Integrating Double Robustness into Causal Latent Factor Models

$$uv - \hat{u}\hat{v} = O(|u - \hat{u}| + |v - \hat{v}|) \quad \Longrightarrow \quad uv - ?? = O(|u - \hat{u}| \times |v - \hat{v}|)$$

Raaz Dwivedi

CORNELL TECH

Online Causal Inference Seminar, May 7, 2024

Talk outline

1. **Causal Latent Factor Models:** Data-rich enviroments
2. **Double Robustness:** A Layman’s Perspective
3. **Integrating:** Two Vignettes

1. Causal Latent Factor Models:

Inference for modern data-rich settings

Decision-making in data-rich environments

Decision-making in data-rich environments

Data: N units with T measurements under (finitely) many interventions

Decision-making in data-rich environments

Data: N units with T measurements under (finitely) many interventions

Online platforms

Digital health

Precision medicine

Decision-making in data-rich environments

Data: N units with T measurements under (finitely) many interventions

Goal: Determine counterfactuals—units' outcomes under alternate interventions

Online platforms

Digital health

Precision medicine

Causal panel data: Basic set-up

Causal panel data: Basic set-up

Potential outcome: $Y_{i,t}^{(a)} = \theta_{i,t}^{(a)} + \varepsilon_{i,t}^{(a)}$ – unit i at time t under intervention a

- Neyman-Rubin potential outcome framework

Causal panel data: Basic set-up

Potential outcome: $Y_{i,t}^{(a)} = \theta_{i,t}^{(a)} + \varepsilon_{i,t}^{(a)}$ – unit i at time t under intervention a

- Neyman-Rubin potential outcome framework

Observed data: outcome $Y_{i,t} = Y_{i,t}^{(A_{i,t})}$ and intervention $A_{i,t}$

- **No spill-over** of treatment on future outcomes

Causal panel data: Basic set-up

Potential outcome: $Y_{i,t}^{(a)} = \theta_{i,t}^{(a)} + \varepsilon_{i,t}^{(a)}$ – unit i at time t under intervention a

- Neyman-Rubin potential outcome framework

Observed data: outcome $Y_{i,t} = Y_{i,t}^{(A_{i,t})}$ and intervention $A_{i,t}$

- **No spill-over** of treatment on future outcomes

Goals: Estimate

- Average treatment effect (ATE): $\mathrm{ATE}_t = \frac{1}{N}\sum_{i=1}^{N}(\theta_{i,t}^{(1)} - \theta_{i,t}^{(0)})$ ⟵ Analog of CATE for unobserved confounding

Causal panel data: Basic set-up

Potential outcome: $Y_{i,t}^{(a)} = \theta_{i,t}^{(a)} + \varepsilon_{i,t}^{(a)}$ – unit i at time t under intervention a

- Neyman-Rubin potential outcome framework

Observed data: outcome $Y_{i,t} = Y_{i,t}^{(A_{i,t})}$ and intervention $A_{i,t}$

- **No spill-over** of treatment on future outcomes

Goals: Estimate

- Average treatment effect (ATE): $\mathrm{ATE}_t = \frac{1}{N}\sum_{i=1}^{N}(\theta_{i,t}^{(1)} - \theta_{i,t}^{(0)})$ ⟵ Analog of CATE for unobserved confounding
- Individual treatment effect (ITE): $\mathrm{ITE}_{i,t} = \theta_{i,t}^{(1)} - \theta_{i,t}^{(0)}$

Addressing challenges for counterfactual inference

Addressing challenges for counterfactual inference

1. Confounding

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \not\perp\!\!\!\perp A_{i,t}$$

Addressing challenges for counterfactual inference

1. Confounding

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \not\perp\!\!\!\perp A_{i,t}$$

Sufficient unobserved confounders

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \text{Latent factors } \underbrace{\mathcal{F}}_{\text{Unobserved}}$$

Addressing challenges for counterfactual inference

1. Confounding

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \not\perp\!\!\!\perp A_{i,t}$$

Sufficient unobserved confounders

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \text{Latent factors } \underbrace{\mathscr{F}}_{\text{Unobserved}}$$

quantities of interest

$$\theta_{i,t}^{(a)} \triangleq \mathbb{E}[Y_{i,t}^{(a)} \mid \mathscr{F}]$$

$$p_{i,t} \triangleq \mathbb{E}[A_{i,t} \mid \mathscr{F}]$$

Addressing challenges for counterfactual inference

1. Confounding

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \not\perp\!\!\!\perp A_{i,t}$$

Sufficient unobserved confounders

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \text{Latent factors } \underbrace{\mathcal{F}}_{\text{Unobserved}}$$

quantities of interest

$$\theta_{i,t}^{(a)} \triangleq \mathbb{E}[Y_{i,t}^{(a)} \mid \mathcal{F}]$$

$$p_{i,t} \triangleq \mathbb{E}[A_{i,t} \mid \mathcal{F}]$$

2. Complexity of unknowns

Estimating $2NT$ parameters with NT noisy observations

Addressing challenges for counterfactual inference

1. Confounding

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \not\perp\!\!\!\perp A_{i,t}$$

Sufficient unobserved confounders

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \text{Latent factors } \underbrace{\mathcal{F}}_{\text{Unobserved}}$$

quantities of interest

$$\theta_{i,t}^{(a)} \triangleq \mathbb{E}[Y_{i,t}^{(a)} \mid \mathcal{F}]$$

$$p_{i,t} \triangleq \mathbb{E}[A_{i,t} \mid \mathcal{F}]$$

2. Complexity of unknowns

Estimating $2NT$ parameters with NT noisy observations

Factor model for outcomes

$$\theta_{i,t}^{(a)} \triangleq \langle u_i^{(a)}, v_t^{(a)} \rangle$$

Addressing challenges for counterfactual inference

1. Confounding

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \not\perp\!\!\!\perp A_{i,t}$$

Sufficient unobserved confounders

$$(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \text{Latent factors } \underbrace{\mathcal{F}}_{\text{Unobserved}}$$

quantities of interest

$$\theta_{i,t}^{(a)} \triangleq \mathbb{E}[Y_{i,t}^{(a)} \mid \mathcal{F}]$$

$$p_{i,t} \triangleq \mathbb{E}[A_{i,t} \mid \mathcal{F}]$$

2. Complexity of unknowns

Estimating $2NT$ parameters with NT noisy observations

Factor model for outcomes

$$\theta_{i,t}^{(a)} \triangleq \langle u_i^{(a)}, v_t^{(a)} \rangle$$

$\Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i \in [N], t \in [T]}$ always admits rank $N \wedge T$ factorization.

The real assumption is that the rank is $\ll N \wedge T$.

Causal latent factor model: Common assumptions

Causal latent factor model: Common assumptions

1. Sufficient **unmeasured confounders**: $(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \,|\, \mathcal{F}$

2. Factor model for outcomes: $\theta_{i,t}^{(a)} \triangleq \langle u_i^{(a)}, v_t^{(a)} \rangle$

3. **Positivity**/overlap: $p \leq p_{i,t} \leq 1 - p$

4. Random variables drawn **independently** across (i, t) after conditioning on latent factors $\mathcal{F}$

Can also handle dependent noise,
decaying positivity, non-linear factor model

Causal latent factor model: Common assumptions

1. Sufficient **unmeasured confounders**: $(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \mathcal{F}$

2. Factor model for outcomes: $\theta_{i,t}^{(a)} \triangleq \langle u_i^{(a)}, v_t^{(a)} \rangle$

3. **Positivity**/overlap: $p \leq p_{i,t} \leq 1 - p$

4. Random variables drawn **independently** across (i, t) after conditioning on latent factors $\mathcal{F}$

Estimands

$$\text{ATE}_t = \frac{1}{N} \sum_{i=1}^{N} (\theta_{i,t}^{(1)} - \theta_{i,t}^{(0)})$$

$$\text{ITE}_{i,t} = \theta_{i,t}^{(1)} - \theta_{i,t}^{(0)}$$

Can also handle dependent noise, decaying positivity, non-linear factor model

Popular approach for estimating ATE$_t$: Outcome imputation

$$Y_{i,t}^{(A_{i,t})} = \theta_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}^{(A_{i,t})} \qquad \Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i\in[N],t\in[T]}$$

Popular approach for estimating ATE_t: Outcome imputation

$$Y_{i,t}^{(A_{i,t})} = \theta_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}^{(A_{i,t})} \qquad \Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i\in[N],t\in[T]}$$

$\{Y_{i,t} : A_{i,t} = 1\}$ → **Black-box Matrix Completion** → $\hat{\Theta}^{(1)}$

$\{Y_{i,t} : A_{i,t} = 0\}$ → **Black-box Matrix Completion** → $\hat{\Theta}^{(0)}$

$$\widehat{\text{ATE}_t}^{\text{OI}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{i,t}^{(1)} - \hat{\theta}_{i,t}^{(0)})$$

Popular approach for estimating ATE_t: Outcome imputation

$$Y_{i,t}^{(A_{i,t})} = \theta_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}^{(A_{i,t})} \qquad \Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i\in[N],t\in[T]}$$

$\{Y_{i,t} : A_{i,t} = 1\}$ → **Black-box Matrix Completion** → $\hat{\Theta}^{(1)}$

$\{Y_{i,t} : A_{i,t} = 0\}$ → **Black-box Matrix Completion** → $\hat{\Theta}^{(0)}$

$$\widehat{\mathrm{ATE}_t}^{\mathrm{OI}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{i,t}^{(1)} - \hat{\theta}_{i,t}^{(0)})$$

✓ **Works well if $\Theta^{(1)}$ and $\Theta^{(0)}$ have low-rank**

But what if the outcomes are not low-rank?

But what if the outcomes are not low-rank?

How do we do augmented IPW / doubly robust adjustment with unobserved confounding?

2. Double Robustness:

A Layman’s Perspective

When the estimand is a product

$$\theta^\star = \langle u, v \rangle$$

When the estimand is a product

$$\theta^\star = \langle u, v \rangle$$

- For ITE: $\theta_{i,t}^{(a)} = \langle u_i^{(a)}, v_t^{(a)} \rangle$

When the estimand is a product

$$\theta^\star = \langle u, v \rangle$$

- For ITE: $\theta_{i,t}^{(a)} = \langle u_i^{(a)}, v_t^{(a)} \rangle$

- For ATE$_t$:

$$\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t}^{(1)} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_{i,t}^{(1)} \mid \mathcal{F}] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\theta_{i,t}^{(1)}\frac{A_{i,t}}{p_{i,t}} \mid \mathcal{F}\right]$$

When the estimand is a product

$$\theta^\star = \langle u, v \rangle$$

- For ITE: $\theta_{i,t}^{(a)} = \langle u_i^{(a)}, v_t^{(a)} \rangle$

$$= \langle u, v \rangle_{\mathbb{P}}$$

$$\downarrow \quad \downarrow$$

- For ATE$_t$:

$$\frac{1}{N}\sum_{i=1}^{N} \theta_{i,t}^{(1)} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[Y_{i,t}^{(1)} \mid \mathcal{F}] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\theta_{i,t}^{(1)} \frac{A_{i,t}}{p_{i,t}} \mid \mathcal{F}\right]$$

When the estimand is a product

$$\theta^\star = \langle u, v \rangle$$

- For ITE: $\theta_{i,t}^{(a)} = \langle u_i^{(a)}, v_t^{(a)} \rangle$

$= \langle u, v \rangle_{\mathbb{P}}$

- For ATE$_t$:

$$\frac{1}{N}\sum_{i=1}^{N} \theta_{i,t}^{(1)} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[Y_{i,t}^{(1)} \mid \mathcal{F}] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\theta_{i,t}^{(1)} \frac{A_{i,t}}{p_{i,t}} \mid \mathcal{F}\right]$$

$$\frac{1}{N}\sum_{i=1}^{N} \theta_{i,t}^{(0)} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[Y_{i,t}^{(0)} \mid \mathcal{F}] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\theta_{i,t}^{(0)} \frac{1 - A_{i,t}}{1 - p_{i,t}} \mid \mathcal{F}\right]$$

When the estimand is a product

$$\theta^\star = \langle u, v \rangle$$

When the estimand is a product

$$\theta^\star = \langle u, v \rangle$$

Similar structure across problems:

- **For ATE with observed confounding:** $\mathbb{E}[Y(1)] \; = \mathbb{E}\left[\mathbb{E}[Y(1) \,|\, X] \cdot \frac{A}{p(X)}\right]$
- **Importance sampling:** $\mathbb{E}_{X\sim\mathbb{Q}}[Y] \; = \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}[Y \,|\, X] \cdot \frac{q(X)}{p(X)}\right]$

  (e.g., off-policy evaluation, covariate shift, …)

Given estimates $\hat{u}$ and $\hat{v}$, what is a good estimator for $\theta^\star = uv$?

Given estimates $\hat{u}$ and $\hat{v}$, what is a good estimator for $\theta^\star = uv$?

Omitting the inner product notation for clarity

Given estimates $\hat{u}$ and $\hat{v}$, what is a good estimator for $\theta^\star = uv$?

Omitting the inner product notation for clarity

- $|uv - \hat{u}\,\hat{v}| \leq |uv - \hat{u}v| + |\hat{u}v - \hat{u}\hat{v}| \;= O(|u - \hat{u}| + |v - \hat{v}|)$

Given estimates $\hat{u}$ and $\hat{v}$, what is a good estimator for $\theta^\star = uv$?

Omitting the inner product notation for clarity

- $|uv - \hat{u}\,\hat{v}| \leq |uv - \hat{u}v| + |\hat{u}v - \hat{u}\hat{v}| \; = O(|u - \hat{u}| + |v - \hat{v}|)$
- $uv -$?? $= (u - \hat{u}) \times (v - \hat{v})$

Given estimates $\hat{u}$ and $\hat{v}$, what is a good estimator for $\theta^\star = uv$?

Omitting the inner product notation for clarity

- $|uv - \hat{u}\,\hat{v}| \leq |uv - \hat{u}v| + |\hat{u}v - \hat{u}\hat{v}| \;\; = O(|u - \hat{u}| + |v - \hat{v}|)$
- $\not{uv} - ?? = (u - \hat{u}) \times (v - \hat{v})$

$$= \not{uv} - \hat{u}v - u\hat{v} + \hat{u}\hat{v}$$

Given estimates $\hat{u}$ and $\hat{v}$, what is a good estimator for $\theta^\star = uv$?

Omitting the inner product notation for clarity

- $|uv - \hat{u}\,\hat{v}| \leq |uv - \hat{u}v| + |\hat{u}v - \hat{u}\hat{v}| \;\; = O(|u - \hat{u}| + |v - \hat{v}|)$
- $\cancel{uv} - ?? = (u - \hat{u}) \times (v - \hat{v})$

$$= \cancel{uv} - \hat{u}v - u\hat{v} + \hat{u}\hat{v}$$

$\Rightarrow \quad ?? = \hat{u}v + u\hat{v} - \hat{u}\hat{v}$

Double robustness, debiased/double machine learning…

[… Cassel+ '77, Robinson '88, Särndal+ '89, Robins+ '94, '95, '08, '09, Newey+ '94, '18, Bickel+ '98, van der Laan+ '03, Lunceford+ '04, Davidian+ '05, Li+ '11, Jiang+ '15, Chernozhukov+ '18, Hirshberg+ '18, Diaz '19, Arkhangelsky+ '21, Dorn+ '21 …]

Simplified view of doubly robust estimator for uv

Simplified view of doubly robust estimator for uv

$\hat{u}\hat{v}$ $\quad\quad$ $O(|\hat{u}-u| + |\hat{v}-v|)$

↓ Estimate $\quad\quad$ Error ↓

$\hat{u}v + u\hat{v} - \hat{u}\hat{v}$ $\quad\quad$ $O(|\hat{u}-u| \times |\hat{v}-v|)$

Simplified view of doubly robust estimator for uv

$$\hat{u}\hat{v} \xrightarrow{\text{Estimate}} \hat{u}v + u\hat{v} - \hat{u}\hat{v}$$

$$O(|\hat{u} - u| + |\hat{v} - v|) \xrightarrow{\text{Error}} O(|\hat{u} - u| \times |\hat{v} - v|)$$

Problem setting	u	v
ATE with observed confounding	conditional outcome mean	propensity function
Off policy evaluation	mean reward	importance ratio
[This talk] ATE with unobserved confounding	outcome matrix	propensity matrix
[This talk] ITE with unobserved confounding	user factor	time factor

$\hat{u}\hat{v}$ → $\hat{u}v + u\hat{v} - \hat{u}\hat{v}$

3. Integrating double robustness with causal latent factor model

$\hat{u}\hat{v}$ → $\hat{u}v + u\hat{v} - \hat{u}\hat{v}$

Part 1: Doubly robust estimation of $\text{ATE}_t = \frac{1}{N}\sum_{i=1}^{N}(\theta_{i,t}^{(1)} - \theta_{i,t}^{(0)})$

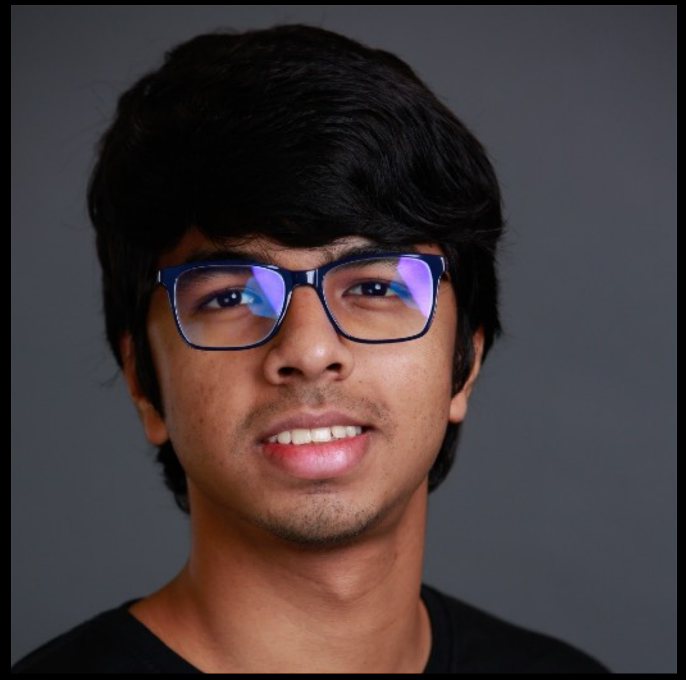
Part 2: Doubly robust estimation of $\text{ITE}_{i,t} = \theta_{i,t}^{(1)} - \theta_{i,t}^{(0)}$

Doubly robust estimation of ATE$_t$ $= \frac{1}{N}\sum_{i=1}^{N}(\theta_{i,t}^{(1)} - \theta_{i,t}^{(0)})$

(CATE for unobserved confounding)

Alberto Abadie

Anish Agarwal



Abhin Shah

https://arxiv.org/abs/2402.11652

How do we do augmented IPW / doubly robust adjustment with unobserved confounding?

$$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}^{(A_{i,t})}$$

$$\Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i\in[N],t\in[T]}$$

How do we do augmented IPW / doubly robust adjustment with unobserved confounding?

$$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}^{(A_{i,t})} \qquad \Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i \in [N], t \in [T]}$$

$\{Y_{i,t} : A_{i,t} = 1\}$ ⟶ **Matrix Completion** ⟶ $\hat{\Theta}^{(1)}$

$\{Y_{i,t} : A_{i,t} = 0\}$ ⟶ **Matrix Completion** ⟶ $\hat{\Theta}^{(0)}$

How do we do augmented IPW / doubly robust adjustment with unobserved confounding?

$$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}^{(A_{i,t})} \qquad \Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i \in [N], t \in [T]}$$

$\{Y_{i,t} : A_{i,t} = 1\}$ ⟶ **Matrix Completion** ⟶ $\hat{\Theta}^{(1)}$

$$\widehat{\text{ATE}_t}^{\text{OI}} = \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_{i,t}^{(1)} - \hat{\theta}_{i,t}^{(0)})$$

$\{Y_{i,t} : A_{i,t} = 0\}$ ⟶ **Matrix Completion** ⟶ $\hat{\Theta}^{(0)}$

How do we do augmented IPW / doubly robust adjustment with unobserved confounding?

$$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}^{(A_{i,t})} \qquad \Theta^{(a)} \triangleq [\theta_{i,t}^{(a)}]_{i\in[N],t\in[T]}$$

$\{Y_{i,t} : A_{i,t} = 1\}$ ⟶ **Matrix Completion** ⟶ $\hat{\Theta}^{(1)}$

$$\widehat{\text{ATE}_t}^{\text{OI}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{i,t}^{(1)} - \hat{\theta}_{i,t}^{(0)})$$

$\{Y_{i,t} : A_{i,t} = 0\}$ ⟶ **Matrix Completion** ⟶ $\hat{\Theta}^{(0)}$

✓ **Works well if $\Theta^{(1)}$ and $\Theta^{(0)}$ have low-ranks**

There is one other matrix that we can leverage!

The Intervention Matrix $\mathbf{A}$!

1	0	1	0	1	0
1	1	0	0	0	1
1	1	1	0	0	1
0	0	1	1	1	1
1	1	0	1	0	1
0	0	1	1	0	1

There is one other matrix that we can leverage!

The Intervention Matrix $\mathbf{A}$!

But its fully observed?

1	0	1	0	1	0
1	1	0	0	0	1
1	1	1	0	0	1
0	0	1	1	1	1
1	1	0	1	0	1
0	0	1	1	0	1

There is one other matrix that we can leverage!

The Intervention Matrix $\mathbf{A}$!

1	0	1	0	1	0
1	1	0	0	0	1
1	1	1	0	0	1
0	0	1	1	1	1
1	1	0	1	0	1
0	0	1	1	0	1

But its fully observed?

What about $P = \mathbb{E}[\mathbf{A} \mid \mathscr{F}]$?

There is one other matrix that we can leverage!

The Intervention Matrix $\mathbf{A}$!

1	0	1	0	1	0
1	1	0	0	0	1
1	1	1	0	0	1
0	0	1	1	1	1
1	1	0	1	0	1
0	0	1	1	0	1

But its fully observed?

What about $P = \mathbb{E}[\mathbf{A} \mid \mathscr{F}]$?

$\mathbf{A}$ → **Matrix *Estimation*** → $\hat{P}$

$$\widehat{\text{ATE}_t}^{\text{IPW}} = \frac{1}{N}\sum_{i=1}^{N} Y_{i,t}\left(\frac{A_{i,t}}{\hat{p}_{i,t}} - \frac{1-A_{i,t}}{1-\hat{p}_{i,t}}\right)$$

There is one other matrix that we can leverage!

The Intervention Matrix $\mathbf{A}$!

1	0	1	0	1	0
1	1	0	0	0	1
1	1	1	0	0	1
0	0	1	1	1	1
1	1	0	1	0	1
0	0	1	1	0	1

But its fully observed?

What about $P = \mathbb{E}[\mathbf{A} \mid \mathcal{F}]$?

$$\widehat{\text{ATE}}_t^{\text{IPW}} = \frac{1}{N}\sum_{i=1}^{N} Y_{i,t}\left(\frac{A_{i,t}}{\hat{p}_{i,t}} - \frac{1 - A_{i,t}}{1 - \hat{p}_{i,t}}\right)$$

✓ **Works well if P is low-rank**

But when is $\mathbf{P}$ low-rank?

smooth f

But when is $\mathbf{P}$ low-rank?

smooth f

$$p_{i,t} = p$$

p	p	p	p
p	p	p	p
p	p	p	p
p	p	p	p

But when is $\mathbf{P}$ low-rank?

smooth f

$p_{i,t} = p$

$p_{i,t} = p_i$ or p_t

p	p	p	p
p	p	p	p
p	p	p	p
p	p	p	p

p1	p1	p1	p1
p2	p2	p2	p2
p3	p3	p3	p3
p4	p4	p4	p4

But when is P low-rank?

$p_{i,t} = p$

p	p	p	p
p	p	p	p
p	p	p	p
p	p	p	p

$p_{i,t} = p_i$ or p_t

p1	p1	p1	p1
p2	p2	p2	p2
p3	p3	p3	p3
p4	p4	p4	p4

smooth f

$p_{i,t} = u_i^T f(X_j)$

p11	p12	p13	p14
p21	p22	p23	p24
p31	p32	p33	p34
p41	p42	p43	p44

So..can we design estimators that are robust to either outcome matrix or propensity matrix being low rank?

So..can we design estimators that are robust to either outcome matrix or propensity matrix being low rank?

$\mathbf{Y} \circ \mathbf{A}$ → Black-box Matrix Completion → $\hat{\Theta}^{(1)}$

$\mathbf{Y} \circ (\mathbf{1} - \mathbf{A})$ → Black-box Matrix Completion → $\hat{\Theta}^{(0)}$

$\mathbf{A}$ → Black-box Matrix ***Estimation*** → $\hat{P}$

Yes…let's make an attempt

Formula

$$\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t}^{(1)}$$

Yes…let's make an attempt

Formula

$$\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t}^{(1)}$$

$\longrightarrow$

Estimate

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{i,t}^{(1)}$$

OI estimate

Yes…let's make an attempt

Formula		Estimate	
$\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t}^{(1)}$	$\longrightarrow$	$\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{i,t}^{(1)}$	OI estimate
$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_{i,t}^{(1)} \mid \mathscr{F}]$			
$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[Y_{i,t}\frac{A_{i,t}}{p_{i,t}} \mid \mathscr{F}\right]$	$\longrightarrow$	$\frac{1}{N}\sum_{i=1}^{N}Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}$	IPW estimate

Yes...let's make an attempt

Formula

$$\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t}^{(1)}$$

$$=\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_{i,t}^{(1)}\,|\,\mathscr{F}]$$

$$=\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[Y_{i,t}\frac{A_{i,t}}{p_{i,t}}\,|\,\mathscr{F}\right]$$

$$=\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\theta_{i,t}^{(1)}\frac{A_{i,t}}{p_{i,t}}\,|\,\mathscr{F}\right]$$

$$=\langle u, v\rangle_{\mathbb{P}}$$

Estimate

$\longrightarrow$ $$\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{i,t}^{(1)}$$ OI estimate

$\longrightarrow$ $$\frac{1}{N}\sum_{i=1}^{N}Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}$$ IPW estimate

Yes…let's make an attempt

Formula		Estimate	
$\frac{1}{N}\sum_{i=1}^{N}\theta_{i,t}^{(1)}$	$\longrightarrow$	$\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{i,t}^{(1)}$	OI estimate
$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_{i,t}^{(1)} \mid \mathcal{F}]$			
$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[Y_{i,t}\frac{A_{i,t}}{p_{i,t}} \mid \mathcal{F}\right]$	$\longrightarrow$	$\frac{1}{N}\sum_{i=1}^{N}Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}$	IPW estimate
$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\theta_{i,t}^{(1)}\frac{A_{i,t}}{p_{i,t}} \mid \mathcal{F}\right]$	$\longrightarrow$	$\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{i,t}^{(1)} \cdot \frac{A_{i,t}}{\hat{p}_{i,t}}$	
$= \langle u, v\rangle_{\mathbb{P}}$		$\approx \langle \hat{u}, \hat{v}\rangle_{\mathbb{P}}$	

error $O(\ \|\hat{u} - u\| + \|\hat{v} - v\|\ )$

Need to identify the terms from the observed data...

