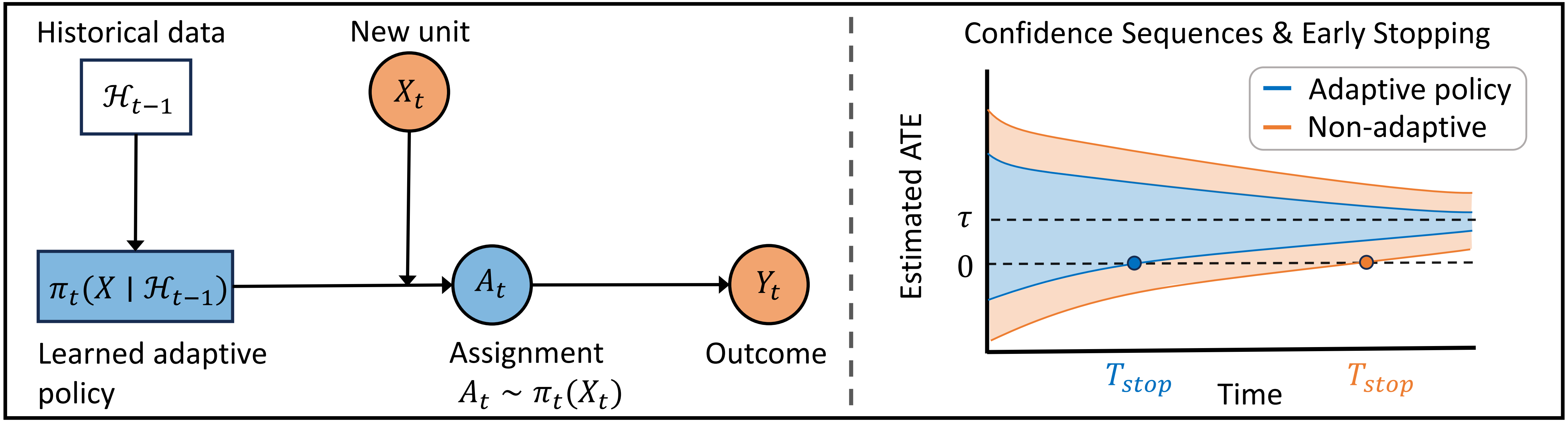
Efficient Adaptive Experimentation with Noncompliance

