE
 - TMLE road map
 - Non-parametric theory and empirical efficiency: Influence Curve
 - Machine learning: ensemble learning
- 4 Stata Implementation
 - Simulations
 - Links: online tutorial and GitHub open source eltmle
- 5 eltmle one sample simulation
- 6 Next steps
- 7 References



Rubin and Heckman

- This framework was developed first by statisticians (Rubin, 1983) and econometricians (Heckman, 1978) as a new approach for the estimation of **causal effects** from observational data.
- We will keep separate the **causal framework** (a conceptual issue briefly introduce here) and the "**how to estimate causal effects**" (an statistical issue also introduced here)



Causal effect

Potential Outcomes

We only observe:

$$Y_i(1) = Y_i(A = 1) \text{ and } Y_i(0) = Y_i(A = 0)$$

However we would like to know what would have happened if:

Treated $Y_i(1)$ would have been non-treated $Y_i(A = 0) = Y_i(0)$.

Controls $Y_i(0)$ would have been treated $Y_i(A = 1) = Y_i(1)$.

Identifiability

- How we can identify the effect of the potential outcomes Y^a if they are not observed?
- How we can estimate the expected difference between the potential outcomes $E[Y(1) - Y(0)]$, namely the ATE.

Notation and definitions

Observed Data

- Treatment **A**.
Often, $A = 1$ for treated and $A = 0$ for control.
- Confounders **W**.
- Outcome **Y**.

Potential Outcomes

- For patient i $Y_i(1)$ and $Y_i(0)$ set to $A = a$ $Y^{(a)}$, namely $A = 1$ and $A = 0$.

Causal Effects

- Average Treatment Effect: $E[Y(1) - Y(0)]$.

ASSUMPTIONS for Identification

- Rosebaum & Rubin, 1983: **The Ignorable Treatment Assignment** (A.K.A Ignorability, Unconfoundedness or Conditional Mean Independence).
- **POSITIVITY.**
- **SUTVA.**



Causal effect with OBSERVATIONAL data

IGNORABILITY

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp A_i \mid W_i$$

POSITIVITY

POSITIVITY: $P(A = a \mid W) > 0$ for all a, W

SUTVA

- We have assumed that there is **only one version of the treatment (consistency)** $Y(1)$ if $A = 1$ and $Y(0)$ if $A = 0$.
- The assignment to the treatment to one unit doesn't affect the outcome of another unit (**no interference**) or **IID** random variables.
- The model used to estimate the assignment probability has to **be correctly specified**.

G-Formula, (Robins, 1986)

G-Formula for the **identification** of the ATE with observational data

$$\begin{aligned} E(Y^a) &= \sum_y E(Y^a | W = w) P(W = w) \\ &= \sum_y E(Y^a | A = a, W = w) P(W = w) \text{ by consistency} \\ &= \sum_y E(Y = y | A = a, W = w) P(W = w) \text{ by ignorability} \end{aligned}$$

The **ATE**=

$$\sum_w \left[\sum_y \mathbf{P}(Y = y | A = 1, W = w) - \sum_y \mathbf{P}(Y = y | A = 0, W = w) \right] \mathbf{P}(W = w)$$

$$P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)$$

G-Formula, (Robins, 1986)

G-Formula for the identification of the ATE with observational data

The **ATE**=

$$\sum_w \left[\sum_y P(Y = y | A = 1, W = w) - \sum_y P(Y = y | A = 0, W = w) \right] P(W = w)$$

$$P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)$$

G-Formula

- The sums is generic notation. In reality, likely involves sums and integrals (we are just integrating out the W's).
- The **g-formula** is a **generalization of standardization** and allow to estimate unbiased treatment effect estimates.