All events/expectations conditional on $\mathscr{F}$

$$= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\theta_{i,t}^{(1)} \frac{A_{i,t}}{p_{i,t}} \mid \mathscr{F} \right]$$

$$\uparrow \quad \uparrow$$

$$= \langle u, v \rangle_{\mathbb{P}}$$

$$\langle \hat{u}, v \rangle_{\mathbb{P}} \quad + \quad \langle u, \hat{v} \rangle_{\mathbb{P}} \quad - \quad \langle \hat{u}, \hat{v} \rangle_{\mathbb{P}}$$

error $O(\ \|\hat{u} - u\| \times \|\hat{v} - v\|\ )$

Need to identify the terms from the observed data...

All events/expectations conditional on $\mathscr{F}$

Assuming $\hat{\theta}^{(1)}_{i,t}, \hat{p}_{i,t} \perp\!\!\!\perp A_{i,t}$

$$\mathbb{E}\left[\, \hat{\theta}^{(1)}_{i,t} \frac{A_{i,t}}{\hat{p}_{i,t}} \,\right]$$

$$= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\theta^{(1)}_{i,t} \frac{A_{i,t}}{p_{i,t}} \mid \mathscr{F} \right]$$

$$= \langle u, v \rangle_{\mathbb{P}}$$

$$\langle \hat{u}, v \rangle_{\mathbb{P}} + \langle u, \hat{v} \rangle_{\mathbb{P}} - \langle \hat{u}, \hat{v} \rangle_{\mathbb{P}}$$

error $O(\, \|\hat{u} - u\| \times \|\hat{v} - v\| \,)$

Need to identify the terms from the observed data...

All events/expectations conditional on $\mathscr{F}$

Assuming $\hat{p}_{i,t} \perp\!\!\!\perp Y^{(1)}_{i,t}, A_{i,t}$

$$\mathbb{E}\left[\theta^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right] = \mathbb{E}\left[Y^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right] = \mathbb{E}\left[Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right]$$

Assuming $\hat{\theta}^{(1)}_{i,t}, \hat{p}_{i,t} \perp\!\!\!\perp A_{i,t}$

$$\mathbb{E}\left[\hat{\theta}^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\theta^{(1)}_{i,t}\frac{A_{i,t}}{p_{i,t}} \mid \mathscr{F}\right]$$

$$= \langle u, v\rangle_{\mathbb{P}}$$

$$\langle \hat{u}, v\rangle_{\mathbb{P}} + \langle u, \hat{v}\rangle_{\mathbb{P}} - \langle \hat{u}, \hat{v}\rangle_{\mathbb{P}}$$

error $O(\ \|\hat{u} - u\| \times \|\hat{v} - v\|\ )$

Need to identify the terms from the observed data...

All events/expectations conditional on $\mathscr{F}$

Assuming $\hat{\theta}^{(1)}_{i,t} \perp\!\!\!\perp A_{i,t}$

$$\mathbb{E}\left[\hat{\theta}^{(1)}_{i,t}\frac{A_{i,t}}{p_{i,t}}\right] = \hat{\theta}^{(1)}_{i,t}\mathbb{E}\left[\frac{A_{i,t}}{p_{i,t}}\right] = \hat{\theta}^{(1)}_{i,t}$$

Assuming $\hat{p}_{i,t} \perp\!\!\!\perp Y^{(1)}_{i,t}, A_{i,t}$

$$\mathbb{E}\left[\theta^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right] = \mathbb{E}\left[Y^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right] = \mathbb{E}\left[Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right]$$

Assuming $\hat{\theta}^{(1)}_{i,t}, \hat{p}_{i,t} \perp\!\!\!\perp A_{i,t}$

$$\mathbb{E}\left[\hat{\theta}^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\theta^{(1)}_{i,t}\frac{A_{i,t}}{p_{i,t}} \mid \mathscr{F}\right]$$

$$= \langle u, v\rangle_{\mathbb{P}}$$

$$\langle \hat{u}, v\rangle_{\mathbb{P}} + \langle u, \hat{v}\rangle_{\mathbb{P}} - \langle \hat{u}, \hat{v}\rangle_{\mathbb{P}}$$

error $O(\ \|\hat{u} - u\| \times \|\hat{v} - v\|\ )$

And thus arrives the doubly robust estimate

$$\mathbb{E}\Big[\hat{\theta}^{(1)}_{i,t}\frac{A_{i,t}}{p_{i,t}}\Big] = \hat{\theta}^{(1)}_{i,t}\mathbb{E}\Big[\frac{A_{i,t}}{p_{i,t}}\Big] = \hat{\theta}^{(1)}_{i,t}$$

*Assuming $\hat{p}_{i,t} \perp\!\!\!\perp Y^{(1)}_{i,t}, A_{i,t}$

$$\mathbb{E}\Big[\theta^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\Big] = \mathbb{E}\Big[Y^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\Big] = \mathbb{E}\Big[Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\Big]$$

*Assuming $\hat{\theta}^{(1)}_{i,t}, \hat{p}_{i,t} \perp\!\!\!\perp A_{i,t}$

$$\mathbb{E}\Big[\hat{\theta}^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}\Big]$$

(*in a block sense)

$$\frac{1}{N}\sum_{i=1}^{N} \quad \hat{\theta}^{(1)}_{i,t} \quad + \quad Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}} \quad - \quad \hat{\theta}^{(1)}_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}}$$

$$\langle \hat{u}, v \rangle_{\mathbb{P}} \quad + \quad \langle u, \hat{v} \rangle_{\mathbb{P}} \quad - \quad \langle \hat{u}, \hat{v} \rangle_{\mathbb{P}}$$

error $O(\ \|\hat{u} - u\| \times \|\hat{v} - v\|\ )$

And subsequently the doubly robust estimate for ATE

$$\widehat{\text{ATE}}_t^{\text{DR}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_{i,t}^{(1)} + Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(1)}\frac{A_{i,t}}{\hat{p}_{i,t}}\right) - \left(\hat{\theta}_{i,t}^{(0)} + Y_{i,t}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(0)}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}}\right)$$

$$\langle \hat{u}, v\rangle_{\mathbb{P}} + \langle u, \hat{v}\rangle_{\mathbb{P}} - \langle \hat{u}, \hat{v}\rangle_{\mathbb{P}}$$

And subsequently the doubly robust estimate for ATE

$$\widehat{\mathrm{ATE}}_t^{\mathrm{DR}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_{i,t}^{(1)} + Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(1)}\frac{A_{i,t}}{\hat{p}_{i,t}}\right) - \left(\hat{\theta}_{i,t}^{(0)} + Y_{i,t}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(0)}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}}\right)$$

$$\langle \hat{u}, v\rangle_{\mathbb{P}} + \langle u, \hat{v}\rangle_{\mathbb{P}} - \langle \hat{u}, \hat{v}\rangle_{\mathbb{P}}$$

And subsequently the doubly robust estimate for ATE

$$\widehat{\text{ATE}}_t^{\text{DR}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_{i,t}^{(1)} + Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(1)}\frac{A_{i,t}}{\hat{p}_{i,t}}\right) - \left(\hat{\theta}_{i,t}^{(0)} + Y_{i,t}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(0)}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}}\right)$$

$$\langle \hat{u}, v\rangle_{\mathbb{P}} + \langle u, \hat{v}\rangle_{\mathbb{P}} - \langle \hat{u}, \hat{v}\rangle_{\mathbb{P}}$$

Baselines

$$\widehat{\text{ATE}}_t^{\text{OI}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{i,t}^{(1)} - \hat{\theta}_{i,t}^{(0)}) \approx \langle \hat{u}, v\rangle_{\mathbb{P}}$$

And subsequently the doubly robust estimate for ATE

$$\widehat{\text{ATE}}_t^{\text{DR}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_{i,t}^{(1)} + Y_{i,t}\frac{A_{i,t}}{\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(1)}\frac{A_{i,t}}{\hat{p}_{i,t}}\right) - \left(\hat{\theta}_{i,t}^{(0)} + Y_{i,t}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}} - \hat{\theta}_{i,t}^{(0)}\frac{1-A_{i,t}}{1-\hat{p}_{i,t}}\right)$$

$$\langle \hat{u}, v\rangle_{\mathbb{P}} + \langle u, \hat{v}\rangle_{\mathbb{P}} - \langle \hat{u}, \hat{v}\rangle_{\mathbb{P}}$$

Baselines

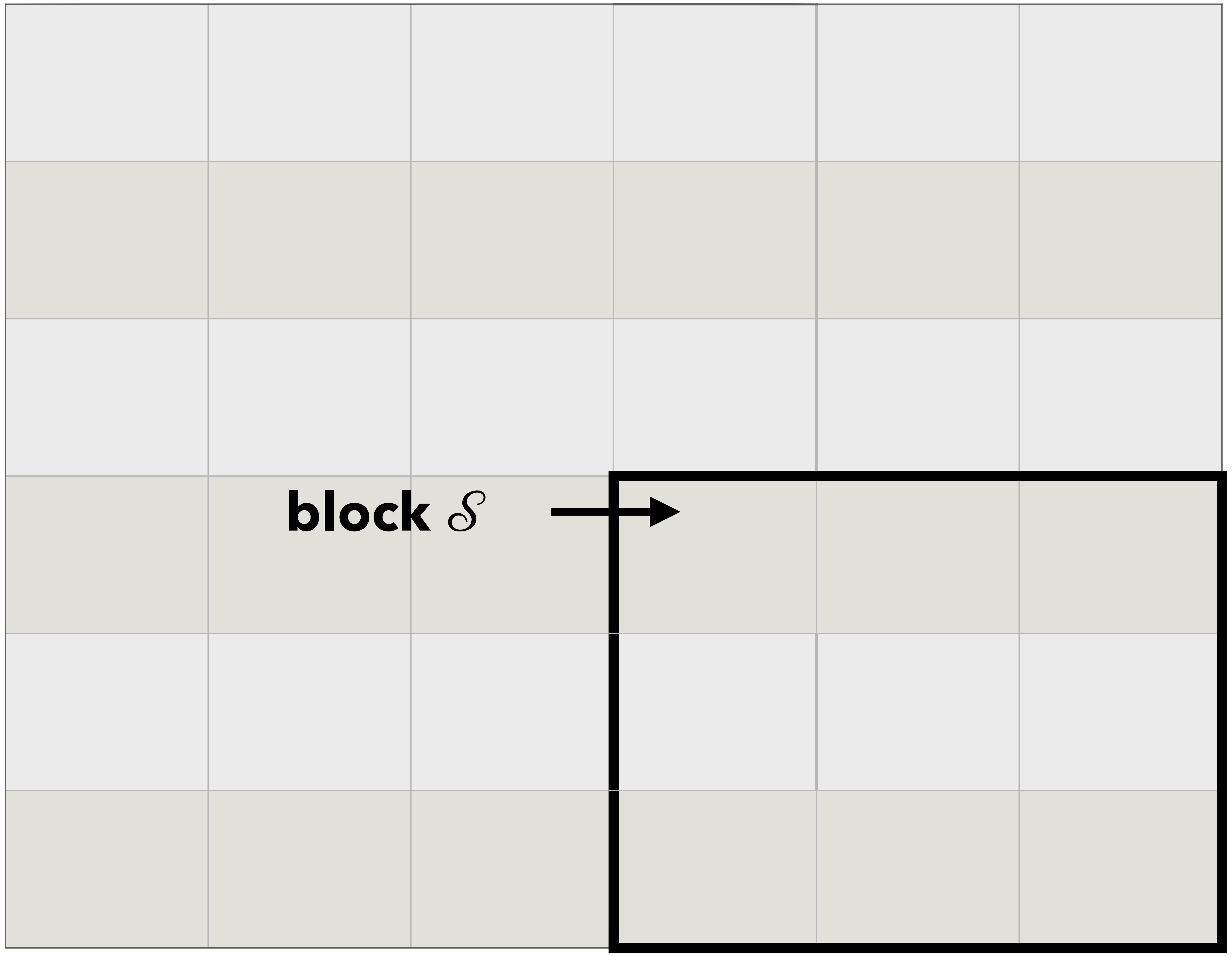
$$\widehat{\text{ATE}}_t^{\text{OI}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_{i,t}^{(1)} - \hat{\theta}_{i,t}^{(0)}) \approx \langle \hat{u}, v\rangle_{\mathbb{P}}$$

$$\widehat{\text{ATE}}_t^{\text{IPW}} = \frac{1}{N}\sum_{i=1}^{N} Y_{i,t}\left(\frac{A_{i,t}}{\hat{p}_{i,t}} - \frac{1-A_{i,t}}{1-\hat{p}_{i,t}}\right) \approx \langle u, \hat{v}\rangle_{\mathbb{P}}$$

Block independence for nuisance estimates

$$\hat{p}_{\mathcal{S}} \perp\!\!\!\perp Y^{(1)}_{\mathcal{S}}, Y^{(0)}_{\mathcal{S}}$$

$$\hat{\Theta}^{(1)}_{\mathcal{S}}, \hat{\Theta}^{(0)}_{\mathcal{S}}, \hat{p}_{\mathcal{S}} \perp\!\!\!\perp A_{\mathcal{S}}$$

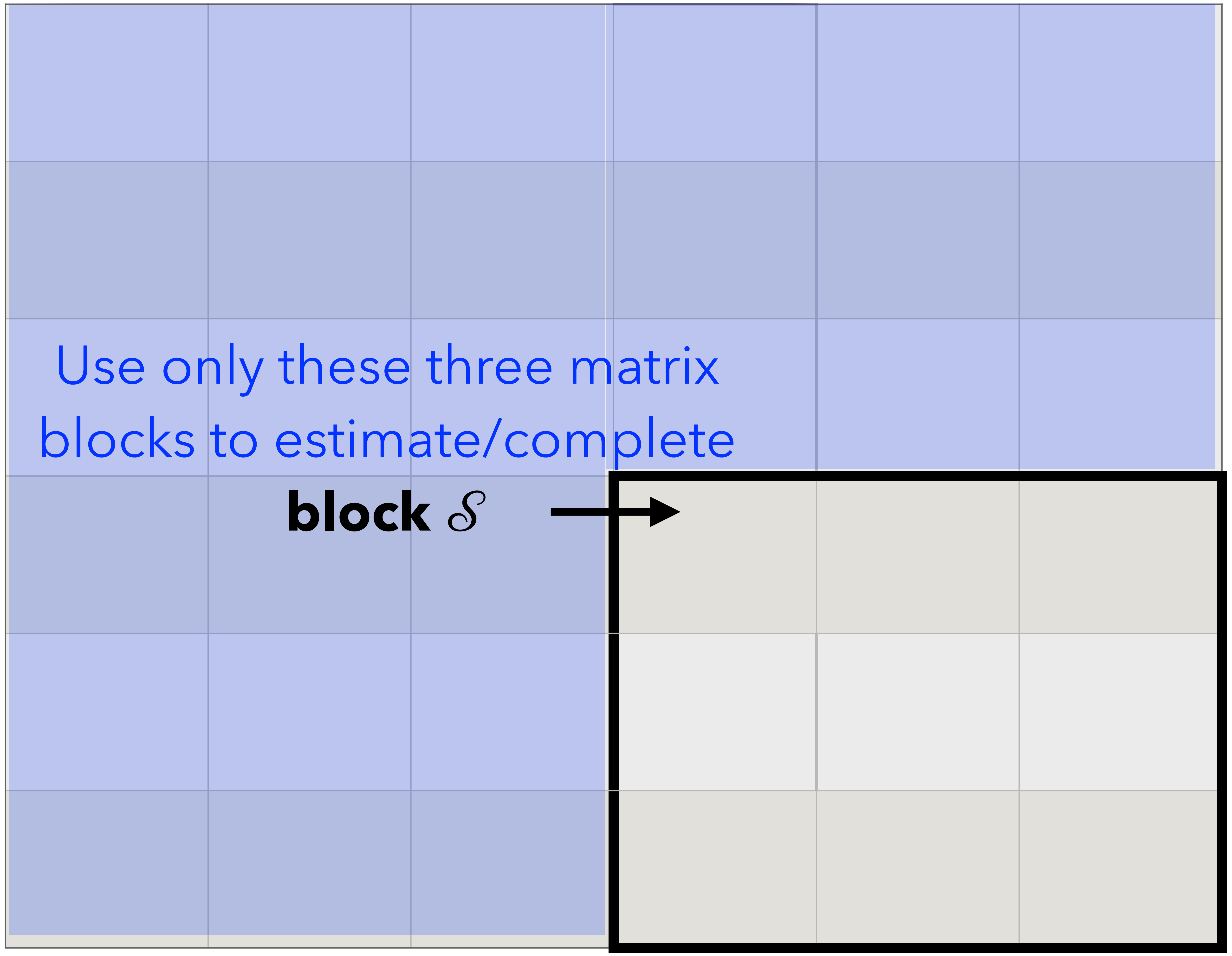


Block independence for nuisance estimates

$$\hat{p}_{\mathcal{S}} \perp\!\!\!\perp Y^{(1)}_{\mathcal{S}}, Y^{(0)}_{\mathcal{S}}$$

$$\hat{\Theta}^{(1)}_{\mathcal{S}}, \hat{\Theta}^{(0)}_{\mathcal{S}}, \hat{p}_{\mathcal{S}} \perp\!\!\!\perp A_{\mathcal{S}}$$

Via Cross-Fitted Matrix Completion



Error guarantees for DR estimate for ATE_t

Error guarantees for DR estimate for ATE$_t$

Informal theorem [Abadie-Agarwal-**Dwivedi**-Shah, '24]

Fix t. If for all i, we have

- unobserved confounding $(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \mathcal{F}$,
- positivity $p \leq P_{i,t} \leq 1 - p$,
- independent noise, and
- block independent matrix estimates,

Error guarantees for DR estimate for ATE_t

Informal theorem [Abadie-Agarwal-**Dwivedi**-Shah, '24]

Fix t. If for all i, we have

- unobserved confounding $(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \mathscr{F}$,
- positivity $p \leq P_{i,t} \leq 1 - p$,
- independent noise, and
- block independent matrix estimates,

then with high probability,

$$| \text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}} | \lesssim \frac{1}{p}\left[\frac{\|\Theta_t^{(a)} - \hat{\Theta}_t^{(a)}\|_2}{\sqrt{N}} \times \frac{\|P_t - \hat{P}_t\|_2}{\sqrt{N}} + \frac{1}{\sqrt{N}}\right]$$

Error guarantees for DR estimate for ATE_t

Informal theorem [Abadie-Agarwal-**Dwivedi**-Shah, '24]

Fix t. If for all i, we have

- unobserved confounding $(Y_{i,t}^{(1)}, Y_{i,t}^{(0)}) \perp\!\!\!\perp A_{i,t} \mid \mathscr{F}$,
- positivity $p \le P_{i,t} \le 1-p$,
- independent noise, and
- block independent matrix estimates,

then with high probability,

$$|\,\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}\,| \lesssim \frac{1}{p}\left[\frac{\|\Theta_t^{(a)} - \hat{\Theta}_t^{(a)}\|_2}{\sqrt{N}} \times \frac{\|P_t - \hat{P}_t\|_2}{\sqrt{N}} + \frac{1}{\sqrt{N}}\right]$$

Typical matrix-completion rates: $\dfrac{poly(rank)}{\sqrt{N \wedge T}}$

Error guarantees for DR estimate for ATE_t

$$|\,\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}\,| \lesssim \frac{1}{p}\left[\frac{\|\Theta_t^{(a)} - \hat{\Theta}_t^{(a)}\|_2}{\sqrt{N}} \times \frac{\|P_t - \hat{P}_t\|_2}{\sqrt{N}} + \frac{1}{\sqrt{N}}\right]$$

Error guarantees for DR estimate for ATE_t

$$|\,\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}\,| \lesssim \frac{1}{p}\left[\frac{\|\Theta_t^{(a)} - \hat{\Theta}_t^{(a)}\|_2}{\sqrt{N}} \times \frac{\|P_t - \hat{P}_t\|_2}{\sqrt{N}} + \frac{1}{\sqrt{N}}\right]$$

Bias

Error guarantees for DR estimate for ATE_t

If **bias** $= o_p(N^{-1/2})$

$$\sqrt{N}(\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}) \longrightarrow \mathcal{N}(0, \sigma^2)$$

$$|\,\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}\,| \lesssim \frac{1}{p}\left[\frac{\|\Theta_t^{(a)} - \hat{\Theta}_t^{(a)}\|_2}{\sqrt{N}} \times \frac{\|P_t - \hat{P}_t\|_2}{\sqrt{N}} + \frac{1}{\sqrt{N}}\right]$$

Bias

Error guarantees for DR estimate for ATE_t

Generic matrix completion

- No asymptotic normality
- Slow error rates with large ranks

If **bias** $= o_p(N^{-1/2})$

$$\sqrt{N}(\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}) \longrightarrow \mathcal{N}(0, \sigma^2)$$

$$|\,\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}\,| \lesssim \frac{1}{p}\left[\frac{\|\Theta_t^{(a)} - \hat{\Theta}_t^{(a)}\|_2}{\sqrt{N}} \times \frac{\|P_t - \hat{P}_t\|_2}{\sqrt{N}} + \frac{1}{\sqrt{N}}\right]$$

Bias

Error guarantees for DR estimate for ATE_t

Doubly robust to the rank of outcome & propensity matrices

If **bias** $= o_p(N^{-1/2})$

$$\sqrt{N}(\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}) \longrightarrow \mathcal{N}(0, \sigma^2)$$

Generic matrix completion

- No asymptotic normality
- Slow error rates with large ranks

$$|\,\text{ATE}_t - \widehat{\text{ATE}_t}^{\text{DR}}\,| \lesssim \frac{1}{p}\left[\frac{\|\Theta_t^{(a)} - \hat{\Theta}_t^{(a)}\|_2}{\sqrt{N}} \times \frac{\|P_t - \hat{P}_t\|_2}{\sqrt{N}} + \frac{1}{\sqrt{N}}\right]$$

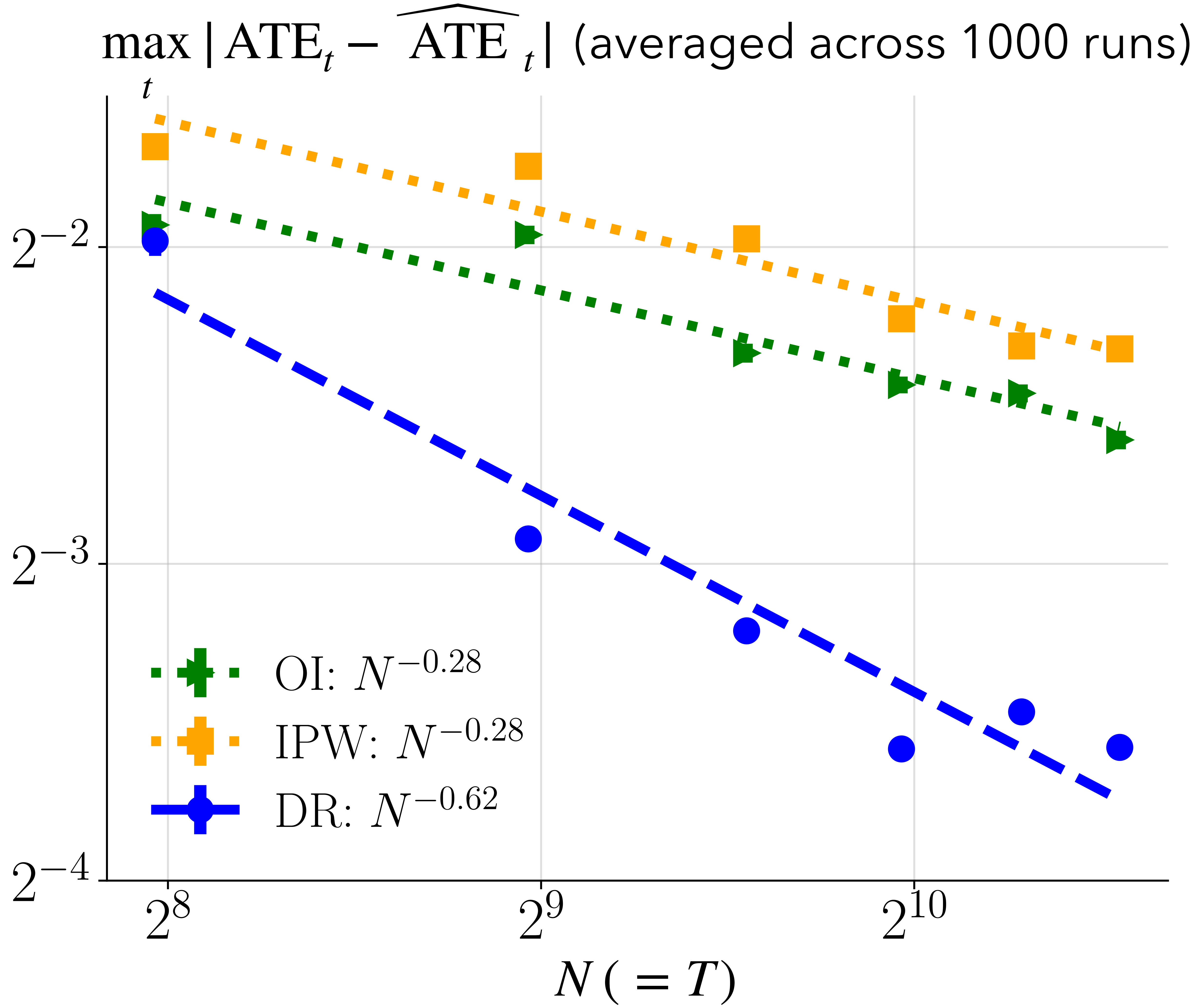
Bias

Simulation results with growing ranks

Uniform factors with $\text{rank}(\Theta^{(a)}) = N^{1/4}$, $\text{Rank}(P) = N^{1/5}$