Miruna Oprescu, Brian M Cho, Nathan Kallus

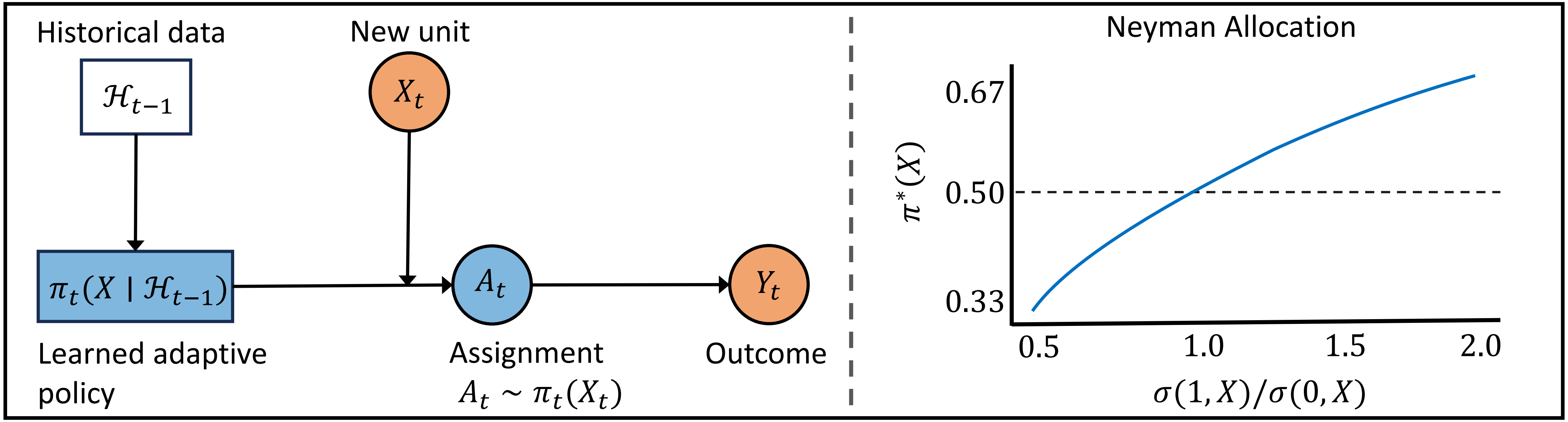
Cornell University, Cornell Tech

Efficient Adaptive Experiments with Direct Treatments



- **Setting:** Binary treatment $A \in \{0, 1\}$ with covariates X; online experiment: observe $\boldsymbol{X_t}$, assign $\boldsymbol{A_t}$ and observe outcome $\boldsymbol{Y_t}$ each round.
- **Goal:** Learn an adaptive policy $\boldsymbol{\pi_t(X \mid \mathcal{H}_{t-1})}$ at time t that minimizes the asymptotic variance of the ATE and provide and estimator that achieves it.
- **Motivation:** Enable reliable *early stopping* by driving faster variance reduction.

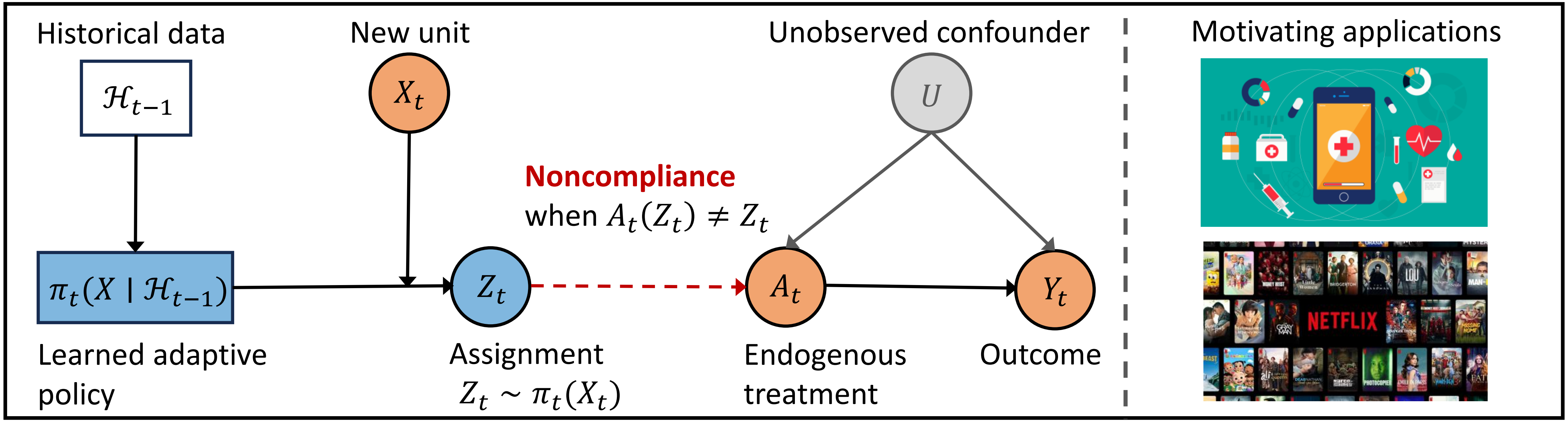
Efficient Adaptive Experiments with Direct Treatments



- **Classical Result:** *Neyman allocation* — assign more where (conditional) outcome variance is larger.

$$\pi^*(X) = \frac{\sqrt{\mathrm{Var}(Y \mid A=1, X)}}{\sqrt{\mathrm{Var}(Y \mid A=0, X)} + \sqrt{\mathrm{Var}(Y \mid A=1, X)}} := \frac{\sigma(1,X)}{\sigma(0,X) + \sigma(1,X)}$$

Efficient Adaptive Experiments with ~~Direct Treatments~~ Noncompliance



- **Noncompliance:** We can assign an *encouragement (instrumental variable)*, but *cannot enforce* the treatment (e.g. ethical considerations, feasibility).
- **Issue:** A_t is *endogenous* (affected by unobserved confounding) ⇒ naive A/B on A_t is biased; only the instrumental variable Z_t is randomized.
- **IV Fix:** Use Z_t to identify the ATE and adapt the instrument policy instead.

