Counterfactual inference in sequential experimental design

Raaz Dwivedi

Ongoing joint work with

Susan Murphy

Harvard John A. Paulson School of Engineering and Applied Sciences

Devavrat Shah

MIT EECS

Learning from interventions workshop, Simons Institute, Feb 14, 2022

Sequential decision making problems

- Online education: Enhance teaching strategies for better learning
- Online advertising: Update ads / placements to increase revenue
- **Mobile health**: Personalized app messages to promote healthy behavior

Physical activity

Wearable/trackers

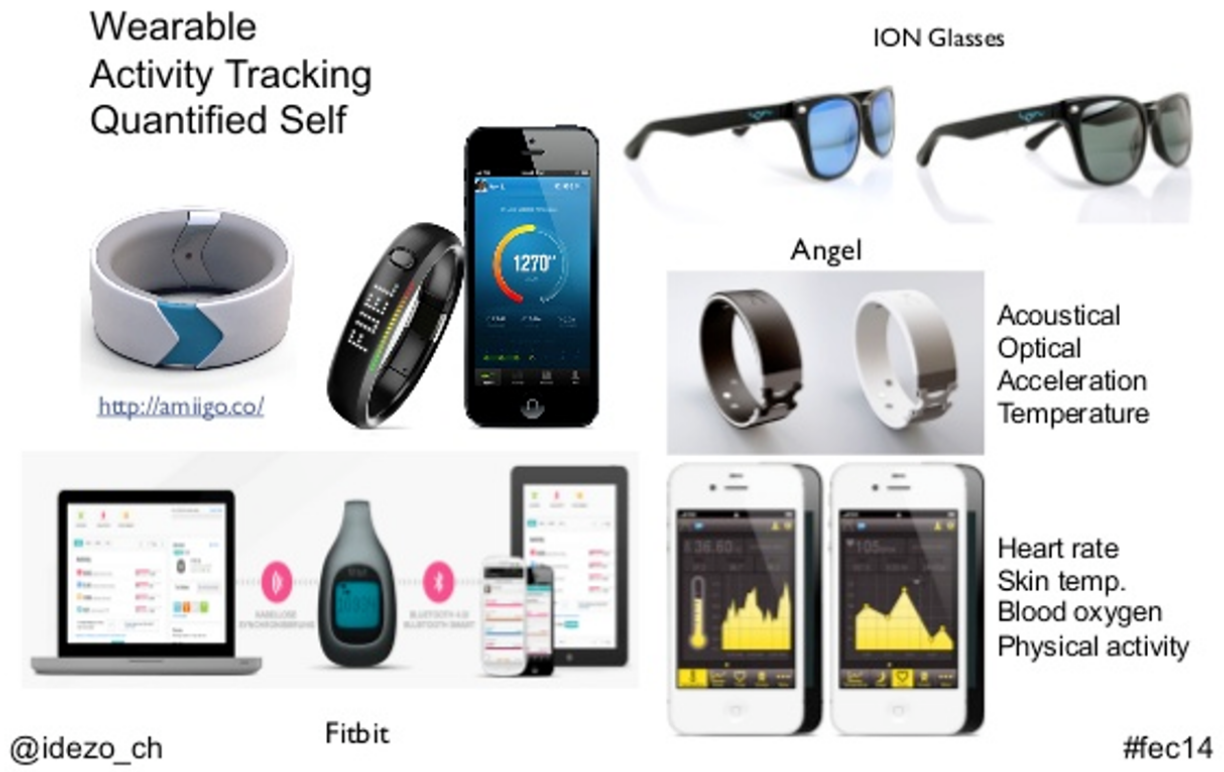


Image credits: Susan Murphy

Mobile health trial: A simplified but representative set-up

$Y_{i,t}^{(a)}$ = potential outcome of user i at time t under treatment a

Neyman-Rubin framework

Mobile health trial: A simplified but representative set-up

$Y_{i,t}^{(a)}$ = potential outcome of user i at time t under treatment a

For time $t = 1,2,\ldots,T$

For user $i = 1,\ldots,N$

1. Randomly send a notification $(A_{i,t} = 1)$ or not $(A_{i,t} = 0)$ using a sampling policy $\pi_t \in [0,1]^N$ with $\pi_t(i) = \mathbb{P}(A_{i,t} = 1 \mid \text{history}_{t-1})$

Mobile health trial: A simplified but representative set-up

$Y_{i,t}^{(a)}$ = potential outcome of user i at time t under treatment a

For time $t = 1,2,\ldots,T$

For user $i = 1,\ldots,N$

1. Randomly send a notification $(A_{i,t} = 1)$ or not $(A_{i,t} = 0)$ using a sampling policy $\pi_t \in [0,1]^N$ with $\pi_t(i) = \mathbb{P}(A_{i,t} = 1 \mid \text{history}_{t-1})$
2. Observe outcome $Z_{i,t} = Y_{i,t}^{(A_{i,t})} + \varepsilon_{i,t} \longleftarrow$ noise

Mobile health trial: A simplified but representative set-up

$Y_{i,t}^{(a)}$ = potential outcome of user i at time t under treatment a

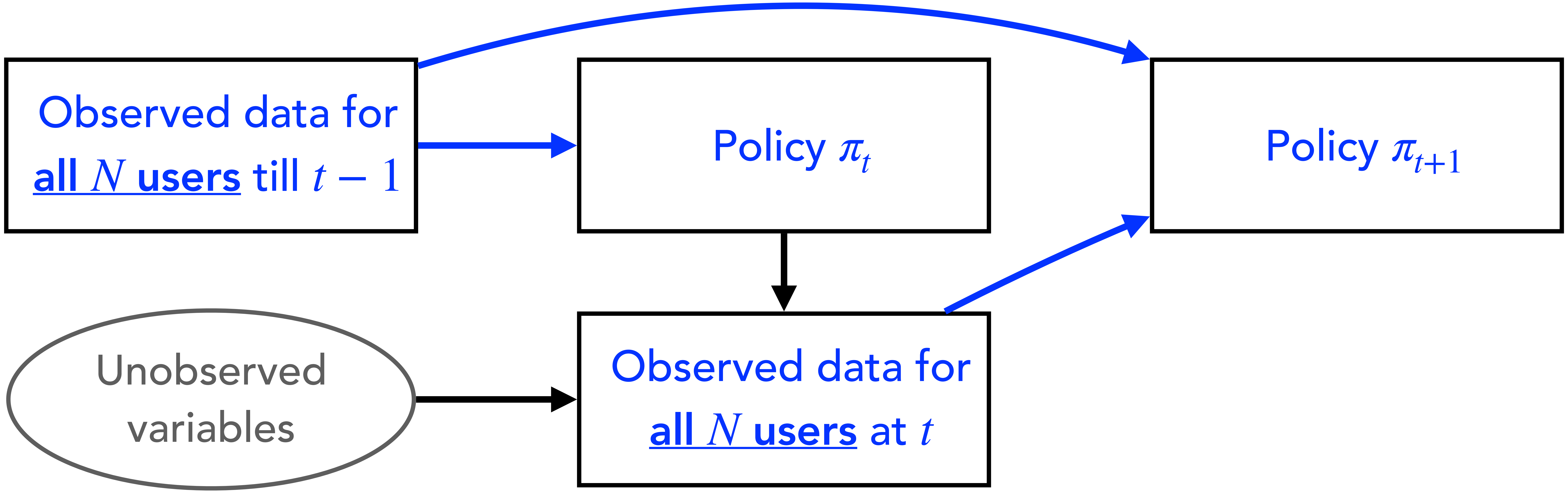
For time $t = 1,2,\ldots,T$

For user $i = 1,\ldots,N$

1. Randomly send a notification $(A_{i,t} = 1)$ or not $(A_{i,t} = 0)$ using a sampling policy $\pi_t \in [0,1]^N$ with $\pi_t(i) = \mathbb{P}(A_{i,t} = 1 \mid \text{history}_{t-1})$
2. Observe outcome $Z_{i,t} = Y_{i,t}^{(A_{i,t})} + \varepsilon_{i,t}$ ⟵ noise

Update policy to π_{t+1} using data till time t

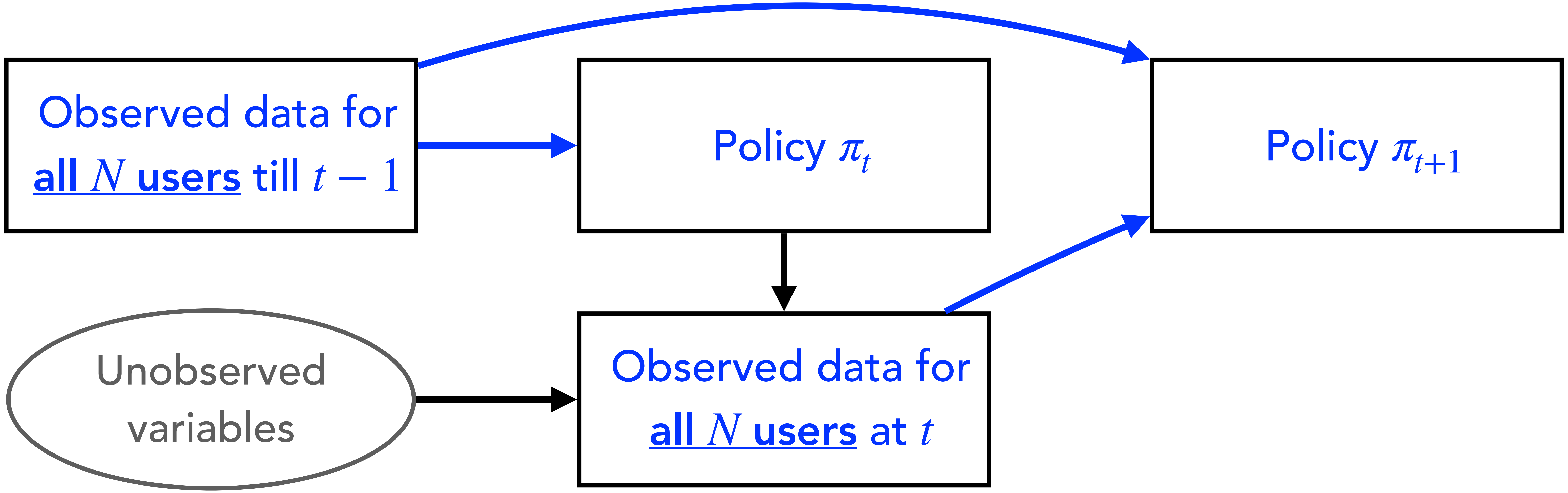
Sequential experimental design



- **Adaptive** policy to personalize to users
- Data **pooled during study** learn a good policy quickly

[…, Yom-Tov 2017, …, Tomkins et al. 2020]

Sequential experimental design: After-study inference



- **Adaptive** policy to personalize to users
- Data **pooled during study** learn a good policy quickly
- After-study questions: **Was the treatment effective? On average? Heterogeneity across users? …**

[…, Yom-Tov 2017, …, Tomkins et al. 2020]

Our goal: Counterfactual inference, namely, estimation of all missing potential outcomes—a hard task due to **heterogeneity across users and time**

- can be used for generic after-study analyses, e.g., individual treatment effect $Y_{i,t}^{(1)} - Y_{i,t}^{(0)}$

Our goal: Counterfactual inference, namely, estimation of all missing potential outcomes—a hard task due to **heterogeneity across users and time**