Simulation results with growing ranks

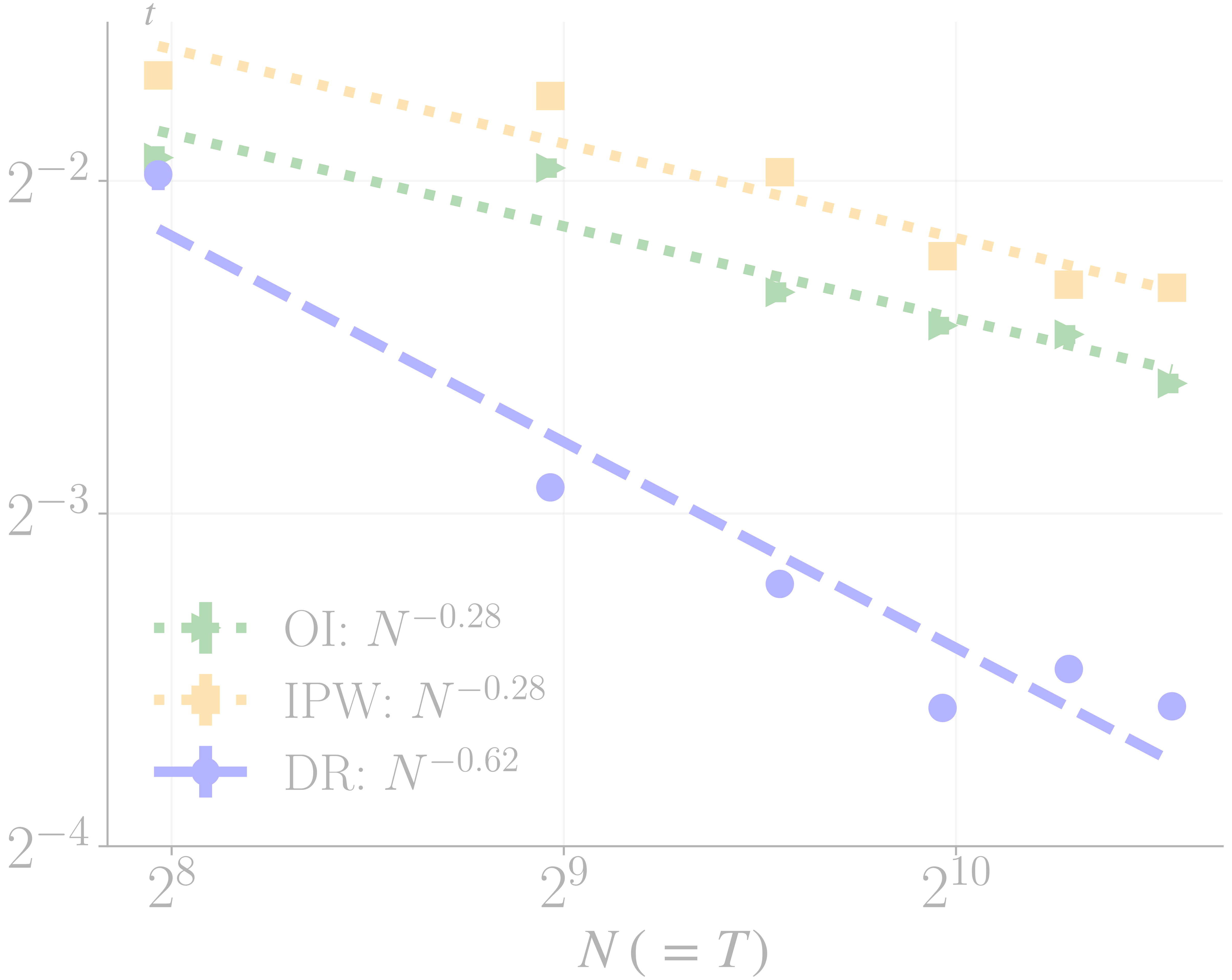
Uniform factors with rank$(\Theta^{(a)}) = N^{1/4}$, Rank$(P) = N^{1/5}$

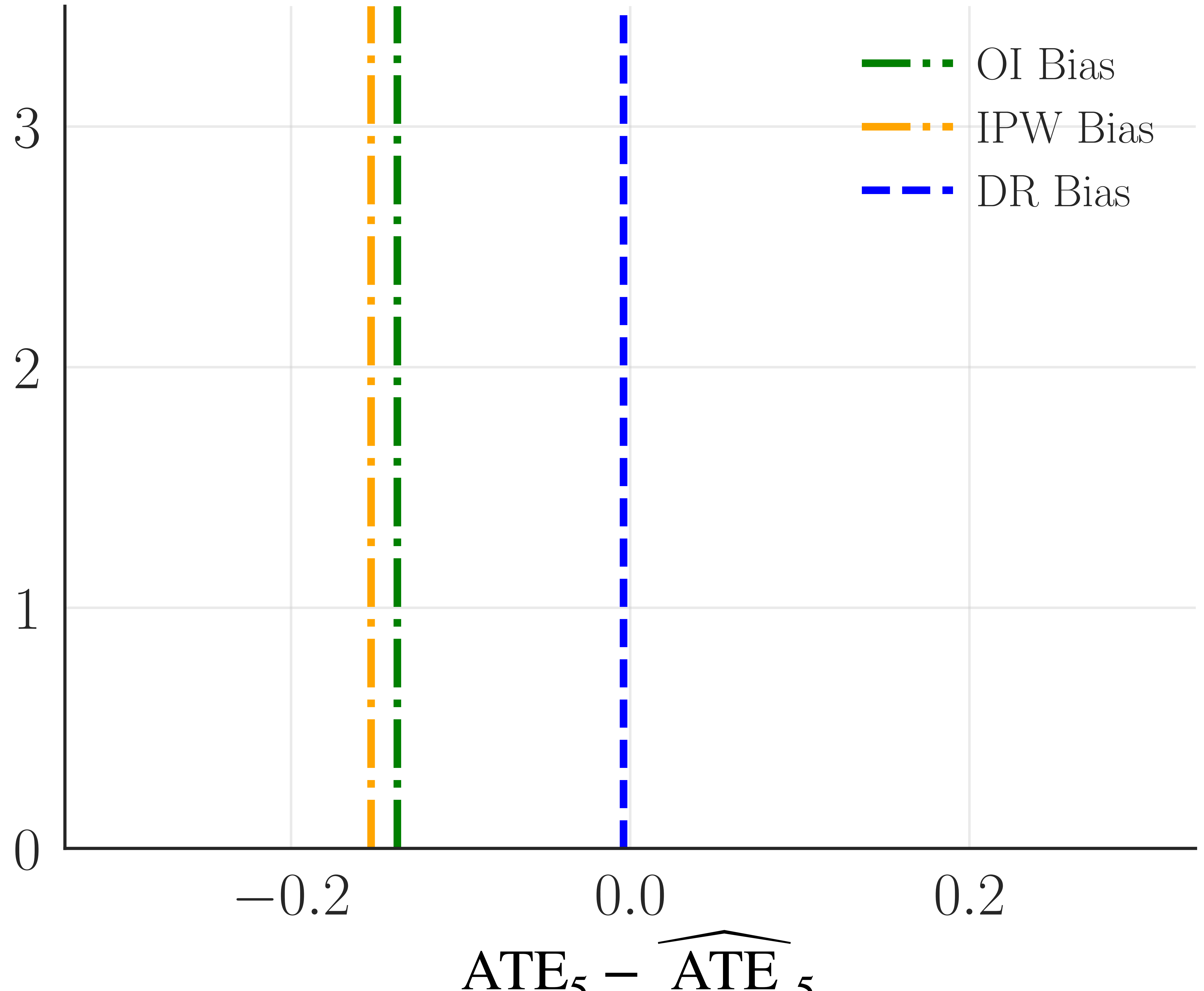


Simulation results with growing ranks

Uniform factors with $\text{rank}(\Theta^{(a)}) = N^{1/4}$, $\text{Rank}(P) = N^{1/5}$

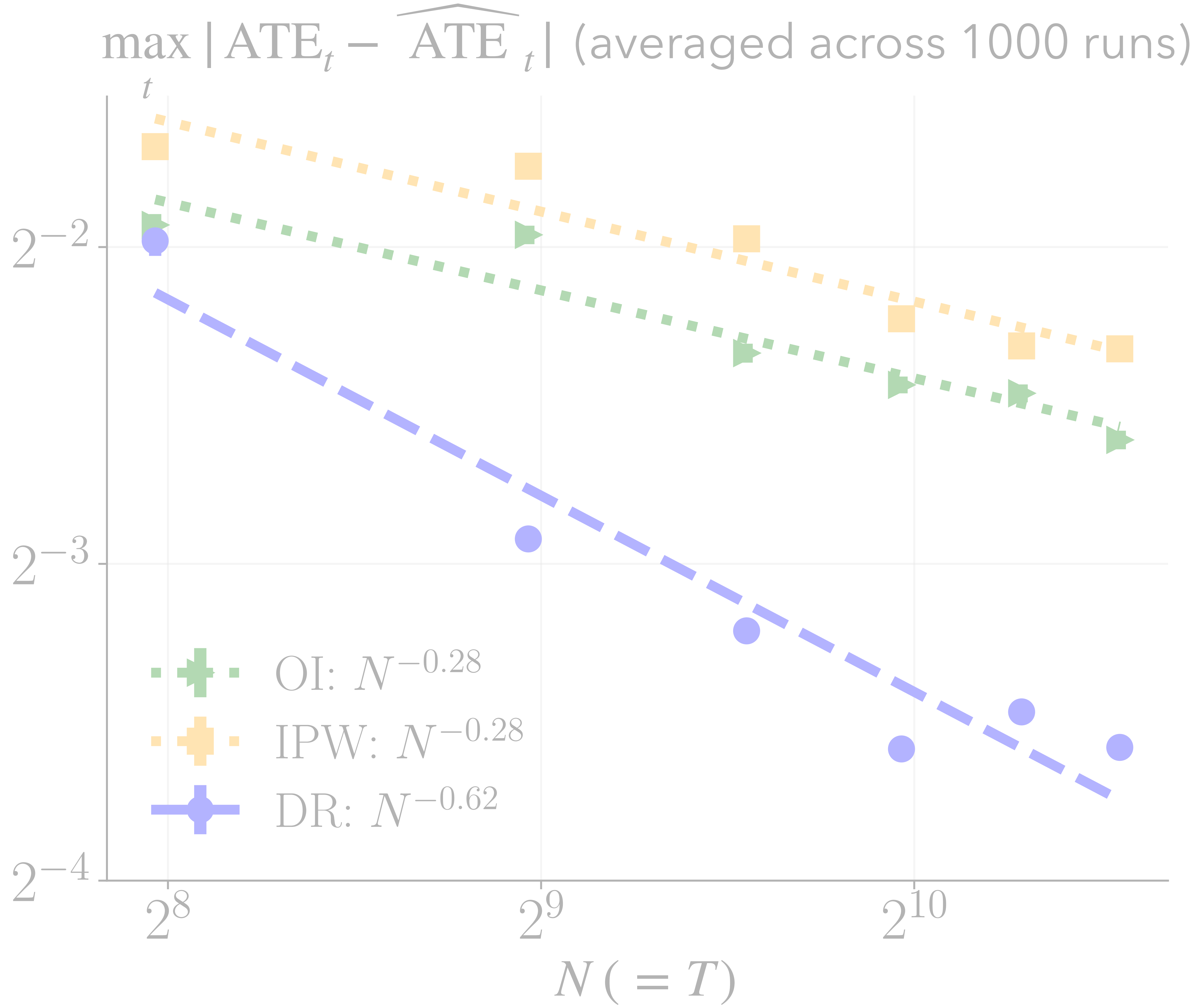
$\max_t |\text{ATE}_t - \widehat{\text{ATE}}_t|$ (averaged across 1000 runs)

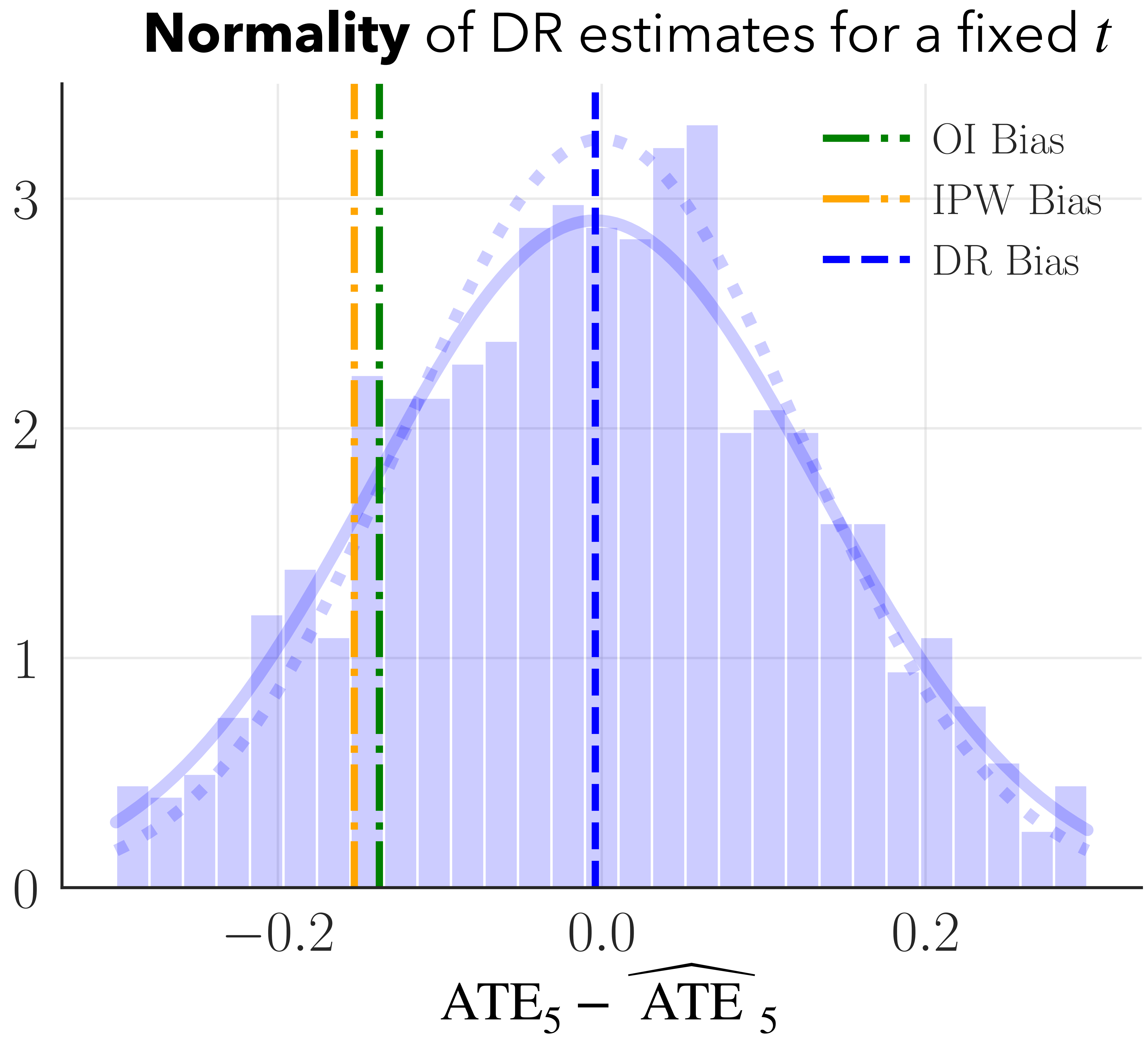




Simulation results with growing ranks

Uniform factors with $\text{rank}(\Theta^{(a)}) = N^{1/4}$, $\text{Rank}(P) = N^{1/5}$

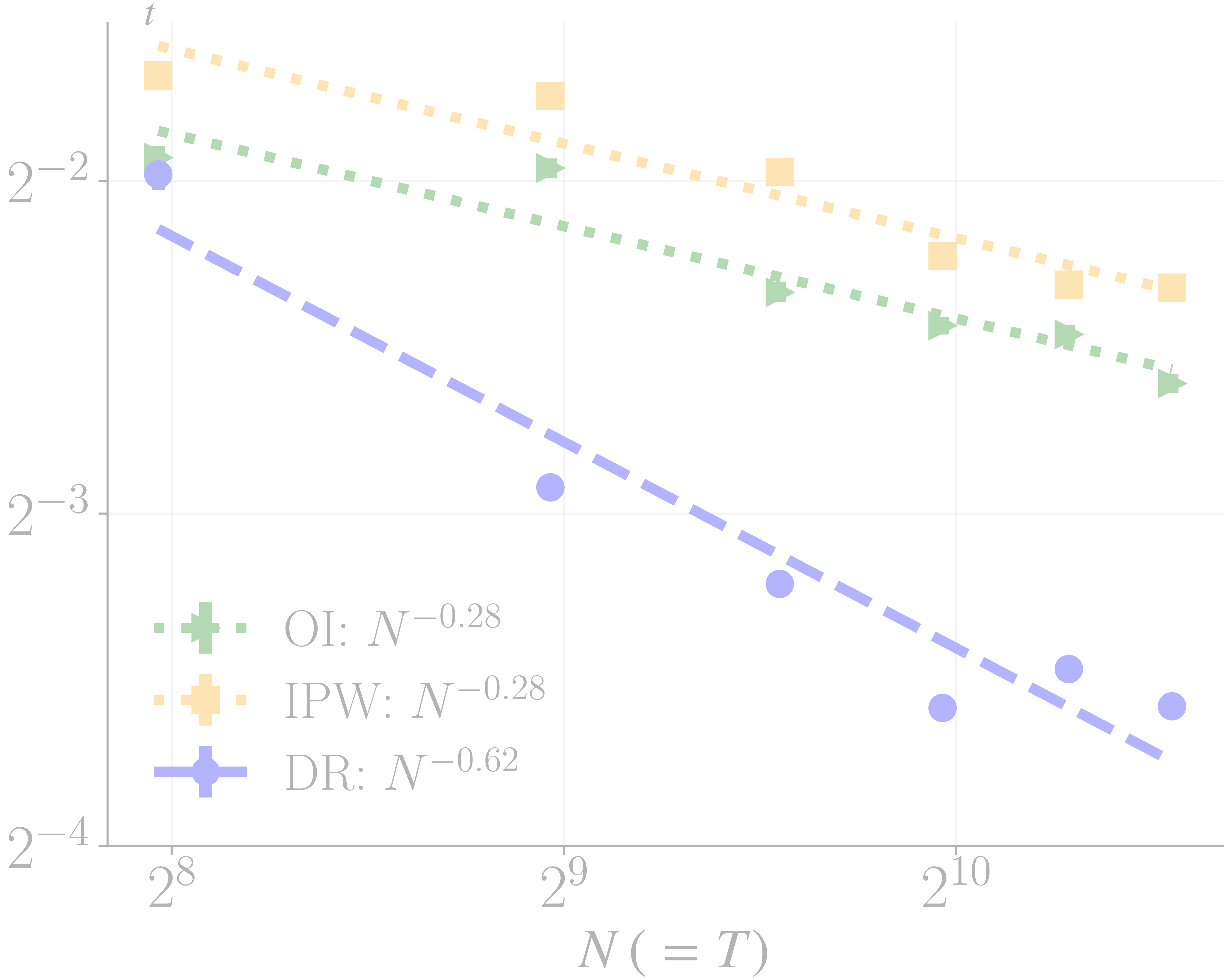




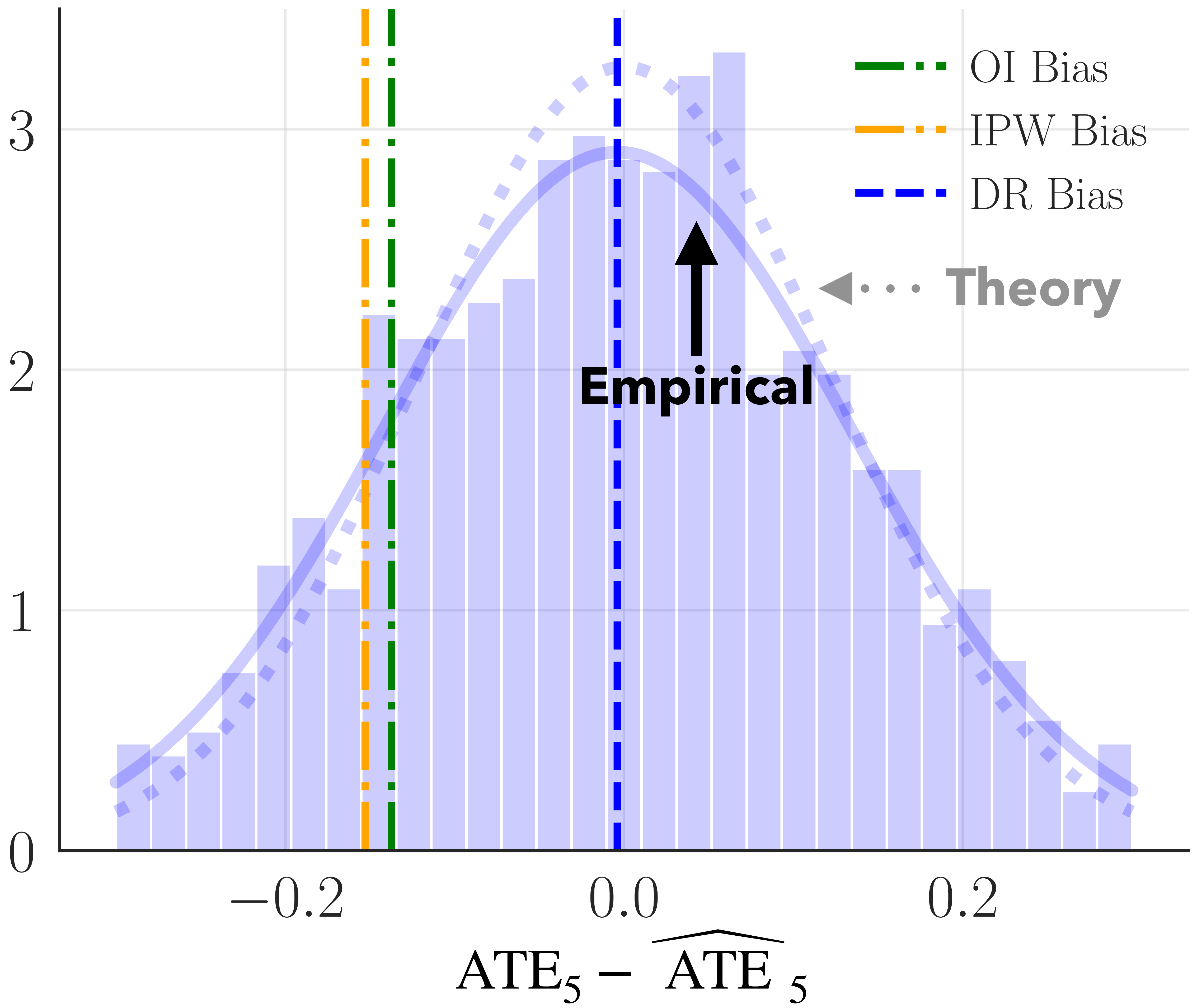
Simulation results with growing ranks

Uniform factors with $\text{rank}(\Theta^{(a)}) = N^{1/4}$, $\text{Rank}(P) = N^{1/5}$

$\max_t |\text{ATE}_t - \widehat{\text{ATE}}_t|$ (averaged across 1000 runs)



Normality of DR estimates for a fixed t



But, what if the outcomes **do have** a low-rank structure?

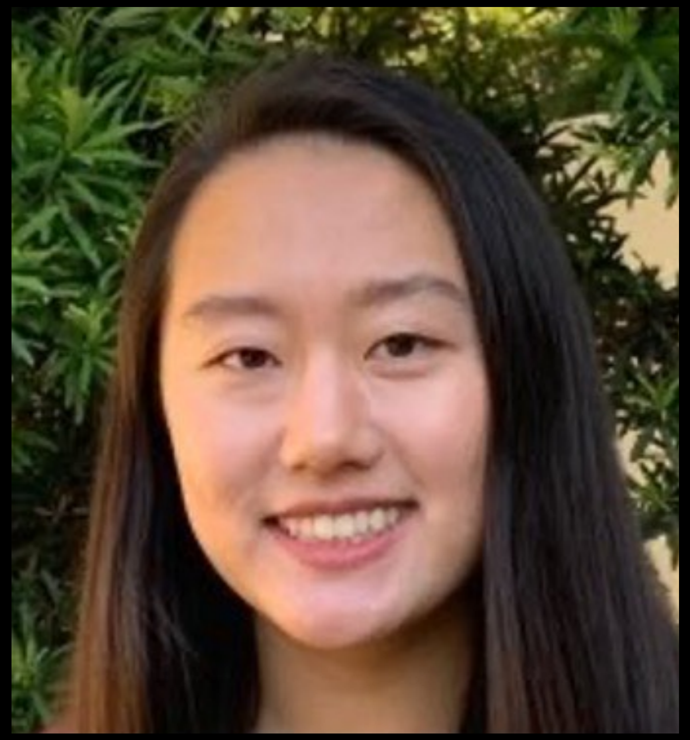
Can we hope to estimate $\mathrm{ITE}_{i,t} = \theta_{i,t}^{(1)} - \theta_{i,t}^{(0)}$?

But, what if the outcomes **do have** a low-rank structure?

Can we hope to estimate $\mathrm{ITE}_{i,t} = \theta_{i,t}^{(1)} - \theta_{i,t}^{(0)}$?

