Cornell Bowers Statistics + Data Science

Counterfactual Forecasting in Panel Data

Navonil Deb* Raaz Dwivedi† Sumanta Basu*

* Statistics and Data Science, Cornell University

† Operations Research and Information Engineering, Cornell Tech

Contact information

Navonil Deb
Department of Statistics and Data Science
Cornell University
Email: nd329@cornell.edu

Counterfactual forecasting problem with N users and T time points

- Binary treatment: $w \in \{0,1\}$. (No nudge and nudge)
- Counterfactual outcome (e.g. 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

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

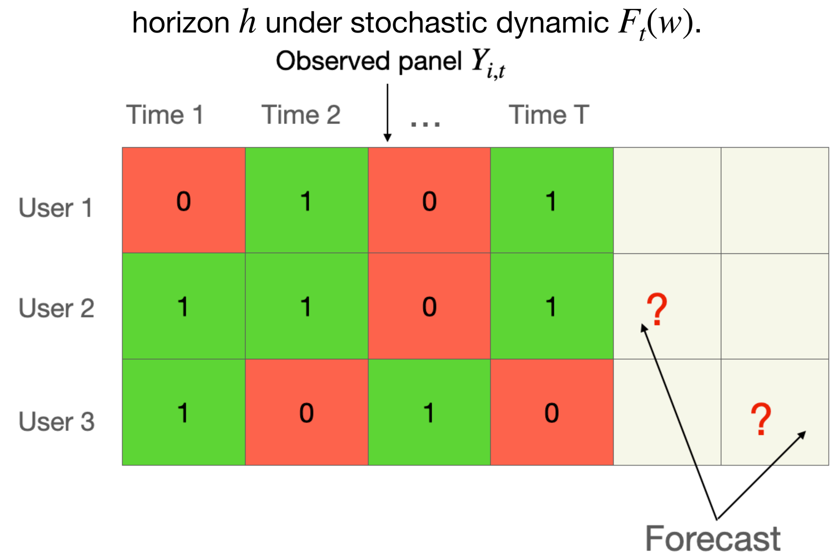
— $W_{i,t}$: Treatment variable

- Low rank structure: $\theta_{i,t}(w) = \Lambda_i^\top(w)\, F_t(w)$.

— **Factors** $F_t(w) \in \mathbb{R}^r$: shared across the users (e.g. shared dependence of walking behavior)

— **Loadings** $\Lambda_i(w) \in \mathbb{R}^r$: shared across time E.g. Association among observed steps and the latent factors

- The latent factors are often stochastic, time varying in nature, with autocorrelation across time.
- Examples of latent factor dynamics: Markov, vector autoregressive (VAR), state space etc.
- **Goal**: Forecasting potential steps $Y_{i,T+h}(w)$ for future horizon h under stochastic dynamic $F_t(w)$.



Overreaching question

Can we accurately forecast counterfactuals by learning the temporal dynamics of the latent factors?

Panel data with dynamic factors and missing entries

- Observed: $(Y_{i,t}, W_{i,t})$, $W_{i,t}$: observation indicator
- Restrict to treated panel: $Y_{i,t} = \Lambda_i^\top F_t + \varepsilon_{i,t}$, $W_{i,t} = 1$
- Assumption: The latent factors F_t follows a stable r -dimensional vector autoregressive model of order 1, i.e. VAR(1) model as

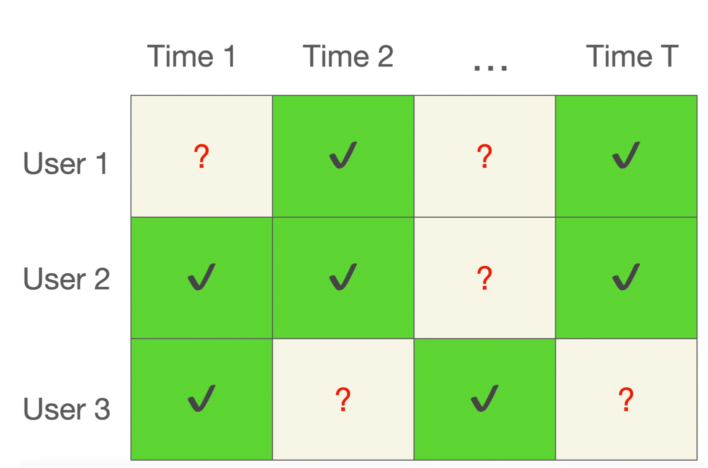
$$\boxed{F_t = AF_{t-1} + \eta_t,}$$

where $A \in \mathbb{R}^{r\times r}$ with $\rho(A) < 1$ and η_t is a noise process with mean 0 and covariance matrix Σ_η.