Regression-adjustment

$$\widehat{ATE}_{RA} = N^{-1} \sum_{i=1}^N [E(Y_i | A = 1, W_i) - E(Y_i | A = 0, W_i)]$$

$$m_A(w_i) = E(Y_i | A_i = A, W_i)$$

$$\widehat{ATE}_{RA} = N^{-1} \sum_{i=1}^N [\hat{m}_1(w_i) - \hat{m}_0(w_i)]$$

IPTW (Inverse probability treatment weighting)

Survey theory (Horvitz-Thompson)

$$\hat{P}_i = E(A_i | W_i) ; \text{So , } \frac{1}{\hat{p}_i} , \text{if } A = 1 \text{ and , } \frac{1}{(1 - \hat{p}_i)} , \text{if } A = 0$$

Average over the total number of individuals

$$\widehat{ATE}_{IPTW} = N^{-1} \sum_{i=1}^N \frac{A_i Y_i}{\hat{p}_i} - N^{-1} \sum_{i=1}^N \frac{(1 - A_i) Y_i}{(1 - \hat{p}_i)}$$

AIPTW (Augmented Inverse probability treatment weighting)

Solving Estimating Equations

$$\widehat{ATE}_{AIPTW} =$$

$$N^{-1} \sum_{i=1}^N [(Y(1) | A_i = 1, W_i) - (Y(0) | A_i = 0, W_i)] + \\ N^{-1} \sum_{i=1}^N \left(\frac{(A_i = 1)}{P(A_i = 1 | W_i)} - \frac{(A_i = 0)}{P(A_i = 0 | W_i)} \right) [Y_i - E(Y | A_i, W_i)]$$



ATE estimators

Nonparametric

- G-formula plug-in estimator (generalization of standardization).

Parametric

- Regression adjustment ([RA](#)).
- Inverse probability treatment weighting ([IPTW](#)).
- Inverse-probability treatment weighting with regression adjustment ([IPTW-RA](#)) (Kang and Schafer, 2007).

Semi-parametric Double robust (DR) methods

- Augmented inverse-probability treatment weighting (Estimation Equations) ([AIPTW](#)) (Robins, 1994).
- Targeted maximum likelihood estimation ([TMLE](#)) (**van der Laan, 2006**).

Nonparametric

- Course of dimensionality (sparsity: zero empty cell)

Parametric

- Parametric models are misspecified (all models are wrong but some are useful, Box, 1976), and break down for high-dimensional data.
- (RA) Issue: extrapolation and biased if misspecification, no information about treatment mechanism.
- (IPTW) Issue: sensitive to course of dimensionality, inefficient in case of extreme weights and biased if misspecification. Non information about the outcome.

Double-robust (DR) estimators

Prons: Semi-parametric Double-Robust Methods

- DR methods give **two chances at consistency** if any of two nuisance parameters is consistently estimated.
- DR methods are **less sensitive** to course of dimensionality.

Cons: Semi-parametric Double-Robust Methods

- DR methods are unstable and inefficient if the propensity score (PS) is small (**violation of positivity assumption**) (vand der Laan, 2007).
- AIPTW and IPTW-RA do not respect the **limits of the boundary space of Y**.
- **Poor performance if dual misspecification** (Benkeser, 2016).

Targeted Maximum Likelihood Estimation (TMLE)

Pros: TMLE

- (TMLE) is a general algorithm for the construction of **double-robust**, **semiparametric** MLE, efficient **substitution** estimator (Van der Laan, 2011)
- Better performance than competitors has been largely documented (Porter, et. al., 2011).
- (TMLE) **Respects bounds on Y**, less sensitive to **misspecification** and to **near-positivity** violations (Benkeser, 2016).
- (TMLE) **Reduces bias** through **ensemble learning** if misspecification, even dual misspecification.
- For the ATE, **Inference** is based on the **Efficient Influence Curve**. Hence, the **CLT** applies, making inference easier.

Cons: TMLE

- The procedure is only available in R: **tmle** package (Gruber, 2011).

Targeted learning

Springer Series in Statistics

Targeted Learning

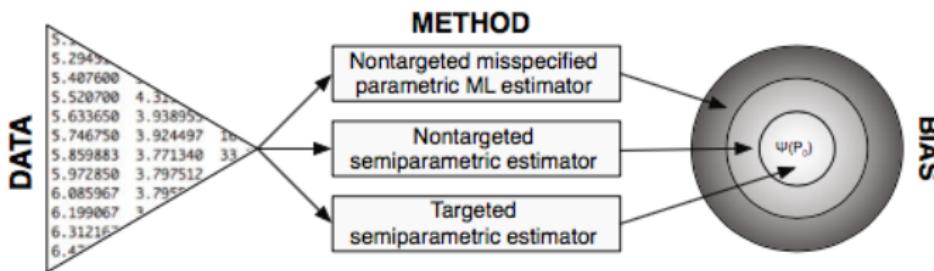
Causal Inference for Observational
and Experimental Data

 Springer

Source: Mark van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer Series in Statistics, 2011.