In this talk, henceforth assume $p_{i,t} \equiv p$

Part 2:
Doubly robust nearest neighbors for estimating ITE



Katherine Tian

Sabina Tomkins

Predrag Klasnja

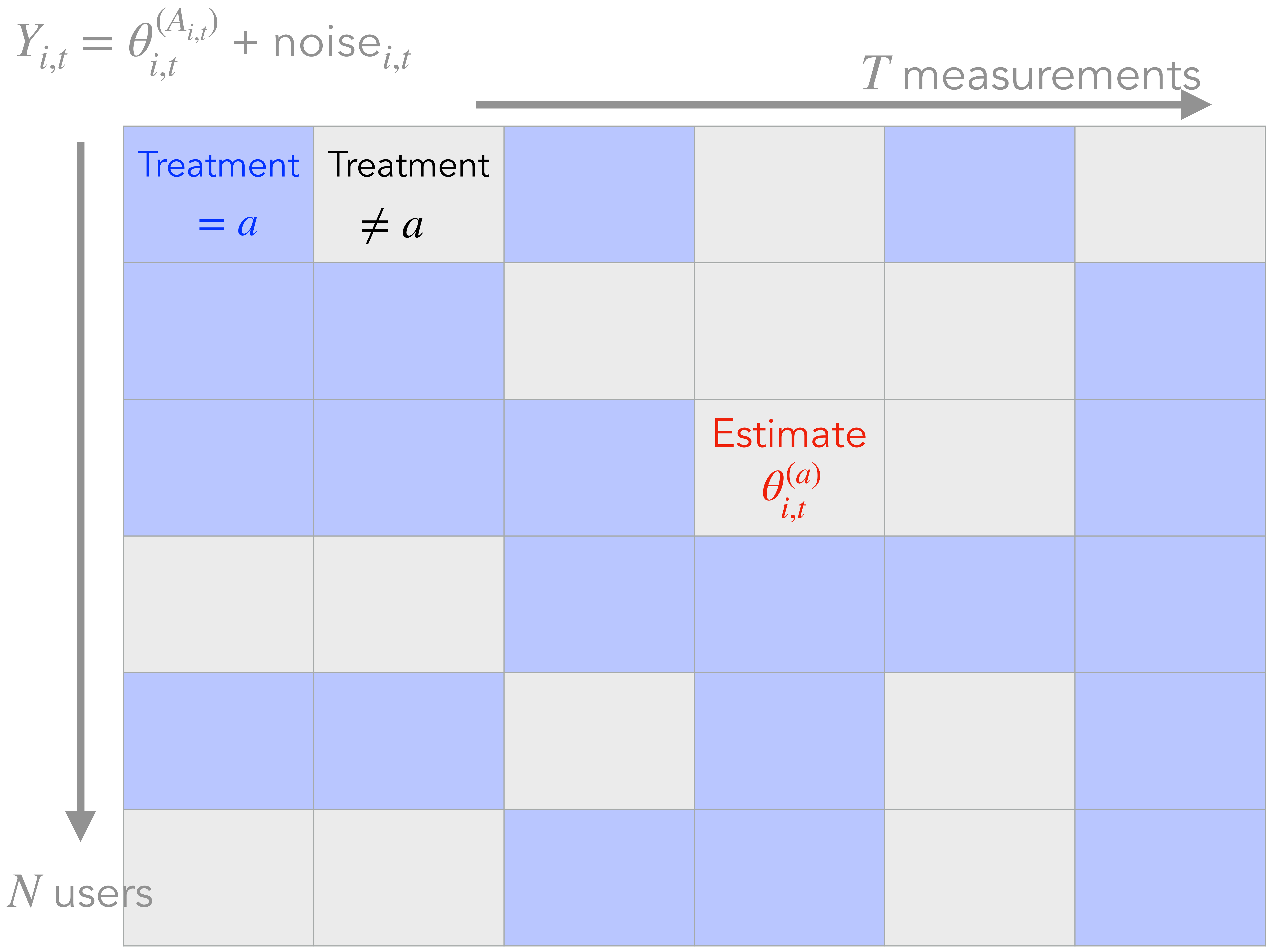
Susan Murphy

Devavrat Shah

https://arxiv.org/abs/2202.06891

https://arxiv.org/abs/2211.14297

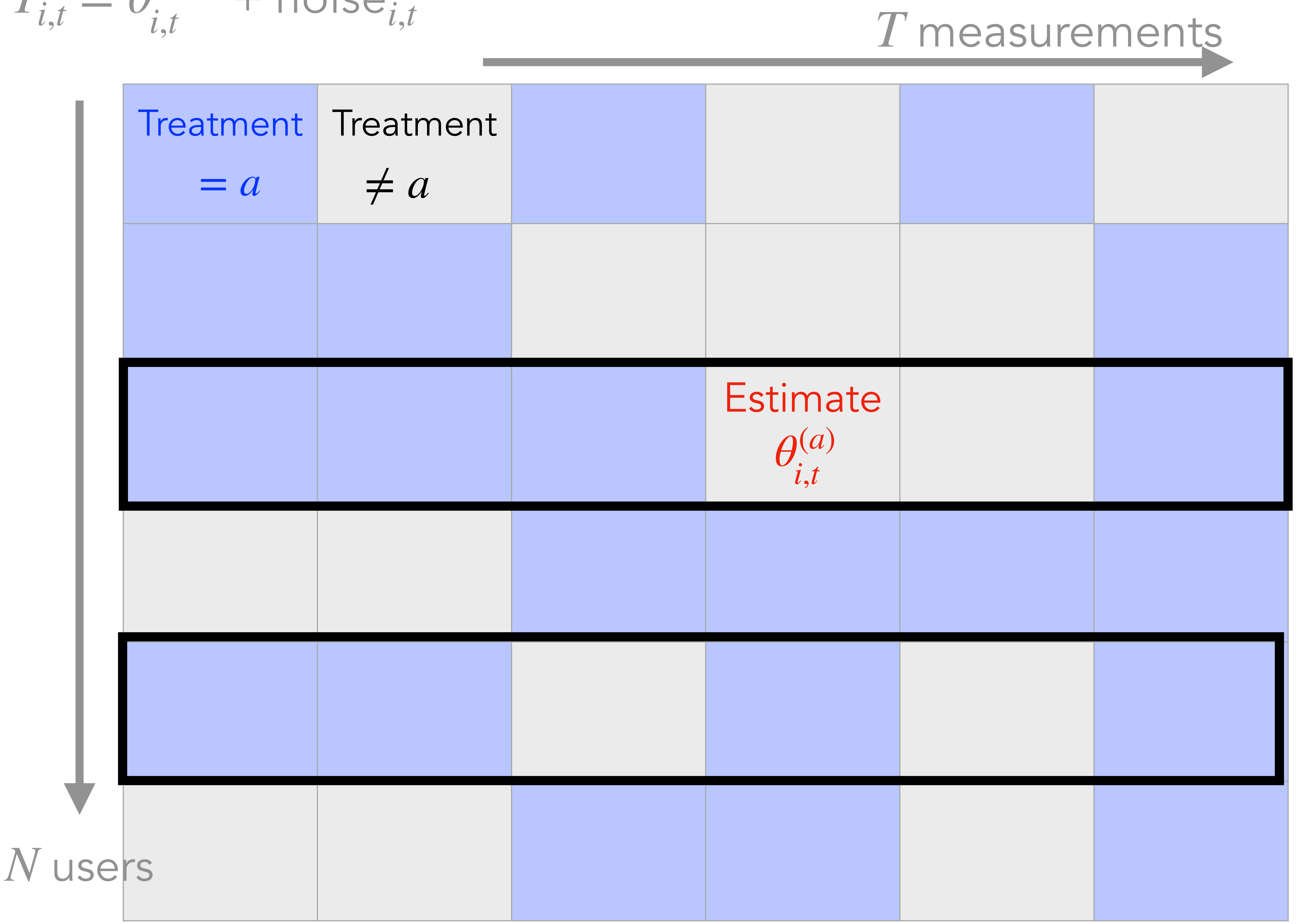
A common approach for ITE: Nearest neighbors



A common approach for ITE: Nearest neighbors

$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \text{noise}_{i,t}$

1. Compute distance b/w users i and j

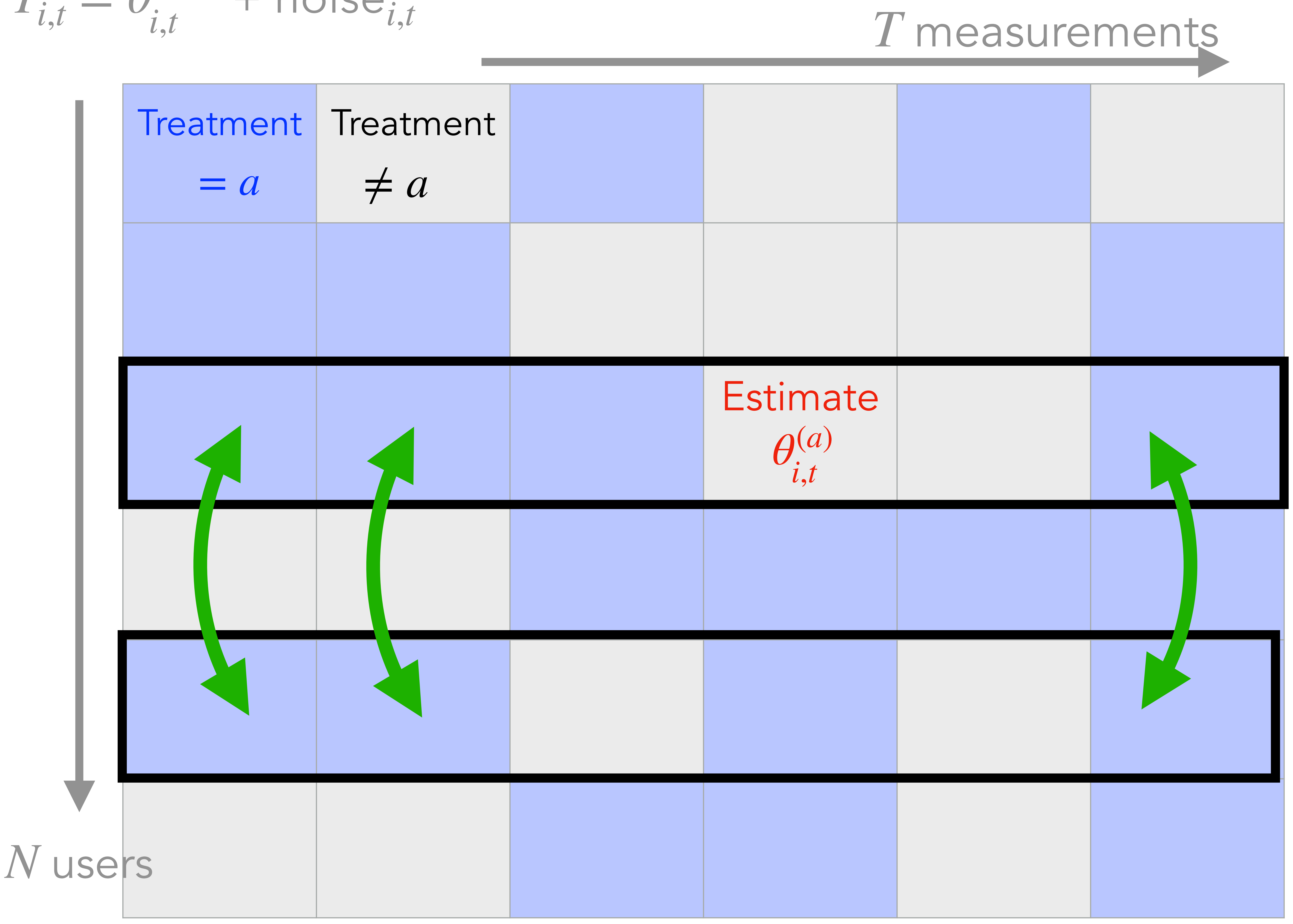


A common approach for ITE: Nearest neighbors

$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \text{noise}_{i,t}$

1. Compute distance b/w users i and j

$$\rho_{i,j}^{(a)} = \frac{\sum_{t' \neq t} (Y_{i,t'} - Y_{j,t'})^2 \cdot \mathbf{1}(A_{i,t'} = A_{j,t'} = a)}{\sum_{t' \neq t} \mathbf{1}(A_{i,t'} = A_{j,t'} = a)}$$



A common approach for ITE: Nearest neighbors

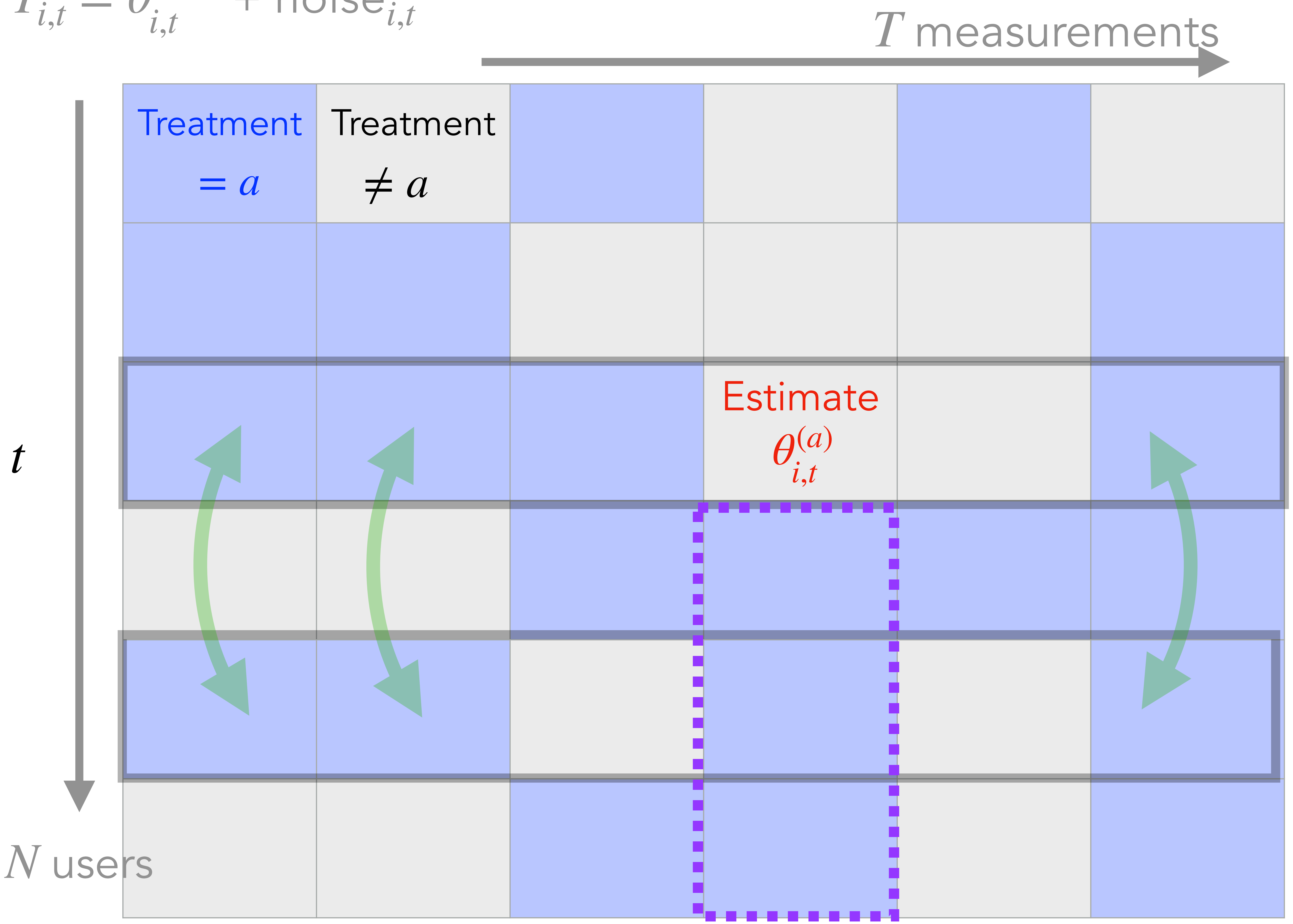
$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \text{noise}_{i,t}$

1. Compute distance b/w users i and j

$$\rho_{i,j}^{(a)} = \frac{\sum_{t' \neq t} (Y_{i,t'} - Y_{j,t'})^2 \cdot \mathbf{1}(A_{i,t'} = A_{j,t'} = a)}{\sum_{t' \neq t} \mathbf{1}(A_{i,t'} = A_{j,t'} = a)}$$

2. Average neighbor outcomes at time t

$$\widehat{\theta}_{i,t,\text{user-NN}}^{(a)} = \frac{\sum_{j=1}^{N} Y_{j,t} \mathbf{1}(\rho_{i,j}^{(a)} \leq \eta, A_{j,t} = a)}{\sum_{j=1}^{N} \mathbf{1}(\rho_{i,j}^{(a)} \leq \eta, A_{j,t} = a)}$$



A common approach for ITE: Nearest neighbors

$Y_{i,t} = \theta_{i,t}^{(A_{i,t})} + \text{noise}_{i,t}$

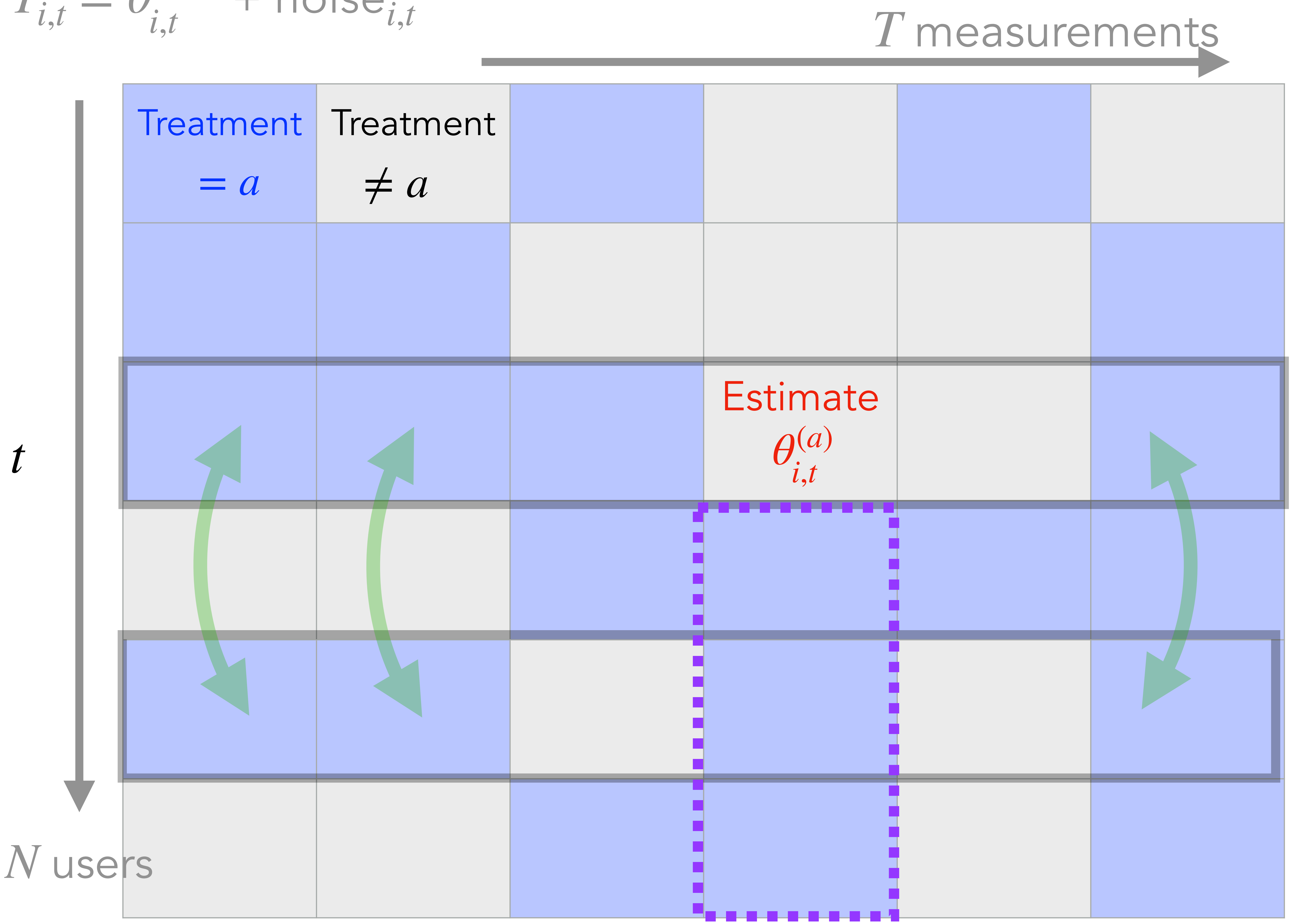
1. Compute distance b/w users i and j

$$\rho_{i,j}^{(a)} = \frac{\sum_{t' \neq t} (Y_{i,t'} - Y_{j,t'})^2 \cdot \mathbf{1}(A_{i,t'} = A_{j,t'} = a)}{\sum_{t' \neq t} \mathbf{1}(A_{i,t'} = A_{j,t'} = a)}$$

2. Average neighbor outcomes at time t

$$\hat{\theta}_{i,t,\text{user-NN}}^{(a)} = \frac{\sum_{j=1}^{N} Y_{j,t} \mathbf{1}(\rho_{i,j}^{(a)} \leq \eta, A_{j,t} = a)}{\sum_{j=1}^{N} \mathbf{1}(\rho_{i,j}^{(a)} \leq \eta, A_{j,t} = a)}$$

3. Do this procedure for $a = 0$ and 1.



Entry-wise guarantees for User-NN

Entry-wise guarantees for User-NN

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t}\,| \lesssim \sqrt{\eta} + \frac{1}{(\#overlap)^{1/4}} + \frac{1}{\sqrt{p \cdot \#Row\ Neighbors\ within\ \eta}}$$

Entry-wise guarantees for User-NN

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{user–NN}} - \theta^{(a)}_{i,t}\,| \lesssim \sqrt{\eta} + \frac{1}{(\#overlap)^{1/4}} + \frac{1}{\sqrt{p \cdot \#Row\ Neighbors\ within\ \eta}}$$

$$\text{overlap} = \sum_{t' \neq t} \mathbf{1}(A_{i,t'} = A_{j,t'} = a) \ (\approx p^2 T \ \text{ for } p_{i,t} \equiv p)$$

Entry-wise guarantees for User-NN

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{user–NN}} - \theta^{(a)}_{i,t}| \;\lesssim\; \sqrt{\eta} \;+\; \frac{1}{(\#overlap)^{1/4}} \;+\; \frac{1}{\sqrt{p \cdot \#Row\ Neighbors\ within\ \eta}}$$

overlap = $\sum_{t' \neq t} \mathbf{1}(A_{i,t'} = A_{j,t'} = a)$ ($\approx p^2 T$ for $p_{i,t} \equiv p$)

Row factor distribution

uniform over a finite set of size M → $\dfrac{1}{(p^2T)^{1/4}} + \dfrac{1}{\sqrt{pN/M}}$

Error rates after tuning η

Entry-wise guarantees for User-NN

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t}| \;\lesssim\; \sqrt{\eta} \;+\; \frac{1}{(\#overlap)^{1/4}} \;+\; \frac{1}{\sqrt{p \cdot \#Row\ Neighbors\ within\ \eta}}$$

overlap $= \sum_{t' \neq t} \mathbf{1}(A_{i,t'} = A_{j,t'} = a)$ ($\approx p^2 T$ for $p_{i,t} \equiv p$)

Row factor distribution

uniform over a finite set of size M → $\frac{1}{(p^2T)^{1/4}} + \frac{1}{\sqrt{pN/M}}$

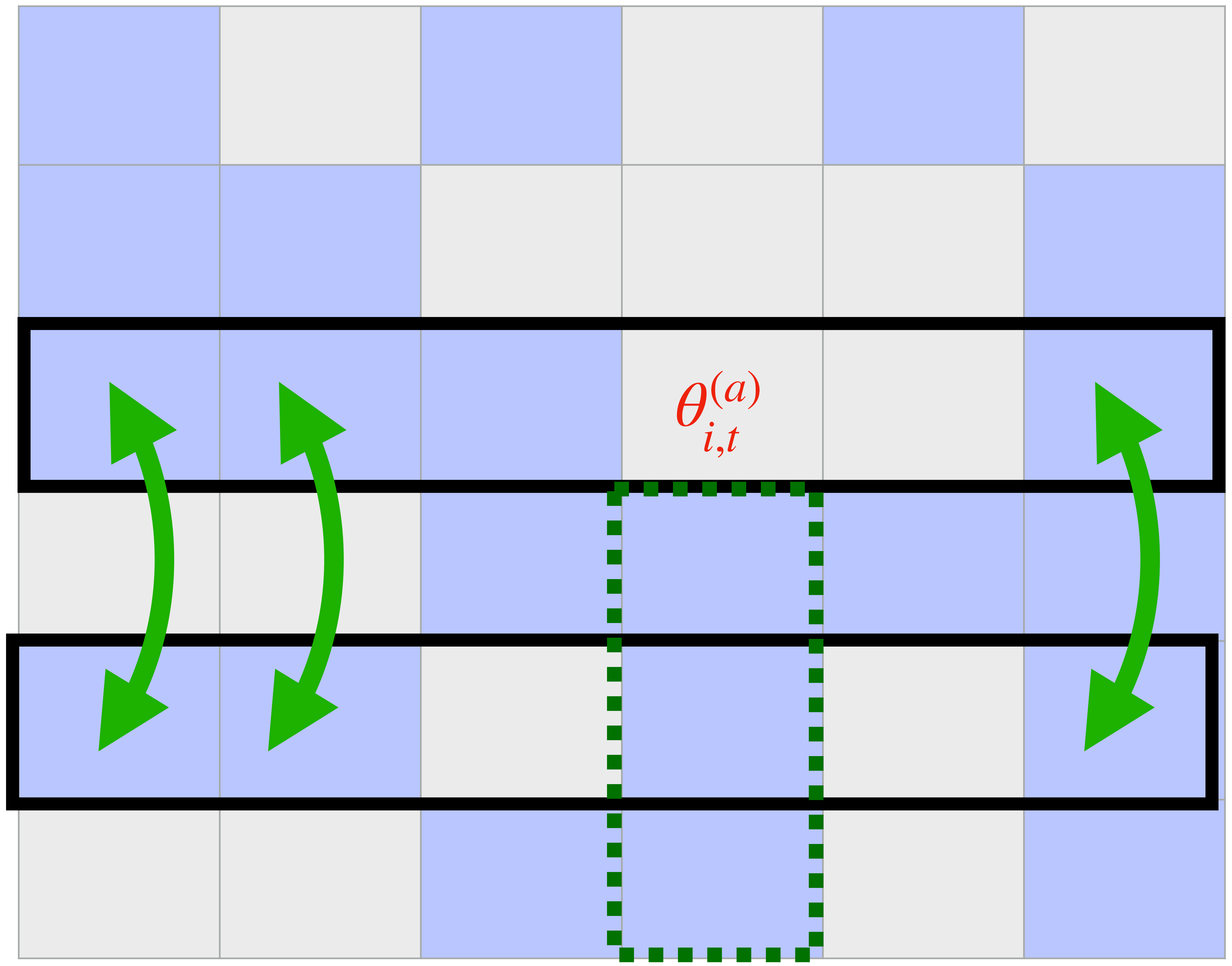
Uniform in $[-1,1]^d$ → $\frac{1}{(p^2T)^{1/4}} + \frac{1}{\sqrt{pN^{2/(d+2)}}}$