Adaptivity and pooling of data across N users for policy π_t makes after-study inference even more challenging

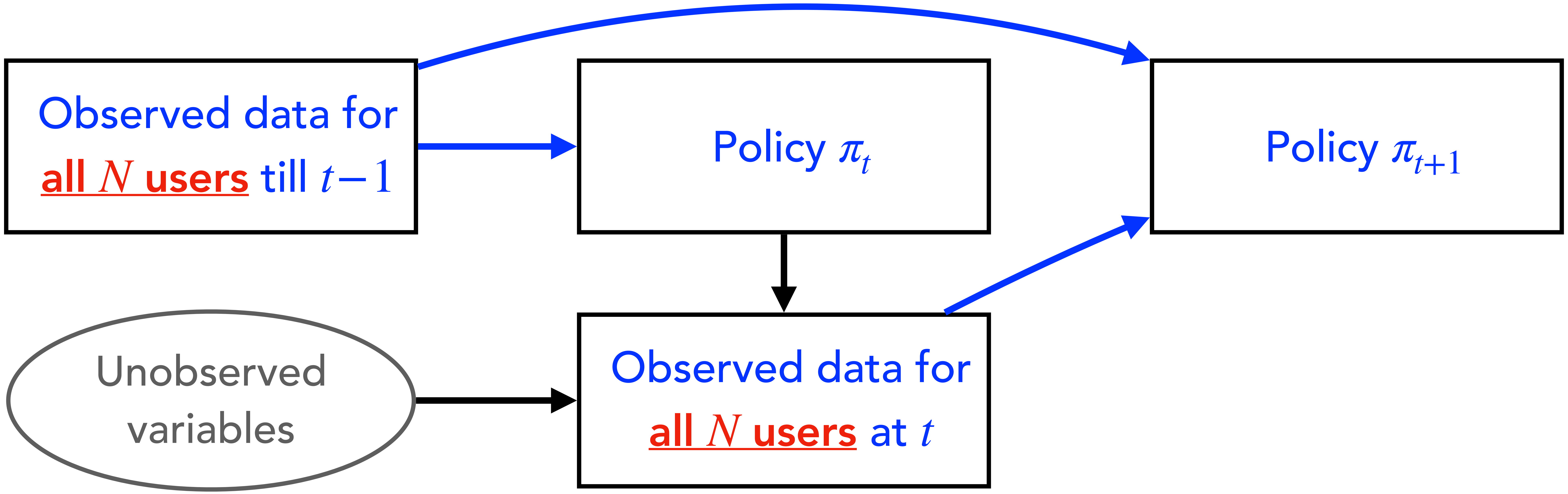
- can be used for generic after-study analyses, e.g., individual treatment effect $Y_{i,t}^{(1)} - Y_{i,t}^{(0)}$

Coming up next

- Settings considered in prior work
- Our contributions

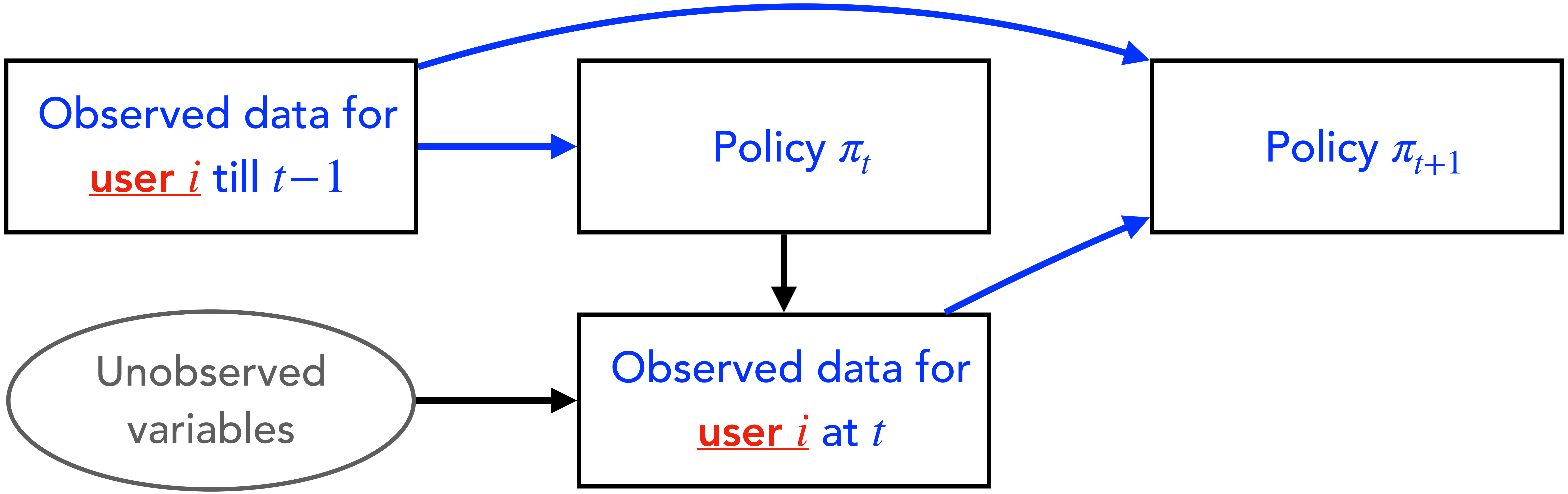
    1. A non-parametric strategy for counterfactual estimation
    2. Distribution-free consistency under bilinear latent factor model
    3. Central limit theorem under a non-linear latent factor model

After-study inference:
Our setting of sequential experiments



After-study inference: Dynamic treatment regime in longitudinal studies

Policy adaptive for *a single fixed user over time* based on prior data, e.g., personalized clinical treatment […, Robins 1986, 1987,…,Murphy 2003 ,…]

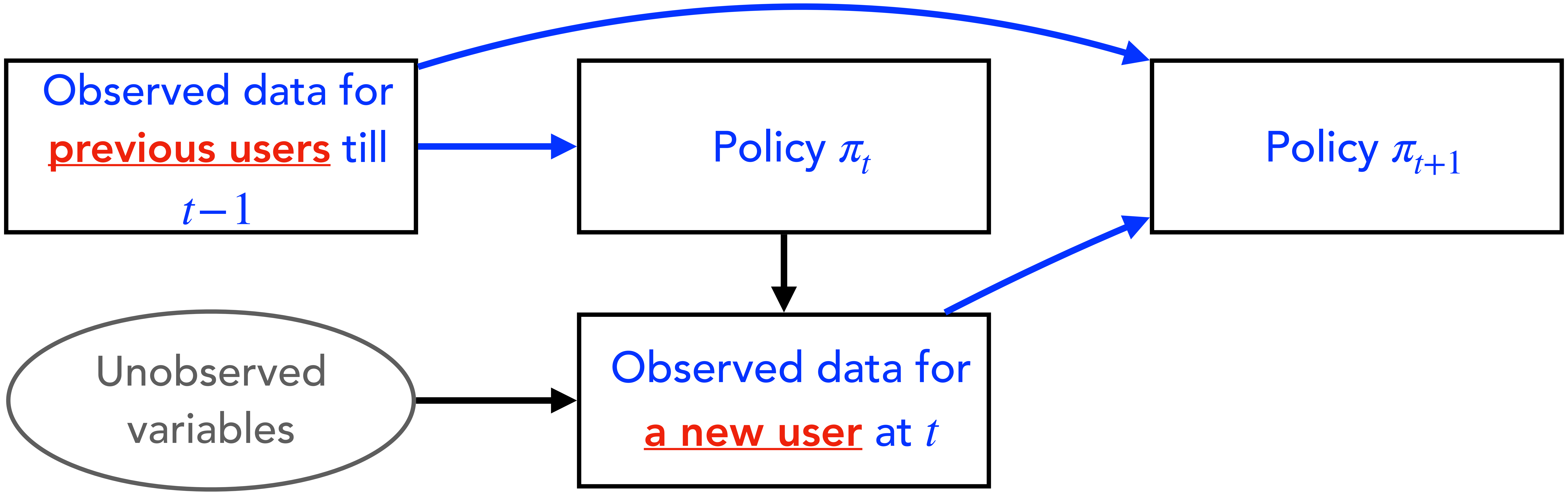


Experiments independent across users

After-study inference:
Policy evaluation for bandits

Adaptive policy with i.i.d. users at each time, e.g., online ads

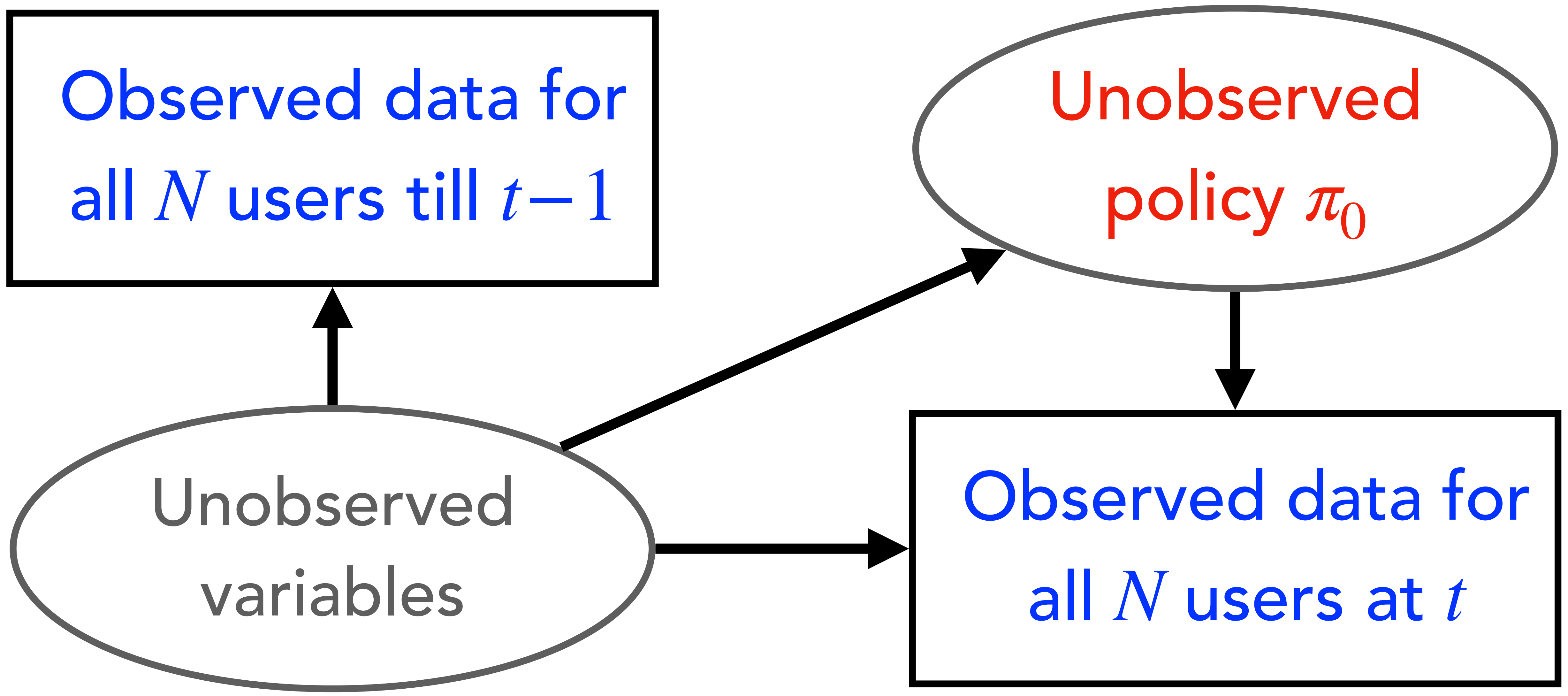
[…, Zhang et al 2021, Hadad et al. 2021, Bibaut et al. 2021,…]



After-study inference:
Causal panel data

Observational settings with multiple users and multiple time, treatment assigned only once per user (no real use of policy)

[… Abadie et al. 2003, 2005, Chernozhukov et al. 2017, Athey et al. 2018, Agarwal et a. 2021 …]



So what do we do?