Why Targeted learning?



Source: Mark van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer Series in Statistics, 2011.

TMLE ROAD MAP

MC simulations: Luque-Fernandez et al, 2017 (in press, American Journal of Epidemiology)

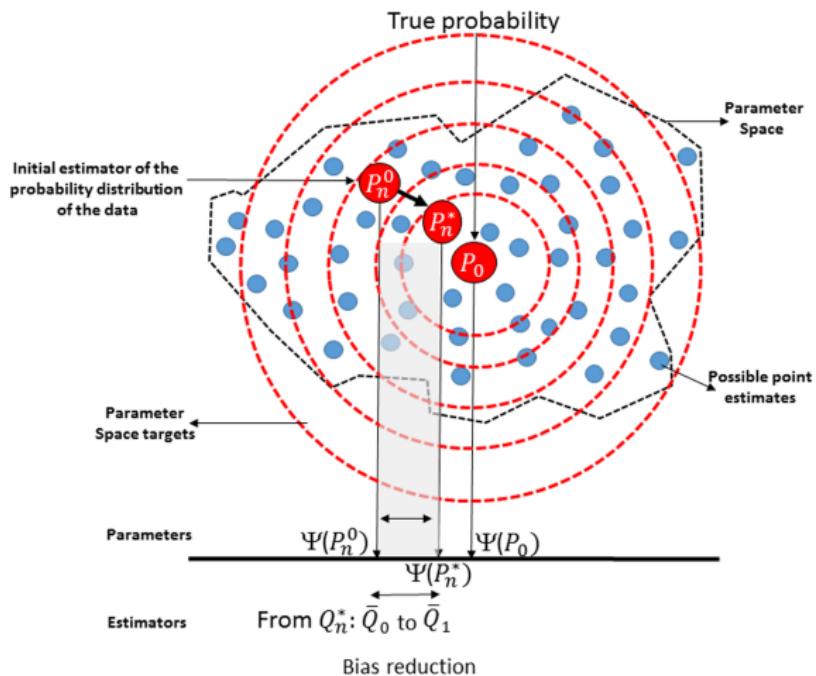
	ATE		BIAS (%)		RMSE		95%CI coverage (%)	
	N=1,000	N=10,000	N=1,000	N=10,000	N=1,000	N=10,000	N=1,000	N=10,000
First scenario* (correctly specified models)								
True ATE	-0.1813							
Naïve	-0.2234	-0.2218	23.2	22.3	0.0575	0.0423	77	89
AIPTW	-0.1843	-0.1848	1.6	1.9	0.0534	0.0180	93	94
IPTW-RA	-0.1831	-0.1838	1.0	1.4	0.0500	0.0174	91	95
TMLE	-0.1832	-0.1821	1.0	0.4	0.0482	0.0158	95	95
Second scenario ** (misspecified models)								
True ATE	-0.1172							
Naïve	-0.0127	-0.0121	89.2	89.7	0.1470	0.1100	0	0
BFit AIPTW	-0.1155	-0.0920	1.5	11.7	0.0928	0.0773	65	65
BFit IPTW-RA	-0.1268	-0.1192	8.2	1.7	0.0442	0.0305	52	73
TMLE	-0.1181	-0.1177	0.8	0.4	0.0281	0.0107	93	95

*First scenario : correctly specified models and near-positivity violation

**Second scenario: misspecification, near-positivity violation and adaptive model selection



TMLE ROAD MAP



Substitution estimation: $\hat{E}(Y | A, W)$

- First compute the outcome regression $E(Y | A, W)$ using the **Super-Learner** to then derive the Potential Outcomes and compute $\Psi^{(0)} = E(Y(1) | A = 1, W) - E(Y(0) | A = 0, W)$.
- Estimate the exposure mechanism $P(A=1|W)$ using the **Super-Learner** to predict the values of the propensity score.
- Compute $HAW = \left(\frac{\mathbb{I}(A_i=1)}{P(A_i=1|W_i)} - \frac{\mathbb{I}(A_i=0)}{P(A_i=0|W_i)} \right)$ for each individual, named the **clever covariate H**.