Error rates after tuning η

Two variants of nearest neighbors

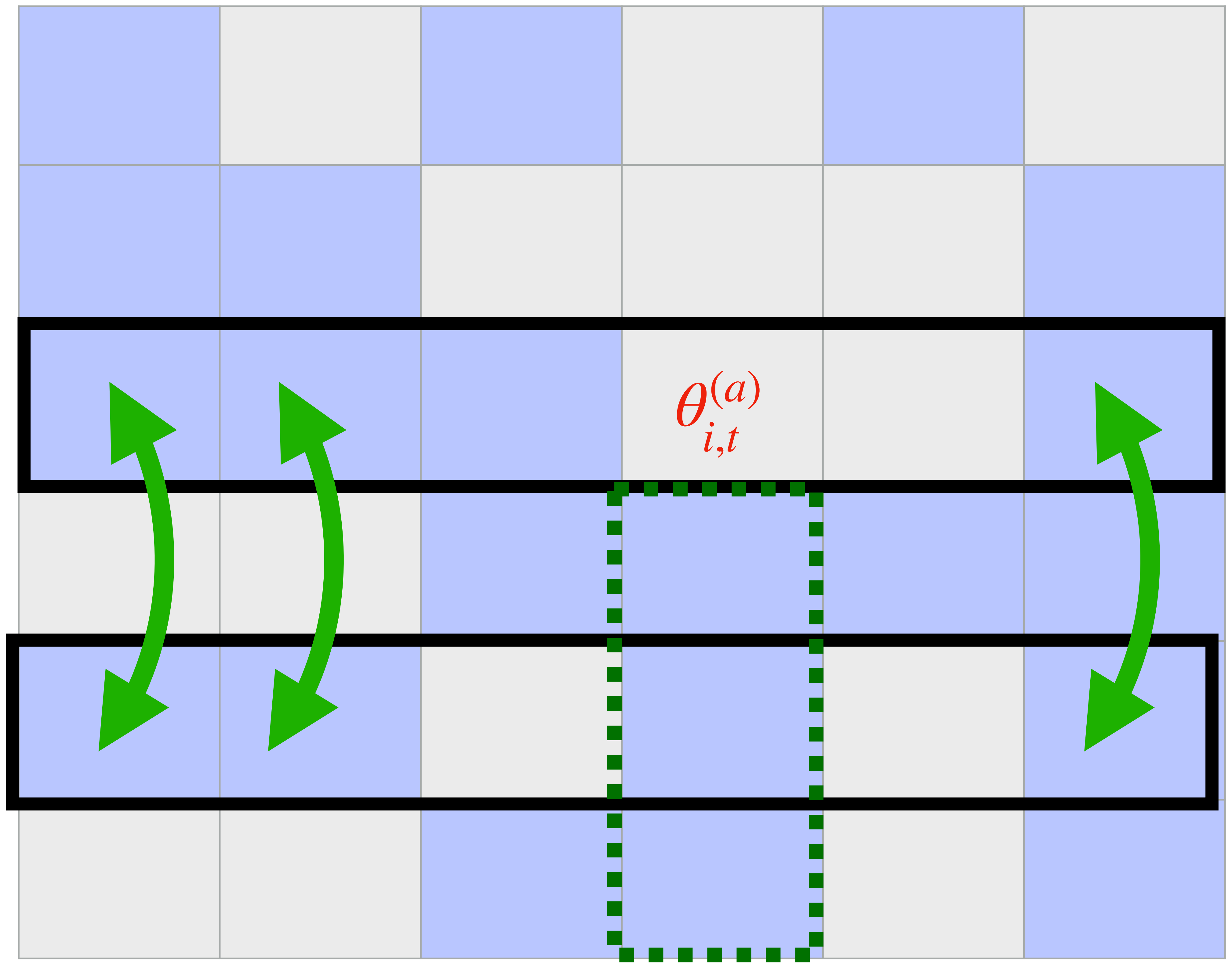
Two variants of nearest neighbors

User-NN

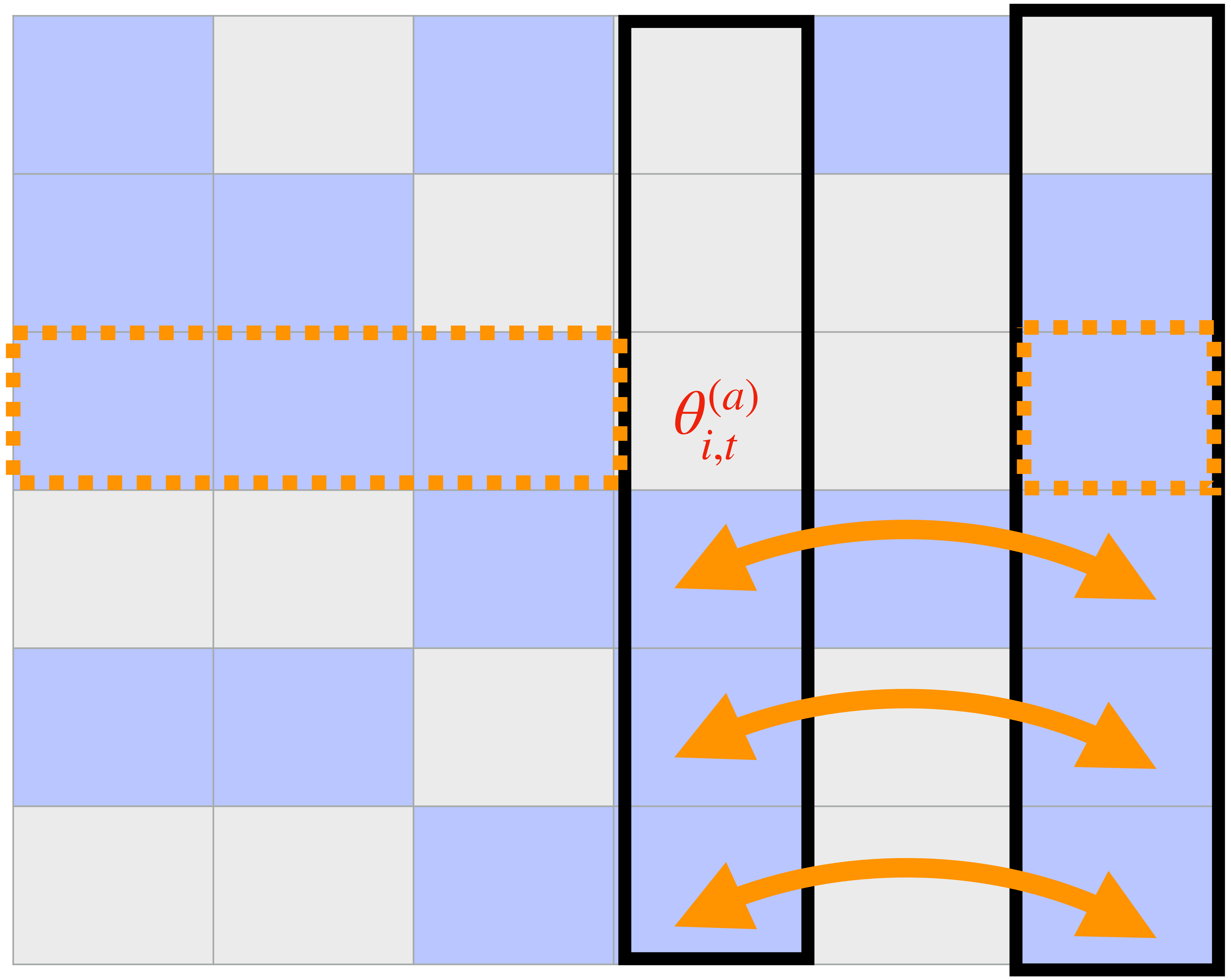


Two variants of nearest neighbors

User-NN



Time-NN



$$| \widehat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t} | = \tilde{O}\left(\frac{1}{T^{1/4}} + \frac{1}{\sqrt{N}}\right)$$

↓

$$| \widehat{\theta}^{(a)}_{i,t,\text{time-NN}} - \theta^{(a)}_{i,t} | = \tilde{O}\left(\frac{1}{N^{1/4}} + \frac{1}{\sqrt{T}}\right)$$

How do we improve the slow error rates?

$$|\widehat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t}| = \tilde{O}\left(\frac{1}{T^{1/4}} + \frac{1}{\sqrt{N}}\right)$$

↓

$$|\widehat{\theta}^{(a)}_{i,t,\text{time-NN}} - \theta^{(a)}_{i,t}| = \tilde{O}\left(\frac{1}{N^{1/4}} + \frac{1}{\sqrt{T}}\right)$$

How do we improve the slow error rates?

Simulation results with N=T

Uniform factors on $[-0.5,0.5]^4$, Gaussian noise,

pooled ε-greedy policy ($\varepsilon = 0.5$)

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t}\,| = \tilde{O}\left(\frac{1}{T^{1/4}} + \frac{1}{\sqrt{N}}\right)$$

↓

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{time-NN}} - \theta^{(a)}_{i,t}\,| = \tilde{O}\left(\frac{1}{N^{1/4}} + \frac{1}{\sqrt{T}}\right)$$

How do we improve the slow error rates?

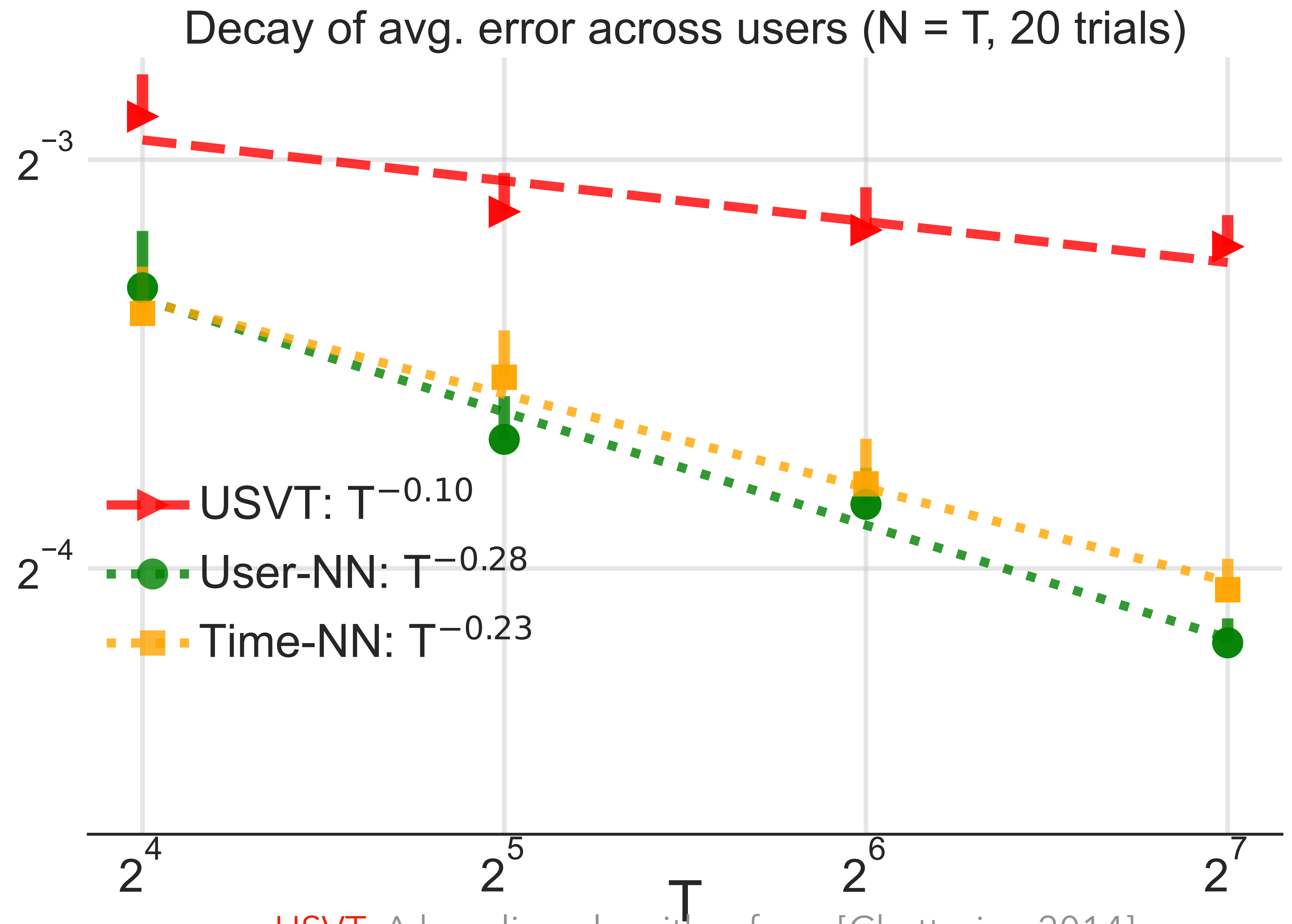
$$|\,\widehat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t}\,| = \tilde{O}\left(\frac{1}{T^{1/4}} + \frac{1}{\sqrt{N}}\right)$$

↓

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{time-NN}} - \theta^{(a)}_{i,t}\,| = \tilde{O}\left(\frac{1}{N^{1/4}} + \frac{1}{\sqrt{T}}\right)$$

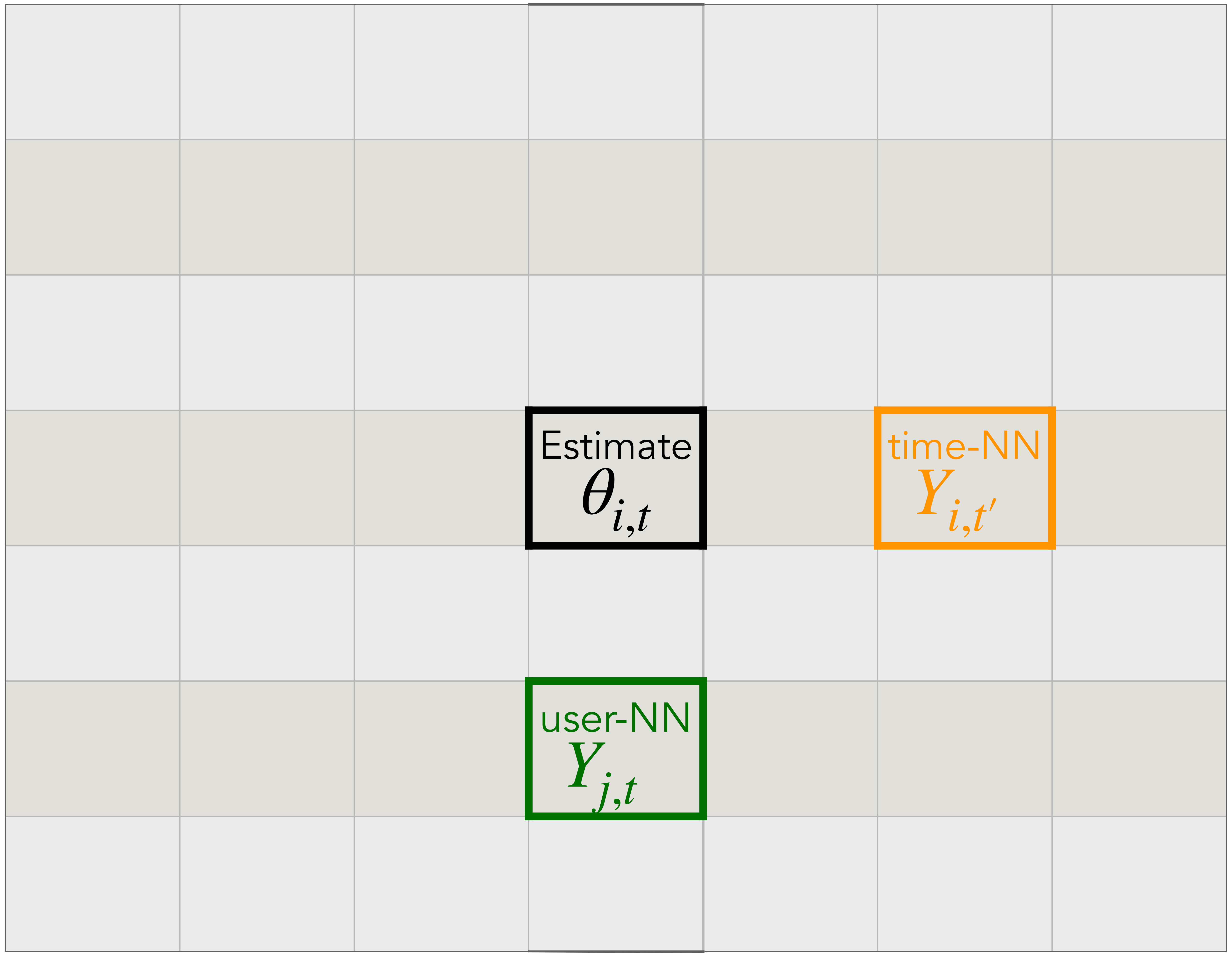
Simulation results with N=T

Uniform factors on $[-0.5,0.5]^4$, Gaussian noise, pooled ε-greedy policy ($\varepsilon = 0.5$)



USVT: A baseline algorithm from [Chatterjee 2014]

Integrating **double robustness** with nearest neighbors



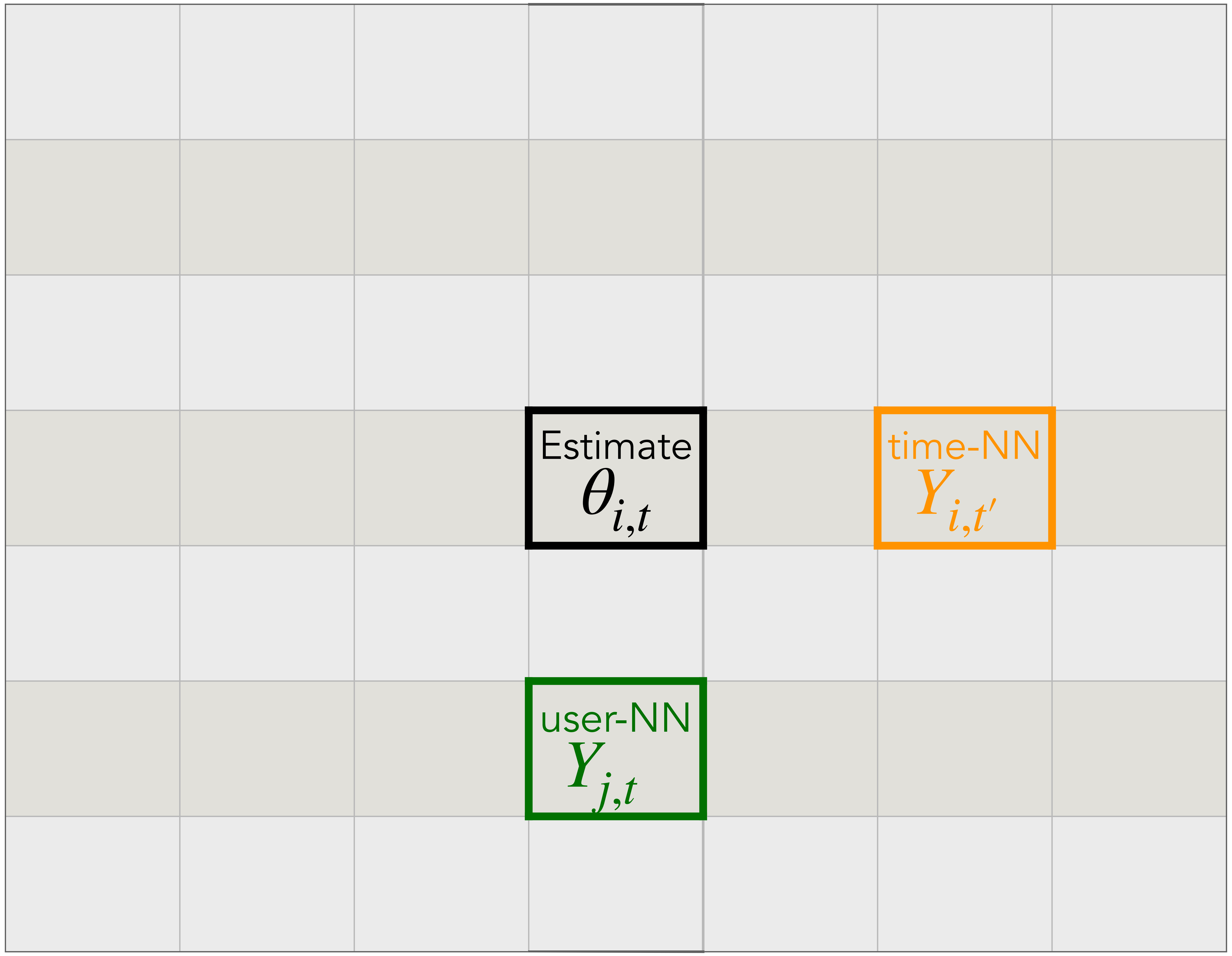
$$\langle \hat{u}_i, v_t \rangle + \langle u_i, \hat{v}_t \rangle - \langle \hat{u}_i, \hat{v}_t \rangle$$

Dropping superscript notation and indicators for simplicity

Integrating double robustness with nearest neighbors

Let j be such that $\rho_{i,j}^{(a)} \leq \eta$ & $A_{j,t} = a$

t' be such that $\rho_{t,t'}^{(a)} \leq \eta$ & $A_{i,t'} = a$



$$\langle \hat{u}_i, v_t \rangle + \langle u_i, \hat{v}_t \rangle - \langle \hat{u}_i, \hat{v}_t \rangle$$

Dropping superscript notation and indicators for simplicity

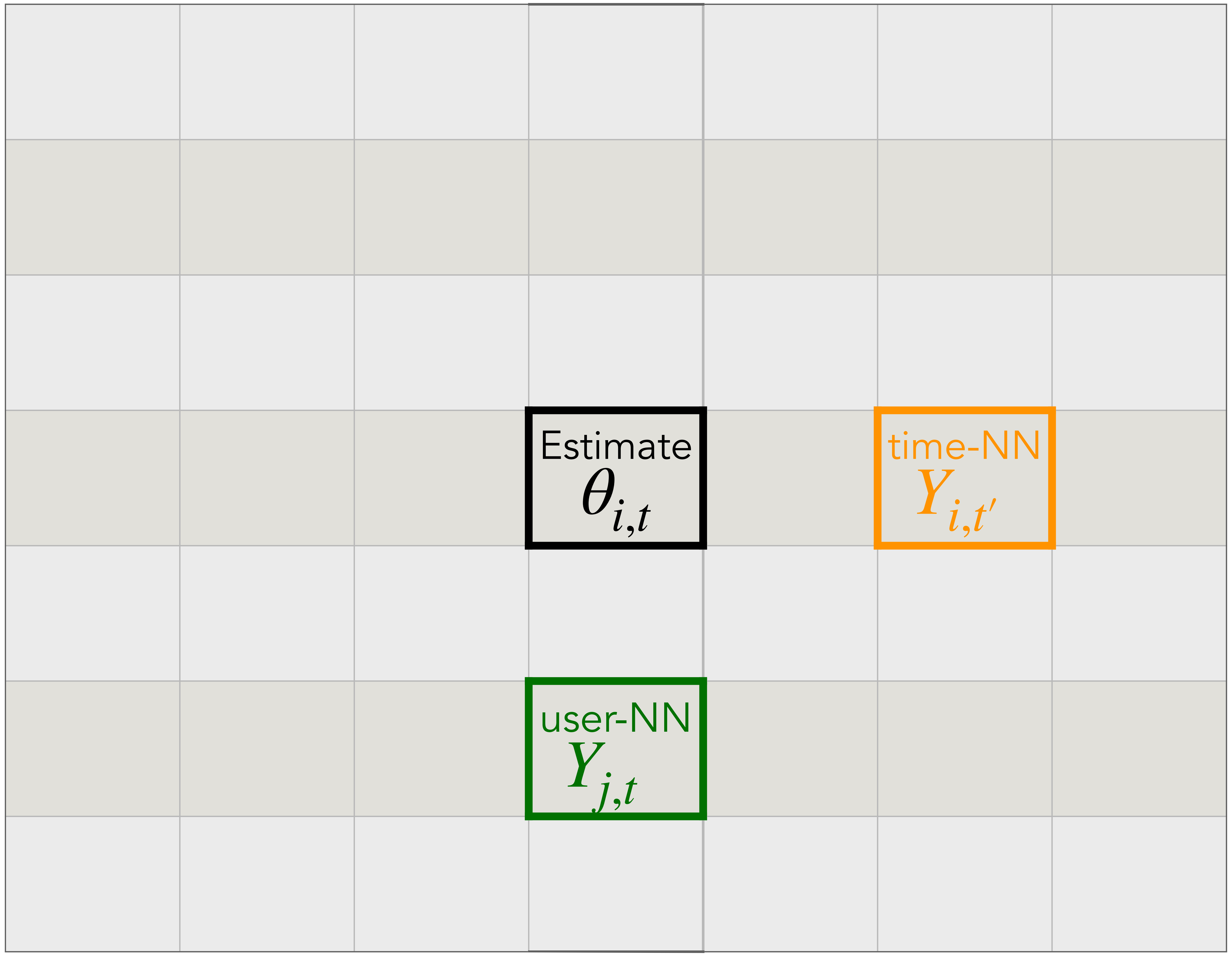
Integrating double robustness with nearest neighbors