So what do we do?
Reduce counterfactual inference to sequential matrix completion

- Fix treatment say 1, with $Y^{\star}_{i,t} \triangleq Y^{(1)}_{i,t}$

- $N \times T$ matrix of potential outcomes with missing at random entries

$$Z_{i,t} = \begin{cases} Y^{\star}_{i,t} + \varepsilon_{i,t} & \text{if} \quad A_{i,t} = 1 \\ \text{unknown} & \text{if} \quad A_{i,t} = 0 \end{cases} \quad \text{where } A_{i,t} = \text{Bernoulli}(\pi_t(i))$$

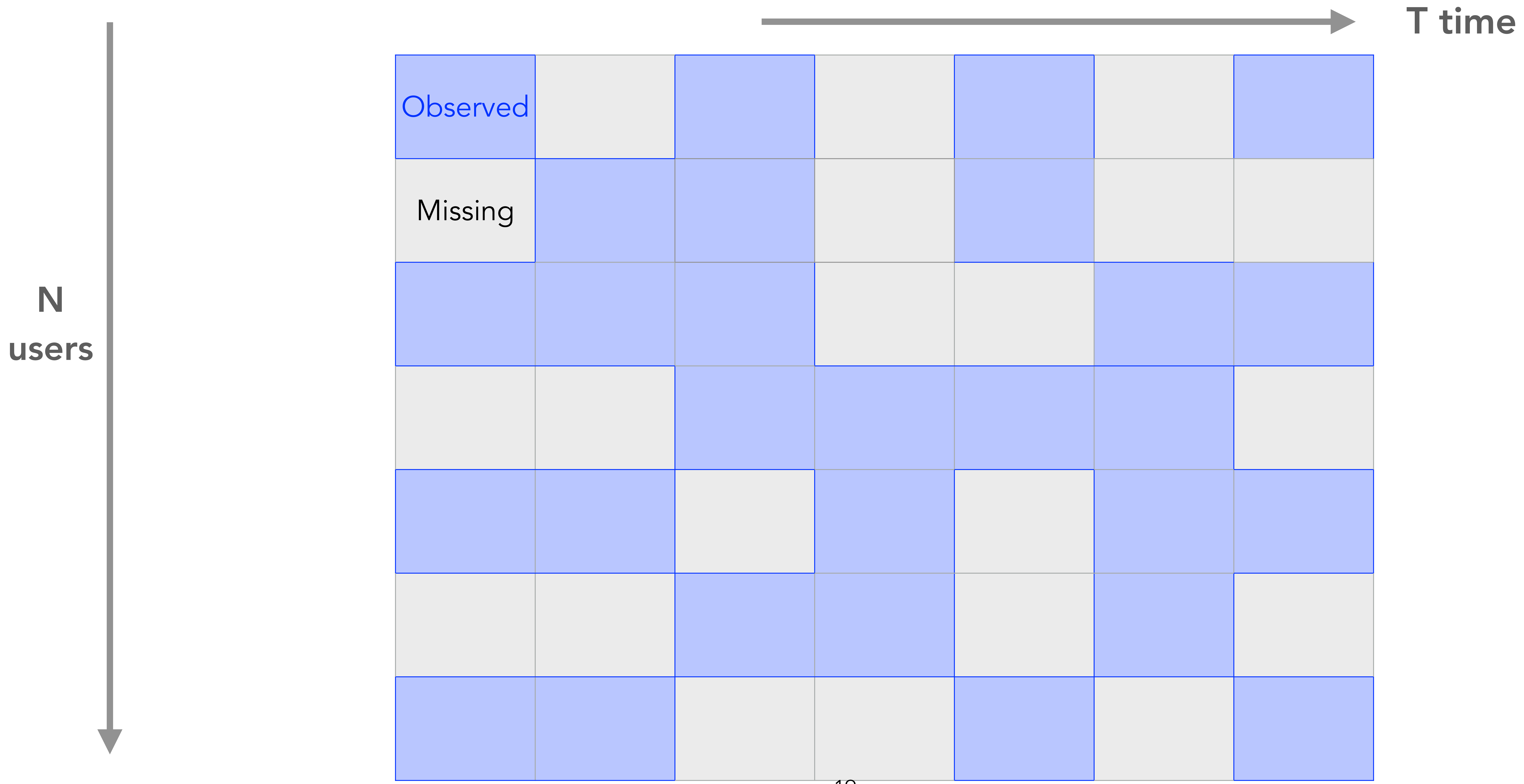
- **New goal**: Estimate missing entries $Y^{\star}_{i,t}$ (separately for each treatment)

Rest of the talk:

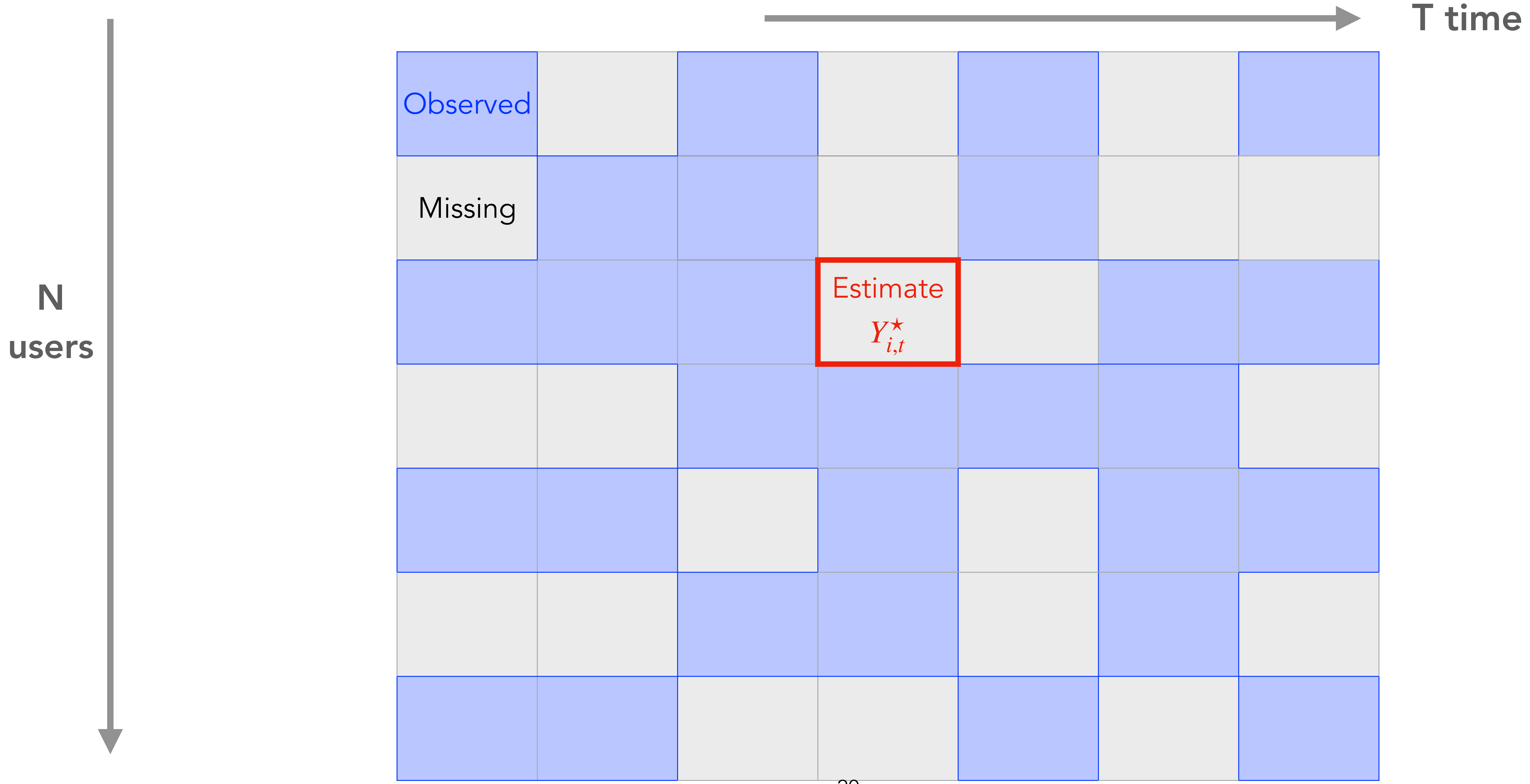
How to estimate $Y^{\star}_{i,t}$?

When will the estimate work?

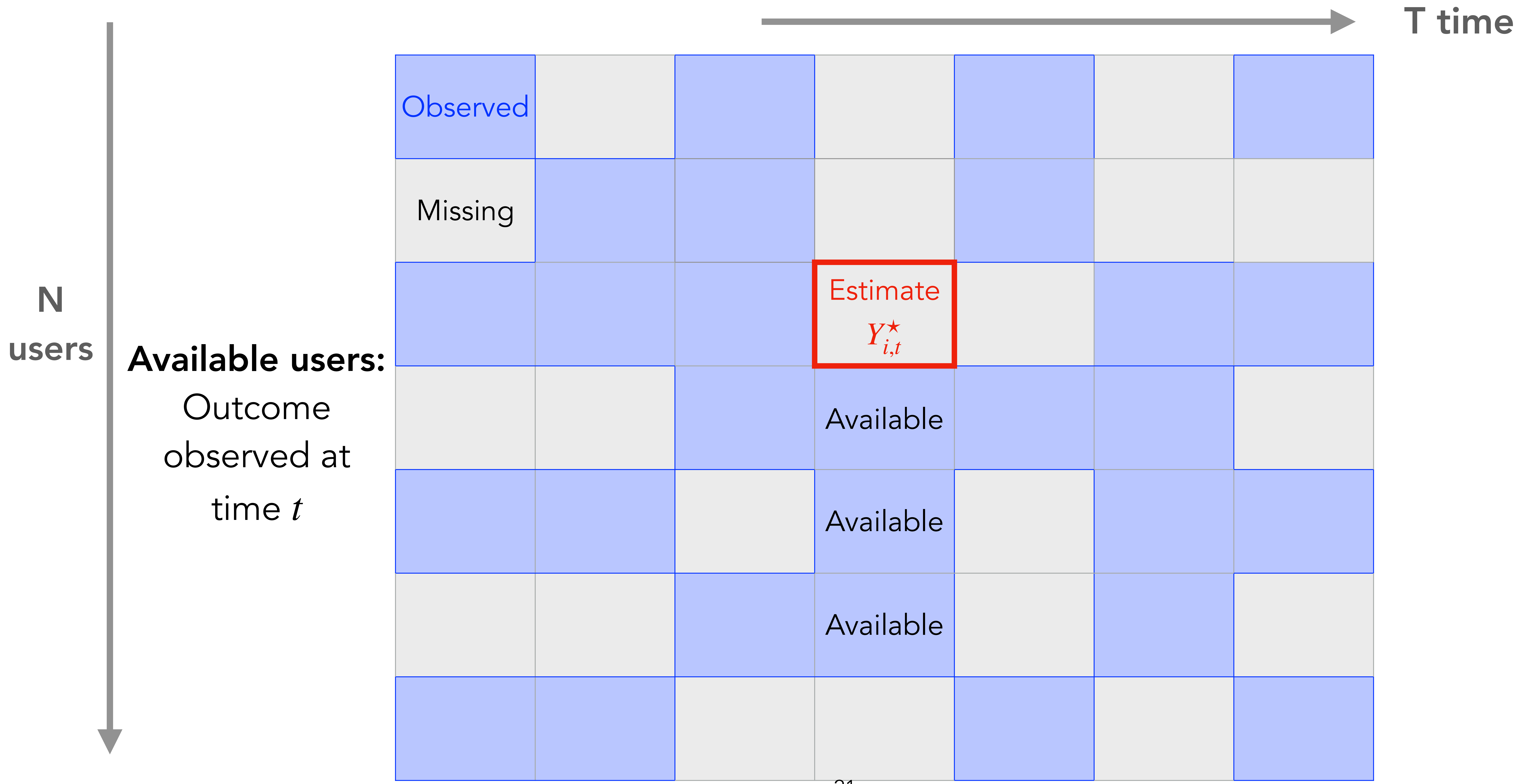
Estimating $Y^{\star}_{i,t}$ via a variant of nearest neighbors