Optimal Policy with Noncompliance

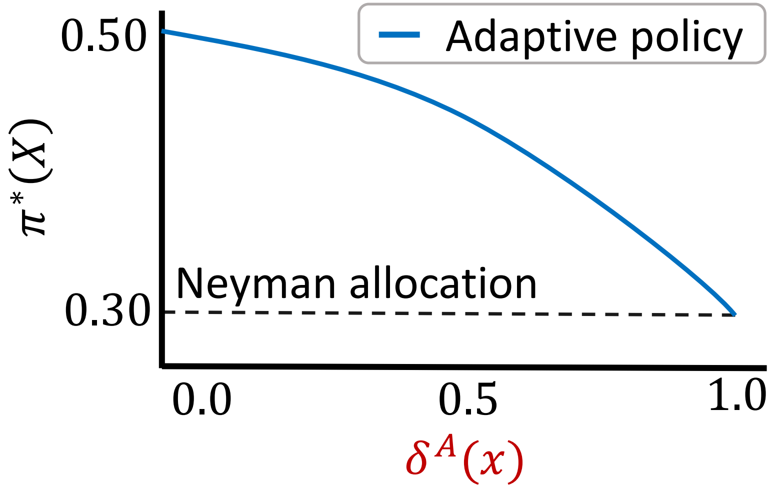
Fixed policy that minimizes asymptotic variance:

$$\boldsymbol{\pi}^*(\boldsymbol{X}) = \frac{\sqrt{\mathrm{Var}(Y - A\delta(X) \mid Z = 1, X)}}{\sqrt{\mathrm{Var}(Y - A\delta(X) \mid Z = 0, X)} + \sqrt{\mathrm{Var}(Y - A\delta(X) \mid Z = 1, X)}}$$

where:

$$\delta(X) = \frac{\delta^Y(X)}{\delta^A(X)} = \frac{\mathbb{E}[\,Y \mid X = x, Z = 1\,] - \mathbb{E}[\,Y \mid X = x, Z = 0\,]}{\underbrace{\mathbb{E}[\,A \mid X = x, Z = 1\,] - \mathbb{E}[\,A \mid X = x, Z = 0\,]}_{\delta^A(x)\text{ (compliance factor)}}}$$

Adaptive policy vs compliance



- **ATE Identification** from Wang & Tchetgen Tchetgen (2018): $\tau = \mathbb{E}[\delta(x)]$.
    - Under IV relevance, exclusion, randomization given X and unconfounded compliance
- **Generalizes Neyman:** balances outcome noise *and* compliance noise.

Introducing the AMRIV

- AMRIV = **A**daptive **M**ultiply-**R**obust estimator for **IV** settings

Burn-in period | Adaptive collection | Estimation

Introducing the AMRIV

- AMRIV = **A**daptive **M**ultiply-**R**obust estimator for **IV** settings

Burn-in period	Adaptive collection	Estimation
For $t = 1 \ldots T_0$: • Get X_t • Assign $Z_t \sim Bern(\pi_0(X_t))$ E.g.: $\pi_0(X) = 1/2$ • Observe A_t, Y_t • Append to history $H_t = \{(X_i, Z_i, A_i, Y_i)\}_{i=1}^{n}$		