Fluctuation step: Epsilon

Fluctuation step ($\hat{\epsilon}_0$, $\hat{\epsilon}_1$)

- Update $\Psi^{(0)}$ through a fluctuation step incorporating the information from the exposure mechanism:

$$\mathbf{H}(1)\mathbf{W} = \frac{\mathbb{I}(A_i=1)}{\hat{P}(A_i=1|W_i)} \text{ and, } \mathbf{H}(0)\mathbf{W} = -\frac{\mathbb{I}(A_i=0)}{\hat{P}(A_i=0|W_i)}.$$

- This step aims to **reduce bias** minimising the mean squared error (MSE) for (Ψ) and considering the **bounds of the limits of Y**.
- The fluctuation parameters ($\hat{\epsilon}_0$, $\hat{\epsilon}_1$) are estimated using maximum likelihood procedures (in Stata):

```
. glm Y HAW, fam(binomial) nocons offset(E(Y| A, W))  
. mat e = e(b),  
. gen double ε = e[1, 1],
```

Targeted estimate of the ATE ($\widehat{\Psi}$)

$\Psi^{(0)}$ update using ϵ (epsilon)

$$\mathbf{E}^*(Y | A = 1, W) = \text{expit} [\text{logit}[E(Y | A = 1, W)] + \hat{\epsilon}_1 H_1(1, W)]$$

$$\mathbf{E}^*(Y | A = 0, W) = \text{expit} [\text{logit}[E(Y | A = 0, W)] + \hat{\epsilon}_0 H_0(0, W)]$$

Targeted estimate of the ATE from $\Psi^{(0)}$ to $\Psi^{(1)}$: $(\widehat{\Psi})$

$$\Psi^{(1)} : \widehat{\Psi} = [\mathbf{E}^*(Y(1) | A = 1, W) - \mathbf{E}^*(Y(0) | A = 0, W)]$$

TMLE inference: INFLUENCE CURVE

M-ESTIMATORS: Semi-parametric and Empirical processes theory

An estimator is **asymptotically linear** with **influence function φ (IC)** if the estimator can be **approximate by an empirical average** in the sense that

$$(\hat{\theta} - \theta_0) = \frac{1}{n} \sum_{i=1}^n (\text{IC}) + O_p(1/\sqrt{n})$$

(Bickel, 1997).

TMLE inference: Bickel (1993); Tsiatis (2007); Van der Laan (2011); Kennedy (2016)

- The **IC** estimation is a more general approach than M-estimation.
- The **Efficient IC** has mean zero $E(\text{IC}_{\hat{\psi}}(y_i, \psi_0)) = 0$ and **finite variance**.
- By the **Weak Law of the Large Numbers**, the **Op** converges to zero in a rate $1/\sqrt{n}$ as $n \rightarrow \infty$ (Bickel, 1993).
- The **Efficient IC** requires **asymptotically linear** estimators.