Estimating $Y^{\star}_{i,t}$ via a variant of nearest neighbors

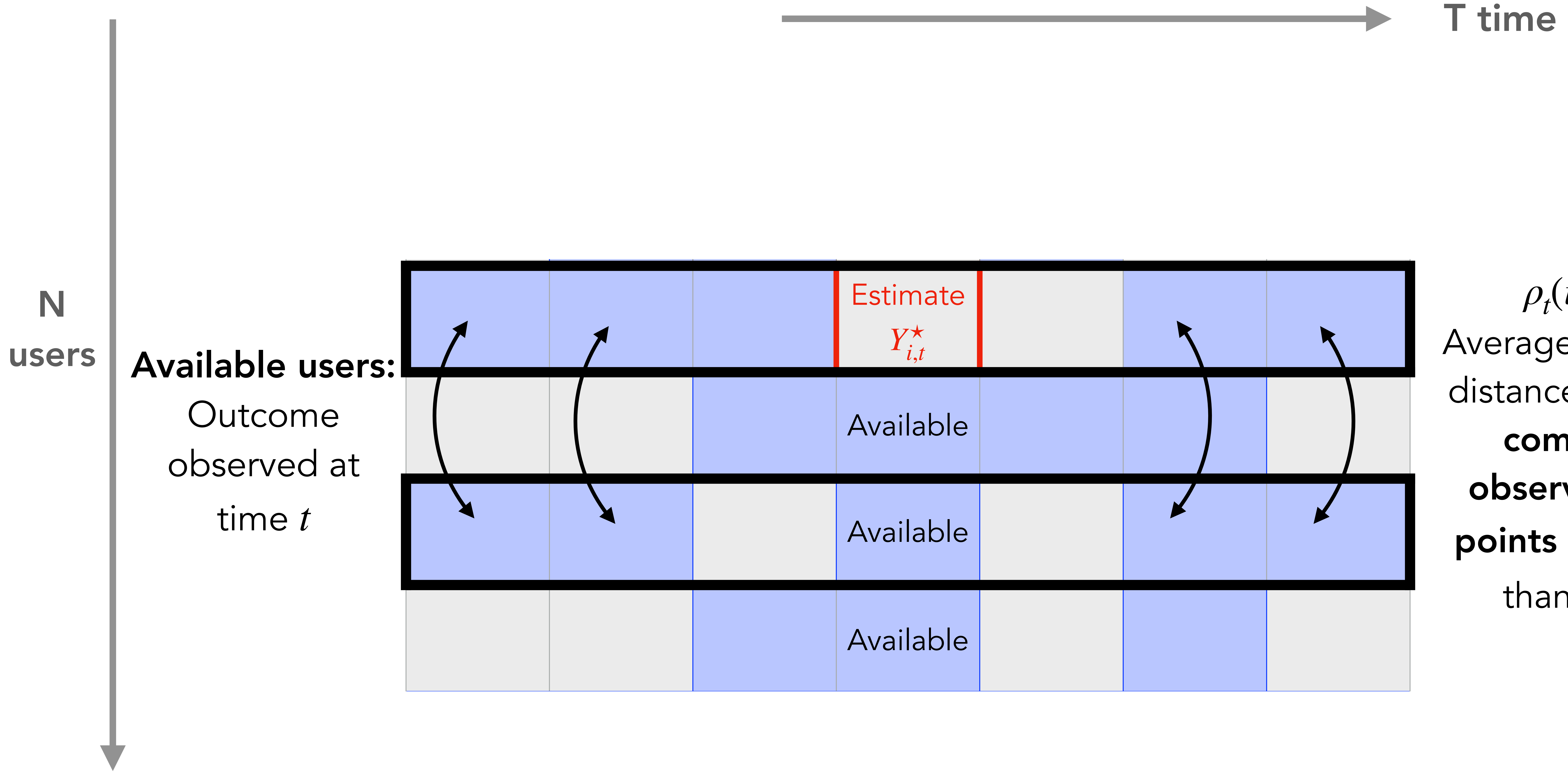


Estimating $Y^{\star}_{i,t}$ via a variant of nearest neighbors



Available users:
Outcome
observed at
time t

Estimating $Y^{\star}_{i,t}$ via a variant of nearest neighbors

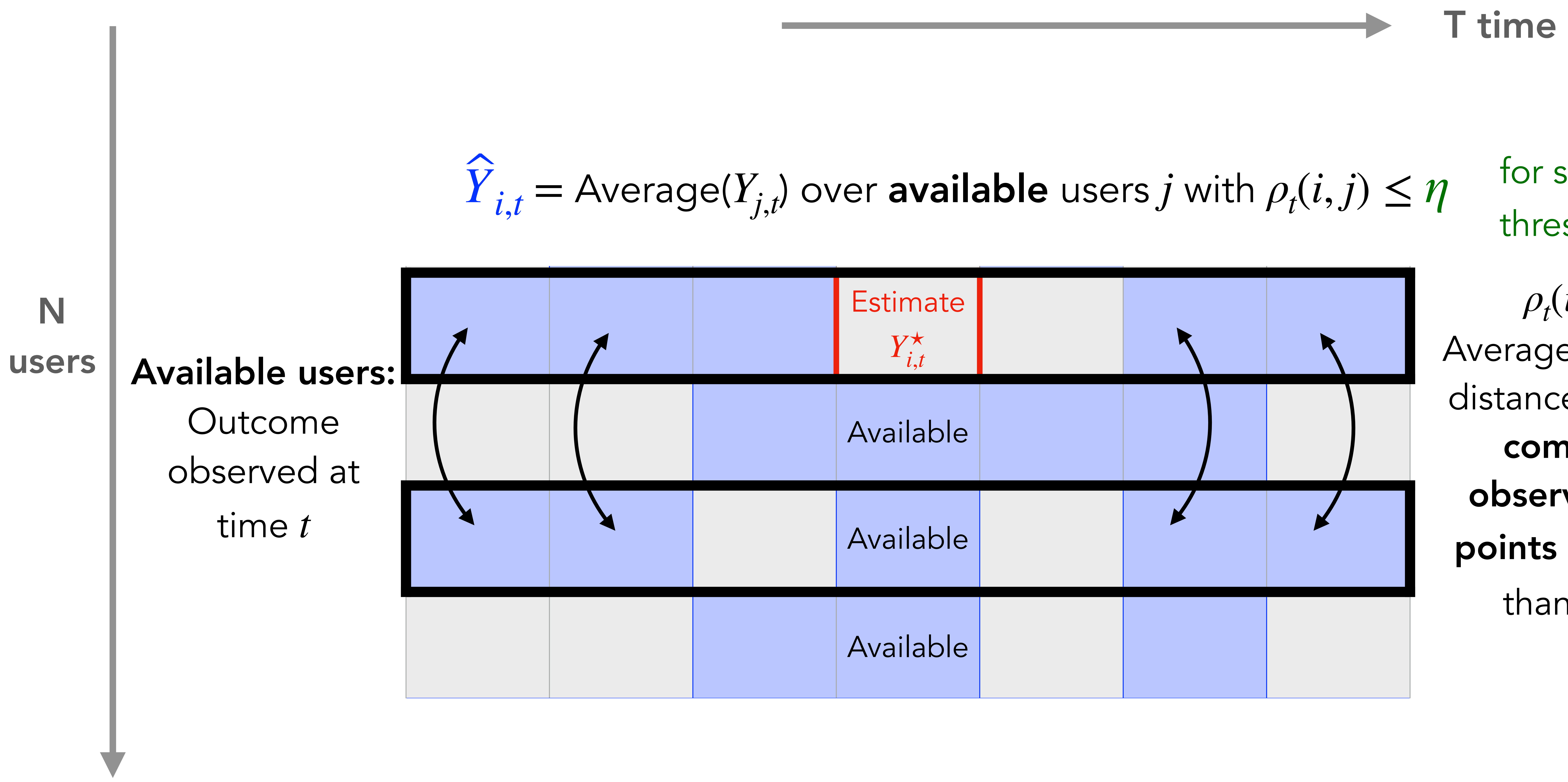


$\rho_t(i, j)$ =
Averaged squared distance between **commonly observed time points** $T_{t,i,j}$ other than time t

Estimating $Y^{\star}_{i,t}$ via a variant of nearest neighbors

$\widehat{Y}_{i,t}$ = Average($Y_{j,t}$) over **available** users j with $\rho_t(i,j) \leq \eta$

for suitable threshold η



$\rho_t(i,j)$ = Averaged squared distance between **commonly observed time points** $T_{t,i,j}$ other than time t

Non-parametric estimate: Agnostic to policy and model!

T time

N users

$\widehat{Y}_{i,t}$ = Average($Y_{j,t}$) over **available** users j with $\rho_t(i,j) \leq \eta$

for suitable threshold η

Available users: Outcome observed at time t

$\rho_t(i,j)$ = Averaged squared distance between **commonly observed time points** $T_{t,i,j}$ other than time t

Will such a non-parametric estimate work for sequential experimental design?