Let j be such that $\rho_{i,j}^{(a)} \le \eta$ & $A_{j,t} = a$

t' be such that $\rho_{t,t'}^{(a)} \le \eta$ & $A_{i,t'} = a$

$$\widehat{\theta}_{i,t,\text{user-NN}}^{(a)} = Y_{j,t} \approx \langle u_j, v_t \rangle = \langle \hat{u}_i, v_t \rangle$$

$$\langle \hat{u}_i, v_t \rangle + \langle u_i, \hat{v}_t \rangle - \langle \hat{u}_i, \hat{v}_t \rangle$$



Dropping superscript notation and indicators for simplicity

Integrating **double robustness** with nearest neighbors

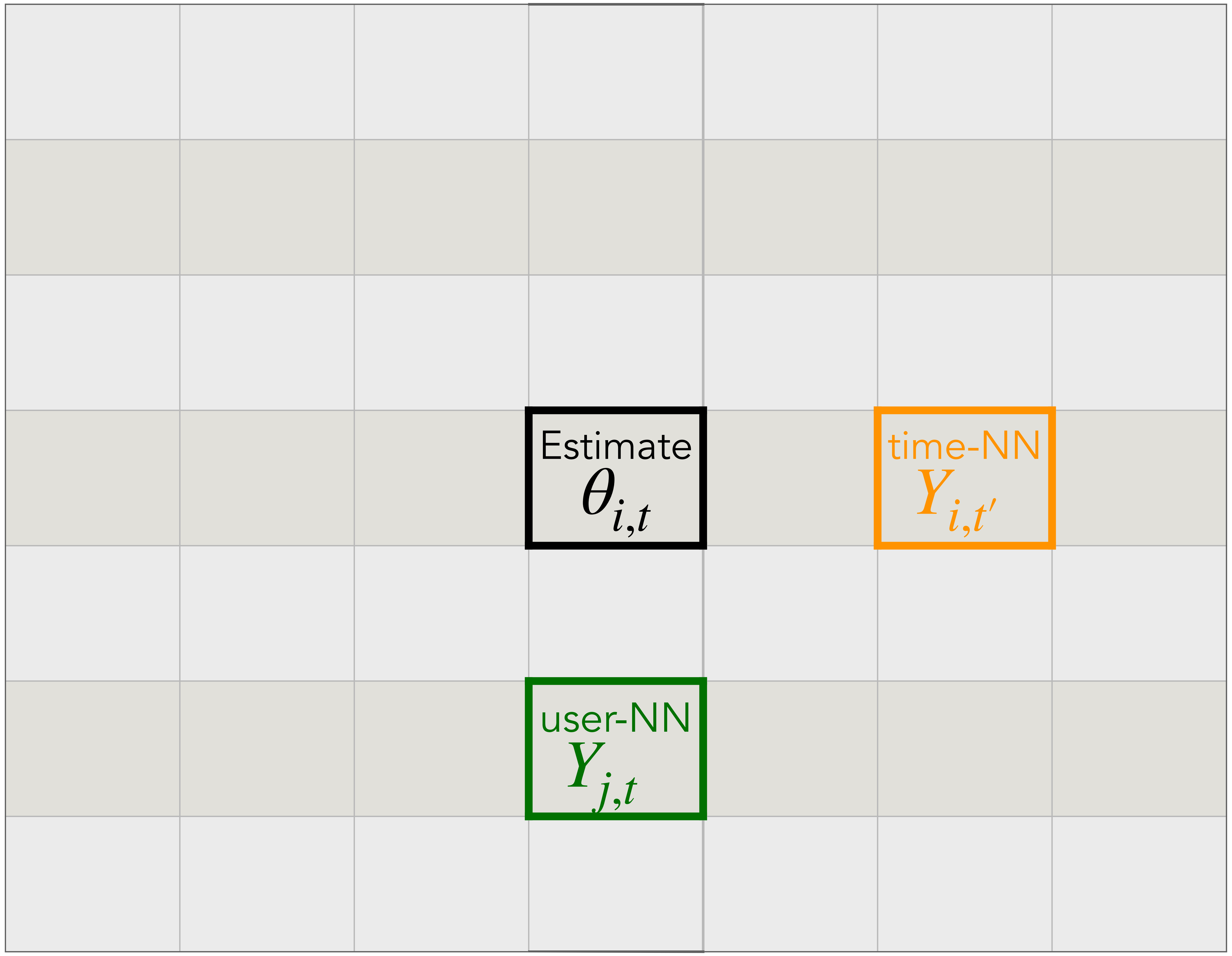
Let j be such that $\rho^{(a)}_{i,j} \leq \eta$ & $A_{j,t} = a$

t' be such that $\rho^{(a)}_{t,t'} \leq \eta$ & $A_{i,t'} = a$

$$\widehat{\theta}^{(a)}_{i,t,\text{user-NN}} = Y_{j,t} \approx \langle u_j, v_t \rangle = \langle \hat{u}_i, v_t \rangle$$

$$\widehat{\theta}^{(a)}_{i,t,\text{time-NN}} = Y_{i,t'} \approx \langle u_i, v_{t'} \rangle = \langle u_i, \hat{v}_t \rangle$$

$$\langle \hat{u}_i, v_t \rangle + \langle u_i, \hat{v}_t \rangle - \langle \hat{u}_i, \hat{v}_t \rangle$$



Dropping superscript notation and indicators for simplicity

Integrating double robustness with nearest neighbors

Let j be such that $\rho_{i,j}^{(a)} \leq \eta$ & $A_{j,t} = a$

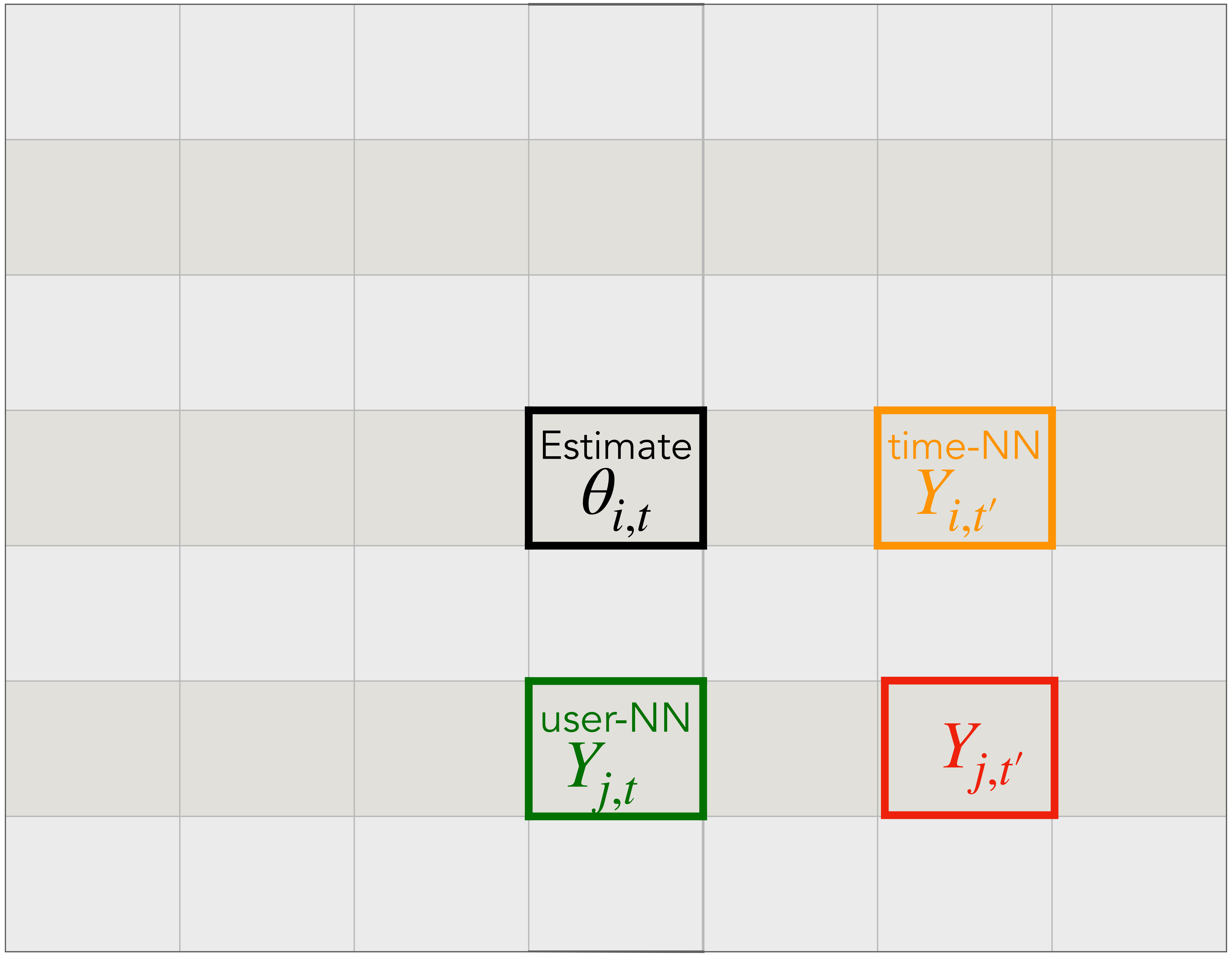
t' be such that $\rho_{t,t'}^{(a)} \leq \eta$ & $A_{i,t'} = a$

$$\widehat{\theta}_{i,t,\text{user-NN}}^{(a)} = Y_{j,t} \approx \langle u_j, v_t \rangle = \langle \hat{u}_i, v_t \rangle$$

$$\widehat{\theta}_{i,t,\text{time-NN}}^{(a)} = Y_{i,t'} \approx \langle u_i, v_{t'} \rangle = \langle u_i, \hat{v}_t \rangle$$

$$\widehat{\theta}_{i,t,\text{DR-NN}}^{(a)} = Y_{j,t} + Y_{i,t'} - Y_{j,t'}$$

$$\langle \hat{u}_i, v_t \rangle + \langle u_i, \hat{v}_t \rangle - \langle \hat{u}_i, \hat{v}_t \rangle$$

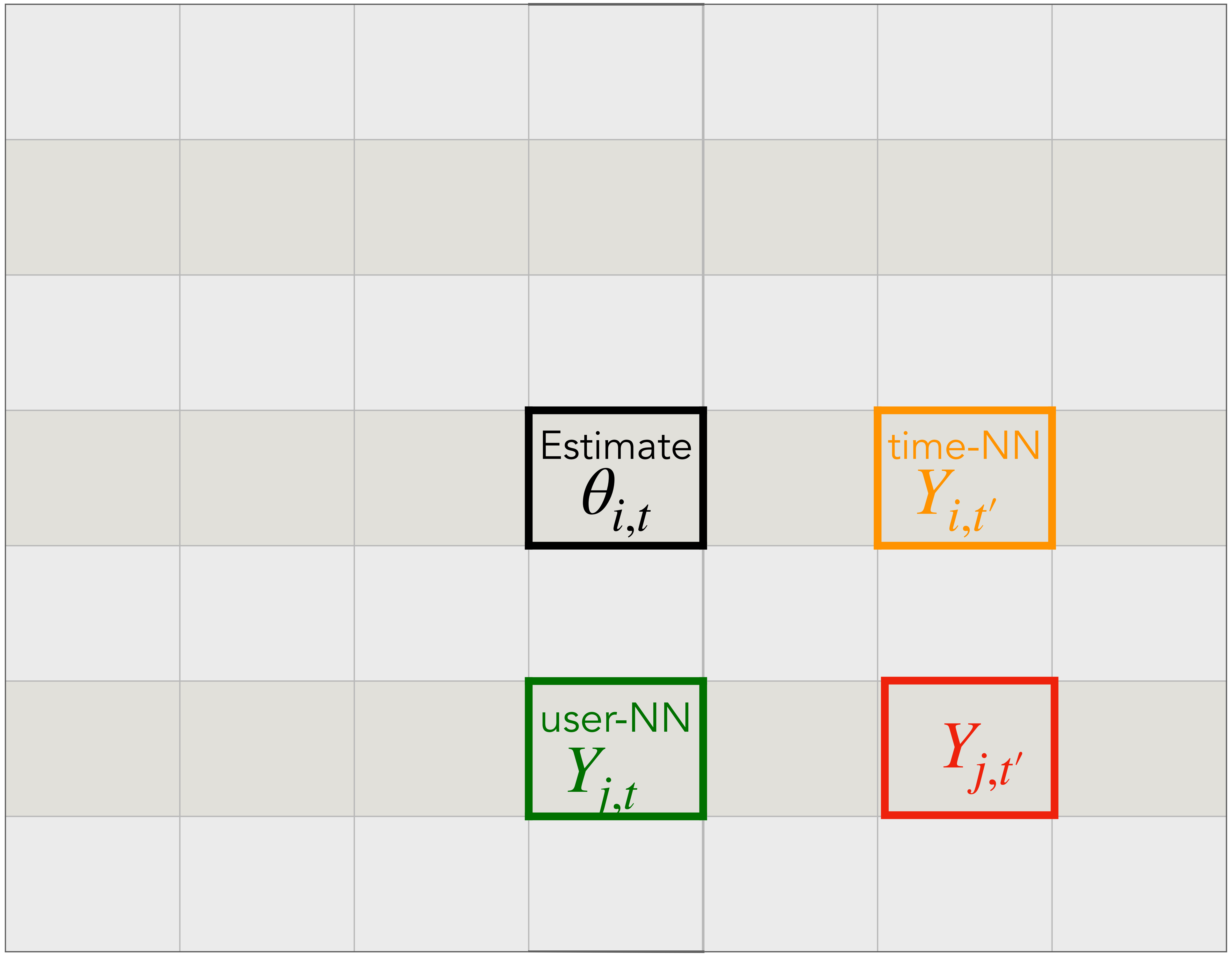


Dropping superscript notation and indicators for simplicity

Integrating double robustness with nearest neighbors

$$\mathbf{1}_{i,t,j,t'} = \mathbf{1}(\rho^{(a)}_{i,j} \le \eta,\ \rho^{(a)}_{t,t'} \le \eta\ , A_{j,t} = A_{i,t'} = A_{j,t'} = a)$$

$$\hat{\theta}^{(a)}_{i,t,\text{DR-NN}} = \frac{1}{\sum_{j\neq i, t'\neq t} \mathbf{1}_{i,t,j,t'}} \sum_{j\neq i, t'\neq t} (Y_{j,t} + Y_{i,t'} - Y_{j,t'}) \cdot \mathbf{1}_{i,t,j,t'}$$



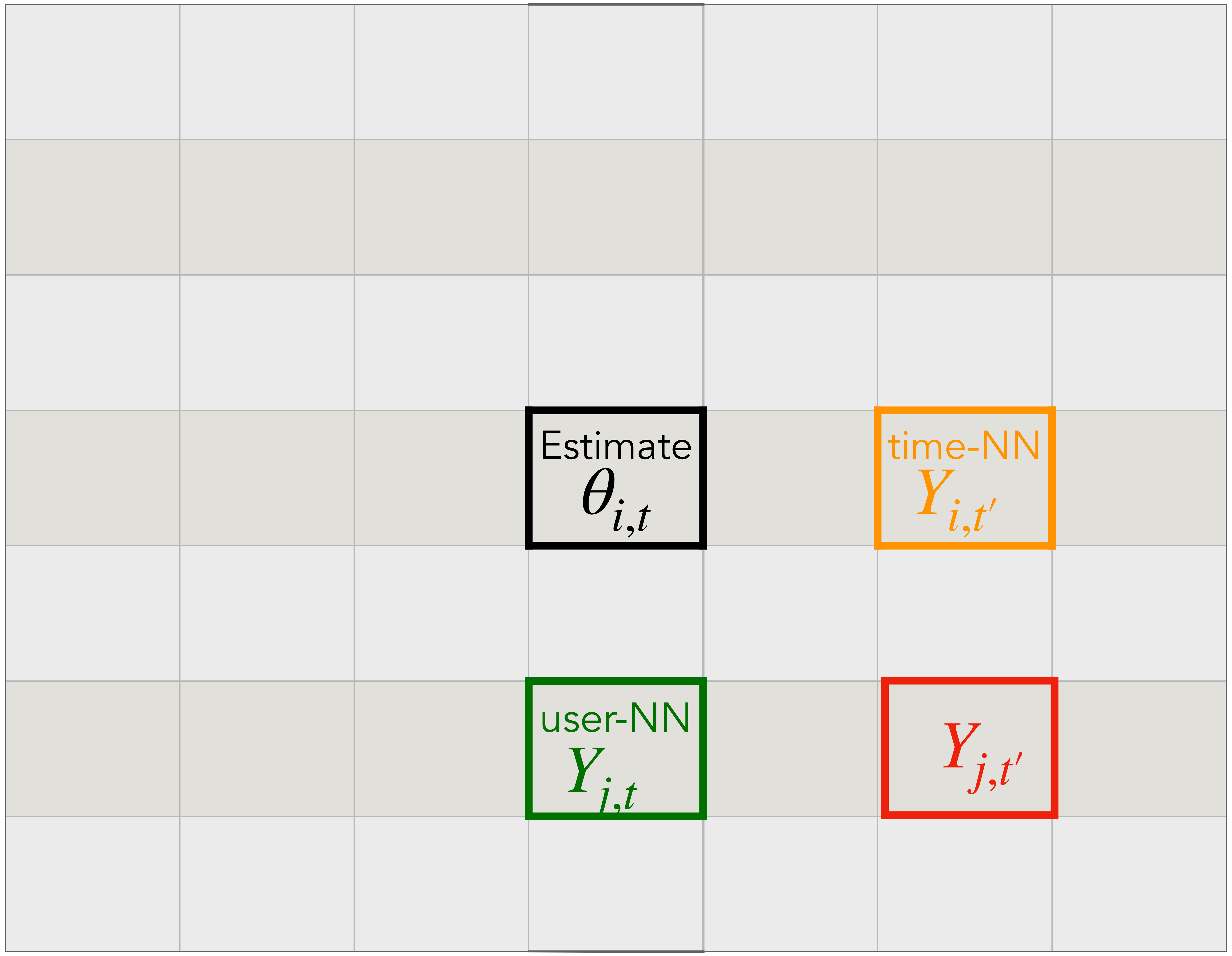
$$\langle \hat{u}_i, v_t \rangle + \langle u_i, \hat{v}_t \rangle - \langle \hat{u}_i, \hat{v}_t \rangle$$

Sample-split for doubly robust nearest neighbors

$$\mathbf{1}_{i,t,j,t'} = \mathbf{1}(\rho^{(a)}_{i,j} \leq \eta,\ \rho^{(a)}_{t,t'} \leq \eta\ , A_{j,t}{=}A_{i,t'}{=}A_{j,t'}{=}a)$$

$$\hat{\theta}^{(a)}_{i,t,\text{DR–NN}} = \frac{1}{\sum_{j\neq i,t'\neq t} \mathbf{1}_{i,t,j,t'}} \sum_{j\neq i,t'\neq t} (Y_{j,t} + Y_{i,t'} - Y_{j,t'}) \cdot \mathbf{1}_{i,t,j,t'}$$

*Nuisance estimates should be fitted independently of terms used for debiasing

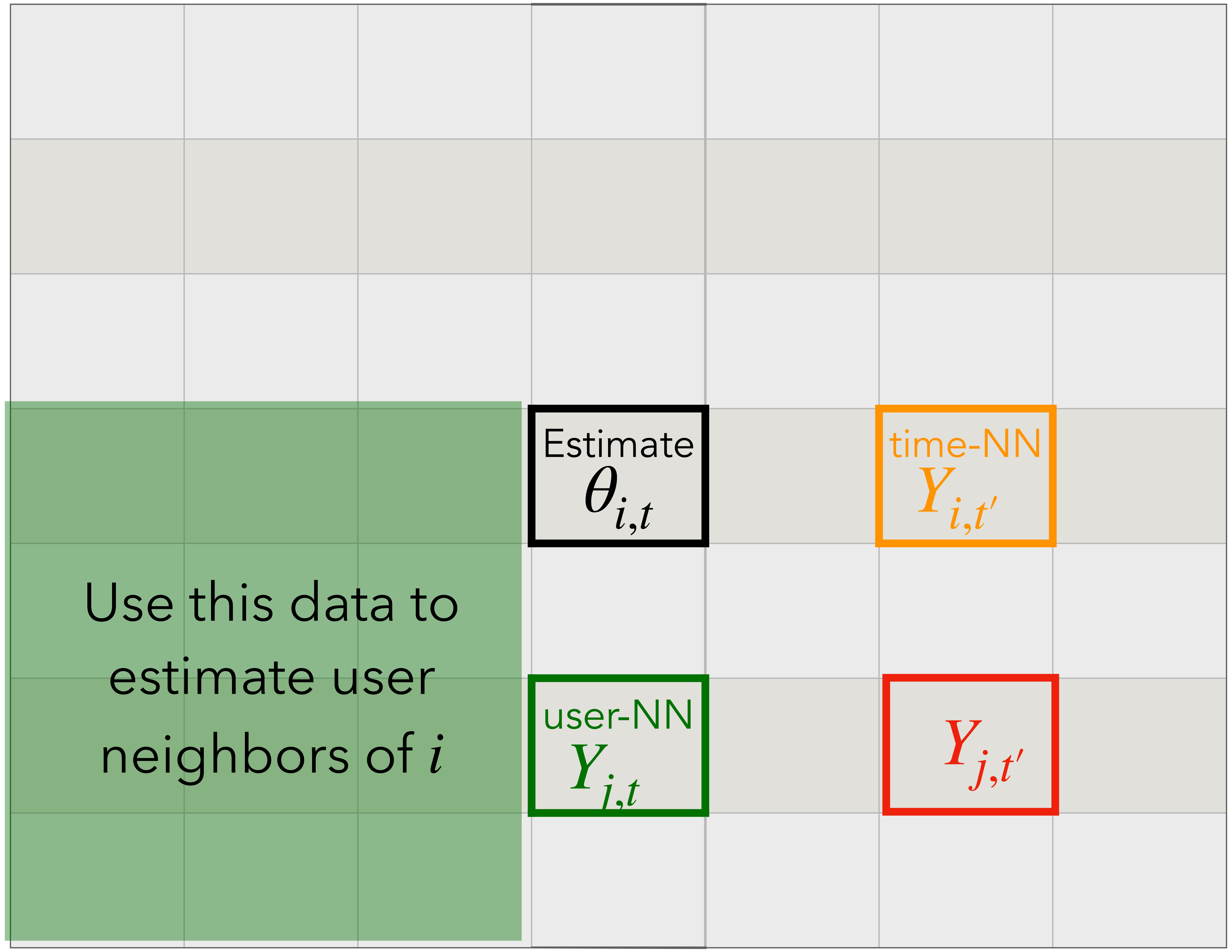


Sample-split for doubly robust nearest neighbors

$$\mathbf{1}_{i,t,j,t'} = \mathbf{1}(\rho^{(a)}_{i,j} \leq \eta,\ \rho^{(a)}_{t,t'} \leq \eta\ , A_{j,t}=A_{i,t'}=A_{j,t'}=a)$$

$$\widehat{\theta}^{(a)}_{i,t,\mathrm{DR\text{-}NN}} = \frac{1}{\sum_{j\neq i, t'\neq t} \mathbf{1}_{i,t,j,t'}} \sum_{j\neq i, t'\neq t} (Y_{j,t} + Y_{i,t'} - Y_{j,t'}) \cdot \mathbf{1}_{i,t,j,t'}$$

*Nuisance estimates should be fitted independently of terms used for debiasing

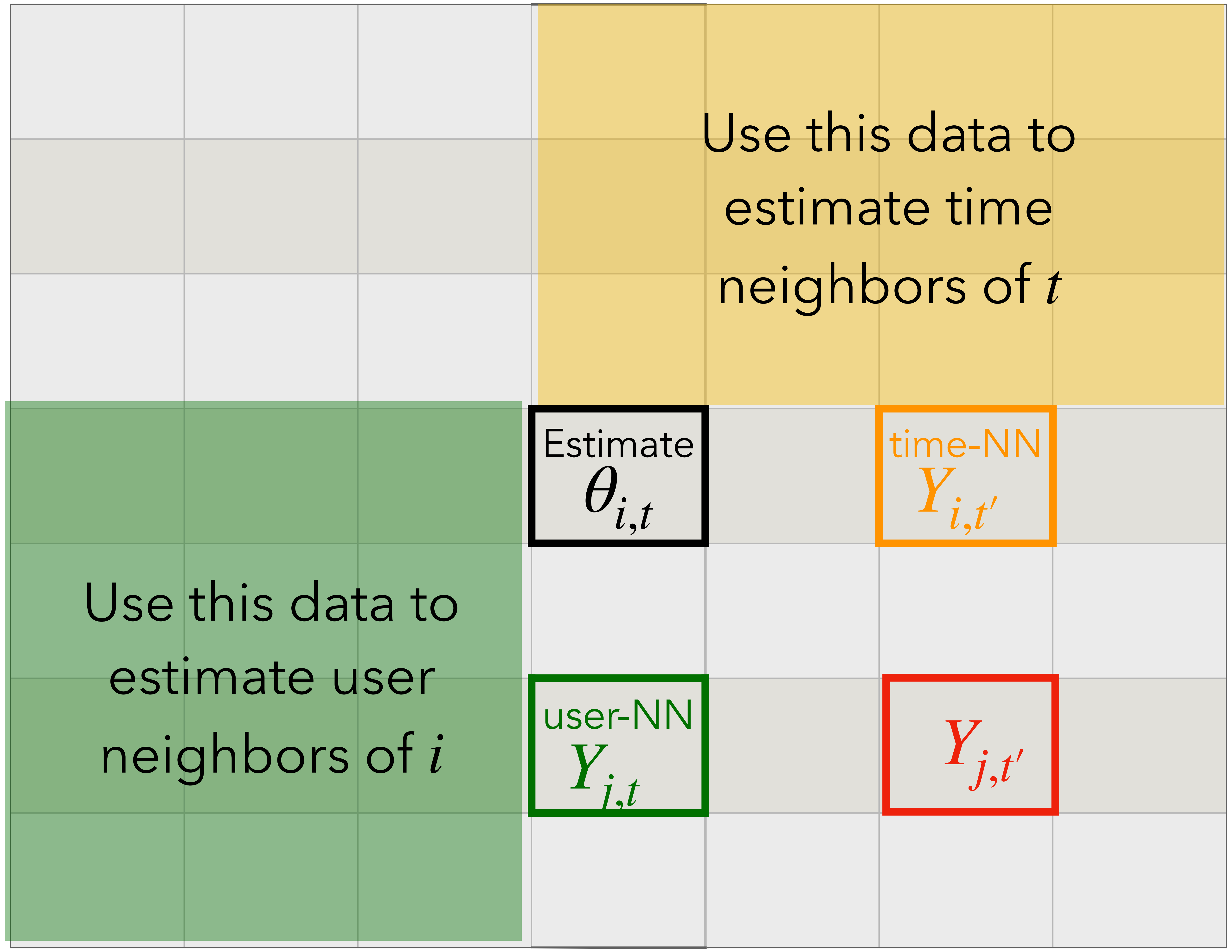


Sample-split for doubly robust nearest neighbors

$$\mathbf{1}_{i,t,j,t'} = \mathbf{1}(\rho^{(a)}_{i,j} \leq \eta,\ \rho^{(a)}_{t,t'} \leq \eta\ , A_{j,t}{=}A_{i,t'}{=}A_{j,t'}{=}a)$$

$$\widehat{\theta}^{(a)}_{i,t,\mathrm{DR\text{-}NN}} = \frac{1}{\sum_{j\neq i,t'\neq t} \mathbf{1}_{i,t,j,t'}} \sum_{j\neq i,t'\neq t} (Y_{j,t} + Y_{i,t'} - Y_{j,t'}) \cdot \mathbf{1}_{i,t,j,t'}$$

