Counterfactual Forecasting for Panel Data

JSM 2025 Nashville

Navonil Deb

with **Raaz Dwivedi** and **Sumanta Basu**

Motivation: HeartSteps V1

Mobile activity walking coach

Source: Susan Muphy and Pedrag Klasnja's works

Motivation: HeartSteps V1

Mobile activity walking coach

- ☑ HeartSteps promotes walking behavior via mobile app

Source: Susan Muphy and Pedrag Klasnja's works

Motivation: HeartSteps V1

Mobile activity walking coach

- 6 weeks pilot micro randomized trial

Source: Susan Muphy and Pedrag Klasnja’s works

Motivation: HeartSteps V1

Mobile activity walking coach

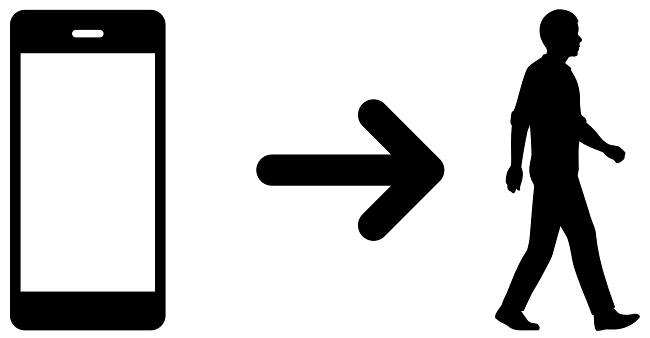
- 37 sedentary participants.

Source: Susan Muphy and Pedrag Klasnja's works

Motivation: HeartSteps V1

Mobile activity walking coach

Activity prompts 5x a day



– morning, afternoon, evening, before dinner, before sleep

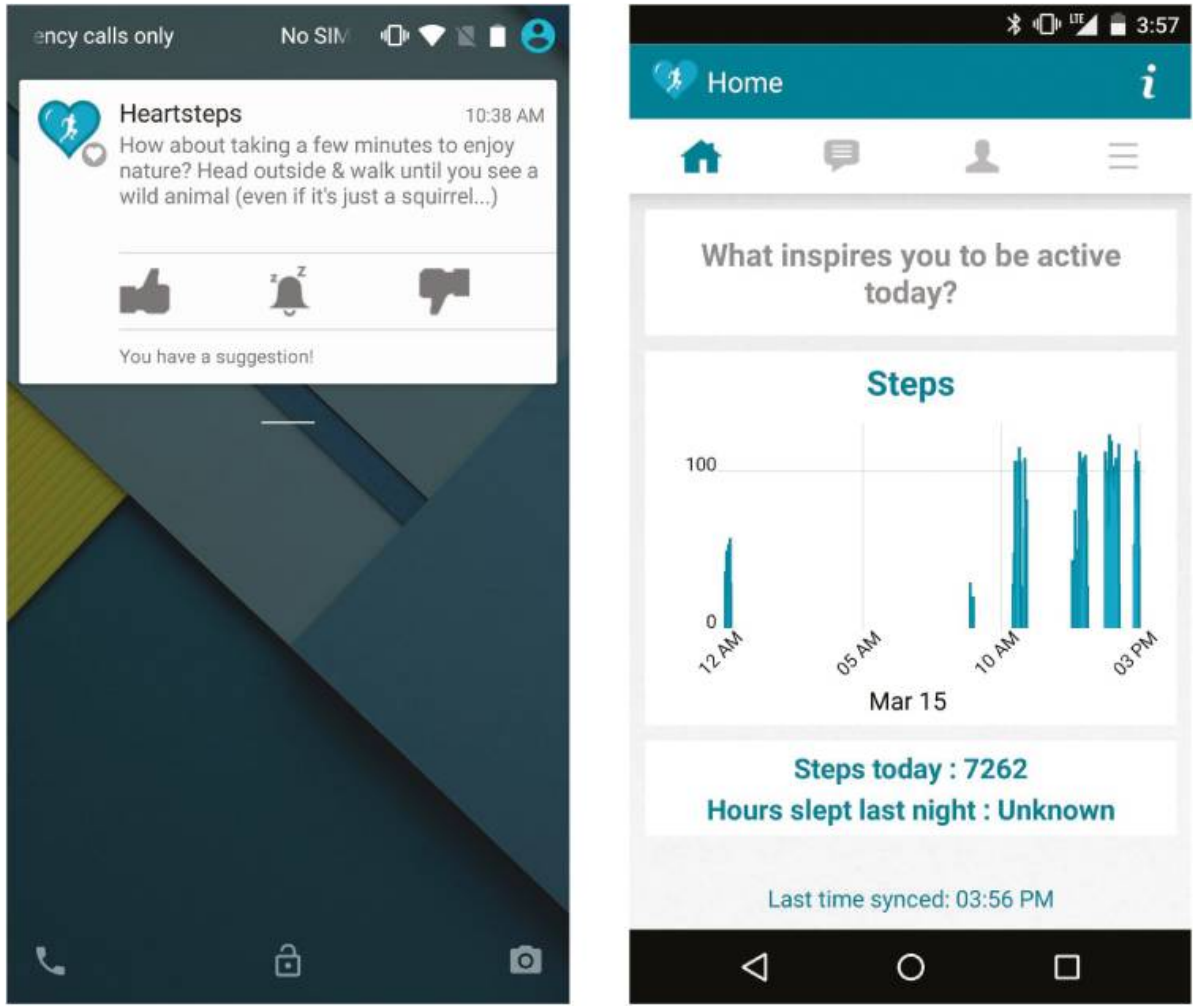
Walking suggestions

– context tailored (sleeping, driving etc.)

Source: Susan Muphy and Pedrag Klasnja's works

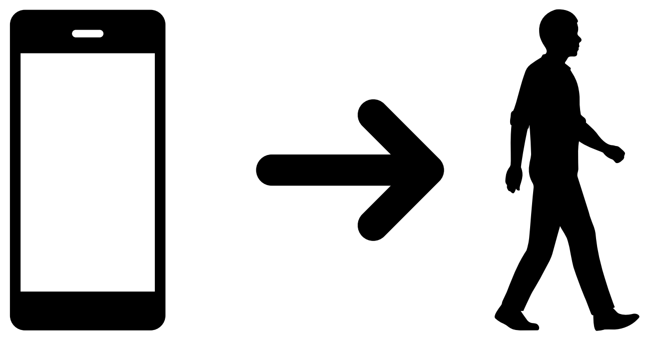
Motivation: HeartSteps V1

Mobile activity walking coach



Source: Susan Muphy and Pedrag Klasnja's works

Activity prompts 5x a day



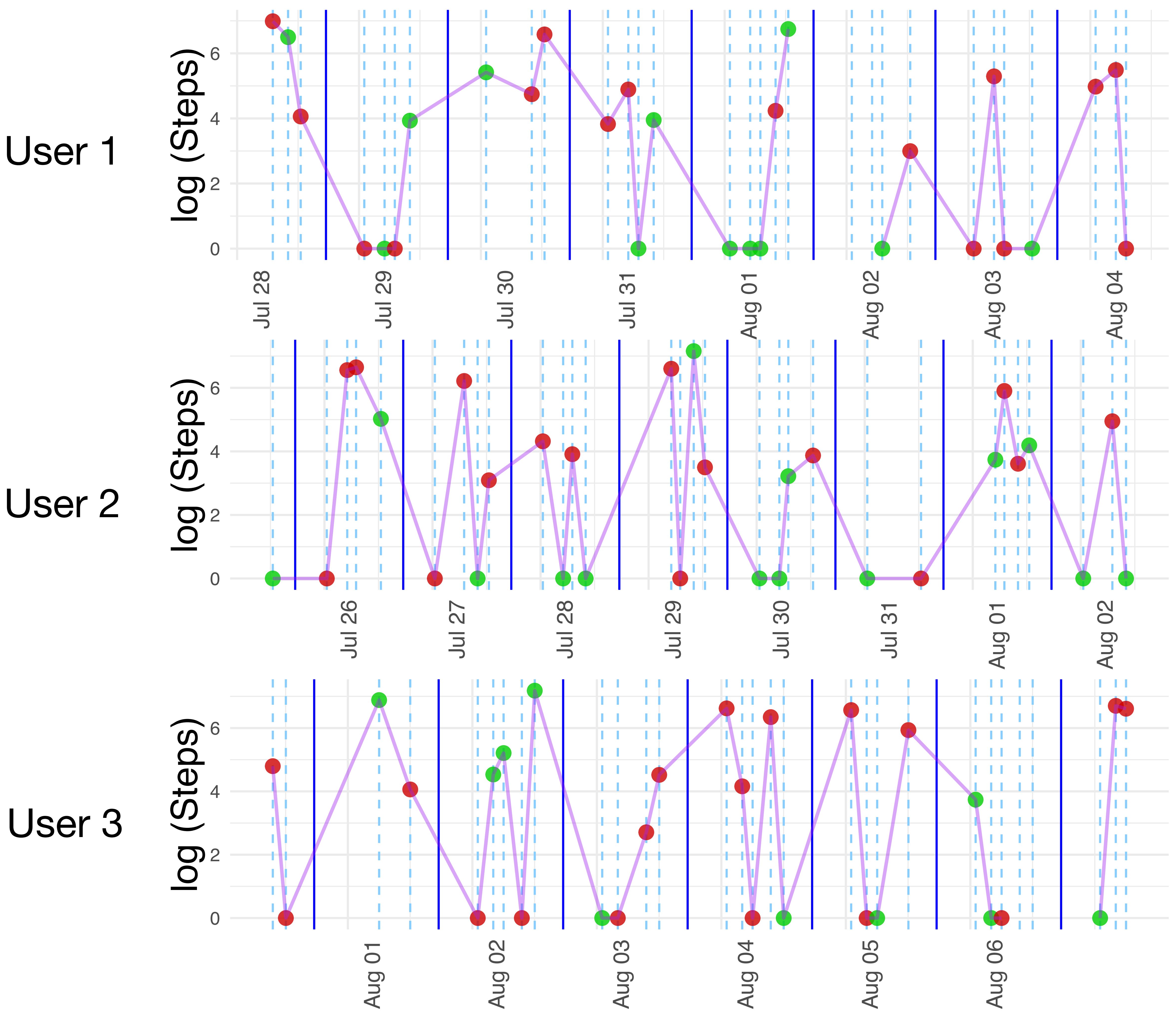
— morning, afternoon, evening, before dinner, before sleep

Walking suggestions

— context tailored (sleeping, driving etc.)