Yes, under a latent factor model!

and suitable conditions

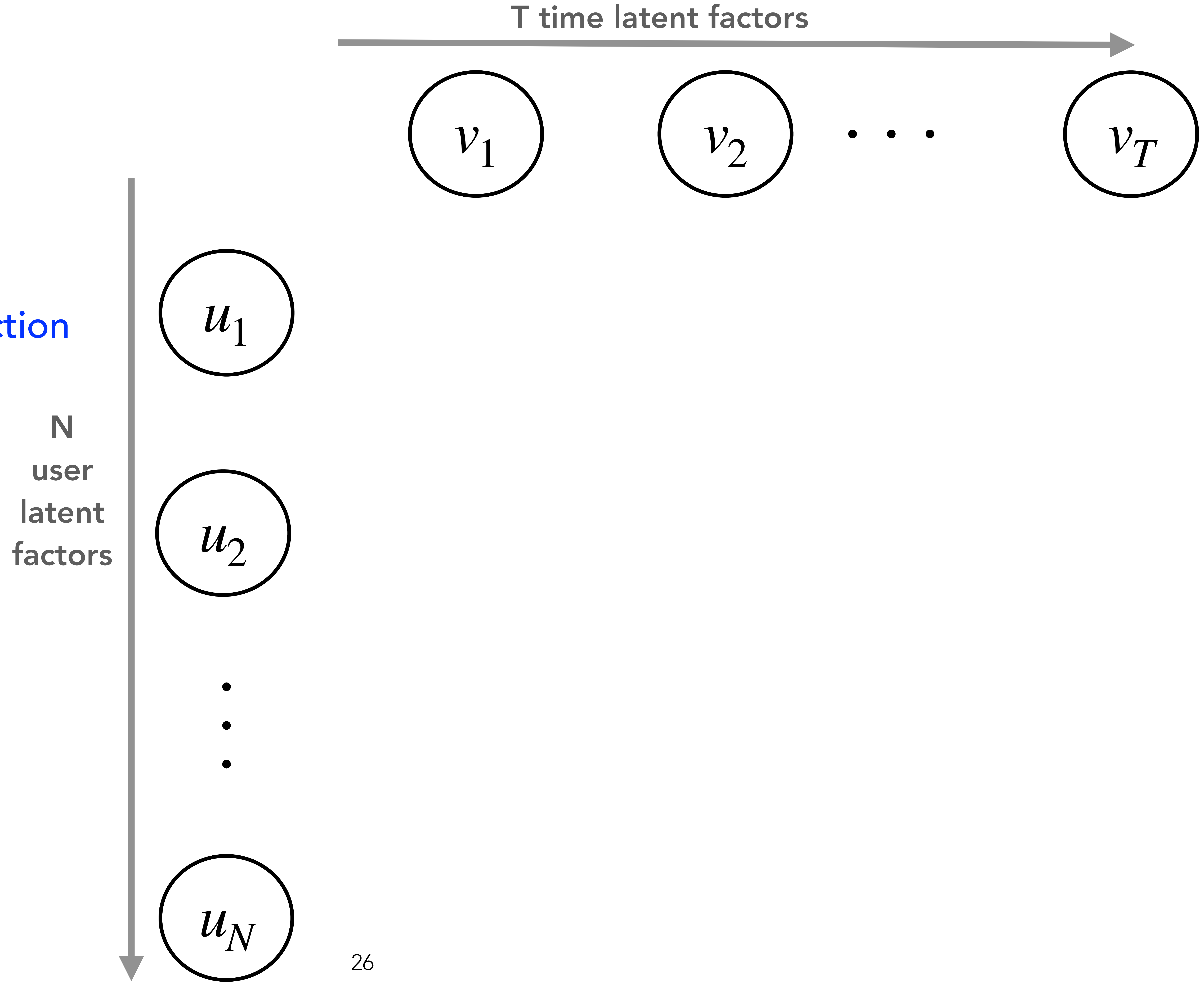
Latent factor model

$$Y^{\star}_{i,t} \triangleq Y^{(1)}_{i,t} \triangleq f(u_i, v_t)$$

u_i : latent factor for user i

v_t : latent factor for time t

f : unknown (non) linear function



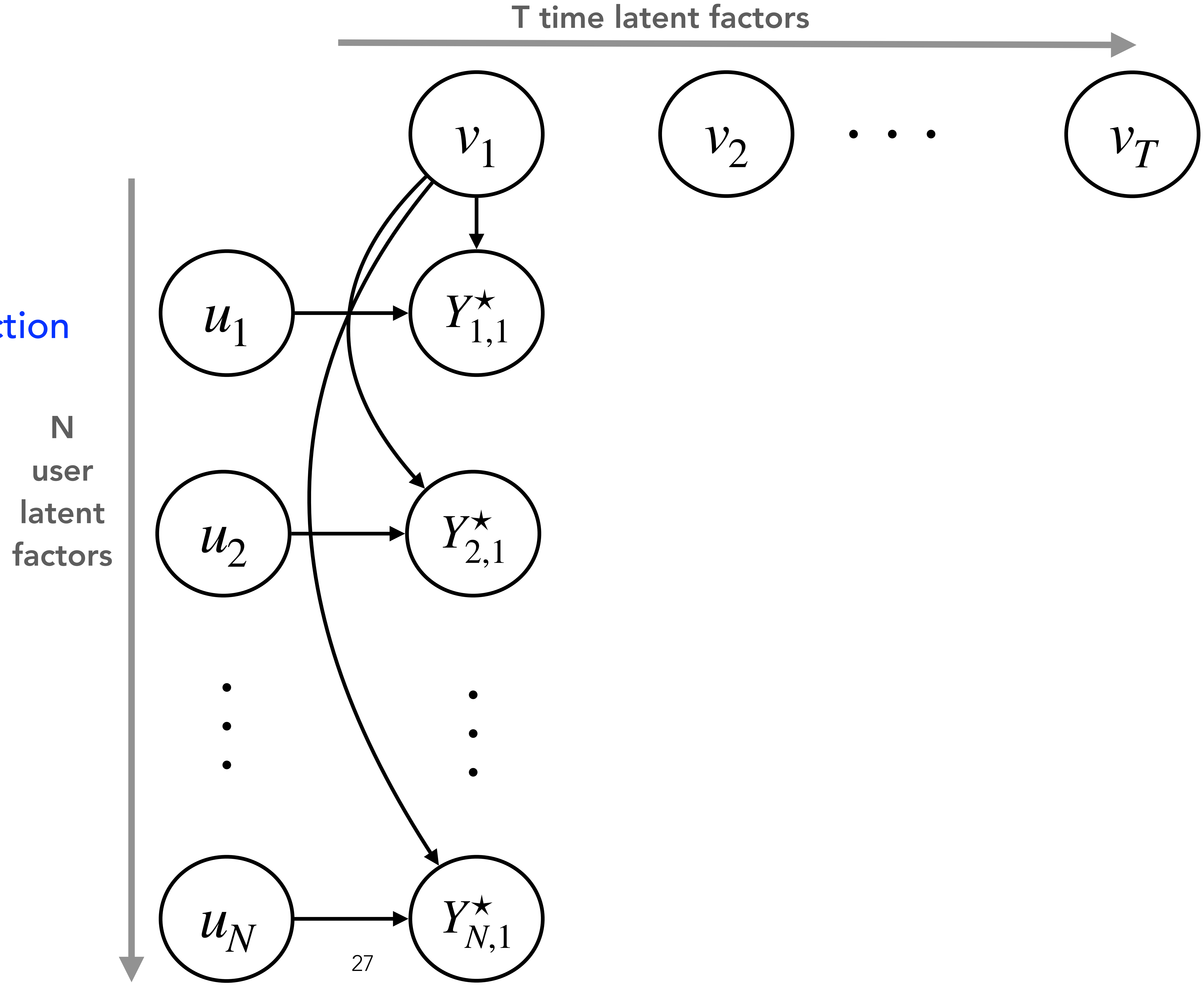
Latent factor model

$$Y^{\star}_{i,t} \triangleq Y^{(1)}_{i,t} \triangleq f(u_i, v_t)$$

u_i : latent factor for user i

v_t : latent factor for time t

f : unknown (non) linear function



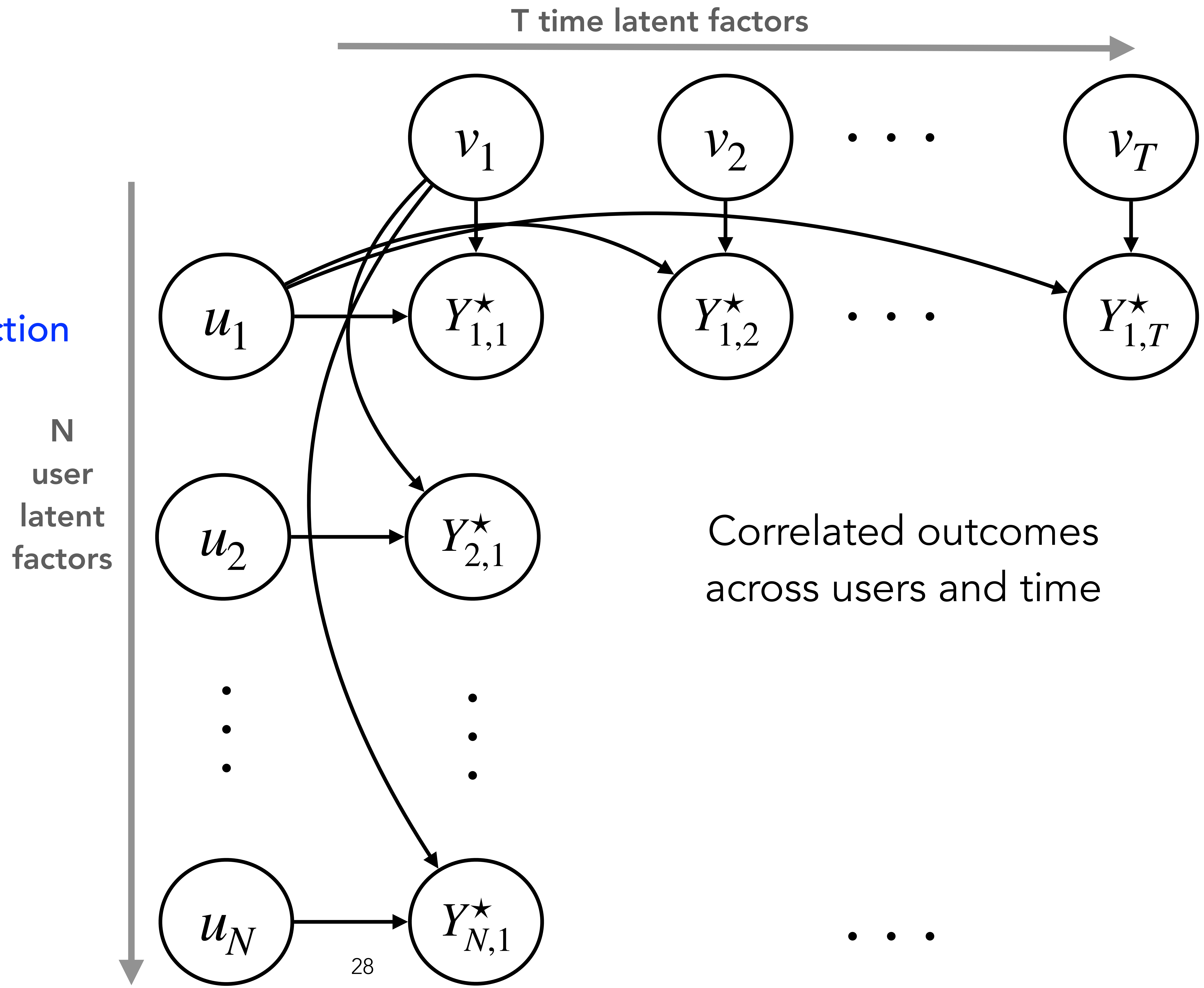
Latent factor model

$$Y^{\star}_{i,t} \triangleq Y^{(1)}_{i,t} \triangleq f(u_i, v_t)$$

u_i : latent factor for user i

v_t : latent factor for time t

f : unknown (non) linear function



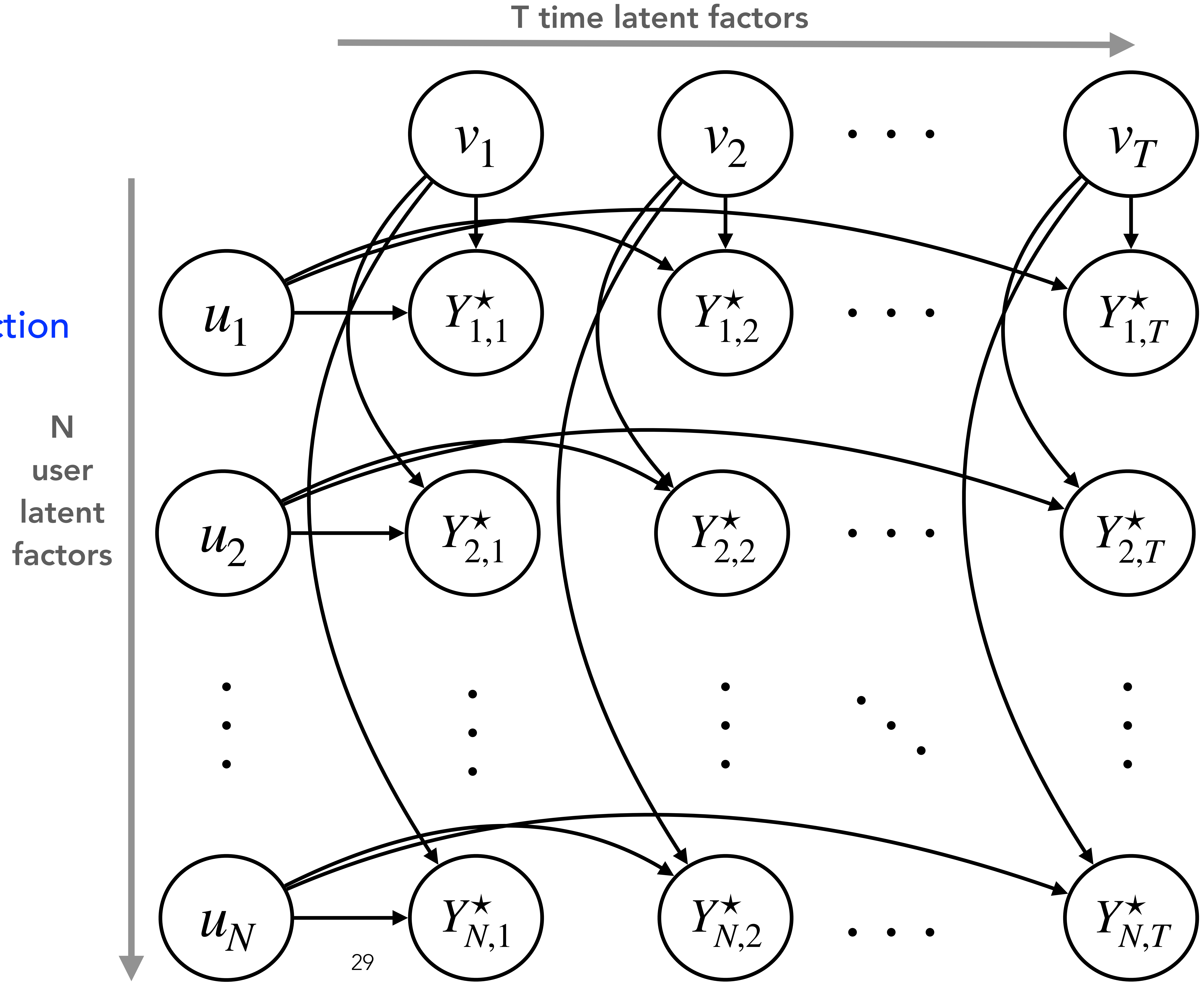
Latent factor model

$$Y^{\star}_{i,t} \triangleq Y^{(1)}_{i,t} \triangleq f(u_i, v_t)$$

u_i : latent factor for user i

v_t : latent factor for time t

f : unknown (non) linear function



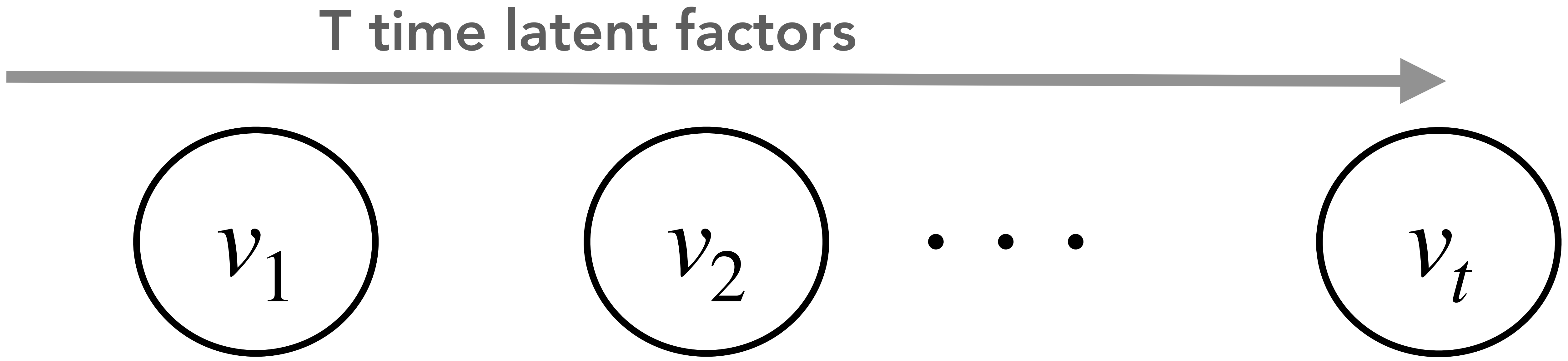
Latent factor model

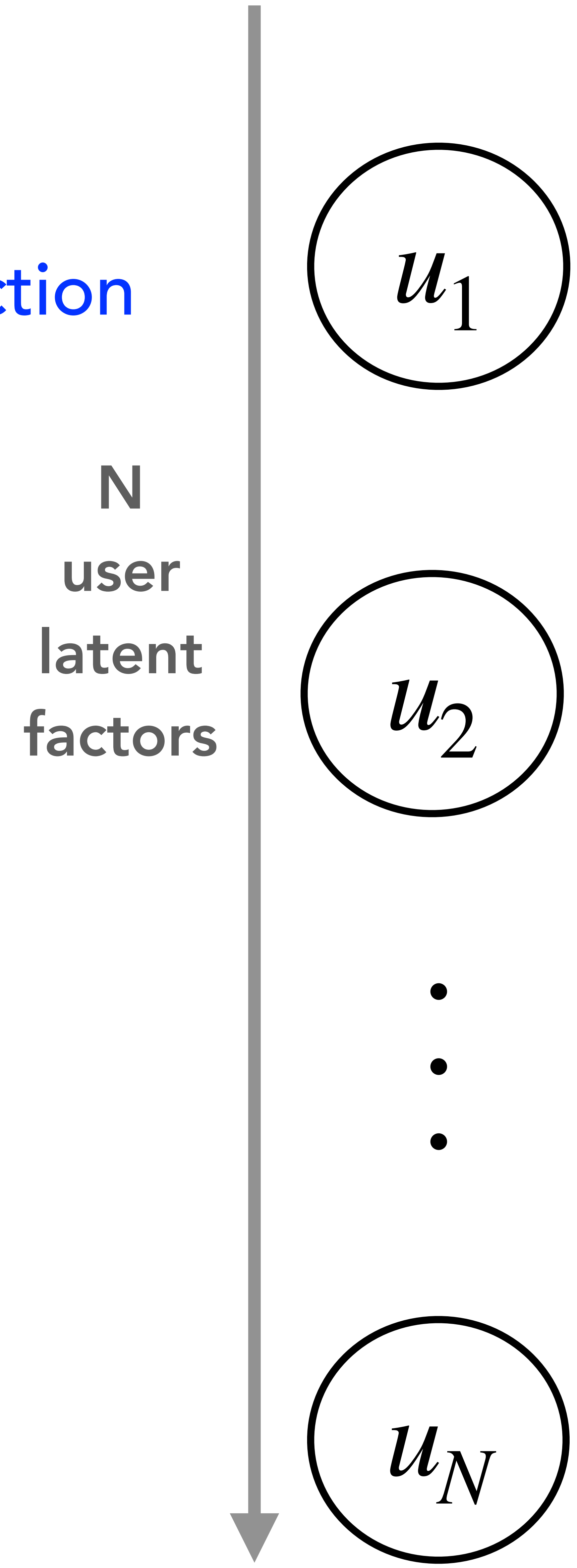
$$Y^{\star}_{i,t} \triangleq Y^{(1)}_{i,t} \triangleq f(u_i, v_t)$$

u_i : latent factor for user i

v_t : latent factor for time t

f : unknown (non) linear function





- Examples include:
  - $Y^{\star}_{i,\cdot} \sim$ Gaussian process with covariance kernel $\mathbf{k}$

    $u_i =$ Gaussian vector

    $v_t =$ Eigenfunctions of $\mathbf{k}$

    $f(u, v) = \langle u, v \rangle$
  - Also a sub-class of exchangeable data

Theoretical guarantees

