Statin Simulation Report

Haodong Li

Jan, 2023

Simulation

DGD:

$$
\begin{aligned}
W_1 &\sim Unif(-1,1) \\
W_2 &\sim Unif(-1,1) \\
A &\sim Bernoulli(p) \text{ where } p = \text{expit}(0.1 * W_1 * W_2 - 0.4 * W_1) \\
\tau &= W_1^2 * (W_1 + 7/5) + (5 * W_2/3)^2 \\
\mu_Y &= A * \tau + W_1 * W_2 + 2 * W_2^2 - W_1 \\
Y &\sim N(\mu_Y, 1)
\end{aligned}
$$

Models:

1) GAM: General Additive Models (Correctly specified based on true DGD of Q and g)
2) Earth: Multivariate Adaptive Regression Splines
3) HAL: Highly Adaptive Lasso

CATE estimation

1) DR-learner: regress pseudo outcome estimates φ_n^0 on W.
2) T-learner: just use $\bar{Q}_n^0(1, W)$ - $\bar{Q}_n^0(0, W)$

Truncation:

$g_n \in [0.025, 0.975]$.

Figure 1: Main performance metrics (T-learner, no CV)

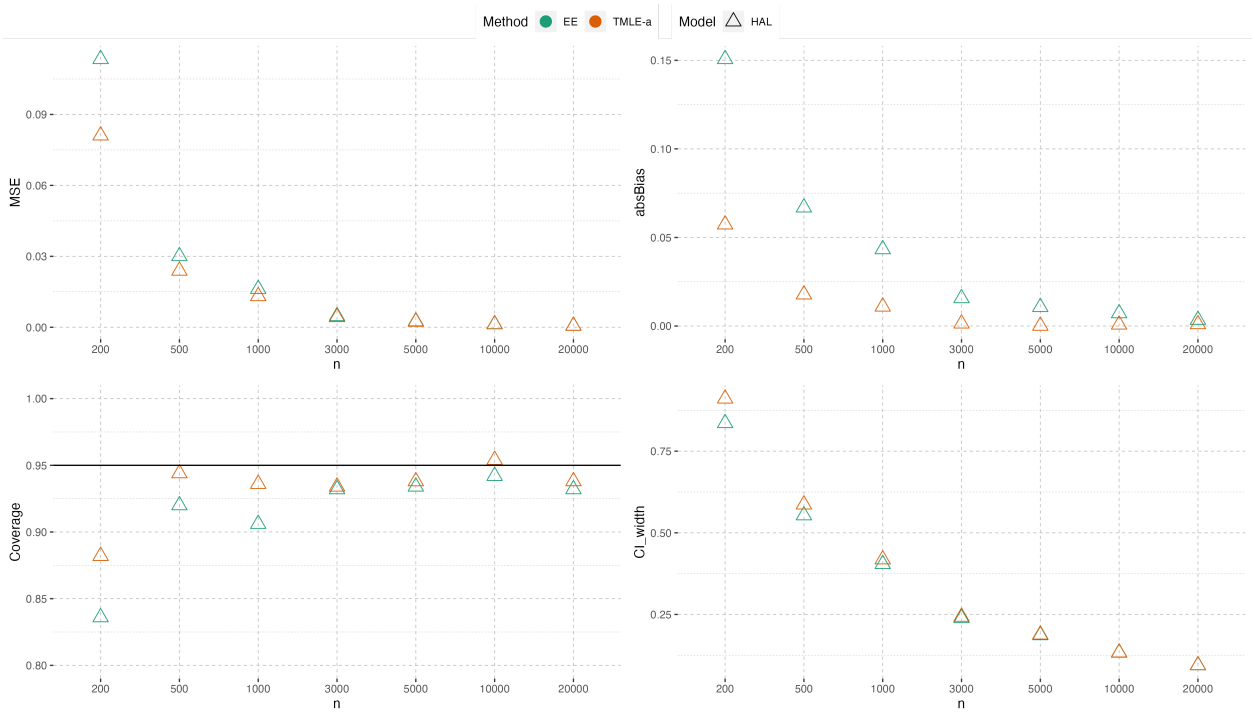

Figure 2: Main performance metrics (DR-learner, no CV)