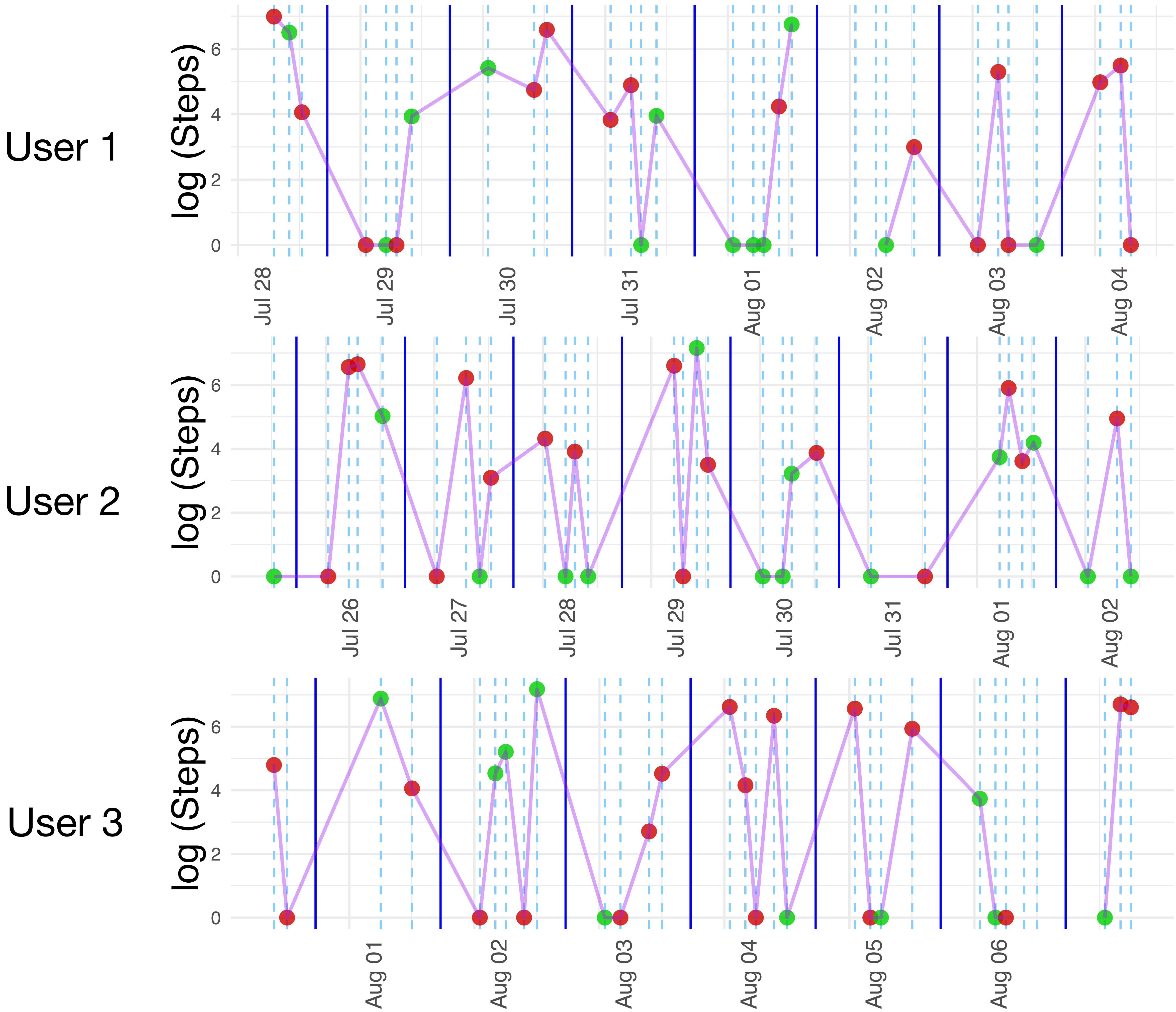
Motivation: HeartSteps V1

Leveraging the temporal walking



Motivation: HeartSteps V1

Leveraging the temporal walking

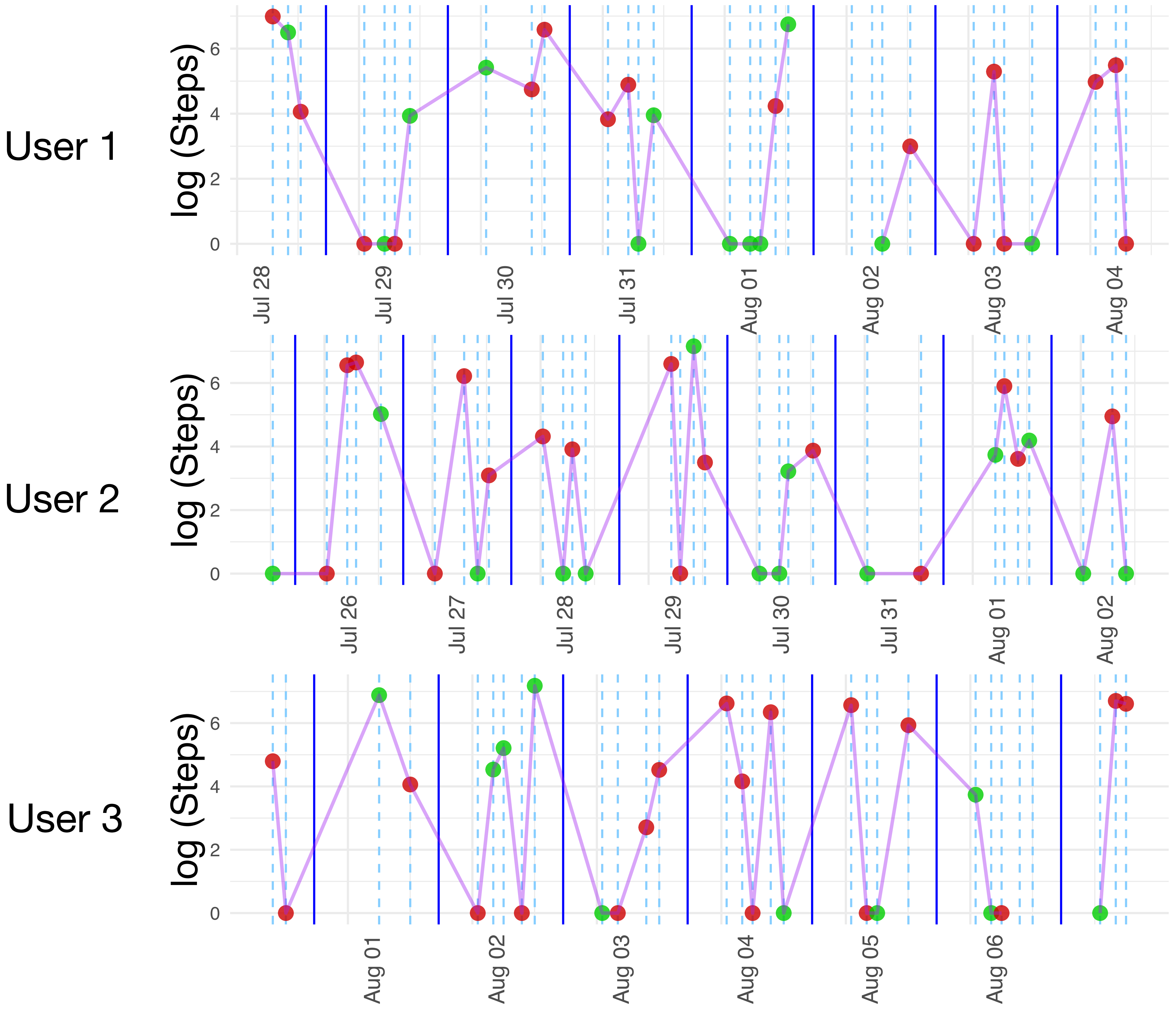


📉 Observed steps under intervention — **Nudge (treatment)** and **no nudge (control)**

Heartsteps

Motivation: HeartSteps V1

Leveraging the temporal walking



Temporal walking behavior

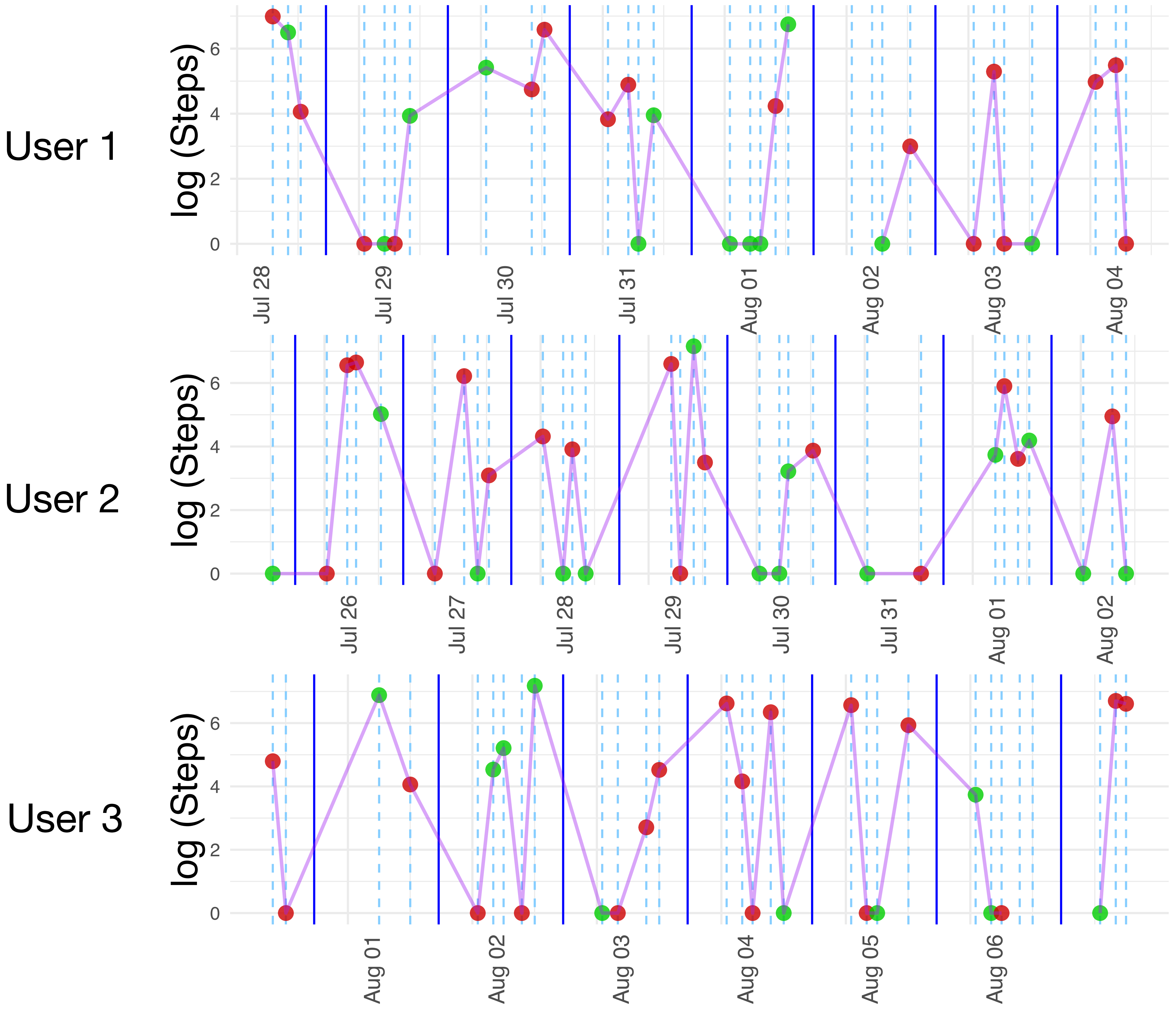
+

Shared pattern across the users

Motivation: HeartSteps V1

Heartsteps

Leveraging the temporal walking



Accurate forecast of potential steps

$\overset{?}{\Longrightarrow}$ More informed decision?

Is the treatment effective?

Counterfactual forecasting for N users, T decision points

Counterfactual forecasting for N users, T decision points

📳 Treatment $w \in \{0,1\}$ (**no prompt** vs **prompt**)

Counterfactual forecasting for N users, T decision points

Counterfactual steps under treatment w for user i and time t

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

— $\theta_{i,t}(w)$: mean potential outcomes

— $\varepsilon_{i,t}$: Idiosyncratic noise

Counterfactual forecasting for N users, T decision points

🚶 Counterfactual steps under treatment w for user i and time t

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

– $\theta_{i,t}(w)$: mean potential outcomes
– $\varepsilon_{i,t}$: Idiosyncratic noise

	Time 1	Time 2	…	Time T
User 1	0	0	0	0
User 2	0	0	0	0
User 3	0	0	0	0

Counterfactual forecasting for N users, T decision points

Counterfactual steps under treatment w for user i and time t

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

– $\theta_{i,t}(w)$: mean potential outcomes
– $\varepsilon_{i,t}$: Idiosyncratic noise

	Time 1	Time 2	…	Time T
User 1	1	1	1	1
User 2	1	1	1	1
User 3	1	1	1	1

Counterfactual forecasting for N users, T decision points

 Observed: $Y_{i,t} = \theta_{i,t}(W_{i,t}) + \varepsilon_{i,t}$

Observed panel $Y_{i,t}$

	Time 1	Time 2	...	Time T
User 1	0	1	0	1
User 2	1	1	0	1
User 3	1	0	1	0

Counterfactual forecasting for N users, T decision points

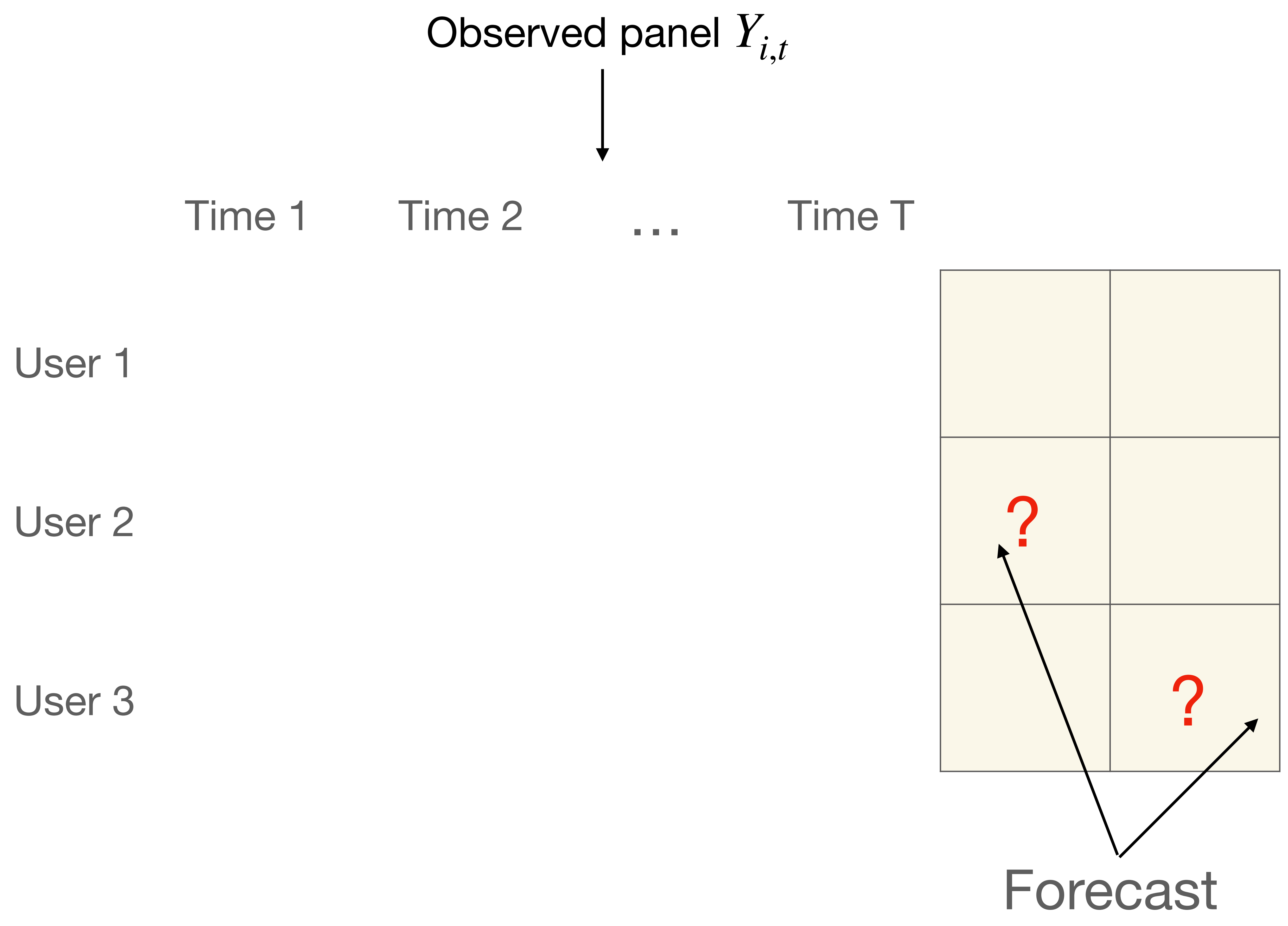
Observed panel $Y_{i,t}$

	Time 1	Time 2	…	Time T
User 1	0	1	0	1
User 2	1	1	0	1
User 3	1	0	1	0

Challenges

- **Missing data:** $2NT$ unknowns, NT observations
- **Stochastic temporal** latent factors

Counterfactual forecasting for N users, T decision points



Challenges

- **Missing data:** $2NT$ unknowns, NT observations
- **Stochastic temporal** latent factors