$f(u, v) = \langle u, v \rangle$

- Consider any (i, t) with enough nearest neighbors j satisfying the conditions
  - ``**diverse latent-time factors**'': $\frac{1}{T_{t,i,j}} \sum_{t' \in T_{t,i,j}} v_{t'} v_{t'}^\top \succeq \lambda I_d$ for $\lambda > 0$

    $T_{t,i,j}$ = commonly observed time points other than t; used to compute distance $\rho_t(i, j)$

$f(u, v) = \langle u, v \rangle$

- Consider any (i, t) with enough nearest neighbors j satisfying the conditions
  - ``**diverse latent-time factors''**: $\frac{1}{T_{t,i,j}} \sum_{t' \in T_{t,i,j}} v_{t'} v_{t'}^\top \succeq \lambda I_d$ for $\lambda > 0$

    $T_{t,i,j} =$ commonly observed time points other than t; used to compute distance $\rho_t(i, j)$
  - ``**non-adversarial noise''** across $T_{t,i,j} \cup \{t\}$: behave roughly like iid/mixing process

$f(u,v) = \langle u, v\rangle$: A distribution-free consistency for any (i,t)

- Consider any (i,t) with enough nearest neighbors j satisfying the conditions
  - **``diverse latent-time factors''**: $\frac{1}{T_{t,i,j}} \sum_{t' \in T_{t,i,j}} v_{t'} v_{t'}^\top \succeq \lambda I_d$ for $\lambda > 0$

    $T_{t,i,j} =$ commonly observed time points other than t; used to compute distance $\rho_t(i,j)$
  - **``non-adversarial noise''** across $T_{t,i,j} \cup \{t\}$: behave roughly like iid/mixing process
- For suitable scaling of threshold η & mild conditions on arbitrarily dependent policy, given user i with latent factor u_i, we have

$$\widehat{Y}_{i,t} \to Y^\star_{i,t} \text{ for any } t \text{ as } N, T \to \infty$$