Introducing the AMRIV

- AMRIV = **A**daptive **M**ultiply-**R**obust estimator for **IV** settings

Burn-in period	Adaptive collection	Estimation
For $\boldsymbol{t = 1 \dots T_0}$: • Get X_t • Assign $Z_t \sim Bern(\pi_0(X_t))$ E.g.: $\pi_0(X) = 1/2$ • Observe A_t, Y_t • Append to history $H_t = \{(X_i, Z_i, A_i, Y_i)\}_{i=1}^{n}$	**For $\boldsymbol{t = T_0 \dots T}$:** **Estimate $\boldsymbol{\pi^*(X)}$** • Estimate $\mathrm{Var}(Y - A\delta(X) \mid Z = z, X)$ as $\hat{\sigma}(z, X)$ from H_{t-1}. $\tilde{\pi}_t(X \mid H_{t-1}) = \dfrac{\hat{\sigma}(1,X)}{\hat{\sigma}(0,X) + \hat{\sigma}(1,X)}$ • Truncate: $\boldsymbol{\pi_t(X \mid H_{t-1})} := \min(1 - \epsilon_t, \max(\epsilon_t, \tilde{\pi}_t))$	

Introducing the AMRIV

- AMRIV = **A**daptive **M**ultiply-**R**obust estimator for **IV** settings

Burn-in period	Adaptive collection	Estimation

Burn-in period

For $\boldsymbol{t = 1 \ldots T_0}$:

- Get X_t
- Assign $Z_t \sim Bern(\pi_0(X_t))$ E.g.: $\pi_0(X) = 1/2$
- Observe A_t, Y_t
- Append to history $H_t = \{(X_i, Z_i, A_i, Y_i)\}_{i=1}^n$

Adaptive collection

For $\boldsymbol{t = T_0 \ldots T}$:

Estimate $\boldsymbol{\pi^*(X)}$

- Estimate $\mathrm{Var}(Y - A\delta(X) \mid Z = z, X)$ as $\hat{\sigma}(z, X)$ from H_{t-1}.

$$\tilde{\pi}_t(X \mid H_{t-1}) = \frac{\hat{\sigma}(1, X)}{\hat{\sigma}(0, X) + \hat{\sigma}(1, X)}$$

- Truncate: $\boldsymbol{\pi_t(X \mid H_{t-1})} := \min(1 - \epsilon_t, \max(\epsilon_t, \tilde{\pi}_t))$

Estimate $\boldsymbol{\phi_t}$

- Get X_t, set $Z_t \sim \pi_t(X_t)$, observe A_t, Y_t.
- Estimate nuisances $\hat{\eta}_t$ from H_{t-1} via cross-fitting.
- Impute the round t effect ϕ_t as

$$\boldsymbol{\phi_t} = \phi(X_t, Z_t, A_t, Y_t, ; \pi_t, \hat{\eta}_t)$$

where ϕ is the **efficient influence function (EIF)** of τ.

Estimation

Introducing the AMRIV

- AMRIV = **A**daptive **M**ultiply-**R**obust estimator for **IV** settings

Burn-in period

For $t = 1 \dots T_0$:

- Get X_t
- Assign

  $Z_t \sim Bern(\pi_0(X_t))$

  E.g.: $\pi_0(X) = 1/2$
- Observe A_t, Y_t
- Append to history

  $H_t = \{(X_i, Z_i, A_i, Y_i)\}_{i=1}^{n}$

Adaptive collection

For $t = T_0 \dots T$:

Estimate $\pi^*(X)$

- Estimate $\mathrm{Var}(Y - A\delta(X) \mid Z = z, X)$ as $\hat{\sigma}(z, X)$ from H_{t-1}.

$$\tilde{\pi}_t(X \mid H_{t-1}) = \frac{\hat{\sigma}(1, X)}{\hat{\sigma}(0, X) + \hat{\sigma}(1, X)}$$

- Truncate: $\boldsymbol{\pi_t}(\boldsymbol{X} \mid \boldsymbol{H_{t-1}}) := \min(1 - \epsilon_t, \max(\epsilon_t, \tilde{\pi}_t))$

Estimate ϕ_t

- Get X_t, set $Z_t \sim \pi_t(X_t)$, observe A_t, Y$_t$.
- Estimate nuisances $\hat{\eta}_t$ from H_{t-1} via cross-fitting.
- Impute the round t effect ϕ_t as

$$\boldsymbol{\phi_t} = \phi(X_t, Z_t, A_t, Y_t, ; \pi_t, \hat{\eta}_t)$$

where ϕ is the **efficient influence function (EIF)** of τ.