TMLE inference: Influence curve

TMLE inference

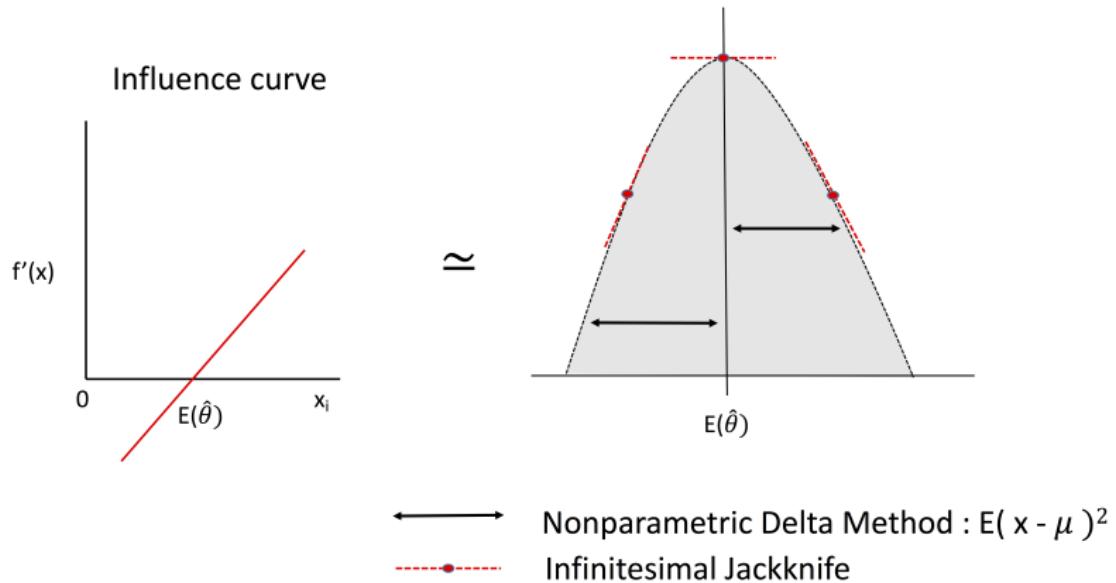
$$\text{IC} = \left(\frac{(A_i = 1)}{P(A_i = 1 | W_i)} - \frac{(A_i = 0)}{P(A_i = 0 | W_i)} \right) [Y_i - E_1(Y | A_i, W_i)] + \\ [E_1(Y(1) | A_i = 1, W_i) - E_1(Y(0) | A_i = 0, W_i)] - \psi$$

$$\text{Standard Error : } \sigma(\psi_0) = \frac{SD(IC_n)}{\sqrt{n}}$$

TMLE inference

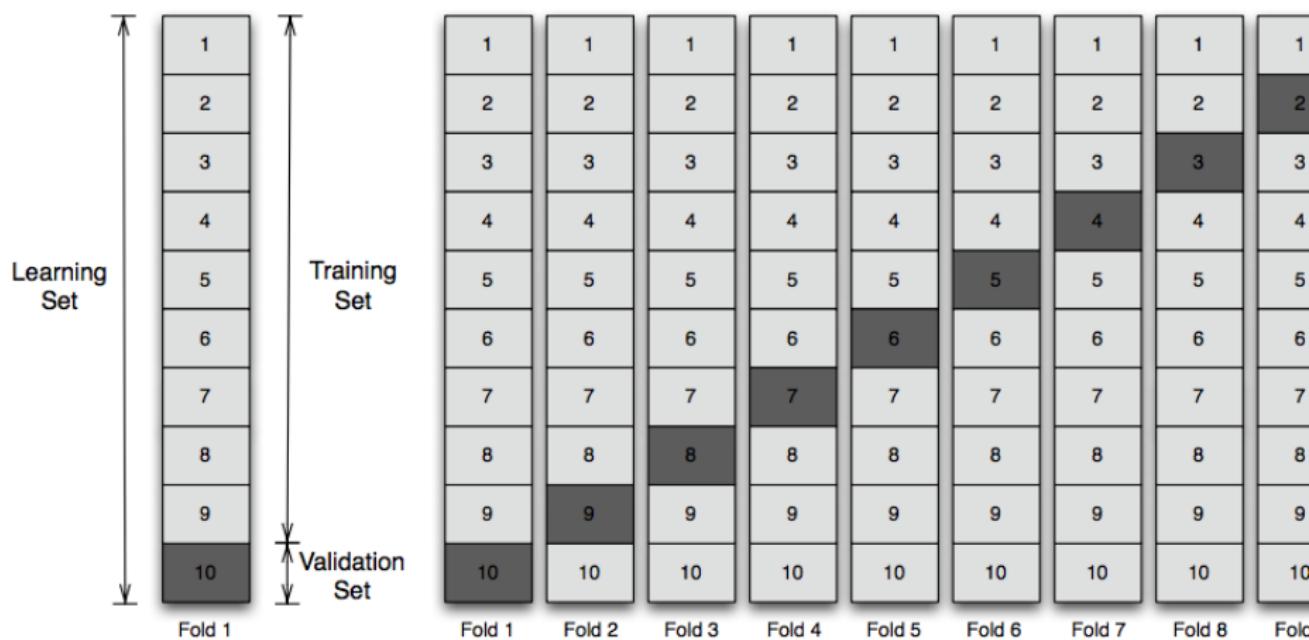
- The **Efficient IC**, first introduced by Hampel (1974), is used to apply readily the **CLT** for statistical inference using TMLE.
- The **Efficient IC** is the same as the infinitesimal jackknife and the **nonparametric delta method**. Also named the "**canonical gradient**" of the pathwise derivative of the target parameter ψ or "**approximation by averages**"(Efron, 1982).