$f(u, v) = \langle u, v \rangle$: Advantage of bilinearity

Key intuition:

If latent-time factors $\{v_t\}$ ``diverse enough'', and noise ``non-adversarial''

$$\rho_t(i, j) \leq \eta \quad \Longrightarrow \quad \|u_i - u_j\|_2 \lesssim \sqrt{\frac{\eta}{\lambda}} \quad \Longrightarrow \quad |Y^\star_{i,t} - Y^\star_{j,t}| \lesssim \|v_t\|_2 \cdot \sqrt{\frac{\eta}{\lambda}}$$

Lipschitz f

Consider a **non-linear Lipschitz** f, and suppose

- $u_j \sim_{iid} \mathbb{P}_{user}$, $v_t \sim_{iid} \mathbb{P}_{time}$,
- $\varepsilon_{j,t} \sim_{iid} \mathbb{P}_{noise}$, $\mathbb{E}[\varepsilon_{j,t}] = 0$, $\mathbb{E}[\varepsilon_{j,t}^2] = \sigma^2$
- π_t depends on *all users'* history till $t - 1$; treatments$\{A_{j,t}\}$ assigned independently given the history

Lipschitz f: Central limit theorem for sequential estimation of $Y^{\star}_{i,T}$

Consider a **non-linear Lipschitz** f, and suppose

- $u_j \sim_{iid} \mathbb{P}_{user}$, $v_t \sim_{iid} \mathbb{P}_{time}$,
- $\varepsilon_{j,t} \sim_{iid} \mathbb{P}_{noise}$, $\mathbb{E}[\varepsilon_{j,t}] = 0$, $\mathbb{E}[\varepsilon^2_{j,t}] = \sigma^2$
- π_t depends on *all users'* history till $t-1$; treatments$\{A_{j,t}\}$ assigned independently given the history

Under regularity conditions, given *any user i at last time T* *with* *number of neighbors $N_{i,T}$*

$$\sqrt{N_{i,T}}(\widehat{Y}_{i,T} - Y^{\star}_{i,T}) \Longrightarrow \mathcal{N}(0,\sigma^2) \quad \text{as } N, T \to \infty \text{ together at suitable rates}$$

Lipschitz f: Non-asymptotic expected squared error bound

$$\mathbb{E}\left[(\widehat{Y}_{i,T} - Y^{\star}_{i,T})^2 \,|\, u_i\right] \lesssim (\eta - 2\sigma^2) + \frac{D^2(1+\gamma_{i,T})}{p^2_{\min,T}\sqrt{T-1}} + \frac{\sigma^2}{p_{\min,T}\,\Phi_i N}$$

Bias due to η

Concentration of neighbor distance

Effective noise variance

Lipschitz f: Non-asymptotic expected squared error bound

$$\mathbb{E}\left[(\widehat{Y}_{i,T} - Y^{\star}_{i,T})^2 \,|\, u_i\right] \lesssim (\eta - 2\sigma^2) + \frac{D^2(1+\gamma_{i,T})}{p^2_{\min,T}\sqrt{T-1}} + \frac{\sigma^2}{p_{\min,T}\Phi_i N}$$

Bias due to η

Concentration of neighbor distance

Effective noise variance

$$p_{\min,T} = \min_{t,j} \pi_t(j)$$

min probability of sampling any entry

Lipschitz f: Non-asymptotic expected squared error bound

$$\mathbb{E}\left[(\widehat{Y}_{i,T} - Y^{\star}_{i,T})^2 \mid u_i\right] \lesssim (\eta - 2\sigma^2) + \frac{D^2(1+\gamma_{i,T})}{p^2_{\min,T}\sqrt{T-1}} + \frac{\sigma^2}{p_{\min,T}\,\Phi_i N}$$

Bias due to η

Concentration of neighbor distance

Effective noise variance

$$p_{\min,T} = \min_{t,j} \pi_t(j)$$ min probability of sampling any entry

$$\Phi_i = \mathbb{P}_u\left(\mathbb{E}_v[f(u_i, v) - f(u,v)^2] \leq \eta/2 - \sigma^2\right)$$ Probability of sampling a nearest neighbor

Lipschitz f: Non-asymptotic expected squared error bound

$$\mathbb{E}\left[(\widehat{Y}_{i,T} - Y^{\star}_{i,T})^2 \,|\, u_i\right] \lesssim (\eta - 2\sigma^2) + \frac{D^2(1+\gamma_{i,T})}{p^2_{\min,T}\sqrt{T-1}} + \frac{\sigma^2}{p_{\min,T}\,\Phi_i N}$$

Bias due to η

Concentration of neighbor distance

Effective noise variance

$$p_{\min,T} = \min_{t,j} \pi_t(j)$$ min probability of sampling any entry

$$\Phi_i = \mathbb{P}_u\left(\mathbb{E}_v[f(u_i, v) - f(u,v))^2] \le \eta/2 - \sigma^2\right)$$ Probability of sampling a nearest neighbor

$$\gamma_{i,T} = \sup_{j\neq i, t<T} \left| \mathbb{E}[\,\Sigma_{t'=t+1}^{T-1} A_{i,t'}A_{j,t'} \,|\, \text{history}_t\,] - \mathbb{E}[\,\Sigma_{t'=t+1}^{T-1} A_{i,t'}A_{j,t'} \,|\, \text{history}_{t-1}\,] \right|$$

Cumulative future dependency of adaptive policies on one column

Lipschitz f: Non-asymptotic expected squared error bound

$$\mathbb{E}\left[(\widehat{Y}_{i,T} - Y^{\star}_{i,T})^2 \mid u_i\right] \lesssim (\eta - 2\sigma^2) + \frac{D^2(1+\gamma_{i,T})}{p^2_{\min,T}\sqrt{T-1}} + \frac{\sigma^2}{p_{\min,T}\,\Phi_i N}$$

Bias due to η

Concentration of neighbor distance

Effective noise variance

- Compared to Li et al 2017's guarantee for matrix completion with missing completely at random (MCAR) entries—-an additional factor of $\gamma_{i,T}/p_{\min,T}$

Regularity conditions needed for the central limit theorem

$$\mathbb{E}\left[(\widehat{Y}_{i,T} - Y^{\star}_{i,T})^2 \,|\, u_i\right] \lesssim (\eta - 2\sigma^2) + \frac{D^2(1+\gamma_{i,T})}{p^2_{\min,T}\sqrt{T-1}} + \frac{\sigma^2}{p_{\min,T}\,\Phi_i N}$$

``Bias" terms go to zero after multiplying by number of neighbors $N_{i,t}$:

- $\eta - 2\sigma^2$ goes to zero fast enough
- $N_{i,t}$ can **not** grow faster than $p^2_{\min,T}\sqrt{T}$ (cap number of nearest neighbors)

The denominator (min number of nearest neighbors) goes to ∞

Summary:
Counterfactual inference in sequential experimental design

For $t = 1,\ldots, T$

For $i = 1,\ldots, N$

$A_{i,t}$ = Bernoulli($\pi_t(i)$)

$$Z_{i,t} = \begin{cases} Y^\star_{i,t} + \varepsilon_{i,t} & \text{if} \quad A_{i,t} = 1 \\ \text{unknown} & \text{if} \quad A_{i,t} = 0 \end{cases}$$

Update policy to π_{t+1} using all users' data

Policy & model agnostic nearest neighbor estimate $\widehat{Y}_{i,t}$ for $Y^\star_{i,t}$

+

Latent factor model $Y^\star_{i,t} = f(u_i, v_t)$

Summary: Counterfactual inference in sequential experimental design

For $t = 1,\ldots,T$

For $i = 1,\ldots,N$

$A_{i,t} =$ Bernoulli($\pi_t(i)$)

$$Z_{i,t} = \begin{cases} Y^\star_{i,t} + \varepsilon_{i,t} & \text{if} \quad A_{i,t} = 1 \\ \text{unknown} & \text{if} \quad A_{i,t} = 0 \end{cases}$$

Update policy to π_{t+1} using all users' data

Policy & model agnostic nearest neighbor estimate $\widehat{Y}_{i,t}$ for $Y^\star_{i,t}$

+

Latent factor model $Y^\star_{i,t} = f(u_i, v_t)$

Bilinear f → Distribution-free **consistency** $\widehat{Y}_{i,t} \to Y^\star_{i,t}$ for any user i **&** any time t

Lipschitz non-linear f → **Central limit theorem** $\sqrt{N_{i,T}}\,(\widehat{Y}_{i,T} - Y^\star_{i,T}) \Rightarrow \mathcal{N}(0,\sigma^2)$ for any user i at last time T

Summary:
Counterfactual inference in sequential experimental design

For $t = 1,\ldots,T$

For $i = 1,\ldots,N$

$A_{i,t} = \text{Bernoulli}(\pi_t(i))$

$$Z_{i,t} = \begin{cases} Y^\star_{i,t} + \varepsilon_{i,t} & \text{if} \quad A_{i,t} = 1 \\ \text{unknown} & \text{if} \quad A_{i,t} = 0 \end{cases}$$

Update policy to π_{t+1} using all users' data

Policy & model agnostic nearest neighbor estimate $\widehat{Y}_{i,t}$ for $Y^\star_{i,t}$

+

Latent factor model $Y^\star_{i,t} = f(u_i, v_t)$

Bilinear f → Distribution-free **consistency** $\widehat{Y}_{i,t} \to Y^\star_{i,t}$ for any user i **&** any time t

Lipschitz non-linear f → **Central limit theorem** $\sqrt{N_{i,T}}\,(\widehat{Y}_{i,T} - Y^\star_{i,T}) \Rightarrow \mathcal{N}(0,\sigma^2)$ for any user i at last time T

Just a starting point…Future work:

- Impact on downstream tasks
- Use of covariates/contexts
- Temporal structure
- Information across treatments

Preprint on arxiv coming today evening! (Title same as this slide's title)

Additional slides

Explicit non-asymptotic bound for the bilinear case

$f(u, v) = \langle u, v \rangle$: A deterministic error bound for any (i, t)

- $T_{t,i,j}$ = commonly observed time points between i and j other than t (used to compute distance $\rho_t(i, j)$)
- $N_{i,t}$ = nearest neighbors for (i, t)
- Deterministic error bound:

$$(\widehat{Y}_{i,t} - Y^{\star}_{i,t})^2 \lesssim \frac{\|v_t\|_2^2}{\lambda}\left(\eta - c\sigma^2 + \frac{\langle \text{noise}, \{v_t\}\rangle}{\min_{j \in N_{i,t}} T_{t,i,j}}\right) + \left(\frac{\sum_{j \in N_{i,t}} \text{noise}_{j,t}}{|N_{i,t}|}\right)^2$$

