

# Supplementary Materials for Learning Optimal Dynamic Treatment Regimes from Longitudinal Data

## A1 Simulation

We considered the following DGM:

$$\begin{aligned}W_1 &\sim \mathcal{U}(-1, 1) \\W_2 &\sim \mathcal{U}(-1, 1) \\P(A_1 = 1) &= \text{expit}(0.5 - 1.3W_1 + 0.4W_2) \\W_3 &\sim \mathcal{U}(-1, 1) \times 1.25A_1 + 0.25 \\P(A_2 = 1) &= \text{expit}(0.5 + 0.4A_1 - 1.5W_3) \\Y &\sim \mathcal{N}(0.4 - 0.4A_1 - A_2W_3 - 4A_1W_1 + 0.08A_1A_2 + A_2W_3 - 4A_1W_1 - 2A_1W_2 - 0.1A_2 + 1.5W_1, 1)\end{aligned}$$

We evaluated estimator performance in terms of absolute bias, absolute bias scaled by  $\sqrt{n}$ , and 95% confidence interval (CI) coverage. We conducted 1000 simulations for sample sizes  $n \in \{500, 1000, 10000\}$ . The true value of  $\psi$  is approximately 0.248.

Table 1: Simulation results.

$n$	$\hat{\psi}$	Bias	$\sqrt{n} \times  \text{Bias} $	95% CI Covr.
500	2.269	0.023	0.514	0.929
1000	2.267	0.021	0.653	0.925
10000	2.250	0.004	0.378	0.875