- Under VAR(1) assumption:

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

- $\mathbb{E}[F_{T+h} \,|\, \sigma(F_t : t \le T)] = A^h F_T$



- **Forecast target:** The best linear predictor of the outcome variable $\theta_{i,T:T+h} := \mathbb{E}[\theta_{i,T+h} \,|\, \sigma(F_t : t \le T)] = \Lambda_i^\top A^h F_T$

Why is this problem significant?

- Counterfactual estimation problem arising in causal inference has important applications in medical sciences and mobile health, recommendation systems, policy evaluation and other disciplines.
- A powerful approach models outcome trajectories via low rank models that capture shared variation over time and units.

  E.g. synthetic controls, difference-in-differences, and matrix completion methods.
- Incorporating temporal evolution of latent factors enables forecasting counterfactual outcomes beyond the observed panel. Improved counterfactual forecasts and facilitate more informed decisions.

Rigor of prior works

- Matrix completion and causal effect estimation via low-rank factor models
  - **Factor model approaches:** Tall-wide algorithm (Bai and Ng, 2021), EM-based method (Jin et al., 2021), Tall-project algorithm (Cahan et al., 2023), PCA-based method (Xiong and Pelger, 2023).
    — Do not exploit temporal dynamics of the factors.
  - **Neural network + synthetic control approach:** SyN-BEATS (Goldin et al., 2022)
    — Do not exploit temporal dynamics of the factors. No theoretical backing.
  - **Multivariate singular spectrum analysis** (mSSA, Agarwal et. al., 2020) and its variants
    — Assumes that the time-varying factors are deterministic i.e. only singular spectra are present in the factor process.
    — Not suitable to forecast in presence of stochastic factors with serial correlation.
- Forecasting with factor models in time series literature:
  - Factor augmented VAR models (e.g. Bernanke et al., 2005) and collapsing techniques in dynamic factor models (e.g. Bräuning and Koopman, 2014).
    — Proportion of missing entries in the panel is significantly smaller than that arising in counterfactual estimation problem (e.g. number of control observations in the treatment panel).

Forecasting Counterfactuals Under Stochastic dynamics (FOCUS)

An algorithm to estimate the target $\theta_{i,T:T+h}$ for a fixed unit i and horizon h

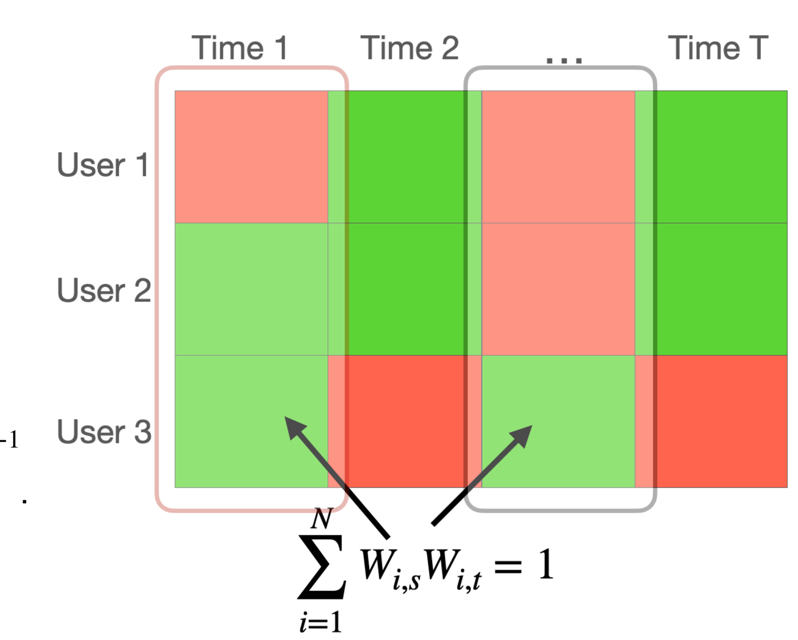
1. **Estimate the factors and loadings using PCA**
   - For each pair of time points $s, t \in [T]$, calculate sample covariance matrix $\hat\Sigma$ (similar to PCA of Xiong and Pelger, 2023)

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

   - Estimated factors: $\hat F = \sqrt{T} \times$ First r eigenvectors of $\frac{1}{T}\hat\Sigma$.
   - Estimated loadings: by regressing $Y_{i,t}$ on $W_{i,t}\hat F_t$

$$\hat\Lambda_i = \left(\sum_{t=1}^{T} W_{i,t}\hat F_t \hat F_t^\top\right)^{-1}\left(\sum_{t=1}^{T} W_{i,t}\hat F_t Y_{i,t}\right).$$