*Nuisance estimates should be fitted independently of terms used for debiasing



$$|\hat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t}| = \tilde{O}\left(\frac{1}{T^{1/4}} + \frac{1}{\sqrt{N}}\right)$$

$$|\hat{\theta}^{(a)}_{i,t,\text{time-NN}} - \theta^{(a)}_{i,t}| = \tilde{O}\left(\frac{1}{N^{1/4}} + \frac{1}{\sqrt{T}}\right)$$

Doubly robust estimate fixes the **slow error rates**

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{user–NN}} - \theta^{(a)}_{i,t}\,| = \tilde{O}\left(\frac{1}{T^{1/4}} + \frac{1}{\sqrt{N}}\right)$$

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{time–NN}} - \theta^{(a)}_{i,t}\,| = \tilde{O}\left(\frac{1}{N^{1/4}} + \frac{1}{\sqrt{T}}\right)$$

$$\downarrow$$

$$|\,\widehat{\theta}^{(a)}_{i,t,\text{DR–NN}} - \theta^{(a)}_{i,t}\,| = \tilde{O}\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}}\right)$$

[**Dwivedi**-Tian-Tomkins-Klasnja-Murphy-Shah '22b]

Doubly robust estimate fixes the slow error rates

$$| \widehat{\theta}^{(a)}_{i,t,\text{user-NN}} - \theta^{(a)}_{i,t} | = \tilde{O}\left(\frac{1}{T^{1/4}} + \frac{1}{\sqrt{N}}\right)$$

$$| \widehat{\theta}^{(a)}_{i,t,\text{time-NN}} - \theta^{(a)}_{i,t} | = \tilde{O}\left(\frac{1}{N^{1/4}} + \frac{1}{\sqrt{T}}\right)$$

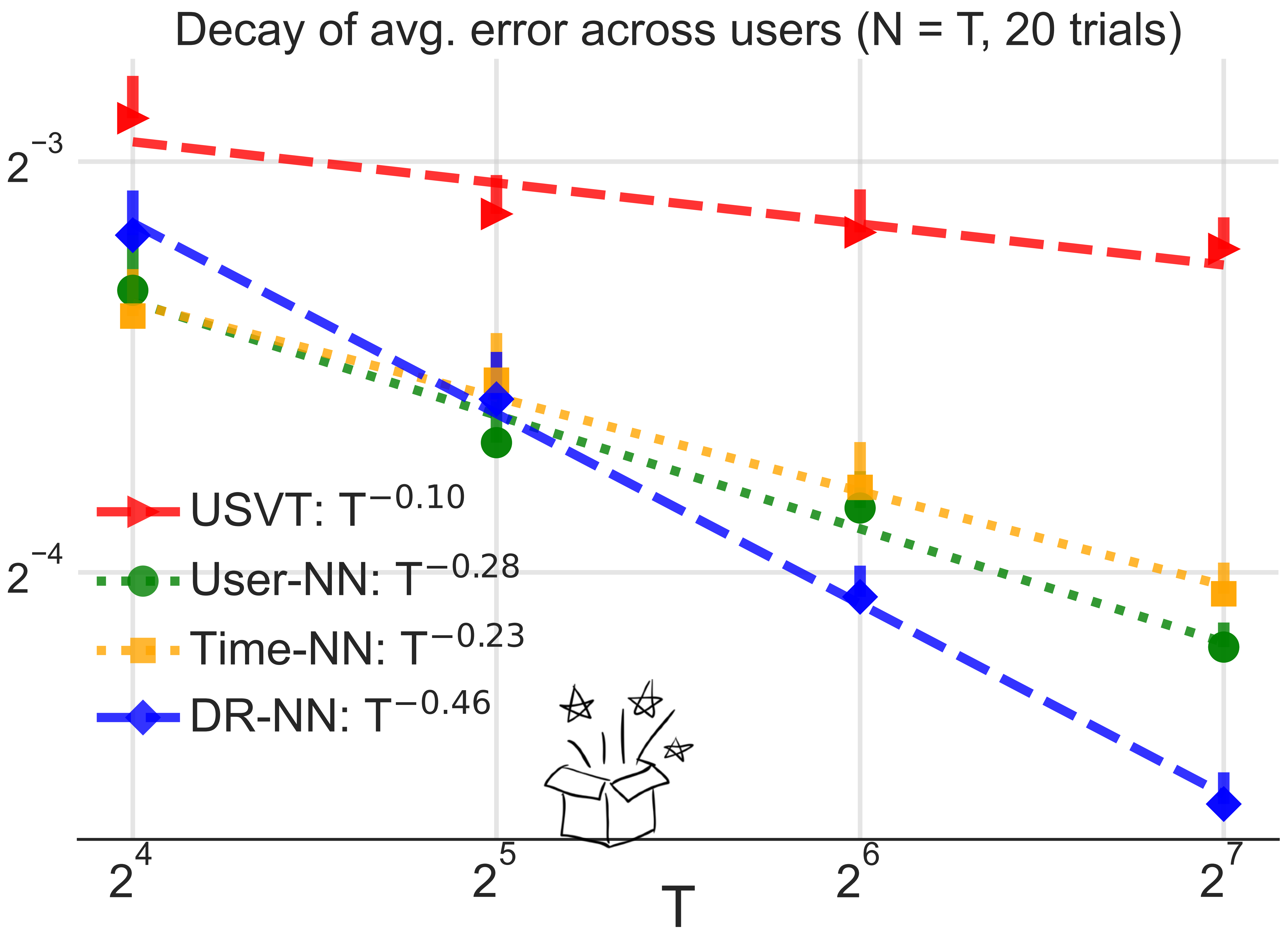
↓

$$| \widehat{\theta}^{(a)}_{i,t,\text{DR-NN}} - \theta^{(a)}_{i,t} | = \tilde{O}\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{N}}\right)$$

[**Dwivedi**-Tian-Tomkins-Klasnja-Murphy-Shah '22b]

Simulation results with N=T

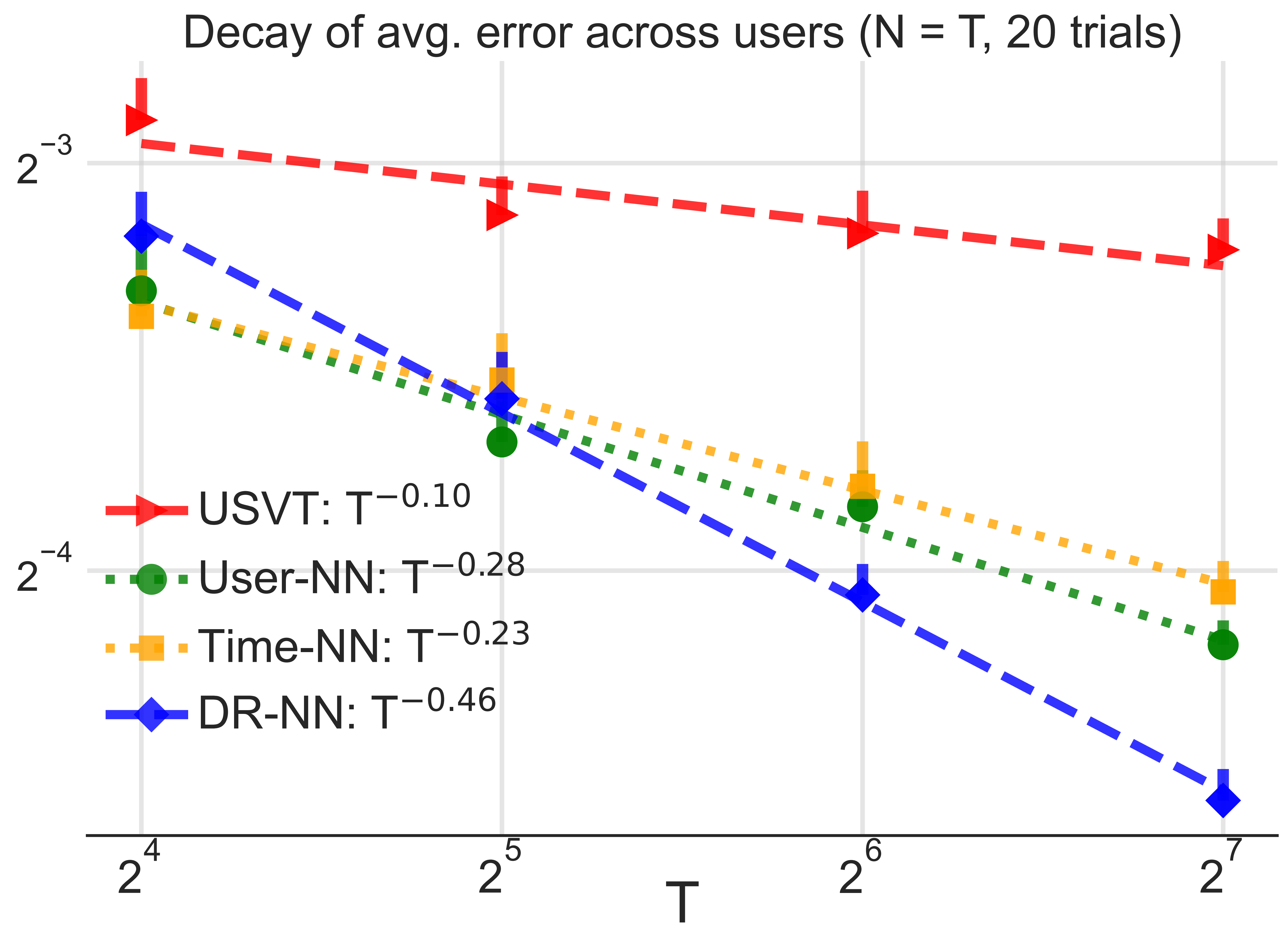
Uniform factors on $[-0.5,0.5]^4$, Gaussian noise, pooled ε-greedy policy ($\varepsilon = 0.5$)



USVT: A baseline algorithm from [Chatterjee 2014]

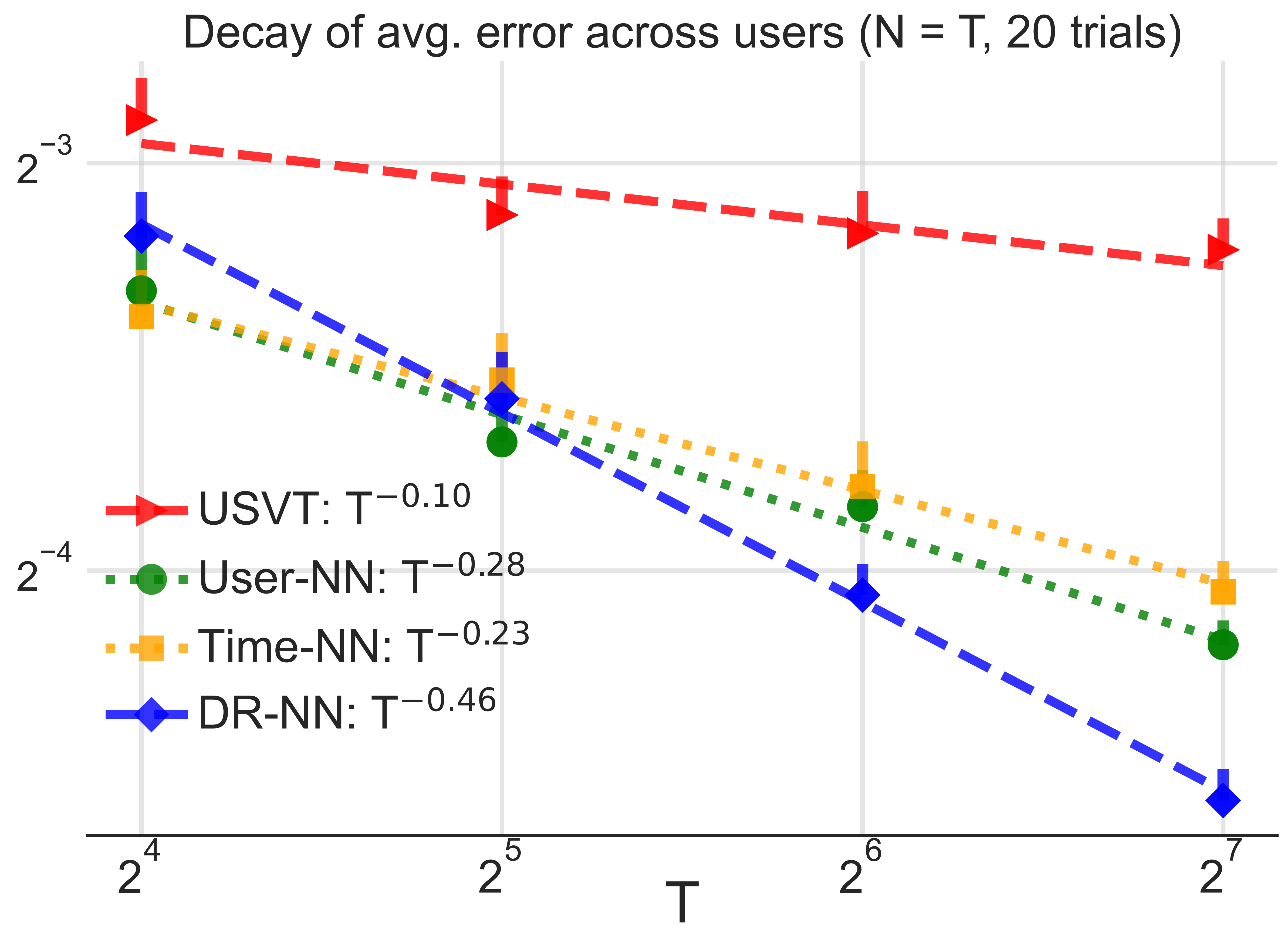
Simulation results

Uniform factors on $[-0.5, 0.5]^4$, Gaussian noise,
pooled ε-greedy policy ($\varepsilon = 0.5$)



Simulation results

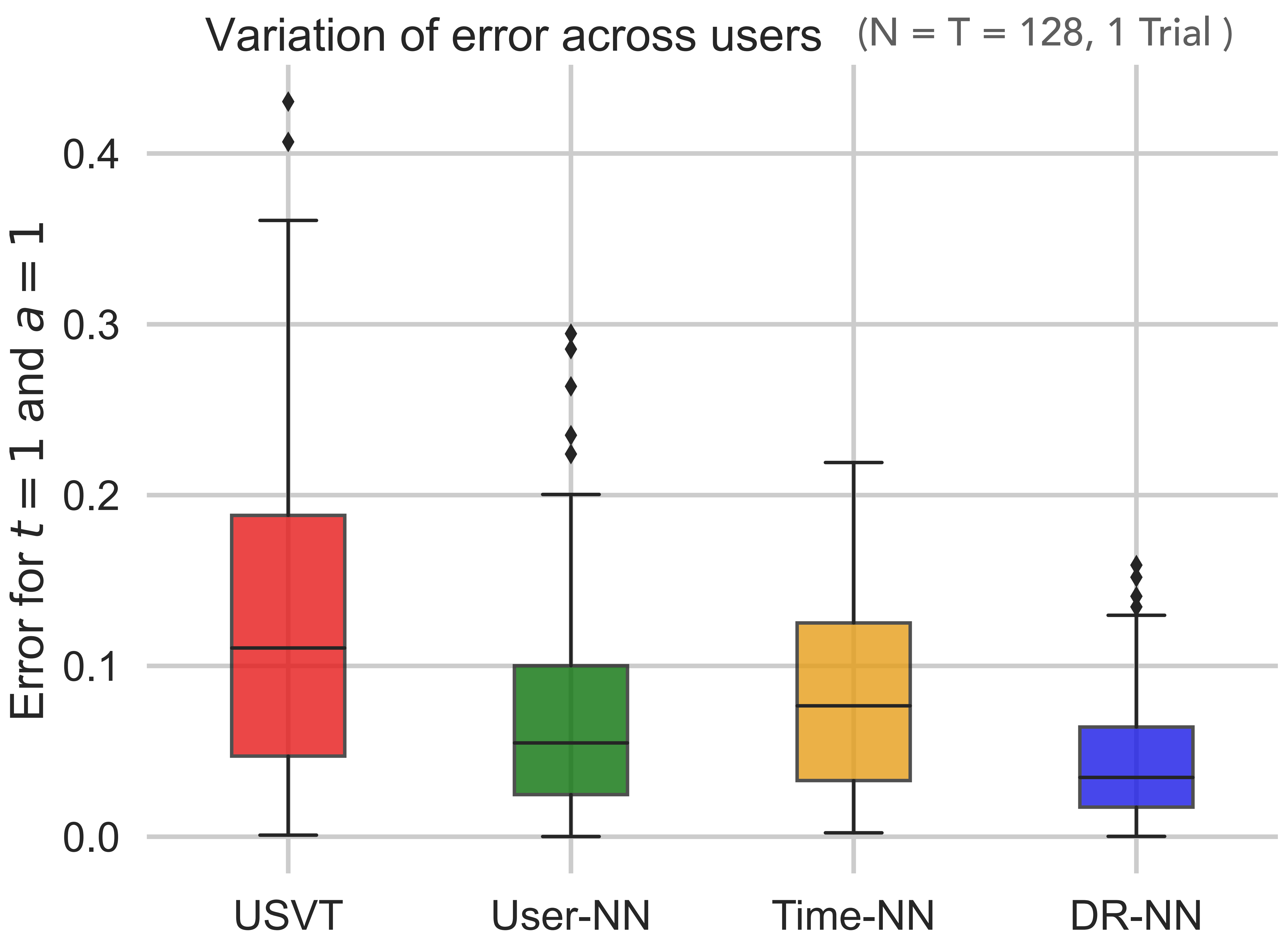
Uniform factors on $[-0.5,0.5]^4$, Gaussian noise,
pooled ε-greedy policy ($\varepsilon = 0.5$)

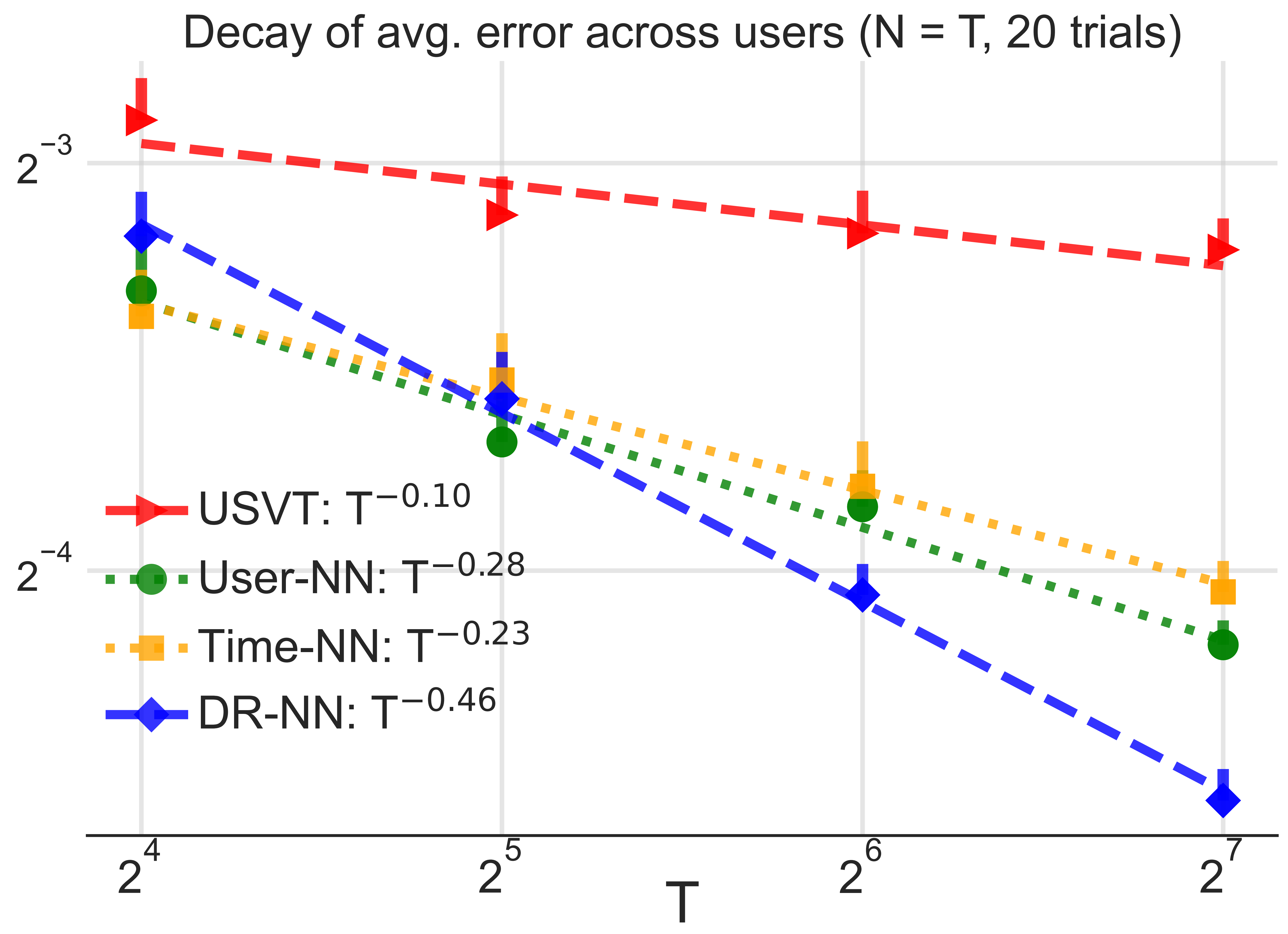


DR-NN error ≪ min { user-NN error, time-NN error }

Simulation results

Uniform factors on $[-0.5,0.5]^4$, Gaussian noise,
pooled ε-greedy policy ($\varepsilon = 0.5$)





A baseline algorithm from [Chatterjee 2014]

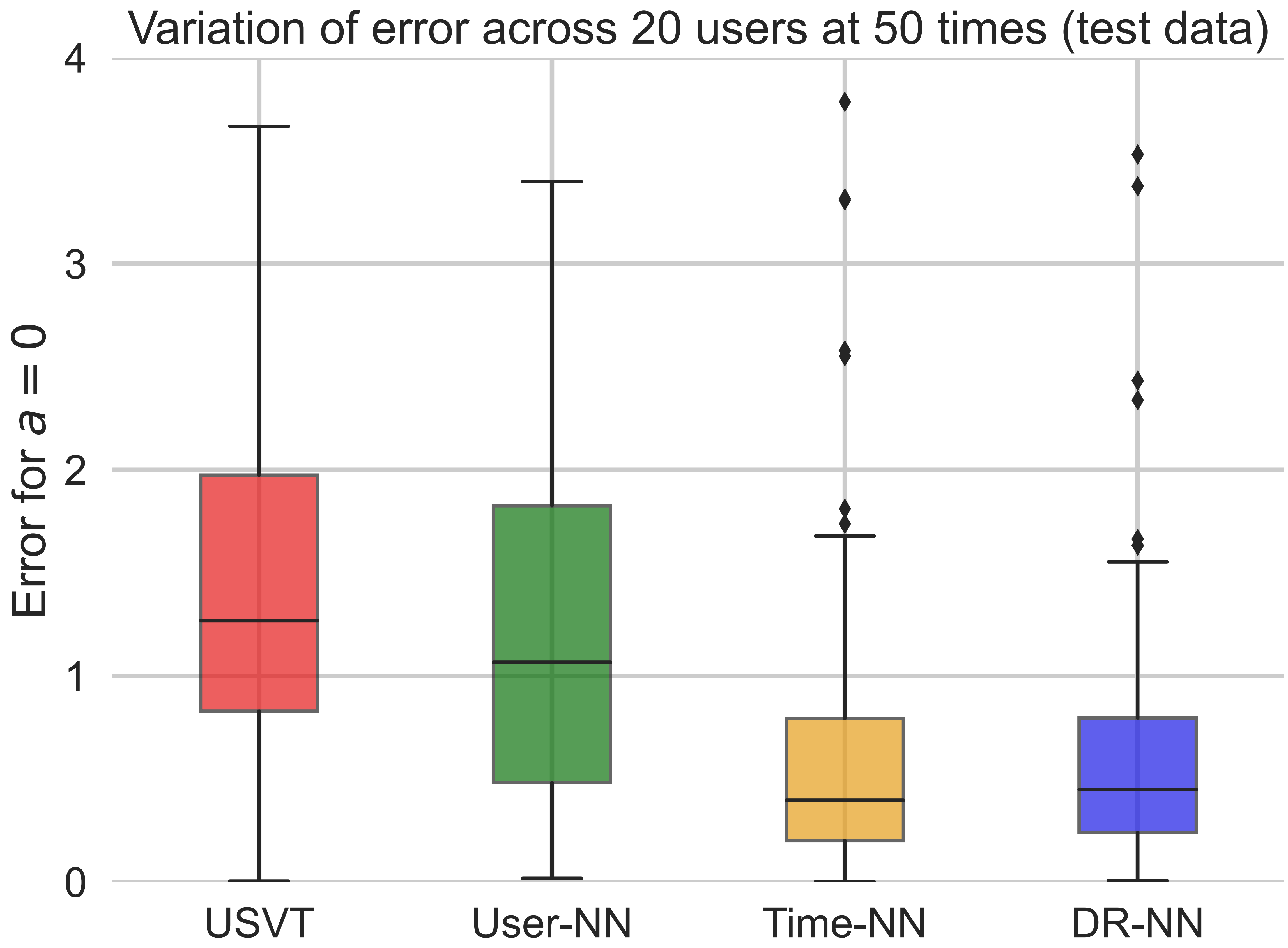
DR-NN error ≪ min { user-NN error, time-NN error }