Gaussian process as a bilinear latent factor model

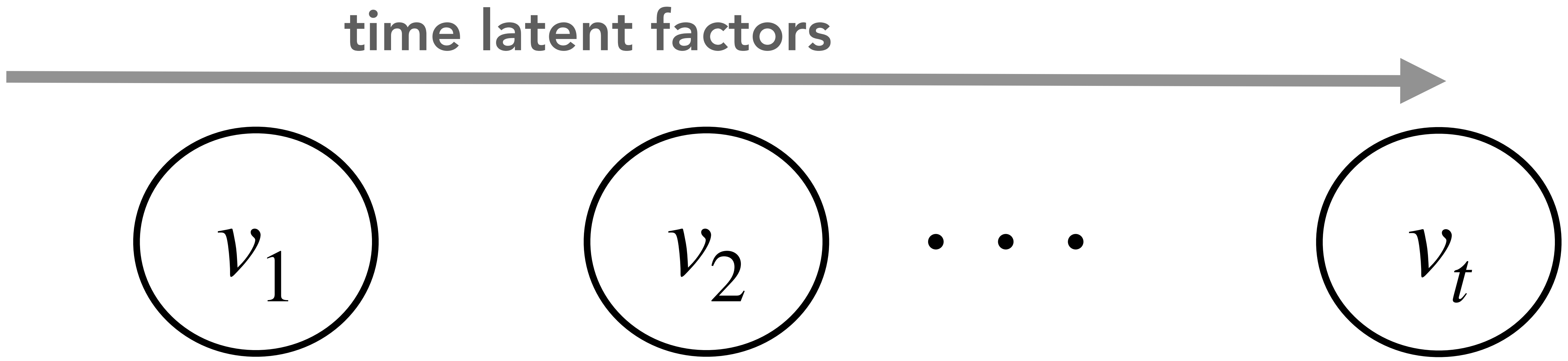
Latent factor model

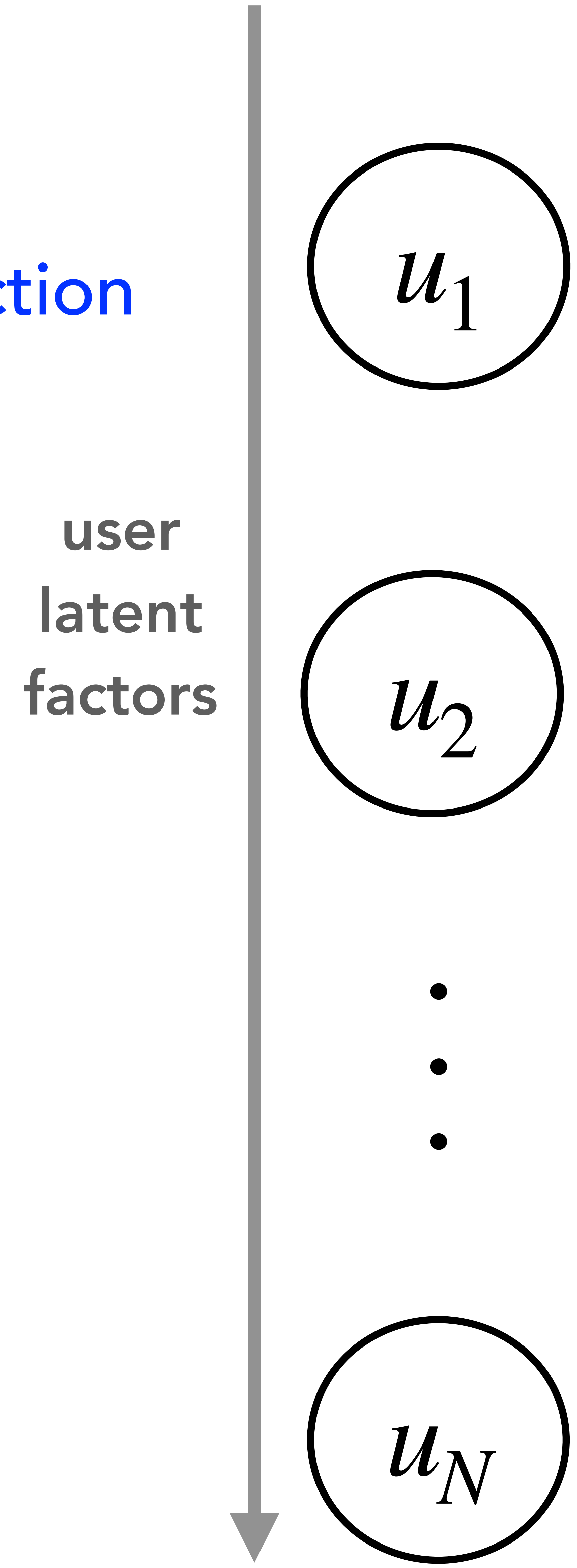
$$Y^{\star}_{i,t} \triangleq Y^{(1)}_{i,t} \triangleq f(u_i, v_t)$$

u_i : latent factor for user i

v_t : latent factor for time t

f : unknown (non) linear function





- $Y^{\star}_{i,\cdot} \sim$ Gaussian process with mean m in the reproducing kernel Hilbert space of the covariance kernel $\mathbf{k}$, which has eigenfunctions ϕ_j then

$$f(u, v) = \langle u, v \rangle,$$

$$u_i = (\langle m, \phi_j \rangle)_{j=1}^{r} + \mathcal{N}(0, I_r), \text{ and}$$

$$v_t = (\phi_j(x_t))_{j=1}^{r}.$$

Gaussian process as a latent factor model

- If each user's data is a sample from $\mathcal{GP}(0,\mathbf{k})$ where $\mathbf{k}$ is Mercer's kernel such that

$$\mathbf{k}(t_1, t_2) = \sum_{\ell=1}^{\infty} \lambda_\ell \phi_\ell(t_1)\phi_\ell(t_2),$$

where λ_ℓ, ϕ_ℓ denote eigenvalue-eigenfunctions with $\{\phi_\ell\}$ orthonormal

- Then for $\xi_{i,\ell} \sim_{iid} \mathcal{N}(0,1)$, we have

$$Y_{i,t} = \sum_{\ell=1}^{\infty} \xi_{i,\ell}\sqrt{\lambda_\ell}\phi_\ell(t) \text{ almost surely } \implies Y_{i,t} = f(u_i, v_t) = \langle u_i, v_t \rangle$$

for $u_i = (a_1, a_2, \ldots) + (\xi_{i,1}, \xi_{i,2}, \ldots)$, and $v_t = (\sqrt{\lambda_1}\phi_1(t), \sqrt{\lambda_2}\phi_2(t), \ldots)$

Example: Exchangeable data

- The latent factor model also holds if the matrix $\{Y^{\star}_{i,t} + \varepsilon_{i,t}, i = 1,\ldots.N, t = 1,\ldots,T\}$ for a sub-class of exchangeable data i.e., exchangeable under row and column permutations

- See Sec II.C Li et al. 2017

Mathematical description of nearest neighbors

Nearest neighbors estimate for $Y^{\star}_{i,t}$

- **Input**: Partially observed matrix; **Output**: Estimate of noiseless entry (i, t)

- **Algorithm**: Compute

    - Available neighbors at time $t = \{j : A_{j,t} = 1\}$

    - Good neighbors for user i at time $t = \{j : \rho_t(i, j) \leq \eta\}$ where

$$\rho_t(i, j) = \frac{1}{\sum_{t' \neq t} A_{i,t'} A_{j,t'}} \sum_{t' \neq t} (Z_{i,t'} - Z_{j,t'})^2 A_{i,t'} A_{j,t'}$$

    - $\widehat{Y}_{i,t}$ = Simple average of $Y_{j,t}$ over $\{j : A_{j,t} = 1 \text{ and } \rho_t(i, j) \leq \eta\}$.

Further details about prior work

Data collected with a <u>fixed</u> policy: Off policy evaluation

- Set-up considers either
  - i.i.d. users, e.g., multi-armed bandits
  - or, one user over time, e.g., Markov decision proces
  - but ***not*** multiple users over multiple time
- **Quantities of interest:** Average reward under alternative policy, estimated using IPW-based estimates, switch estimators etc.

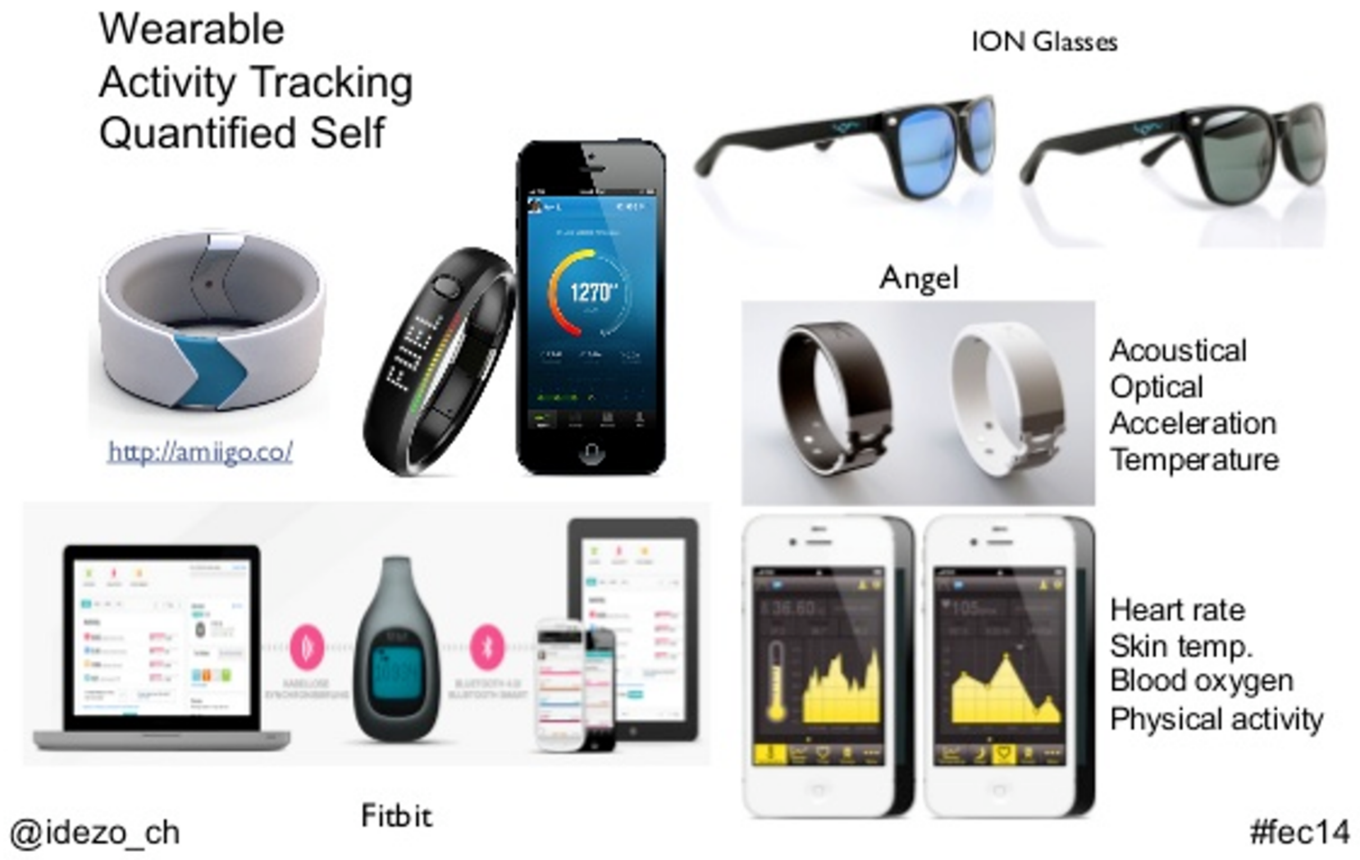
[…, Li et al 2015, Wang et al. 2021, Ma et al. 2021,…]

Diverse mobile health applications

Smoking addiction

Well-being

Wearable/trackers



Binge drinking

Recovery support

Physical activity

Image credits: Susan Murphy