Estimation

At $t = T$:

- Estimate τ as:

$$\hat{\tau}_T^{AMRIV} = \frac{1}{T}\sum_{t=1}^{T} \phi_t$$

- Calculate α-level confidence sequences:

  $[L_t, U_t]$
- Decide whether to continue or stop data collection

Introducing the AMRIV

Theoretical properties:

- **Efficient:**

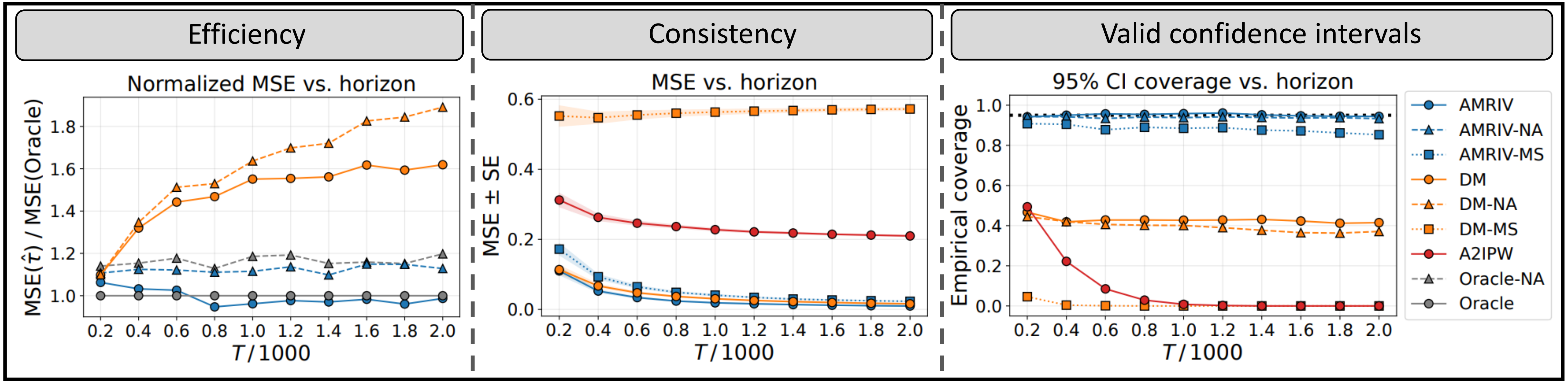
$$\sqrt{T}\left(\hat{\tau}_T^{AMRIV} - \tau\right) \to \mathcal{N}\left(0, V_{eff}(\pi)\right)$$

  with $\pi = \pi^*$ achieving the minimum bound.

- **Multiply-robust:** Consistent if either $\delta(X)$ or $\delta_A(X)$ is learned consistently; AMRIV is $O_p(T^{-1/2})$ if both $\delta(X)$ and $\delta_A(X)$ are $o_p(T^{-1/4})$.

- **Anytime-valid:** Can build anytime valid asymptotic confidence sequences (AsymCS) from online EIF variance ⇒ peek-safe early stopping.

Experimental Results



- **Efficiency:** Adaptivity improves efficiency of all estimators.
- **Consistency:** AMRIV-MS is consistent even when one of the nuisances is misspecified, whereas the direct method DM-MS is not.
- **Valid confidence intervals**: AMRIV achieves nominal (95%) coverage unlike non-robust methods.

Summary of Contributions and Impact

Key Contributions:

- We proposed an **adaptive IV framework** for online experiments with noncompliance and derived an **optimal instrument assignment policy** to minimize asymptotic variance.
- We introduced **AMRIV**, an adaptive IV estimator that provides strong theoretical guarantees: **asymptotic efficiency**, **multiply-robust consistency**, and **time-uniform confidence sequences**.
- We validated our framework through simulations and real-world applications.

Broader Impact:

- We enabled adaptive experimentation **when treatment isn’t assignable**, delivering **more information**, **earlier stopping**, and **valid inference** for digital platforms, personalized medicine, and beyond.