Counterfactual forecasting for N users, T decision points

Counterfactual forecasting for N users, T decision points

Low rank:

$$\theta_{i,t}(w) = \Lambda_i^\top(w)\; F_t(w),$$

- $\Lambda_i(w), F_t(w)$ has dimension $r \leq \min\{N, T\}$.
- $(N + T)r$ unknowns

Counterfactual forecasting for N users, T decision points

- [x] **Low rank:**

$$\theta_{i,t}(w) = \Lambda_i^\top(w)\ F_t(w),$$

  - $\Lambda_i(w), F_t(w)$ has dimension $r \leq \min\{N, T\}$.
  - $(N + T)r$ unknowns

- [ ] Factors $F_t(w)$: shared across the users
  - **Shared dependence** of walking behavior
  - **Temporal dynamics** (e.g. Markov, VAR, state space…)

Counterfactual forecasting for N users, T decision points

☑ **Low rank:**

$$\theta_{i,t}(w) = \Lambda_i^\top(w)\ F_t(w),$$

- $\Lambda_i(w), F_t(w)$ has dimension $r \leq \min\{N, T\}$.
- $(N + T)r$ unknowns

☐ Factors $F_t(w)$: shared across the users

- **Shared dependence** of walking behavior
- **Temporal dynamics** (e.g. Markov, VAR, state space…)

☐ Loadings $\Lambda_i(w)$: shared across time

- **Association** among observed steps and the latent structure

Counterfactual forecasting for N users, T decision points

- ☑ **Low rank:**

$$\theta_{i,t}(w) = \Lambda_i^\top(w)\ F_t(w),$$

  - $\Lambda_i(w), F_t(w)$ has dimension $r \leq \min\{N, T\}$.
  - $(N + T)r$ unknowns
- ☐ Factors $F_t(w)$: shared across the users
  - **Shared dependence** of walking behavior
  - **Temporal dynamics** (e.g. Markov, VAR, state space…)
- ☐ Loadings $\Lambda_i(w)$: shared across time
  - **Association** among observed steps and the latent structure
- ☑ **This talk**:
  Forecasting potential steps $Y_{i,T+h}(w)$ ($h =$ horizon) under stochastic dynamic $F_t(w)$.

Prior research

Prior research

- **Matrix completion literature**

Bai and Ng 2021 (Tall-Wide algorithm), Jin et al 202 (EM), Cahan et al. 2023 (Tall-Project algorithm), Xiong and Pelger 2023 (PCA), Goldin et al. 2022 (SyN-BEATS), **Agarwal et al 2020 (Multiple singular spectrum analysis, mSSA)**, Alomar et al 2024 (SAMoSSA) …

Prior research

😇 **Missingness**

Prior research

☹️ **Stochastic temporal factors**

Prior research

Time series literature

Bräuning and Koopman 2014, Jungbacker and Koopman 2008, Doz et al. 2011, Poncela et al. 2021, …

Prior research

😇 Leverages **stochastic dynamics** of factors by filtering/smoothing approaches

Prior research

☹️ Large scale **missingness**

Bridging the gap

Bridging the gap

- **Our contribution:**
  Focus (**Fo**recasting **C**ounterfactuals **u**nder **S**tochastic dynamics)

  😇 **Stochastic temporal** dynamics of latent factors

  😇 General **missing** patterns

Bridging the gap

- **Our contribution:**

  Focus (**Fo**recasting **C**ounterfactuals **u**nder **S**tochastic dynamics)

  😇 **Stochastic temporal** dynamics of latent factors

  😇 General **missing** patterns

✤ Counterfactual forecasting methodology

— Empirical validation against mSSA (benchmark)
— Accurate forecast on HeartSteps data set
— Theoretical guarantee

Panel data with dynamic factors and missing entries

	Time 1	Time 2	…	Time T
User 1				
User 2				
User 3				

Panel data with dynamic factors and missing entries

- Observed data: $(Y_{i,t}, W_{i,t})$, $W_{i,t}$ = observation indicator.

	Time 1	Time 2	...	Time T
User 1	0	1	0	1
User 2	1	1	0	1
User 3	1	0	1	0

Panel data with dynamic factors and missing entries

- Observed data: $(Y_{i,t}, W_{i,t})$, $W_{i,t} =$ observation indicator.
- Simplified model : restrict to $w = 1$

$$Y_{i,t} = \begin{cases} \Lambda_i^\top F_t + \varepsilon_{i,t} & \text{if } W_{i,t} = 1, \\ ? & \text{if } W_{i,t} = 0\,. \end{cases}$$

	Time 1	Time 2	…	Time T
User 1	?	✓	?	✓
User 2	✓	✓	?	✓
User 3	✓	?	✓	?

Panel data with dynamic factors and missing entries

- Observed data: $(Y_{i,t}, W_{i,t})$, $W_{i,t}$ = observation indicator.
- Simplified model : restrict to $w = 1$

$$Y_{i,t} = \begin{cases} \Lambda_i^\top F_t + \varepsilon_{i,t} & \text{if } W_{i,t} = 1, \\ ? & \text{if } W_{i,t} = 0\,. \end{cases}$$

- Vector autoregressive model of order 1 ≡ **VAR(1)** factors:

$$F_t = AF_{t-1} + \eta_t$$

  - A: Transition matrix (F_t is stationary, stable for $\rho(A) < 1$)
  - η_t: Stationary noise process

	Time 1	Time 2	…	Time T
User 1	?	✓	?	✓
User 2	✓	✓	?	✓
User 3	✓	?	✓	?

Panel data with dynamic factors and missing entries

What is the forecast estimand?

$$F_{T+h} = A^h F_T + \sum_{j=1}^{h} A^{h-j} \eta_{T+j}$$

Panel data with dynamic factors and missing entries

What is the forecast estimand?

- Future realizations: $F_T = AF_{T-1} + \eta_T = A^2F_{T-2} + A\eta_{T-1} + \eta_T = \cdots$

$$F_{T+h} = \quad A^h F_T \quad + \quad \sum_{j=1}^{h} A^{h-j}\eta_{T+j}$$

Filter on information until T, i.e. project on $\sigma(F_1, \ldots, F_T)$

Panel data with dynamic factors and missing entries

What is the forecast estimand?

- Future realizations: $F_T = AF_{T-1} + \eta_T = A^2 F_{T-2} + A\eta_{T-1} + \eta_T = \cdots$

$$F_{T+h} = A^h F_T + \sum_{j=1}^{h} A^{h-j} \eta_{T+j}$$

Filter on information until T, i.e. project on $\sigma(F_1, \ldots, F_T)$

- Expected mean outcome

$$\theta_{i,T:T+h} := \mathbb{E}[Y_{i,T+h} \mid F_1, \ldots, F_T] = \Lambda_i^\top A^h F_T$$

Panel data with dynamic factors and missing entries

What is the forecast estimand?

- Future realizations: $F_T = AF_{T-1} + \eta_T = A^2 F_{T-2} + A\eta_{T-1} + \eta_T = \cdots$

$$F_{T+h} = A^h F_T + \sum_{j=1}^{h} A^{h-j} \eta_{T+j}$$

Filter on information until T, i.e. project on $\sigma(F_1, \ldots, F_T)$

- Expected mean outcome

$$\theta_{i,T:T+h} := \mathbb{E}[Y_{i,T+h} \mid F_1, \ldots, F_T] = \Lambda_i^\top A^h F_T$$

☑ **Target**: Estimate $\theta_{i,T:T+h}$ using the observed data $(Y_{i,t}, W_{i,t})_{(i,t) \in [N]\times[T]}$

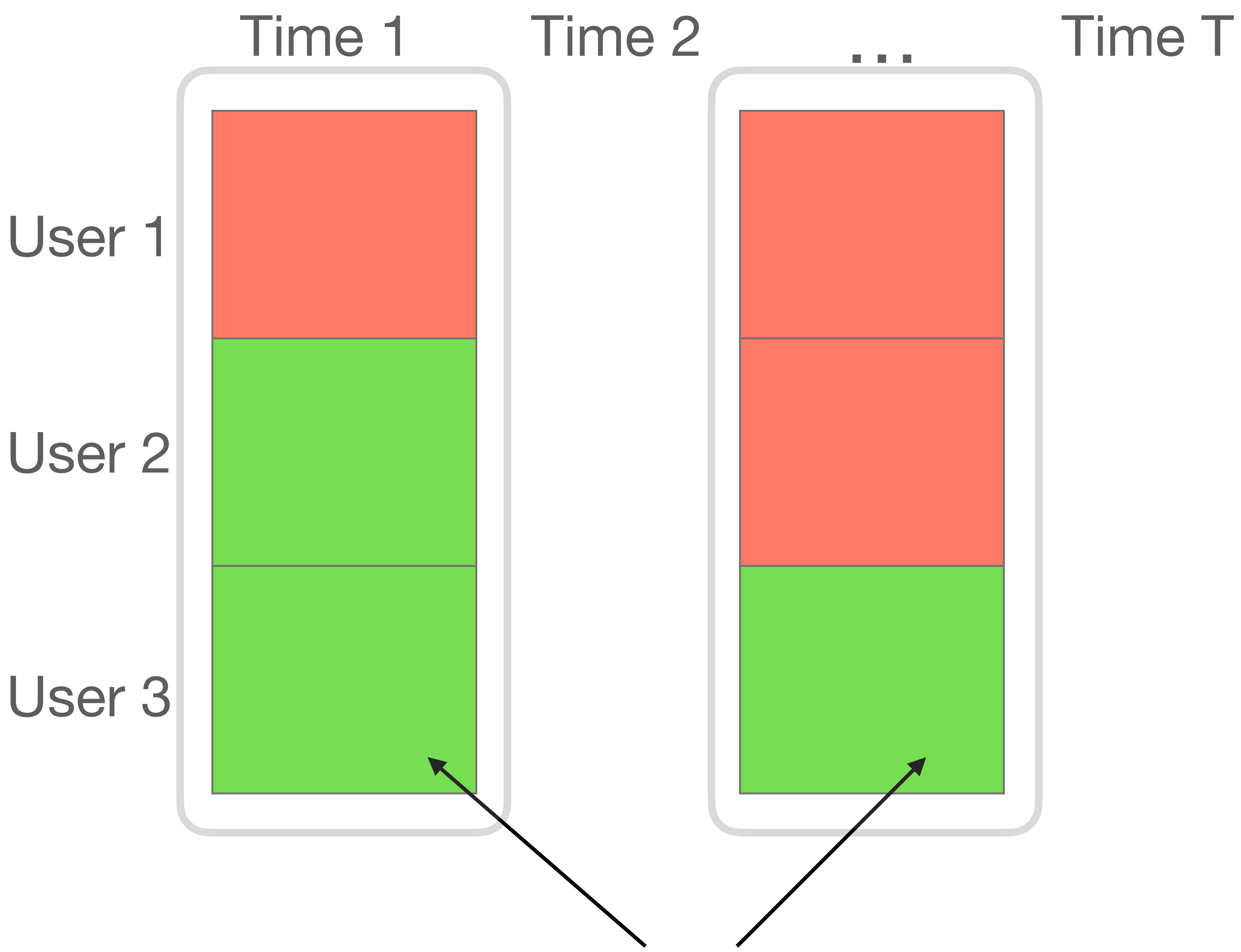
Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 1: Estimate the factors with principal component analysis

Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 1: Estimate the factors with principal component analysis

- PCA algorithm from Xiong and Pelger [2023] : Weight observations with number of treated time points



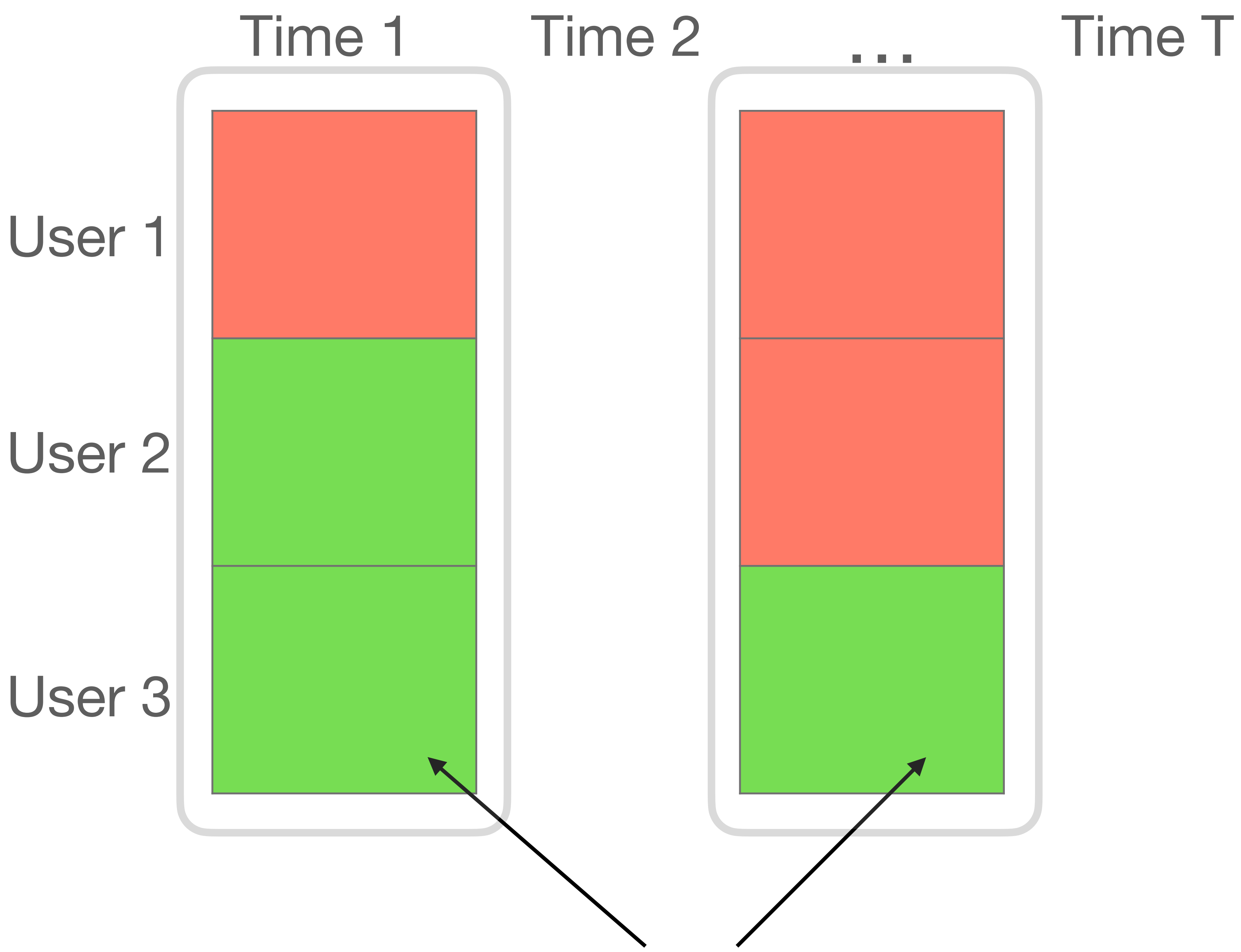
$$\sum_{i=1}^{N} W_{i,s} W_{i,t} = 1$$

Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 1: Estimate the factors with principal component analysis