IC: Geometric interpretation



Estimate of the ψ Standard Error using the efficient Influence Curve.
Image credit: Miguel Angel Luque-Fernandez

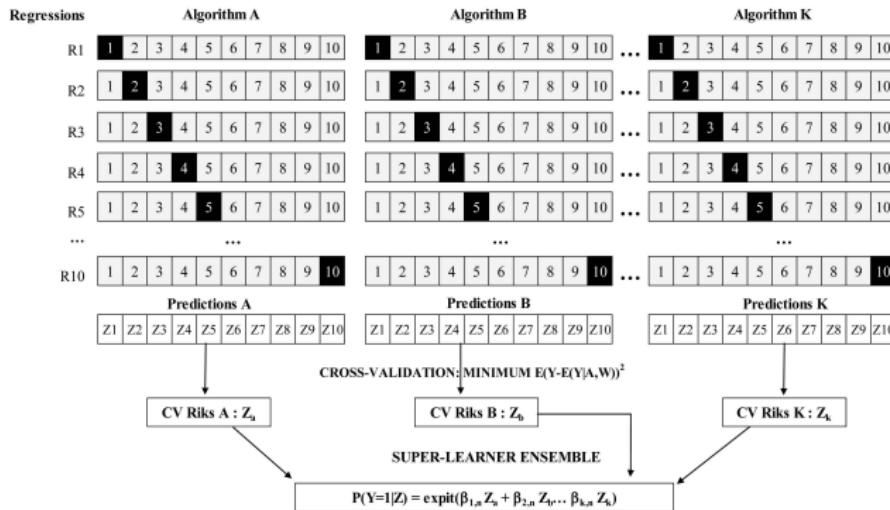
Targeted learning



Source: Mark van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer Series in Statistics, 2011.



Super-Learner: Ensemble learning



To apply the EIC we need data-adaptive estimation for both, the model of the outcome, and the model of the treatment.

Asymptotically, the final weighted combination of algorithms (Super Learner) performs as well as or better than the best-fitting algorithm (van der Laan, 2007).

Luque-Fernandez, MA. 2017. TMLE steps adapted from Van der Laa, 2011.

Ensemble Learning Targeted Maximum Likelihood Estimation

- **eltmle** is a Stata program implementing R-TMLE for the ATE for a binary or continuous outcome and binary treatment.
- **eltmle** includes the use of a **super-learner**(Polley E., et al. 2011).
- I used the default Super-Learner algorithms implemented in the base installation of the tmle-R package v.1.2.0-5 (Susan G. and Van der Laan M., 2007).
- i) stepwise selection, ii) GLM, iii) a GLM interaction.
- Additionally, **eltmle** users will have the option to include Bayes GLM and GAM.



Stata Implementation: overall structure

```
45
46 capture program drop eltmle
47 program define eltmle
48     syntax [varlist] [if] [pw] [, slaipw slaipwbgam tmle tmlebgam]
49     version 13.2
50     marksample touse
51     local var `varlist' if `touse'
52     tokenize `var'
53     local yvar = "`1'"
54     global flag = cond(`yvar'<=1,1,0)
55     qui sum `yvar'
56     global b = `r(max)'
57     global a = `r(min)'
58     qui replace `yvar' = (`yvar' - `r(min)') / (`r(max)' - `r(min)') if `yvar'>1
59     local dir `c(pwd)'
60     cd `dir'
61     qui export delimited `var' using "data.csv", nolabel replace
62     if "`slaipw'" == "" & "`slaipwbgam'" == "" & "`tmlebgam'" == "" {
63         tmle `varlist'
64     }
65     else if "`tmlebgam'" == "tmlebgam" {
66         tmlebgam `varlist'
67     }
68     else if "`slaipw'" == "slaipw" {
69         slaipw `varlist'
70     }
71     else if "`slaipwbgam'" == "slaipwbgam" {
72         slaipwbgam `varlist'
73     }
74 end
```