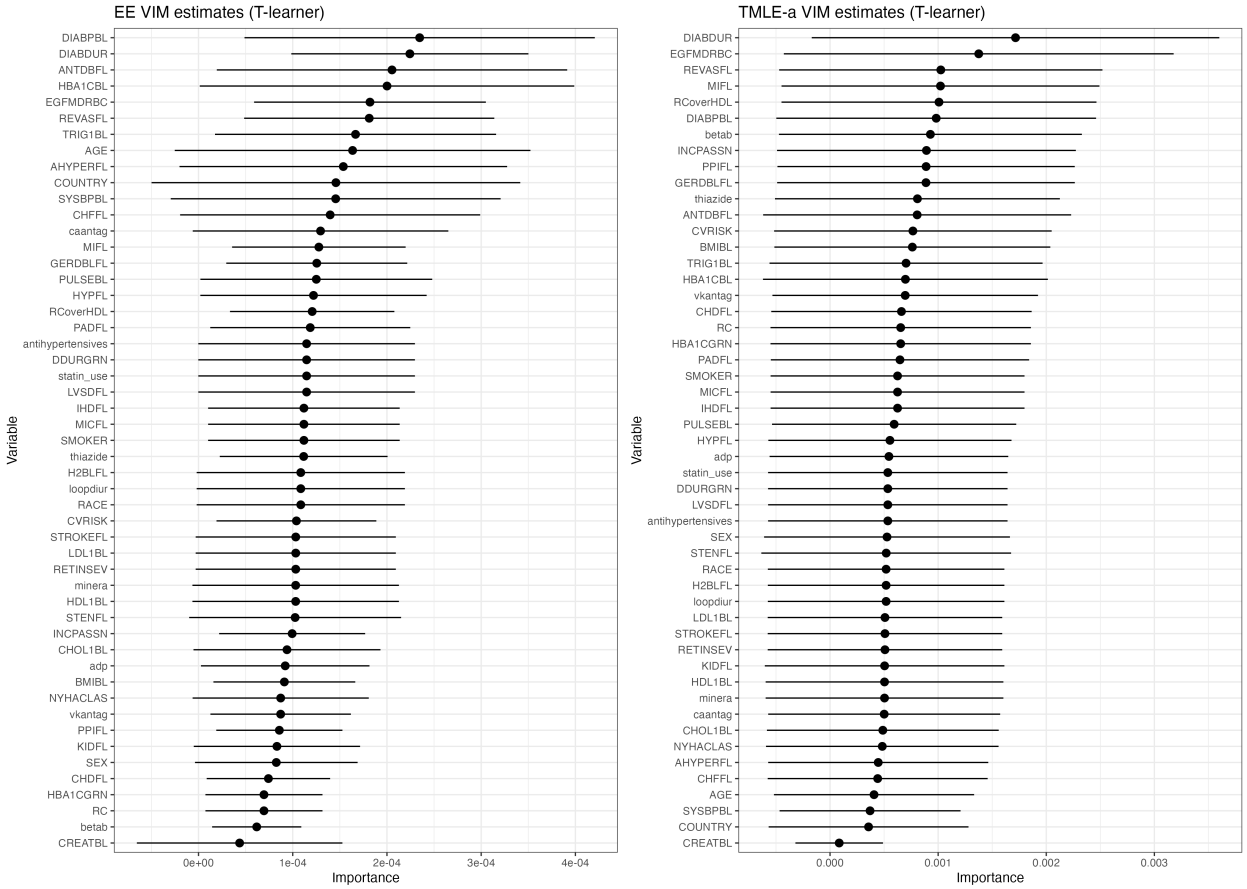

Figure 3: EE and TMLE estimates of variable importance parameter

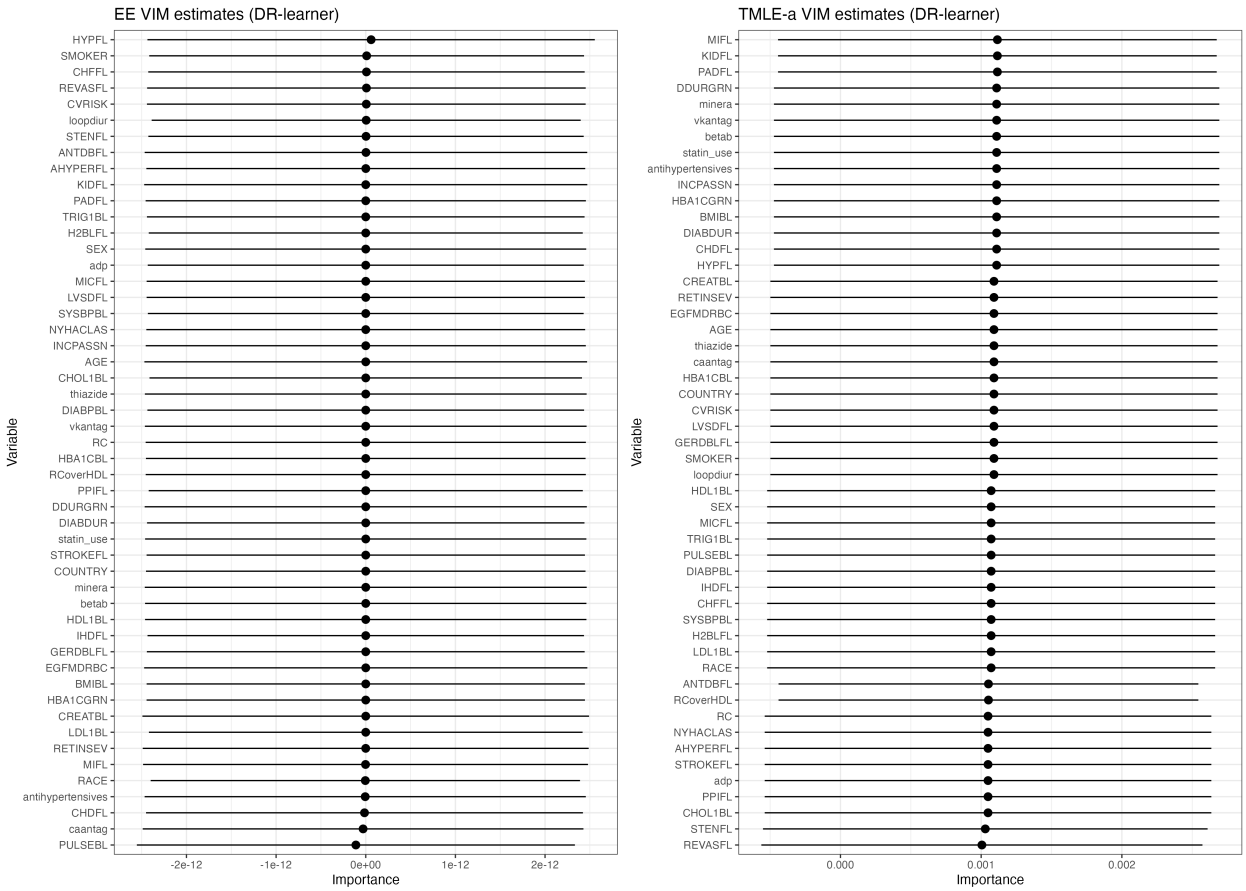

Figure 4: EE and TMLE estimates of variable importance parameter