- PCA algorithm from Xiong and Pelger [2023] : Weight observations with number of treated time points
- Sample covariance $\hat{\Sigma}$, where

$$\hat{\Sigma}_{s,t} = \begin{cases} \frac{\sum_{i=1}^{N} W_{i,s}W_{i,t}Y_{i,s}Y_{i,t}}{\sum_{i=1}^{N} W_{i,s}W_{i,t}} & \text{if } \sum_{i=1}^{N} W_{i,s}W_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$



$$\sum_{i=1}^{N} W_{i,s}W_{i,t} = 1$$

Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 1: Estimate the factors with principal component analysis

- PCA algorithm from Xiong and Pelger [2023] : Weight observations with number of treated time points
- Sample covariance $\hat{\Sigma}$, where

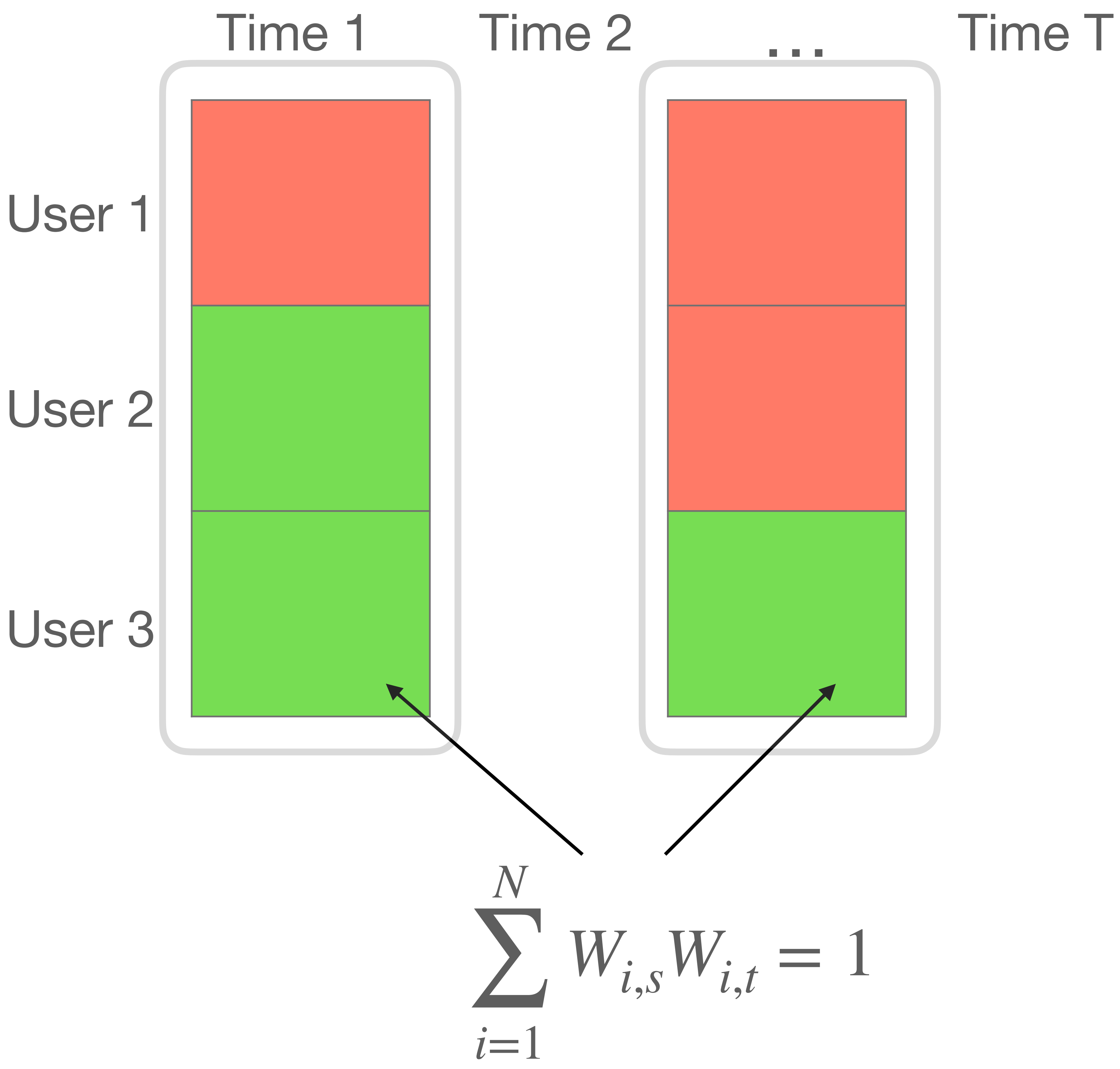
$$\hat{\Sigma}_{s,t} = \begin{cases} \dfrac{\sum_{i=1}^{N} W_{i,s}W_{i,t}Y_{i,s}Y_{i,t}}{\sum_{i=1}^{N} W_{i,s}W_{i,t}} & \text{if } \sum_{i=1}^{N} W_{i,s}W_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Estimated factors:

$$\hat{F} = \sqrt{T} \times \text{First } r \text{ eigenvectors of } \frac{1}{T}\hat{\Sigma}$$

- Estimated loadings by regressing $Y_{i,t}$ on $W_{i,t}\hat{F}_t$

$$\hat{\Lambda}_i = \left(\sum_{t=1}^{T} W_{i,t}\hat{F}_t\hat{F}_t^{\top}\right)^{-1}\left(\sum_{t=1}^{T} W_{i,t}\hat{F}_tY_{i,t}\right)$$



Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 2: Forecast with the estimated factors

Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 2: Forecast with the estimated factors

- Back to the estimand

$$\theta_{i,T:T+h} = \mathbb{E}[\theta_{i,T+h} \mid \mathscr{F}_T] = \Lambda_i^\top A^h F_T$$

Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 2: Forecast with the estimated factors

- OLS estimator of A is

$$\hat{A} = \Big(\sum_{t=1}^{T-1} \hat{F}_{t+1}\hat{F}_t^\top\Big)\Big(\sum_{t=1}^{T-1} \hat{F}_t\hat{F}_t^\top\Big)^{-1}$$

Focus (Forecasting Counterfactuals under Stochastic dynamics)

Step 2: Forecast with the estimated factors

☑ Plug-in estimator:

$$\hat{\theta}_{i,T:T+h} = \hat{\Lambda}_i^\top \hat{A}^h \hat{F}_T$$

Leveraging latent dynamics $\Longrightarrow$ Accurate forecast

Simulation study

- **Benchmark**: mSSA ≡ Multivariate singular spectrum analysis (Agarwal et al. 2020)

  Does not capture stochastic dynamics of the factors

- **Generative model**:

$$Y_{i,t} = \Lambda_i F_t + \varepsilon_{i,t},$$

  $\Lambda_i, \varepsilon_{i,t}$ iid normal

- **Factors**: quadratic + AR(1)

  $F_t = 100t^2/T^2 + F_t^0, \qquad F_t^0 = 0.5F_{t-1}^0 + \eta_t$

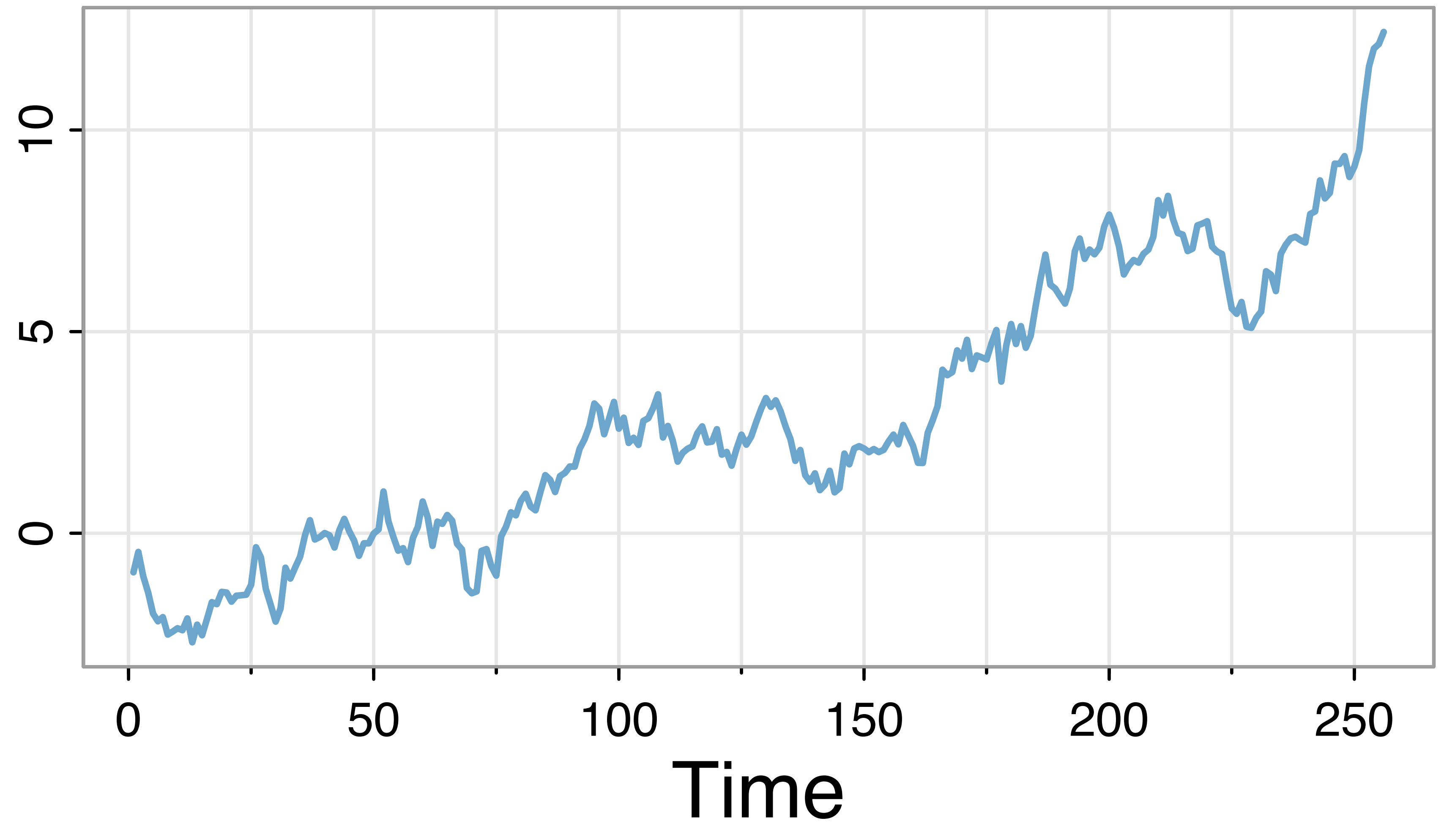
  η_t iid normal

- **Missing pattern:** MCAR, simultaneous adoption

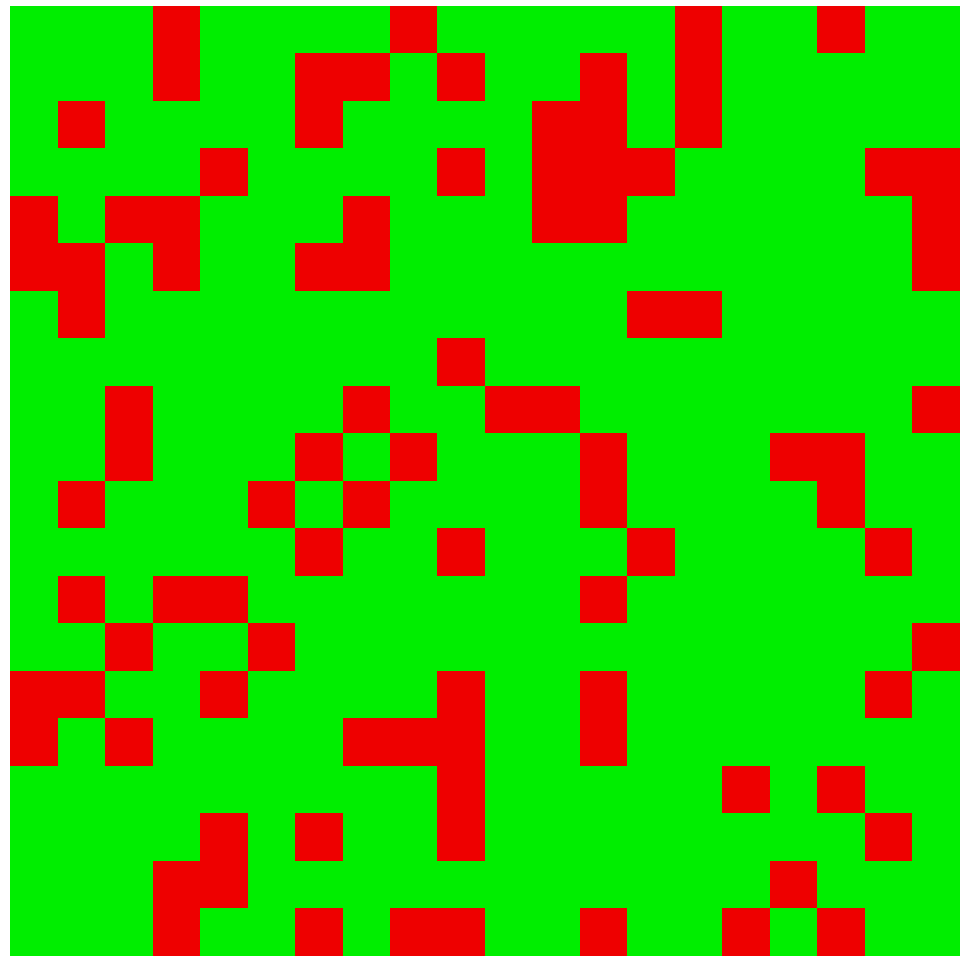
- Use Spline smoothing + Focus on estimated factors