Stata Implementation: calling the SL

```
program tmle
// Write R Code dependencies: foreign Surperlearner
set more off
qui: file close _all
qui: file open rcode using SLS.R, write replace
qui: file write rcode ///
    `"set.seed(123)"' _newline ///
    `"list.of.packages <- c("foreign","SuperLearner")"' _newline ///
    `"new.packages <- list.of.packages[!(list.of.packages %in% installed.packages() [, "Package"])]"' _newline ///
    `"if(length(new.packages)) install.packages(new.packages, repos='http://cran.us.r-project.org')"' _newline ///
    `"library(SuperLearner)"' _newline ///
    `"library(foreign)"' _newline ///
    `"data <- read.csv("data.csv", sep=",")"' _newline ///
    `"attach(data)"' _newline ///
    `"SL.library <- c("SL.glm","SL.step","SL.glm.interaction")"' _newline ///
    `"n <- nrow(data)"' _newline ///
    `"nvar <- dim(data)[2]"' _newline ///
    `"Y <- data[,1]"' _newline ///
    `"A <- data[,2]"' _newline ///
    `"X <- data[,2:nvar]"' _newline ///
    `"W <- data[,3:nvar]"' _newline ///
    `"X1 <- X0 <- X"' _newline ///
    `"X1[,1] <- 1"' _newline ///
    `"X0[,1] <- 0"' _newline ///
    `"newdata <- rbind(X,X1,X0)"' _newline ///
    `"Q <- try(SuperLearner(Y = data[,1], X = X, SL.library=SL.library, family=binomial(), newX=newdata, method="method2"))"' _newline ///
    `"Q <- as.data.frame(Q[[4]])"' _newline ///
    `"QAW <- Q[1:n,]"' _newline ///
    `"QIW <- Q[((n+1):(2*n)),]"' _newline ///
    `"QOW <- Q[((2*n+1):(3*n)),]"' _newline ///
    `"g <- suppressWarnings(SuperLearner(Y = data[,2], X = W, SL.library = SL.library, family = binomial(), method = "method2"))"' _newline ///
    `"ps <- g[[4]]"' _newline ///
    `"ps[ps<0.025] <- 0.025"' _newline ///
    `"ps[ps>0.975] <- 0.975"' _newline ///
    `"data <- cbind(data,QAW,QIW,QOW,ps,Y,A)"' _newline ///
    `"write.dta(data, "data2.dta")"' _newline
qui: file close rcode
```

Stata Implementation: Batch file executing R

```
112 qui: file close rcode
113
114 // Write batch file to find R.exe path and R version
115 set more off
116 qui: file close _all
117 qui: file open bat using setup.bat, write replace
118 qui: file write bat ///
119 `"@echo off" _newline ///
120 `SET PATHROOT=C:\Program Files\R\``_newline ///
121 `echo Locating path of R..."`_newline ///
122 `echo."`_newline ///
123 `if not exist "%PATHROOT%" goto:NO_R``_newline ///
124 `for /f "delims=%" %%r in (' dir /b "%PATHROOT%R*" ') do ("`_newline ///
125 `echo Found %%r" `_newline ///
126 `echo shell "%PATHROOT%&rlbin\x64\R.exe" CMD BATCH SLS.R > runr.do" `_newline ///
127 `echo All set!"`_newline ///
128 `goto:DONE" `_newline ///
129 `")" `_newline ///
130 `:NO_R`_newline ///
131 `echo R is not installed in your system."`_newline ///
132 `echo."`_newline ///
133 `echo Download it from https://cran.r-project.org/bin/windows/base/"`_newline ///
134 `echo Install it and re-run this script" `_newline ///
135 `:DONE" `_newline ///
136 `echo."`_newline ///
137 `"pause"
138 qui: file close bat
139
140 //Run batch
141 shell setup.bat
142 //Run R
143 do runr.do
144
145 // Read Revised Data Back to Stata
146 clear
147 quietly: use "data2.dta", clear
148
149 // Q to logit scale
150 gen logQAW = log(QAW / (1 - QAW))
151 gen logQ1W = log(Q1W / (1 - Q1W))
152 gen logQ0W = log(Q0W / (1 - Q0W))
153
154 // Clever covariate HAW
```



Syntax eltmle Stata command

eltmle Y A W [, slapiw slapwbgam tmle tmlebgam]

Y: Outcome: numeric binary or continuous variable.