HeartSteps study results 

Treatments assigned with Thompson sampling independently for 91 users for 90 days, 5 times a day

HeartSteps study results 

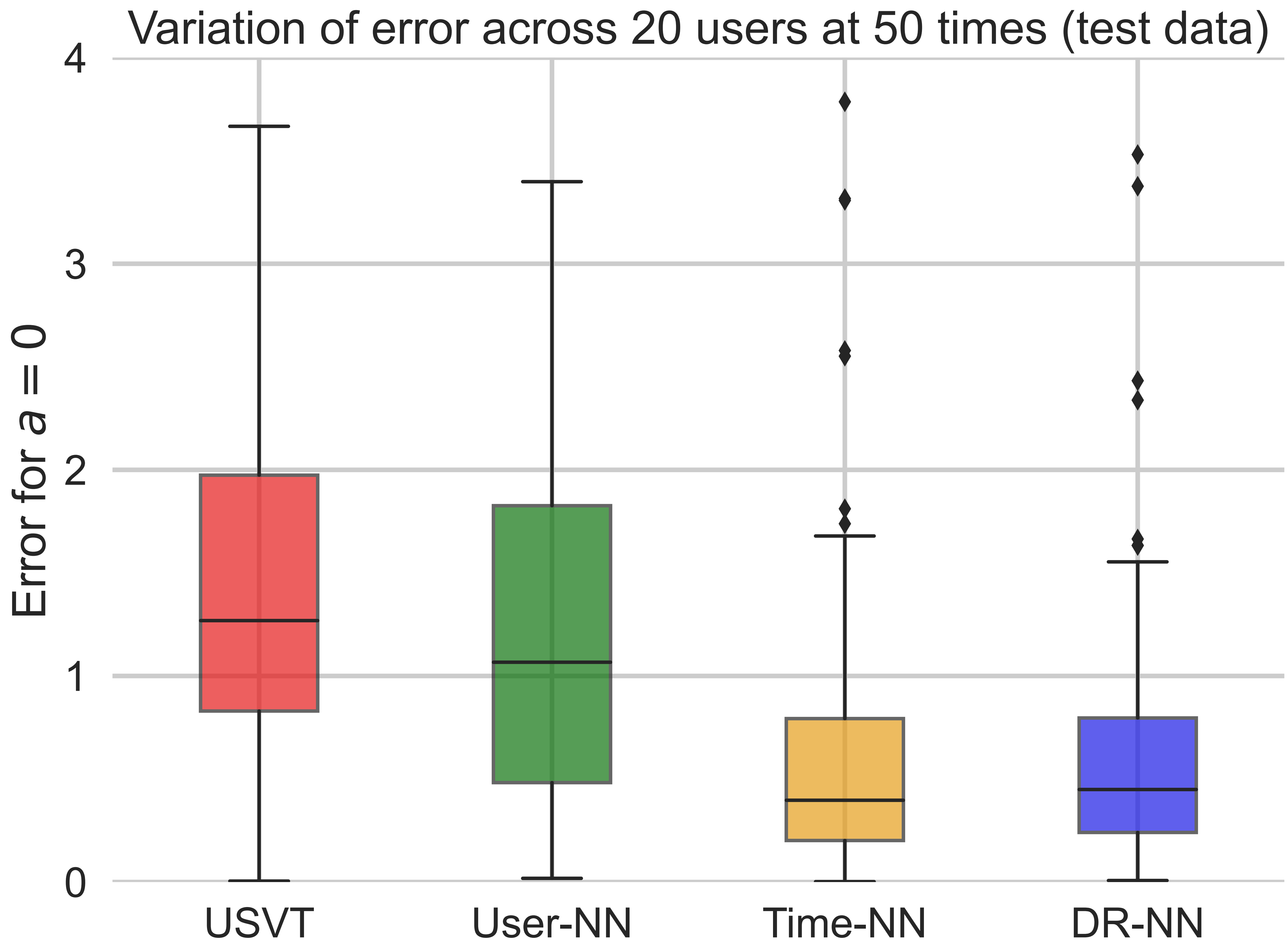
Treatments assigned with Thompson sampling independently for 91 users for 90 days, 5 times a day

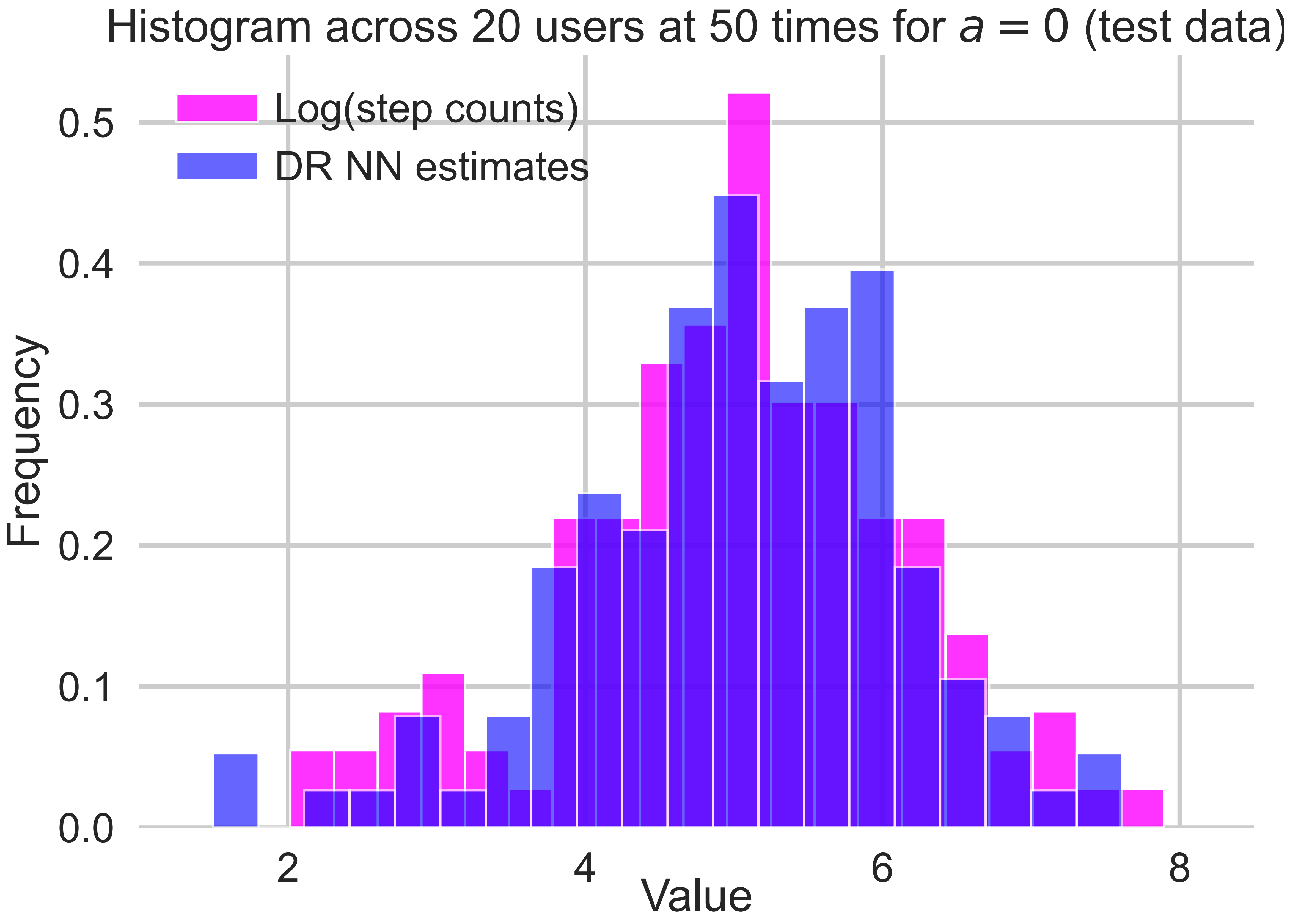


DR-NN error ≈ min { user-NN error, time-NN error }

HeartSteps study results

Treatments assigned with Thompson sampling independently for 91 users for 90 days, 5 times a day





DR-NN error ≈ min { user-NN error, time-NN error }

DR-NN: Robust to heterogeneity in user & time factors

DR-NN: Robust to heterogeneity in user & time factors

DR-NN error $\approx$ user-NN error $\times$ time-NN error

$\lesssim$ min{user-NN error, time-NN error}

As long as, either user factors or time factors
exhibit similarities, DR-NN has a good error

Summary: Integrating Double Robustness into Causal Latent Factor Models

Summary: Integrating Double Robustness into Causal Latent Factor Models

$\hat{u}\hat{v}$ —Estimate→ $\hat{u}v + u\hat{v} - \hat{u}\hat{v}$

$O(|\hat{u} - u| + |\hat{v} - v|)$ —Error→ $O(|\hat{u} - u| \times |\hat{v} - v|)$

	Problem setting	u	v
	ATE with observed confounding	conditional outcome mean	propensity function
	Off policy evaluation	mean reward	importance ratio
This talk	**ATE** with Matrix Completion (**Unobserved confounding**)	outcome matrix	propensity matrix
This talk	**ITE** with Nearest Neighbors (**Unobserved confounding**)	user factor	time factor

Thank you!

raazdwivedi.github.io

Appendix

A popular approach for ITE: Nearest neighbors

\+ Easy to implement and interpretable for checking

\+ Entry-wise guarantees

\+ Robust to interventional patterns

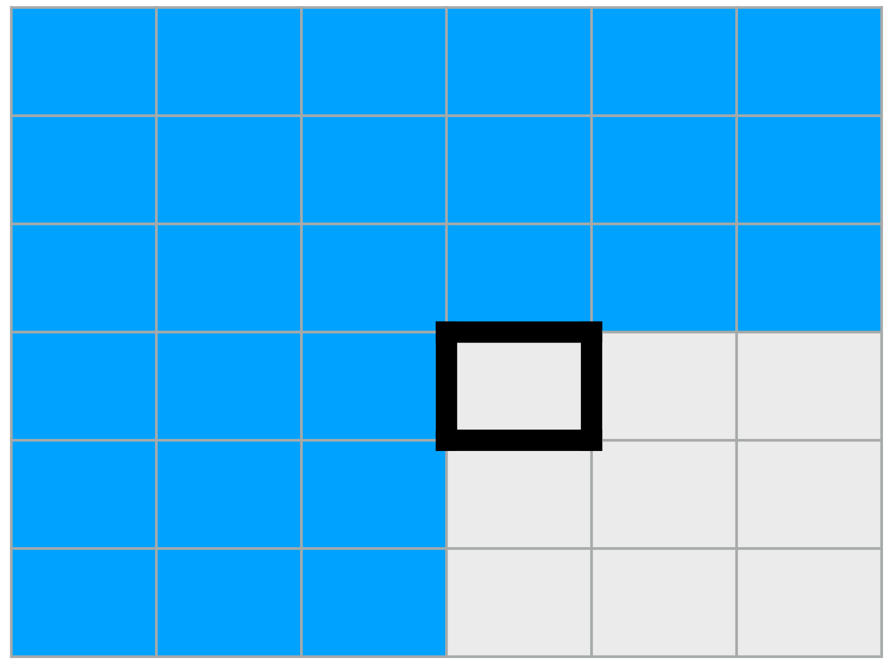
- missing completely at random (MCAR) [Li et al. 2019]
- sequential randomization (MAR) [Dwivedi et al. 2022a]
- unmeasured confounding (MNAR) [ongoing work]

Prior work/other methods

Rich literature

Not so rich literature

entry-wise guarantees essentially assuming →

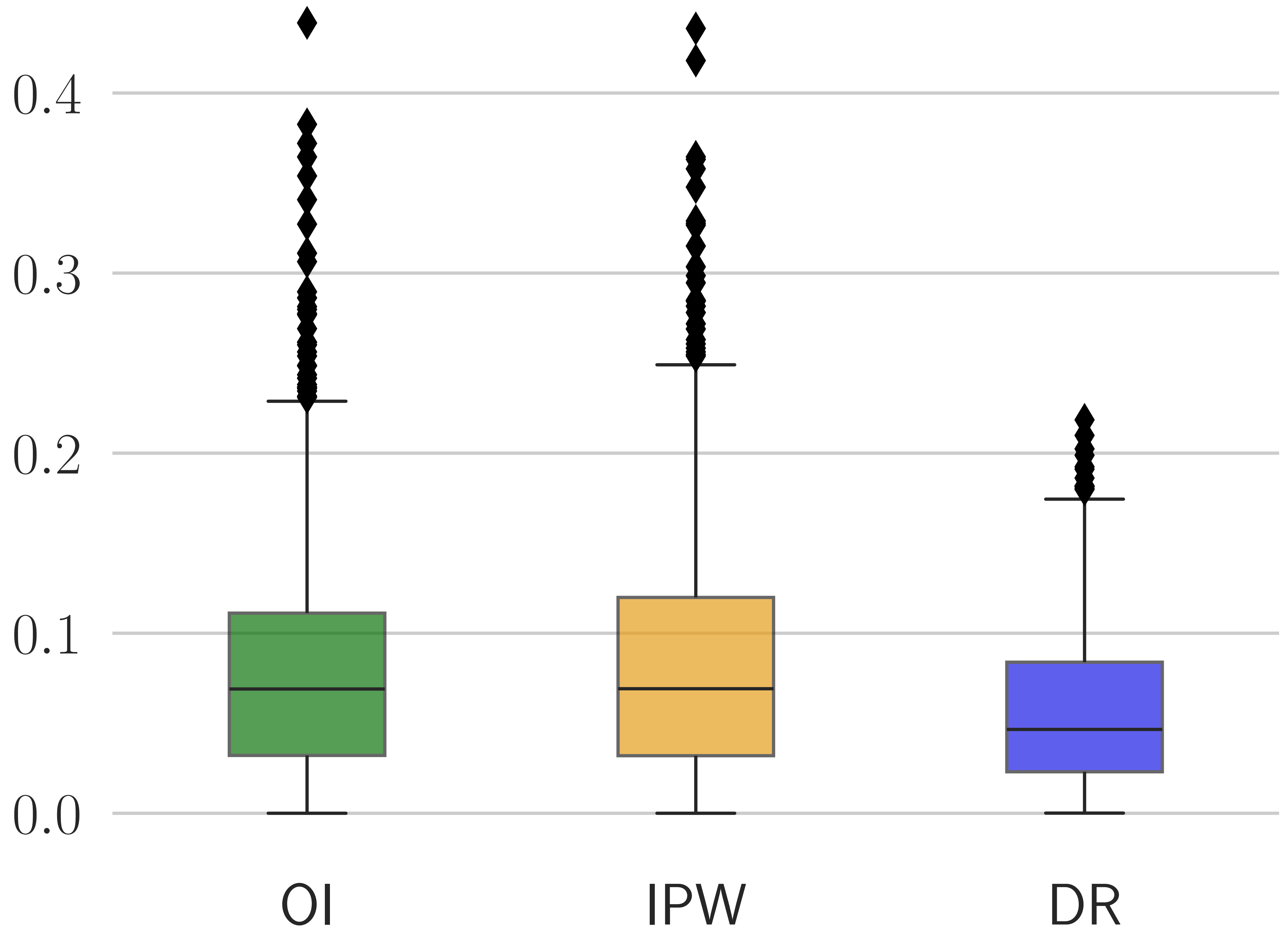


Intervention matrix

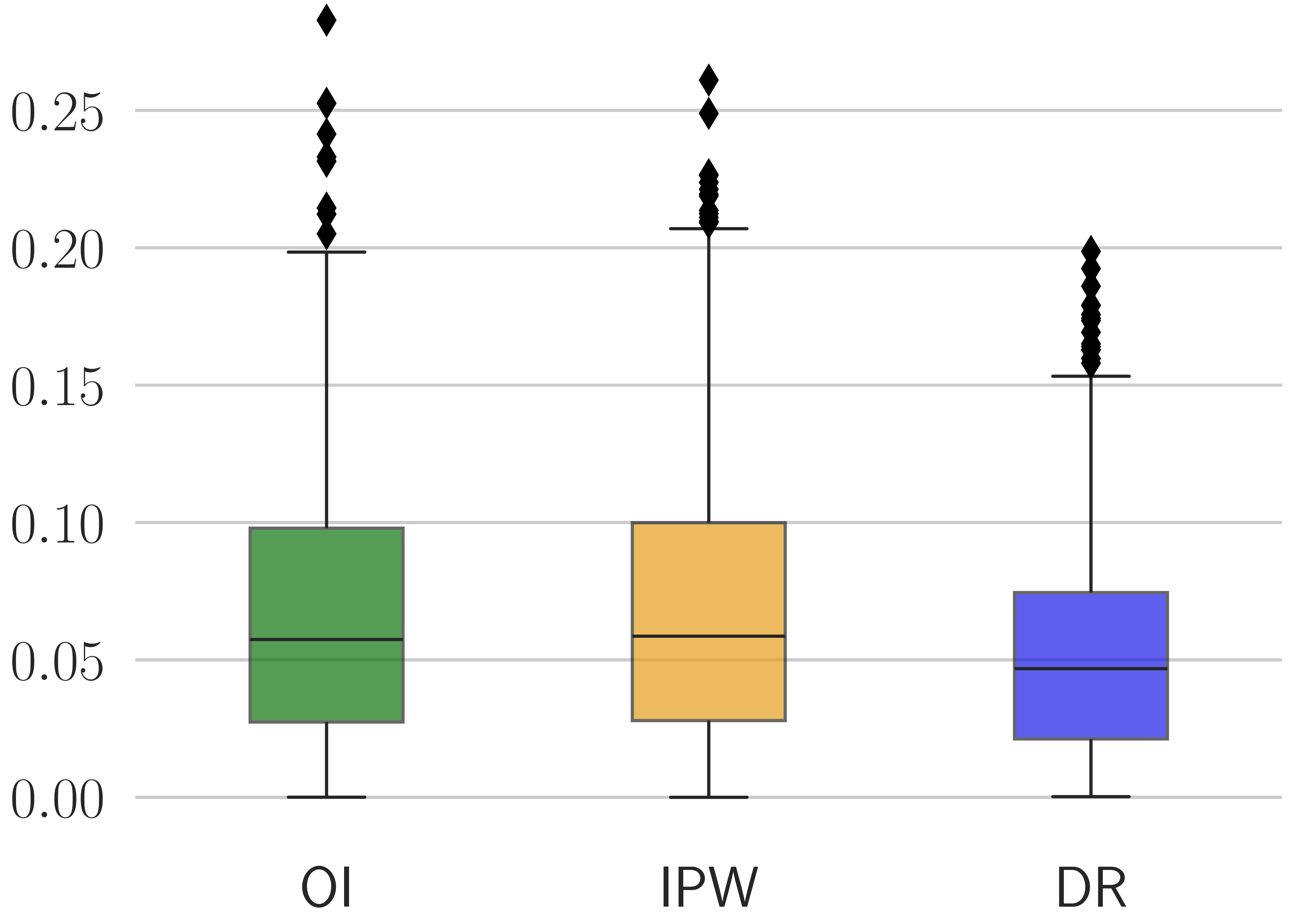
Simulation results with growing ranks

Uniform factors with $\text{rank}(\Theta^{(a)}) = N^{1/4}$, $\text{Rank}(P) = N^{1/5}$

Errors across all t for a fixed trial

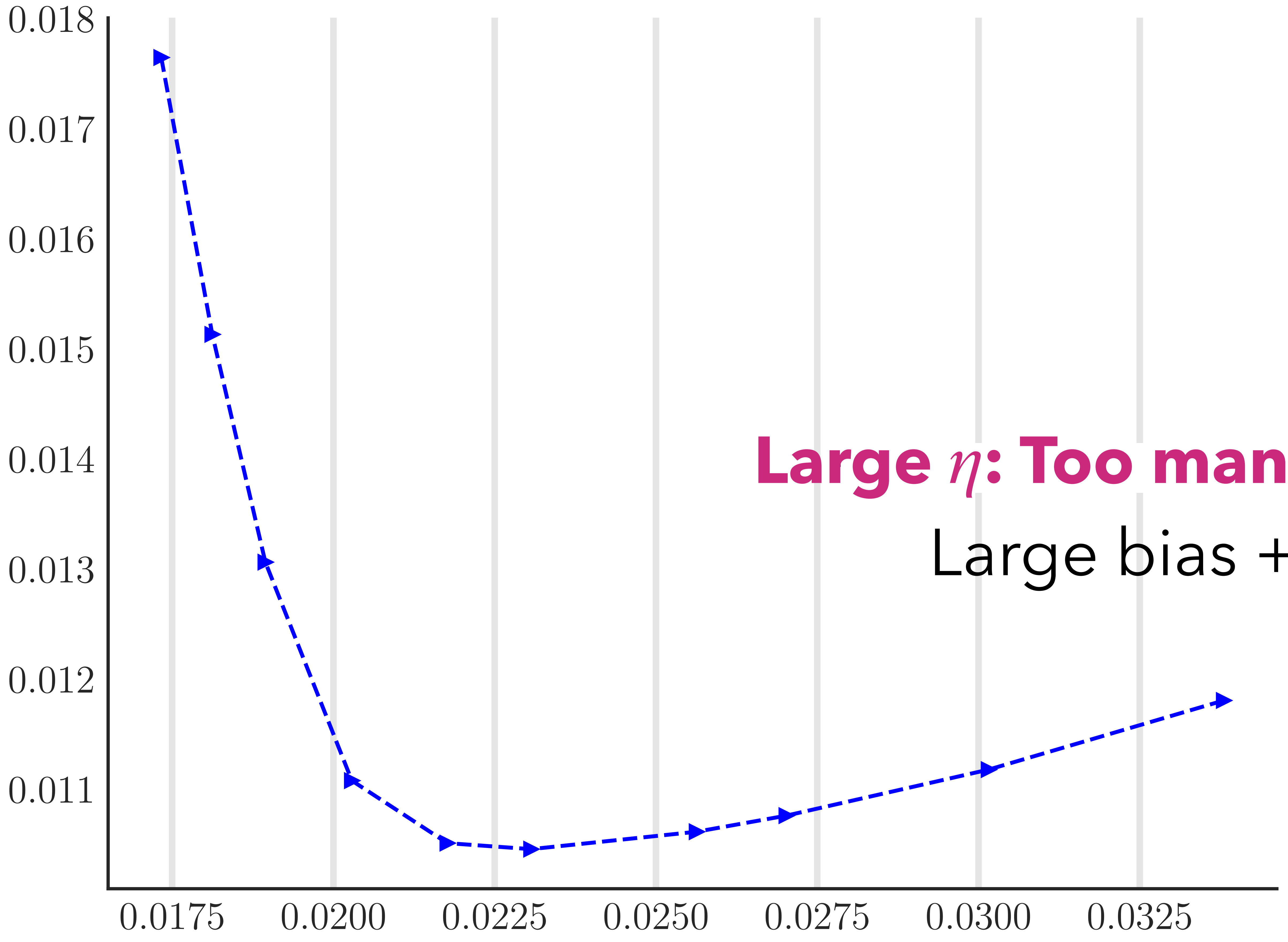


Errors for fixed t across trials



Bias-variance tradeoff for the nearest neighbors with η

MSE for estimates on observed entries



Large η: Too many "noisy" neighbors

Large bias + Small variance

Small η: Few "good" neighbors η

Small bias + Large variance

We prove a general error bound for user NN (with actions sampled by learning policies)

$$(\widehat{\theta}^{(a)}_{i,t,\eta} - \theta^{(a)}_{i,t})^2 \lesssim \frac{1}{\lambda_\star^2}\left(\eta - 2\sigma^2 \quad + \quad \frac{C}{\sqrt{p^2 T}}\right) + \frac{\sigma^2}{pN_{i,\eta'-e_T}} + c_{noise}\left[\frac{N_{i,\eta'+e_T} - N_{i,\eta'-e_T}}{pN_{i,\eta'-e_T}}\right]^2$$

η' — NN bias due to threshold

e_T — Error in NN distance

NN noise variance

NN bias inflation due to **learning** policy

$$\lambda_\star \triangleq \lambda_{\min}(\mathbf{\Sigma}_v) \quad \text{where } \mathbf{\Sigma}_v = \mathbb{E}[v_{t'} v_{t'}^\top]$$

$$N_{i,\gamma} \triangleq |\{j \neq i : (u_i - u_j)^\top \mathbf{\Sigma}_v (u_i - u_j) \leq \gamma\}|$$

Asymptotic intervals

- 95% intervals with asymptotic **coverage** as $N, T \to \infty$ and $\eta \to 2\sigma^2$ with suitable regularity conditions:

$$\hat{\theta}^{(a)}_{i,t,\text{user-NN}} \pm \frac{1.96\,\hat{\sigma}}{\sqrt{\#\text{neighbors}_{i,t,a}}}$$

Non-linear double/squared robustness

- $f(u,0) = f(0,0) + f'_u(0,0)u + \qquad\qquad\qquad + f''_{uu}(\tilde{u},0)u^2$
- $f(0,v) = f(0,0) + \qquad\qquad\qquad + f'_v(0,0)v + f''_{vv}(0,\hat{v})v^2$
- $f(u,v) = f(0,0) + f'_u(0,0)u + f'_v(0,0)v + [u,v]\,\nabla^2 f(\tilde{u},\tilde{v})\begin{bmatrix} u \\ v \end{bmatrix}$
- $f(u,0) + f(0,v) - f(u,v) = f(0,0) + O((u+v)^2) \Longrightarrow \text{Error} = \max\{u^2, v^2\}$