- **Metric**: Median of one-step absolute forecast error

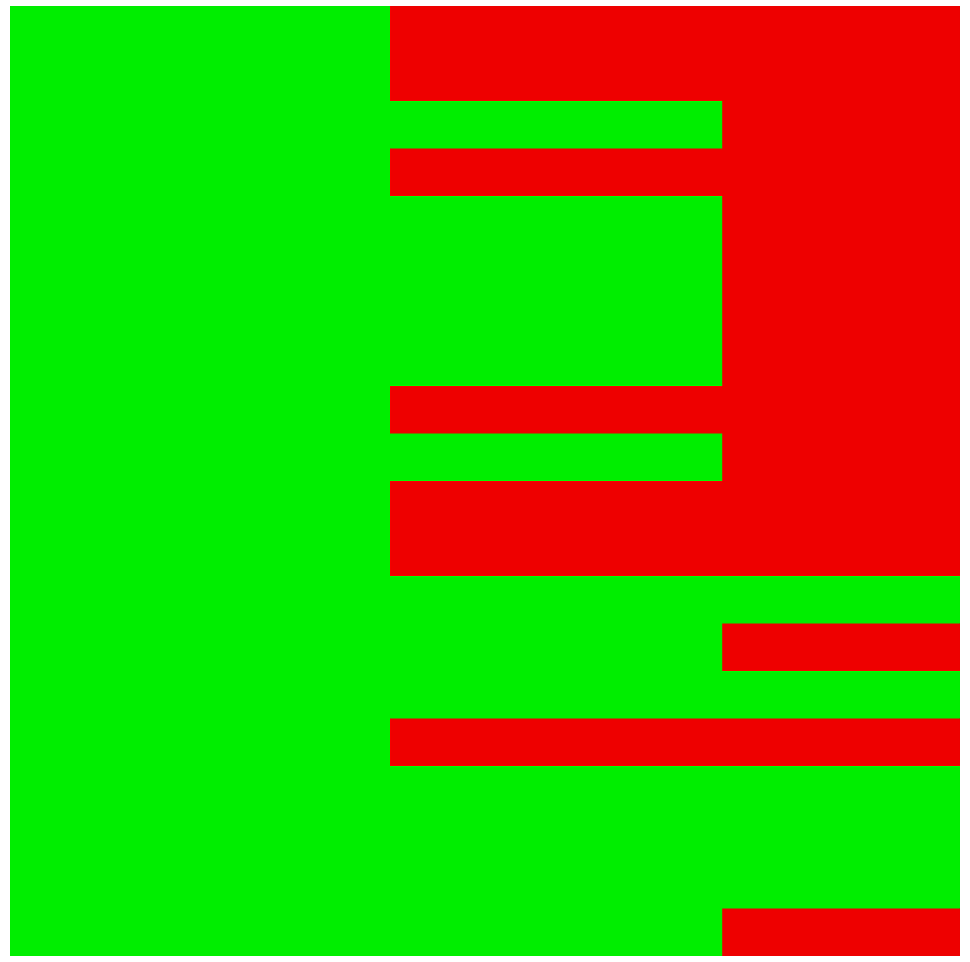
True factors



MCAR(0.8)



Simultaneous adoption

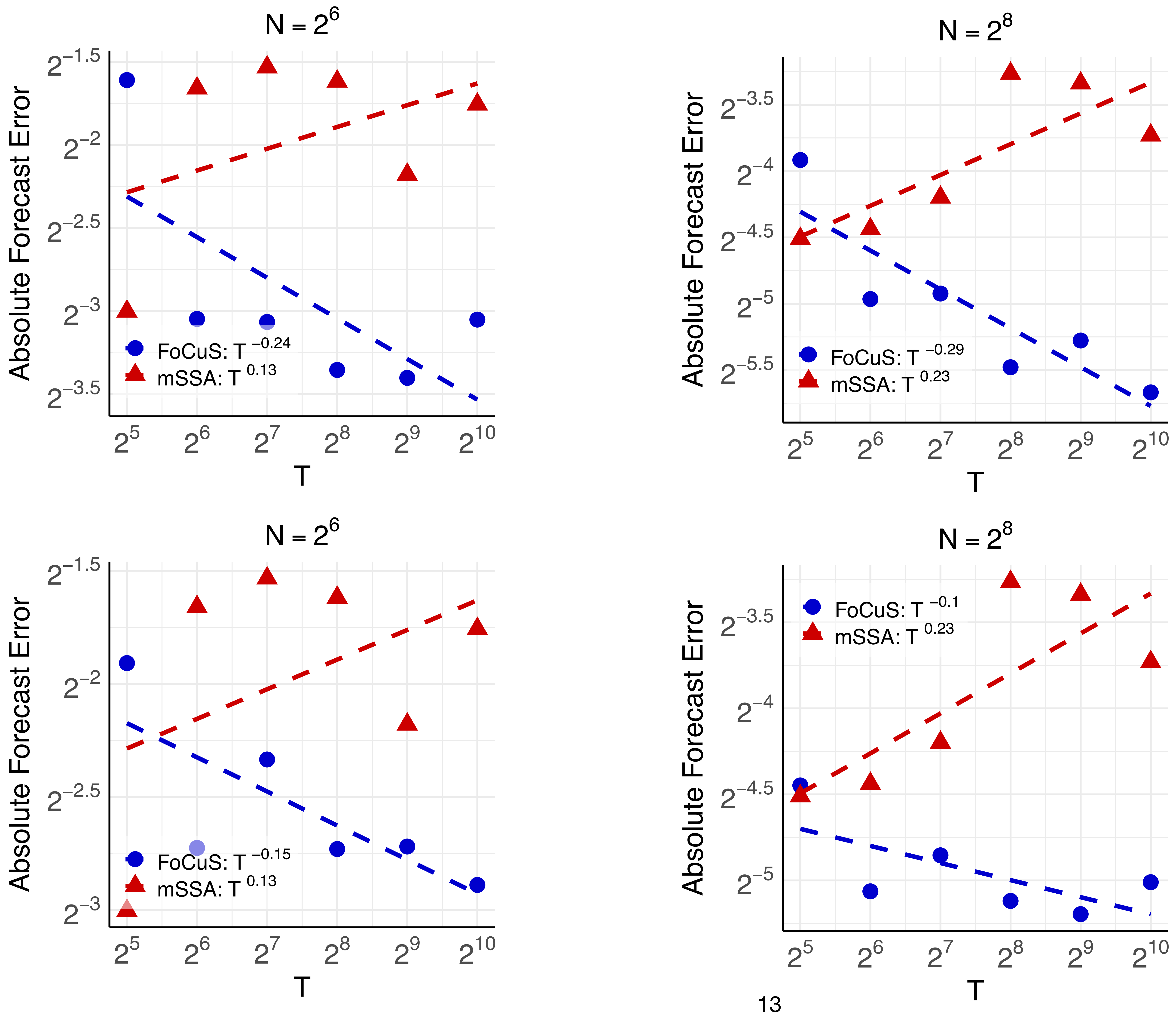


Leveraging latent dynamics ⟹ Accurate forecast

One step forecast error in log2-log2 scale

MCAR (0.8)

Simultaneous Adoption

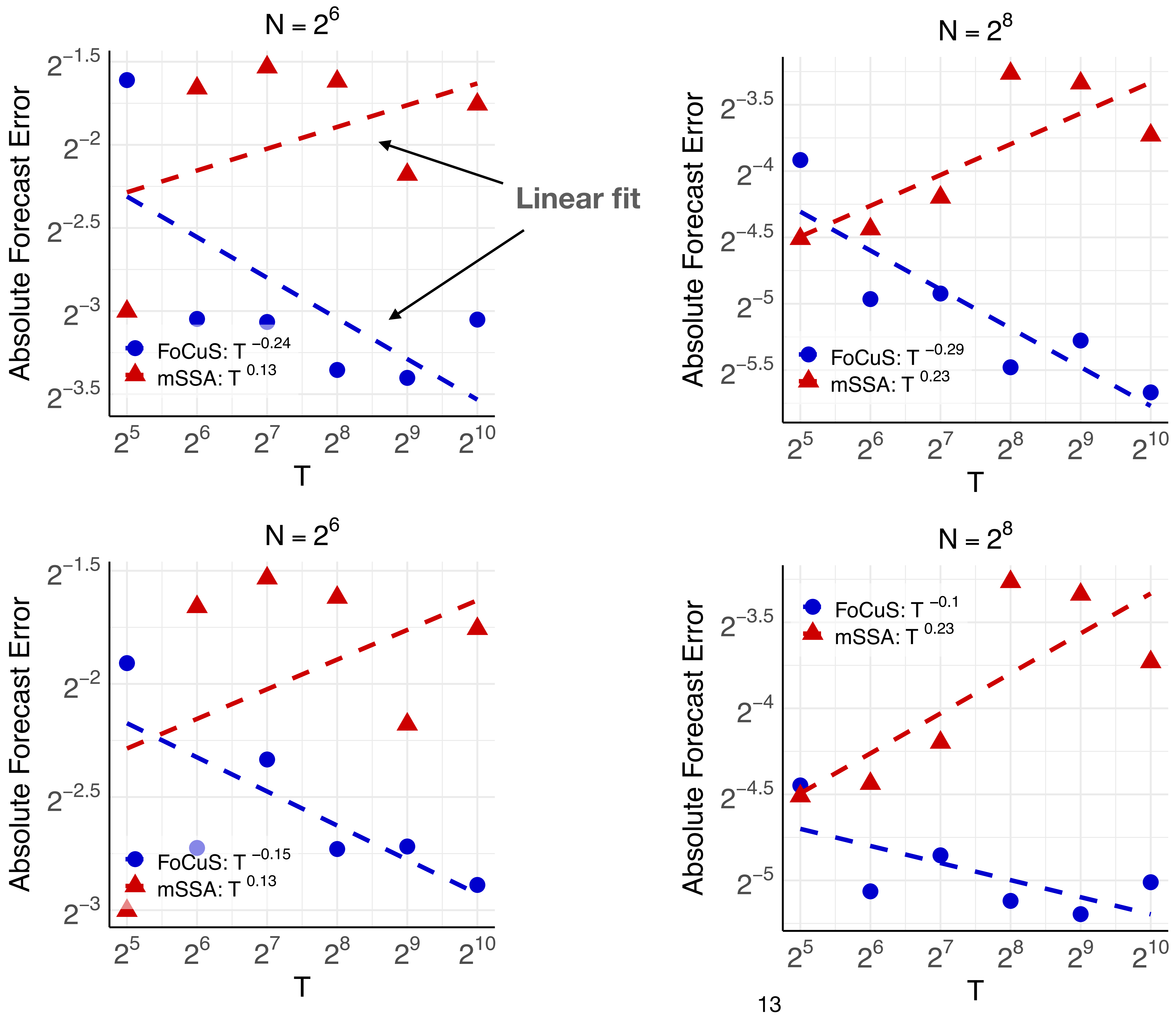


Leveraging latent dynamics ⟹ Accurate forecast

One step forecast error in log2-log2 scale

MCAR (0.8)