A: Treatment or exposure: numeric binary variable.

W: Covariates: vector of numeric and categorical variables.



Output for continuous outcome

```
.use http://www.stata-press.com/data/r14/cattaneo2.dta  
.eltmle bweight mbsmoke mage medu prenatal mmarrried, tmle
```

Variable	Obs	Mean	Std. Dev.	Min	Max
POM1	4,642	2832.384	74.56757	2580.186	2957.627
POM0	4,642	3063.015	89.53935	2868.071	3167.264
WT	4,642	-.0409955	2.830591	-6.644464	21.43709
PS	4,642	.1861267	.110755	.0372202	.8494988

ACE:

Additive Effect: -230.63; Estimated Variance: 600.93; p-value: 0.0000;
95%CI: (-278.68, -182.58)

Risk Differences:-0.0447; SE: 0.0047; p-value: 0.0000;
95%CI:(-0.05, -0.04)



Simulations comparing Stata ELTMLE vs R-TMLE

```
. mean psi aipw slaipw tmle
Mean estimation
Number of obs      = 1,000
-----
|   Mean
+-----+
    True |   .173
    aipw |   .170
    slaipw |   .170
  Stata-tmle |   .170
-----
R-TMLE |   .170
-----
```



ONLINE open free tutorial

Link to the tutorial

<https://migariane.github.io/TMLE.nb.html>

Stata Implementation: source code

<https://github.com/migariane/meltmle> for MAC users

<https://github.com/migariane/weltmle> for Windows users

Stata installation and step by step commented syntax

github install migariane/meltmle (For MAC users)

github install migariane/weltmle (For Windows users)

which eltmle

viewsource eltmle.ado



One sample simulation: TMLE reduces bias

<https://github.com/migariane/SUGML>



Next steps for ELTMLE

Next steps

- Stata Journal manuscript.
- Improving the user interface for **eltmle**.
- Include more machine learning algorithms.
- Implementation of Ensemble Learning in Stata (Super-Learner).
- Recently, we have implemented the **cross-validated AUC**:
<https://github.com/migariane/cvAUROC>. Also available at the **ssc** repository.



References

References

- ① Bickel, Peter J.; Klaassen, Chris A.J.; Ritov, Yaacov; Wellner Jon A. (1997). Efficient and adaptive estimation for semiparametric models. New York: Springer.
- ② Hamble, F.R., (1974). The influence curve and its role in robust estimation. *J Amer Statist Asso.* 69, 375-391.
- ③ Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Amer Statist Assoc.* 1994;89:846866.
- ④ Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics.* 2005;61:962972.
- ⑤ Tsiatis AA. Semiparametric Theory and Missing Data. Springer; New York: 2006
- ⑥ Kang JD, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science.* 2007;22(4):523539
- ⑦ Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology.* 1974;66:688701



References

References

- ① Luque-Fernandez, Miguel Angel. (2017). Targeted Maximum Likelihood Estimation for a Binary Outcome: Tutorial and Guided Implementation.
- ② StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.
- ③ Gruber S, Laan M van der. (2011). Tmle: An R package for targeted maximum likelihood estimation. UC Berkeley Division of Biostatistics Working Paper Series.
- ④ Laan M van der, Rose S. (2011). Targeted learning: Causal inference for observational and experimental data. Springer Series in Statistics.626p.
- ⑤ Van der Laan MJ, Polley EC, Hubbard AE. (2007). Super learner. Statistical applications in genetics and molecular biology 6.
- ⑥ Bickel, Peter J.; Klaassen, Chris A.J.; Ritov, Yaacov; Wellner Jon A. (1997). Efficient and adaptive estimation for semiparametric models. New York: Springer.
- ⑦ E. H. Kennedy. Semiparametric theory and empirical processes in causal inference. In: Statistical Causal Inferences and Their Applications in Public Health Research, in press.



Thank YOU