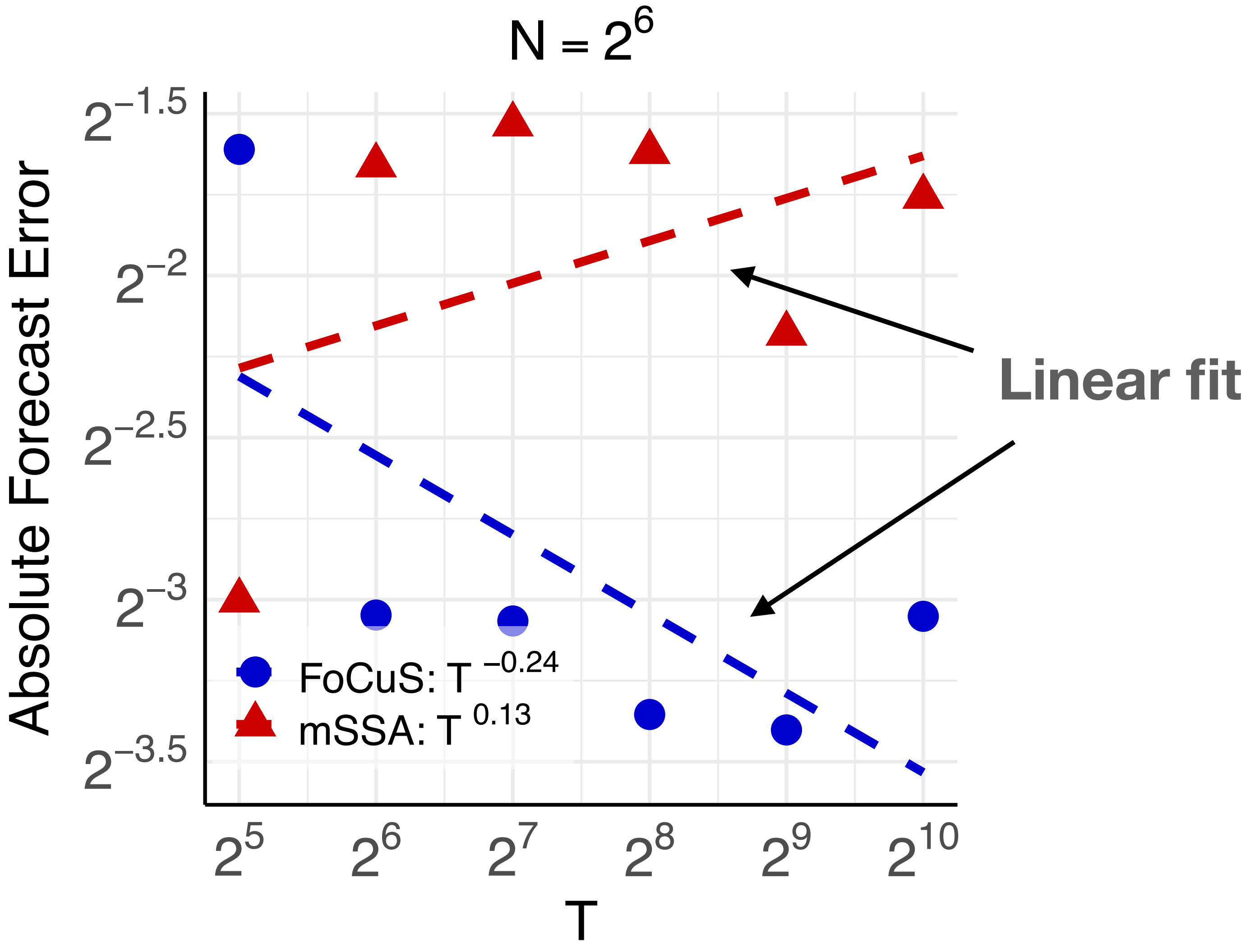
Simultaneous Adoption

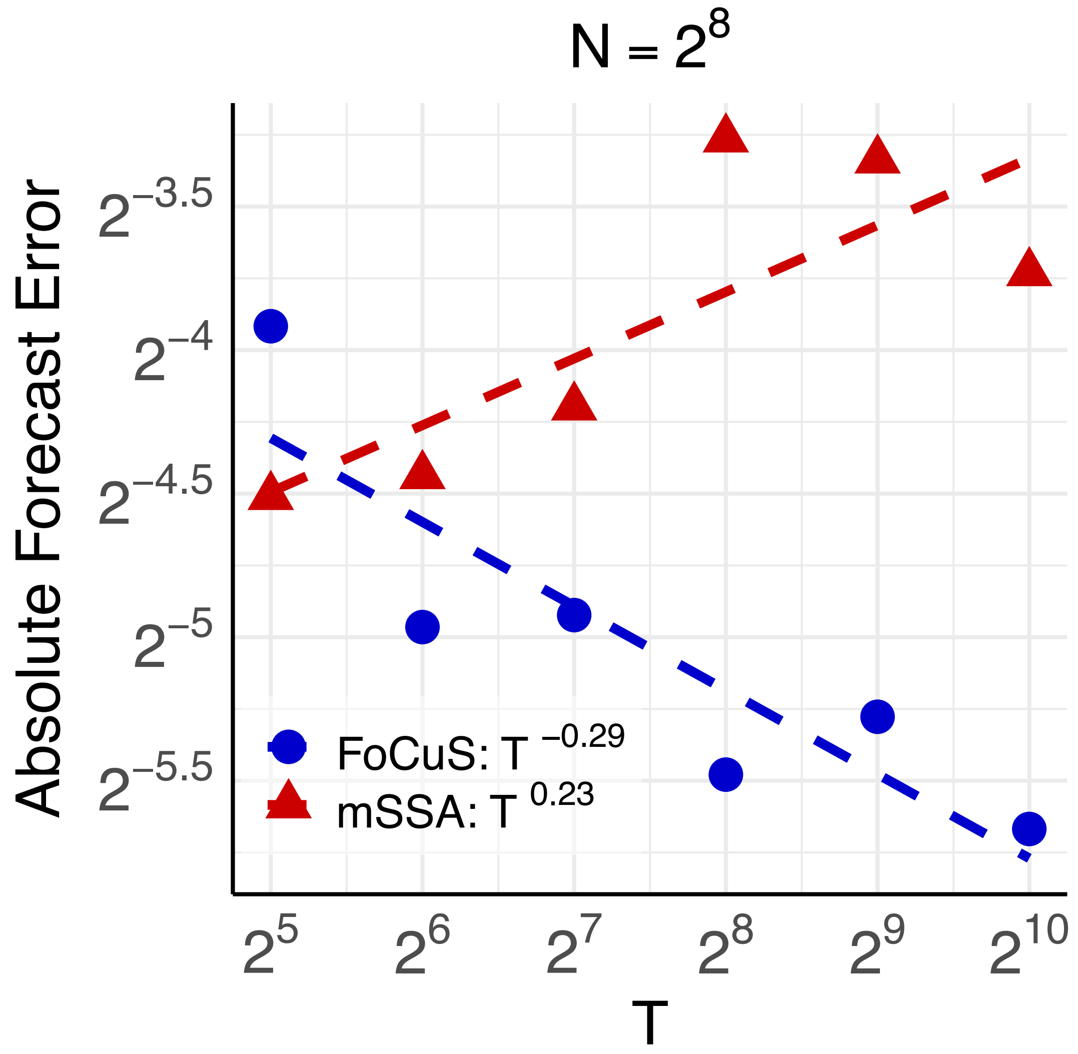


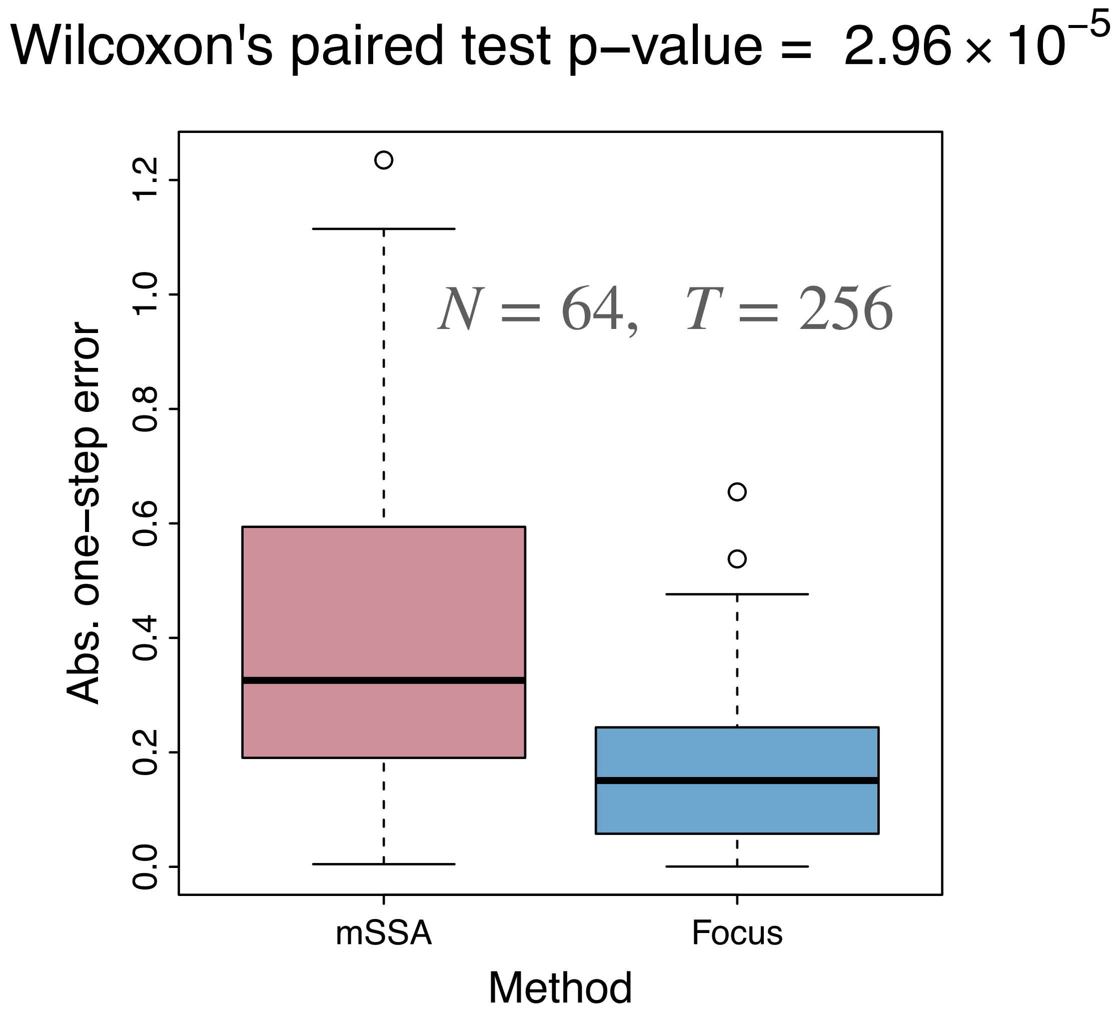
Leveraging latent dynamics $\Longrightarrow$ Accurate forecast

One step forecast error in log2-log2 scale

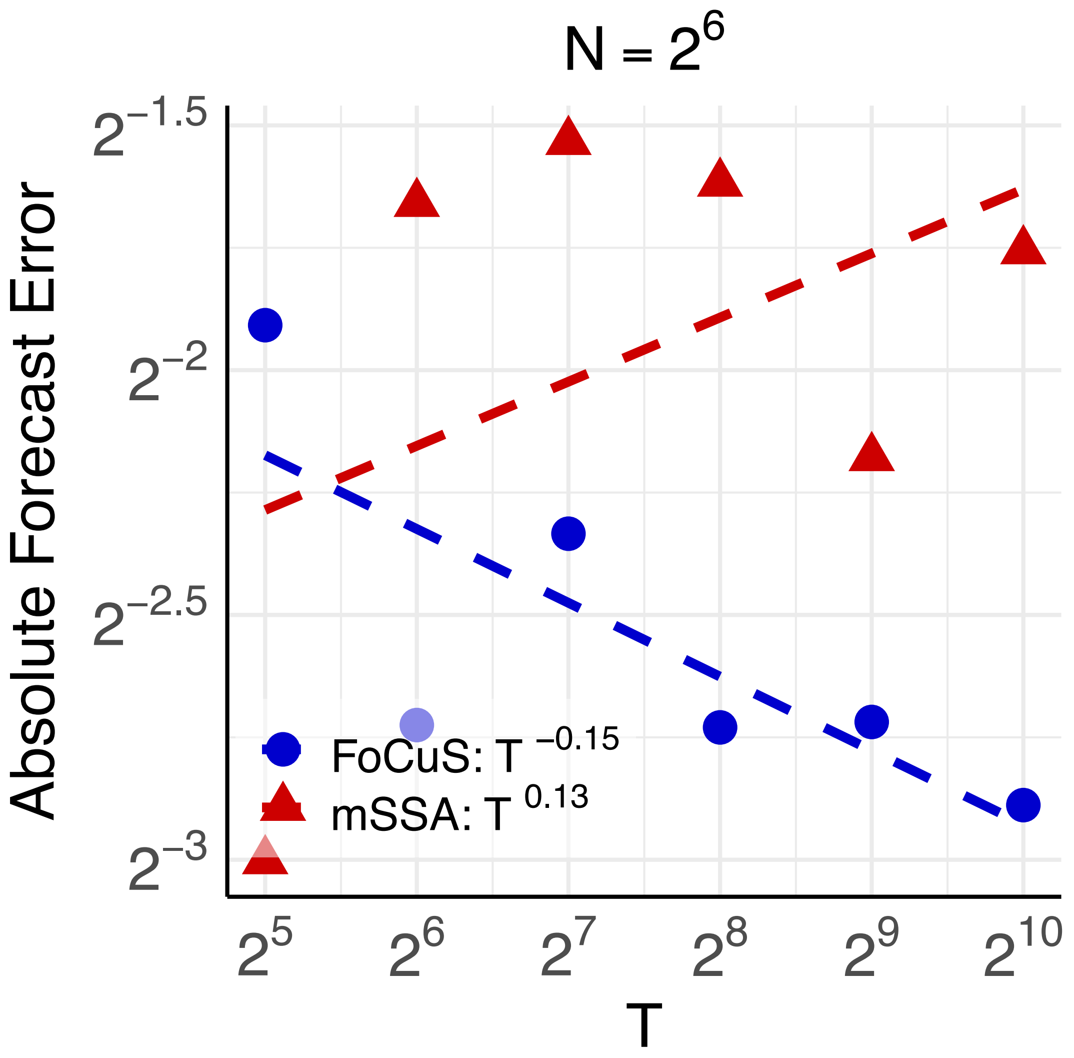
MCAR (0.8)

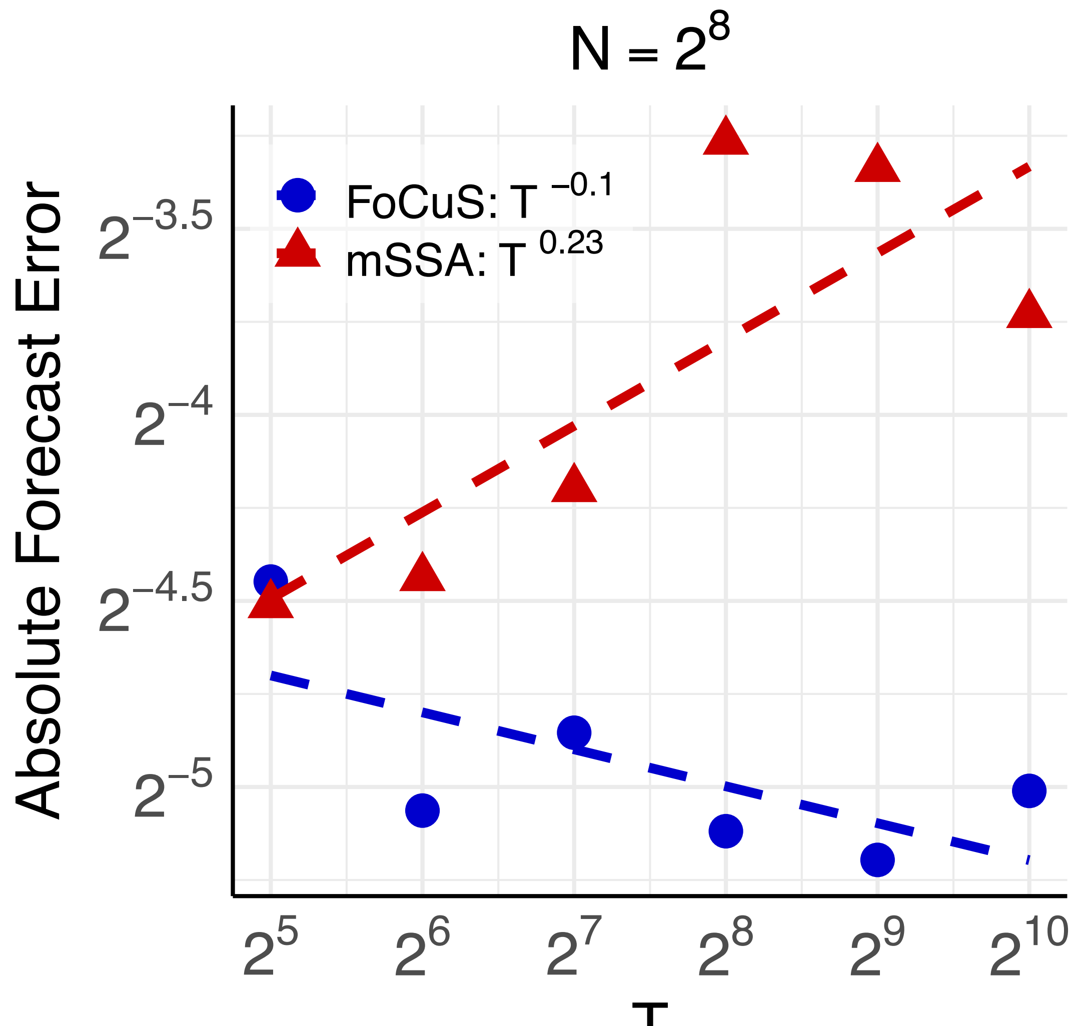


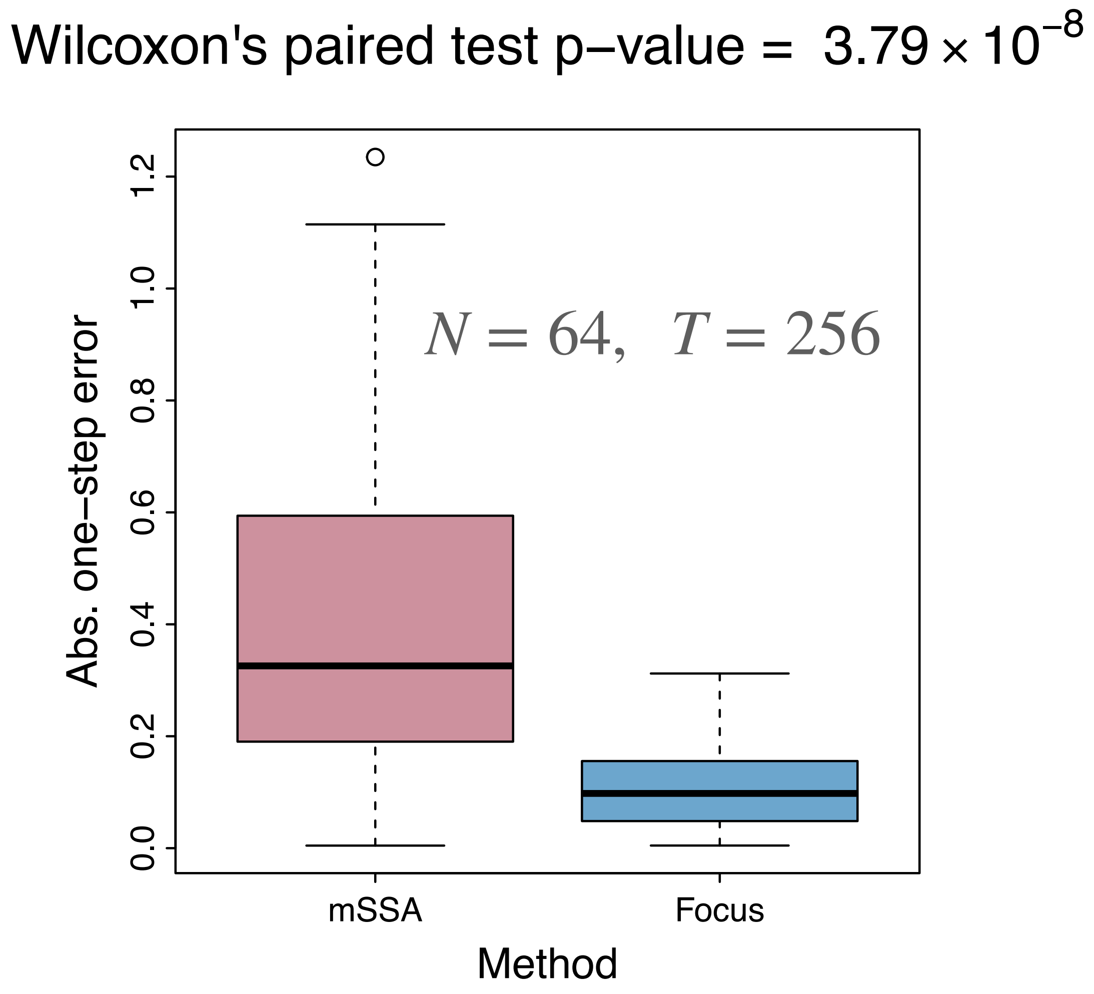




Simultaneous Adoption







Leveraging the temporal walking

Identify informative slot pairs

- $Y = \log(1 + \text{jbsteps30})$,

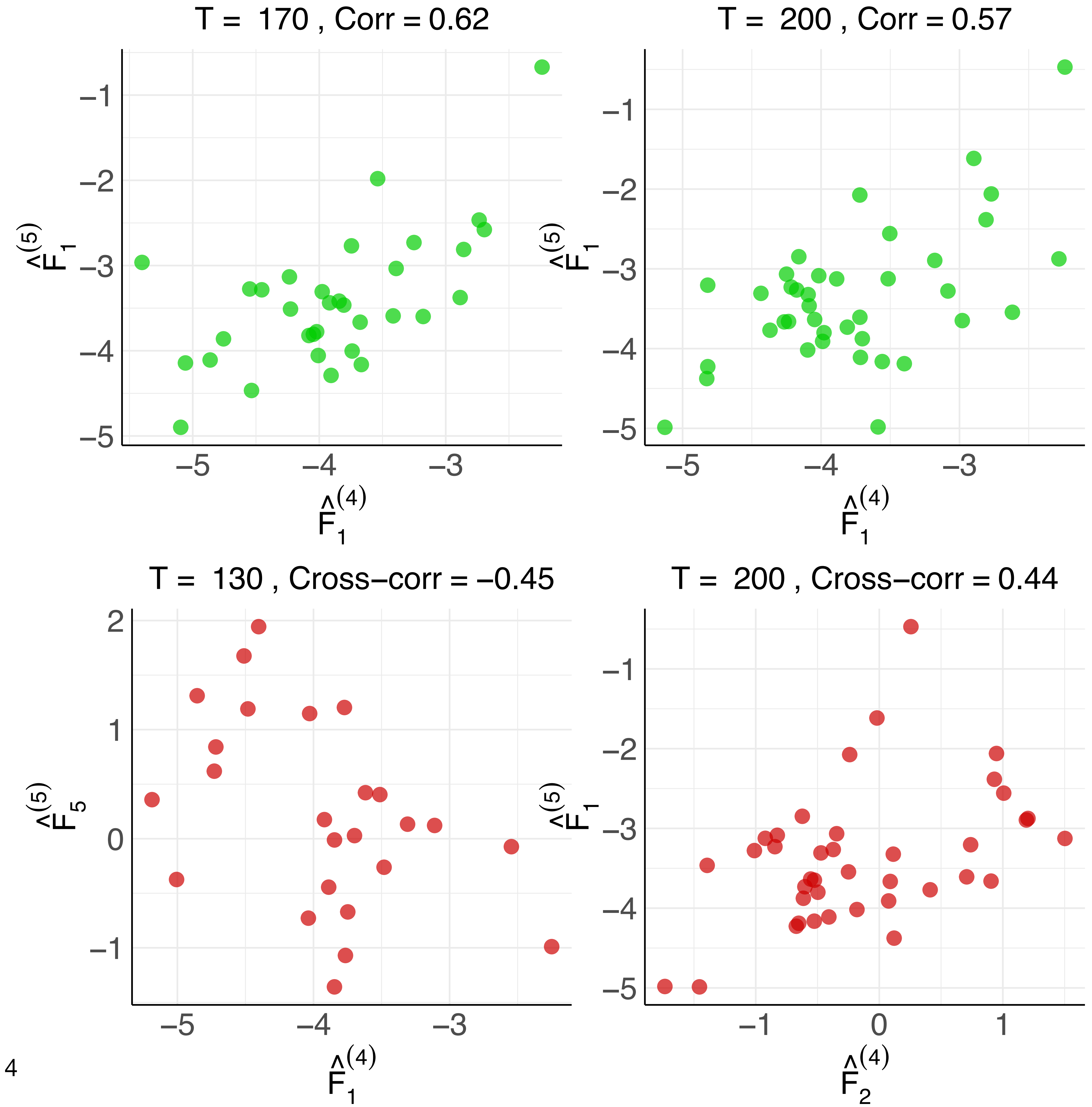
  $W = I\{\text{available, nudged}\}$

  $N = 36, \;\; T \in \{100,\ 110,\ \ldots,\ 200\}$

☑ Identify consecutive and *informative* slot pair(s)

- E.g. Late evening/before dinner (4) before sleep (5)

- Strong correlation and cross-correlation among consecutive slots!



Leveraging the temporal walking

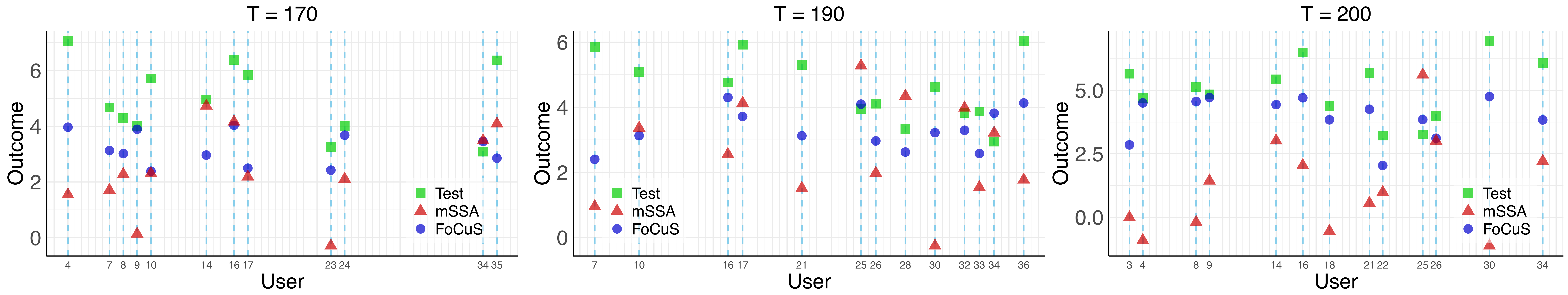
FOCUS more accurately forecasts the steps under nudge

- $\hat{\theta}_{i,T:T+1} = \hat{\Lambda}_i^\top \hat{A} \hat{F}_{T-1},\ T \in$ slot 5 $\qquad \hat{A} = (\hat{F}^{(5)})^\top \hat{F}^{(4)} \left[(\hat{F}^{(4)})^\top \hat{F}^{(4)} \right]^{-1}$

Leveraging the temporal walking

FOCUS more accurately forecasts the steps under nudge

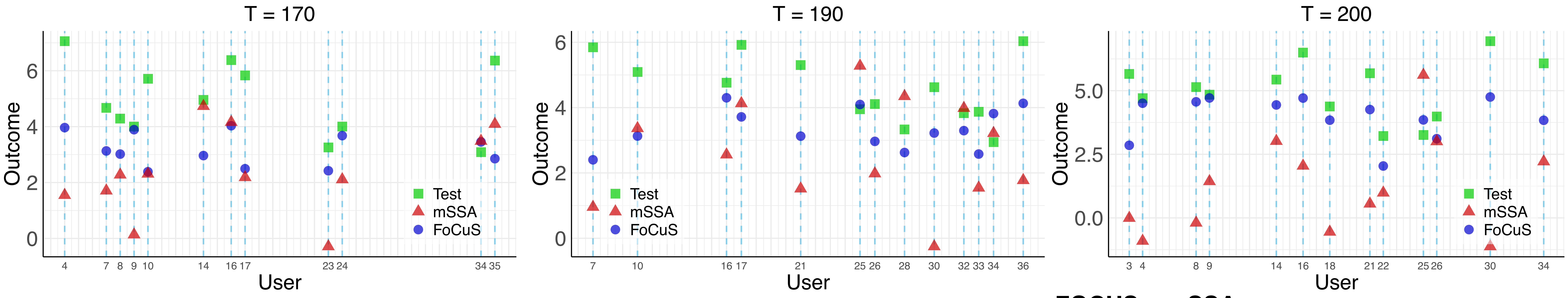
- $\hat{\theta}_{i,T:T+1} = \hat{\Lambda}_i^\top \hat{A} \hat{F}_{T-1},\ T \in$ slot 5 $\qquad \hat{A} = (\hat{F}^{(5)})^\top \hat{F}^{(4)} \left[(\hat{F}^{(4)})^\top \hat{F}^{(4)} \right]^{-1}$



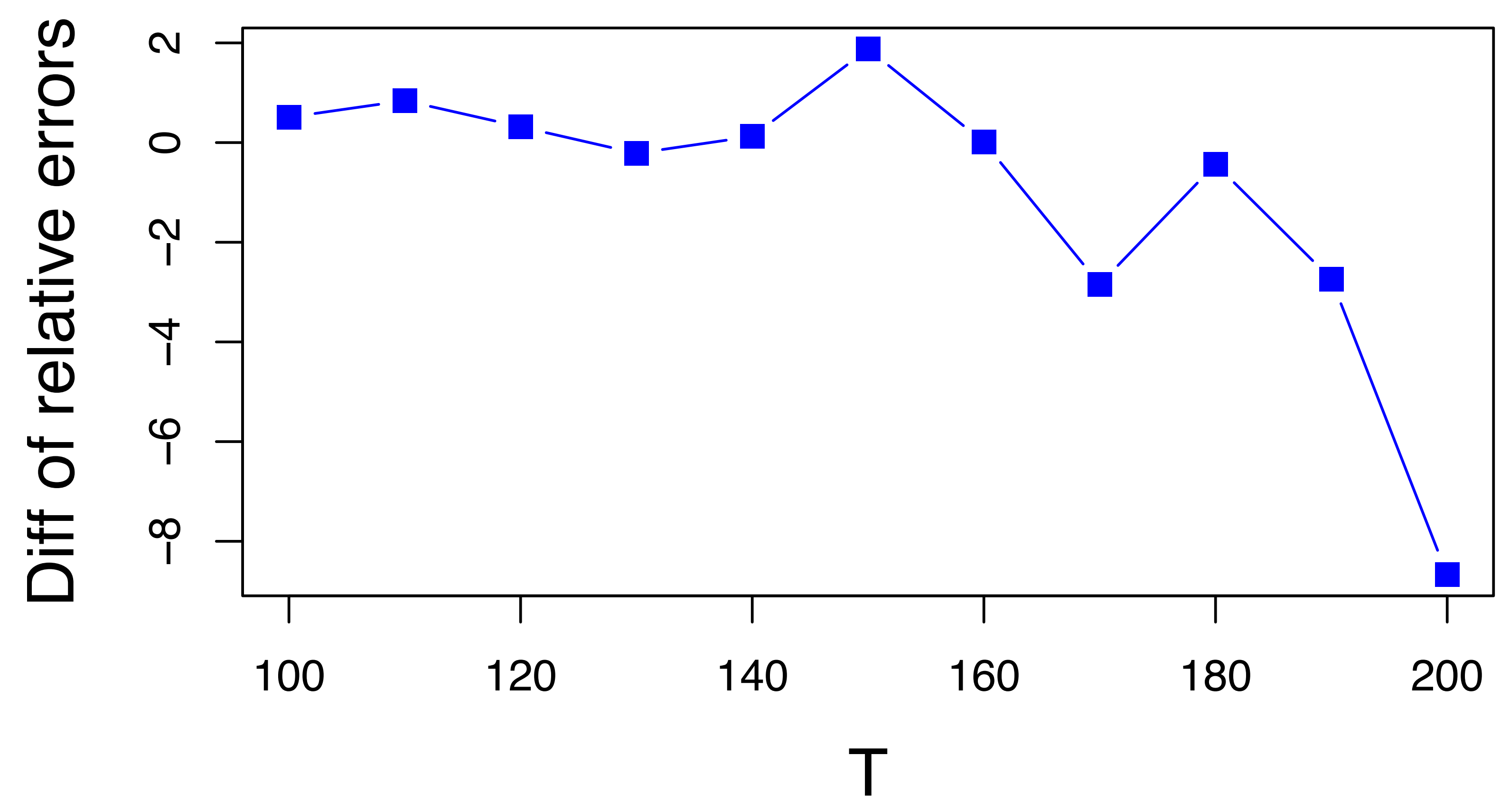
Leveraging the temporal walking

FOCUS more accurately forecasts the steps under nudge

- $\hat{\theta}_{i,T:T+1} = \hat{\Lambda}_i^\top \hat{A} \hat{F}_{T-1}$, $T \in$ slot 5 $\qquad \hat{A} = (\hat{F}^{(5)})^\top \hat{F}^{(4)} \left[(\hat{F}^{(4)})^\top \hat{F}^{(4)} \right]^{-1}$



FOCUS – mSSA



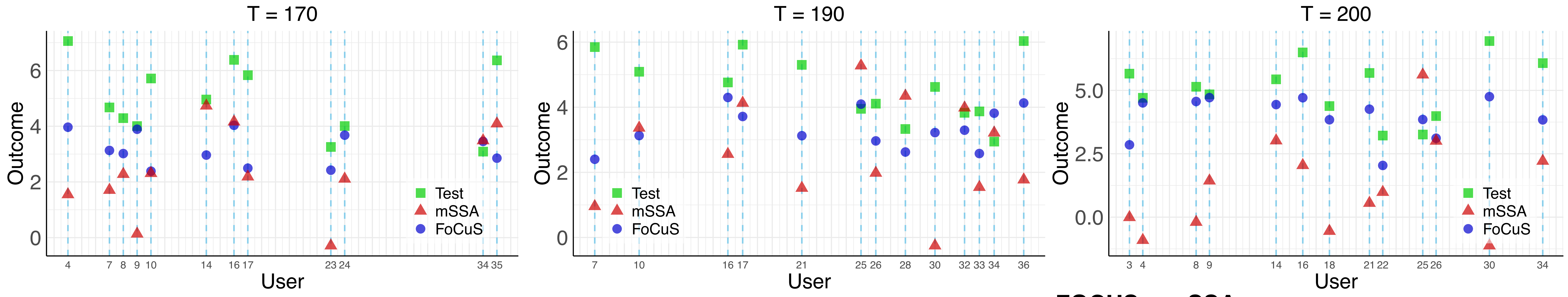
Relative error for non-zero steps

$$\sum_{i:Y_{i,T}>0} \frac{(\hat{\theta}_{i,T-1:T} - Y_{i,T})^2}{Y_{i,T}^2}$$

Leveraging the temporal walking

FOCUS more accurately forecasts the steps under nudge

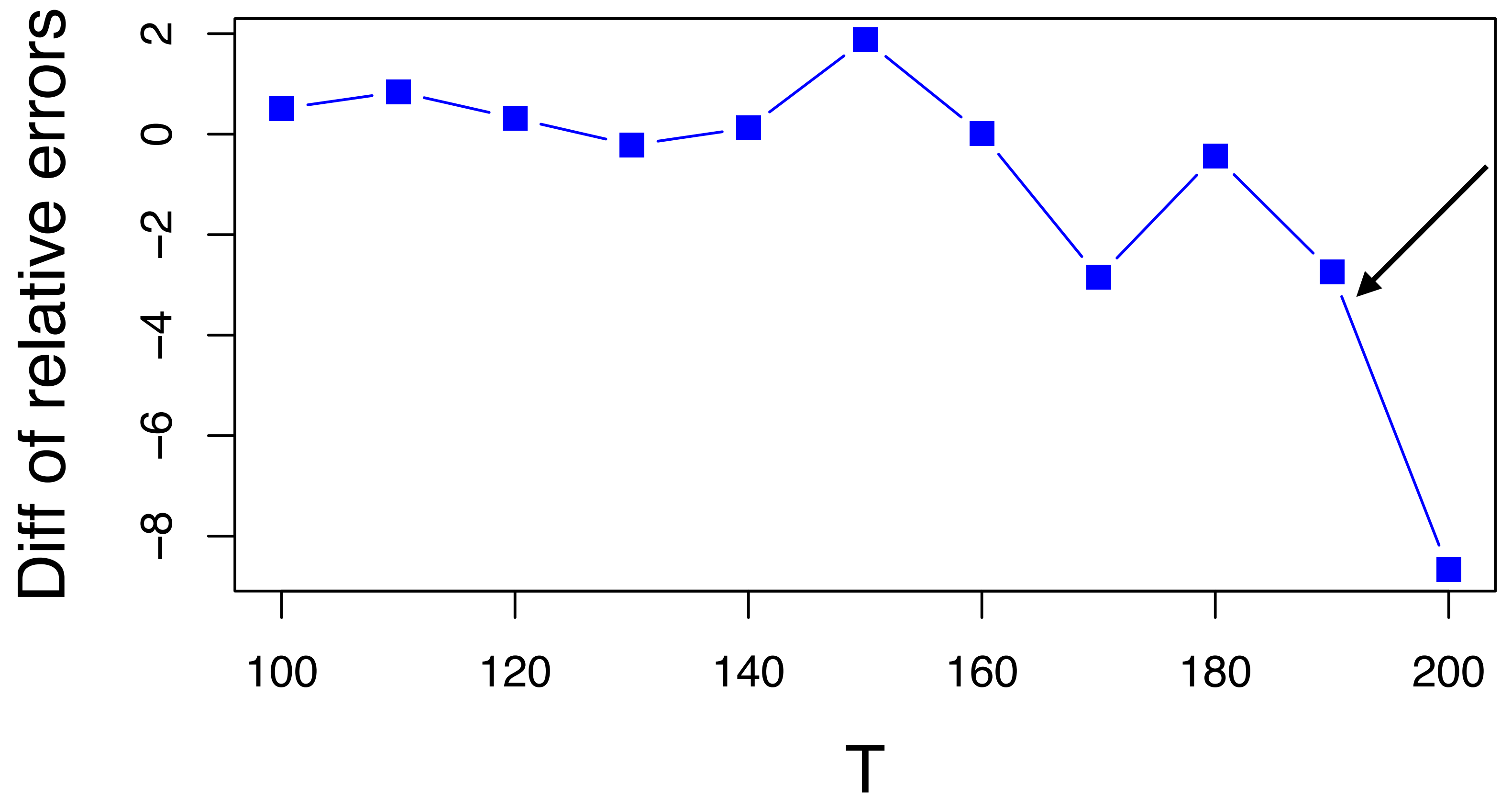
- $\hat{\theta}_{i,T:T+1} = \hat{\Lambda}_i^\top \hat{A} \hat{F}_{T-1},\ T \in$ slot 5 $\qquad \hat{A} = (\hat{F}^{(5)})^\top \hat{F}^{(4)} \left[(\hat{F}^{(4)})^\top \hat{F}^{(4)} \right]^{-1}$



Relative error for non-zero steps

$$\sum_{i:Y_{i,T}>0} \frac{(\hat{\theta}_{i,T-1:T} - Y_{i,T})^2}{Y_{i,T}^2}$$

FOCUS – mSSA



FOCUS predicts more accurately for larger T

Theoretical validation

Regularity conditions 🔐

Theoretical validation

Regularity conditions 🔐

- **Model assumptions**:
  - ☐ **Independent and identically distributed** loadings and outcome noise
  - ☐ Factors process is **Gaussian**

Theoretical validation

Regularity conditions 🔐

- **Model assumptions**:
  - ☐ Independent and identically distributed loadings and outcome noise
  - ☐ Factors process is Gaussian
- **Observation pattern**:
  - Positivity assumption on $W_{i,t}$ for each t and across i.
  - In probability limits for number of users observed across a given time pair , i.e. $\sum_{i=1}^{N} W_{i,t}W_{i,s}$
  - Other similar assumptions …

☑ Holds for several observation patters

— Missing completely at random (MCAR), $W_{i,t} \overset{\text{i.i.d}}{\sim}$ Bernoulli(p)

— Staggered adoption design (often appears in *synthetic controls*)

Theoretical validation

Forecast error bound + forecast intervals

Theoretical validation

Forecast error bound + forecast intervals

- **Forecast error bound**:
  Denote $\delta_{NT} := \min\{\sqrt{N}, \sqrt{T}\}$. Under the regularity conditions,

$$\left| \hat{\theta}_{i,T:T+h} - \theta_{i,T:T+h} \right| = \mathcal{O}_P(\delta_{NT}^{-1}) \ + \ \mathcal{O}_P(h\|A\|^{h-1}N^{-1}) \ + \ \mathcal{O}_P(h\|A\|^{h-1}T^{-1/2})$$

(Bai [2003], Bai and Ng. [2021], Xiong and Pelger [2021])

Theoretical validation

Forecast error bound + forecast intervals

(Bai [2003], Bai and Ng. [2021], Xiong and Pelger [2021])

- **Asymptotic normality**:

$$\delta_{NT}\left(\hat{\theta}_{i,T:T+h} - \theta_{i,T:T+h}\right) / \sigma_{i,T,h} \xrightarrow{d} \mathcal{N}(0,1)$$

$$\sigma^2_{i,T,h} = \sigma^{2,\text{est}}_{i,T,h} + \sigma^{2,\text{for}}_{i,T,h},$$

the variances involves cov(Λ_i), cov(F_t), mixed 4th moments of F_T, Λ_i ...

Theoretical validation

Forecast error bound + forecast intervals

(Bai [2003], Bai and Ng. [2021], Xiong and Pelger [2021])

- HAC estimator $\hat{\sigma}_{i,T,h}$ (Bai 2003) + $100(1-\alpha)\,\%$ confidence interval:

$$\left[\hat{\theta}_{i,T:T+h} \quad \mp \quad z_{1-\alpha/2}\hat{\sigma}_{i,T,h} \Big/ \delta_{NT}\right]$$

Takeaways

Contributions and future scopes

Takeaways

Contributions and future scopes

- Counterfactual forecasting method for panel data under low-rank structure + stochastic dynamic latent factors
- Empirical validation on simulated data
- Reliable forecasting in HeartSteps V1
- Under regularity, error bounds and CI on forecast estimator $\hat{\theta}_{i,T:T+h}$

Takeaways

Contributions and future scopes

- Counterfactual forecasting method for panel data under low-rank structure + stochastic dynamic latent factors
- Empirical validation on simulated data
- Reliable forecasting in HeartSteps V1
- Under regularity, error bounds and CI on forecast estimator $\hat{\theta}_{i,T:T+h}$

⊖ **Limitation:** stationarity of factors

Takeaways

Contributions and future scopes

- Counterfactual forecasting method for panel data under
  low-rank structure + stochastic dynamic latent factors
- Empirical validation on simulated data
- Reliable forecasting in HeartSteps V1
- Under regularity, error bounds and CI on forecast estimator $\hat{\theta}_{i,T:T+h}$

⊖ **Limitation**: stationarity of factors

☑ **Future goals**: more flexible, non-stationary frameworks
E.g. state space models, Markov switching model often appears in RL literature

Thank you!

Questions?

nd329@cornell.edu