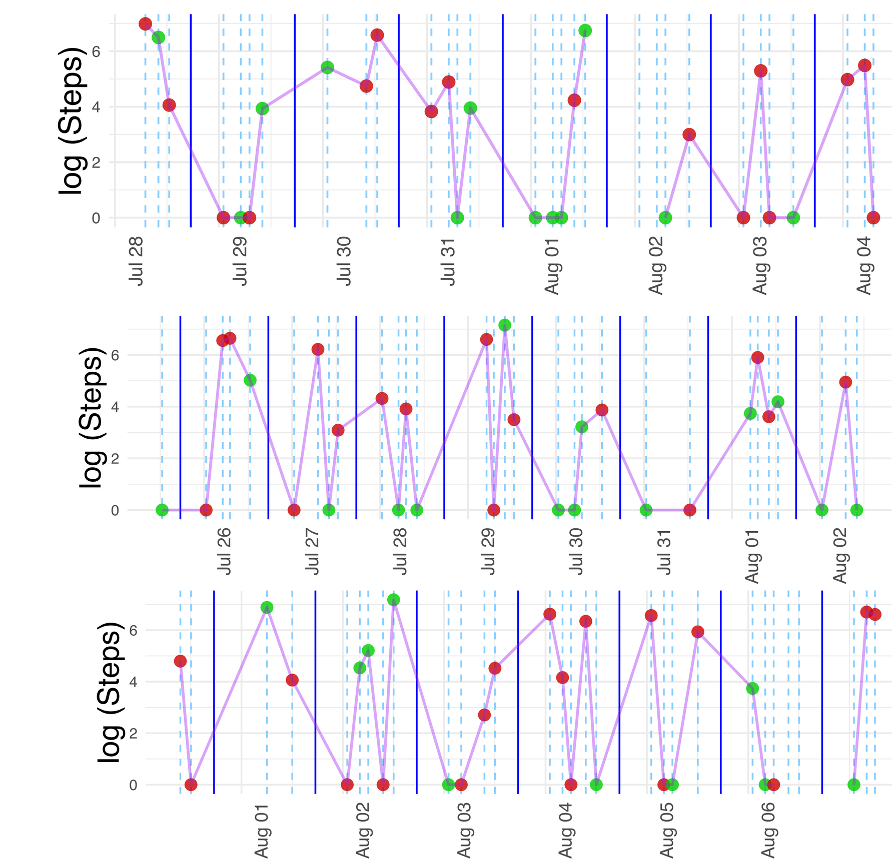
2. **Forecast with the estimated factors and loadings**
   - OLS estimator of A: $\hat A = \left(\sum_{t=1}^{T-1}\hat F_{t+1}\hat F_t^\top\right)\left(\sum_{t=1}^{T-1}\hat F_t \hat F_t^\top\right)^{-1}$.
   - **Plug-in estimator** of $\theta_{i,T:T+h}$: $\hat\theta_{i,T:T+h} = \hat\Lambda_i^\top \hat A^h \hat F_T$.



Motivating example: HeartSteps

- HeartSteps V1 is a 6-week mHealth intervention study with 37 sedentary adults. Participants received a context-tailored walking notifications at 5 times a day via fitness trackers (e.g. jawbone, Google fit).
- Under one of the intervention, the walking behavior has a **temporal pattern shared by users**.
  - In consecutive decision point pairs, users with high step counts in the previous time point tend walk less in the next time point.

    $\Longrightarrow$ A negative autocorrelation
- Accurate forecast of potential steps help practitioners take more informed decision and addressing effectiveness of the nudge.



[Source: Research of Murphy et al.]

Main results

Assumptions on outcome model

1. The VAR(1) noise process η_t is i.i.d $N(0,\Sigma_\eta)$.
2. Λ_i are i.i.d with mean 0, covariance matrix Σ_Λ, and $\mathbb{E}[\|\Lambda_i\|^4] < \infty$.
3. $\varepsilon_{i,t}$ are i.i.d mean 0, variance σ_ε^2, and $\mathbb{E}[\,|\varepsilon_{i,t}|^8] < \infty$.
4. $\Lambda_i, \eta_t, \varepsilon_{i,t}$ are mutually independent.
5. The eigenvalues of $\Sigma_F\Sigma_\Lambda$ are distinct

Assumptions on observation pattern

1. W is independent of F and ε.
2. There is a $\underline{q}$ such that with probability going to 1, $\frac{1}{N}\sum_{i=1}^N W_{i,t}W_{s,t} \ge \underline{q}$, i.e. with high at least one user is observed at any time pairs.
3. $\frac{1}{N}\sum_{i=1}^N W_{i,t}W_{i,s}$ and $\frac{1}{N}\sum_{i=1}^N W_{i,t}W_{i,s}W_{i,u}W_{i,v}$ have almost sure limits with N

Need similar other assumptions on the almost sure limits and the mixed moments of F_t and $W_{i,t}$ as $T \to \infty$.

— Denote $\delta_{NT} := \min\left\{\sqrt{N}, \sqrt{T}\right\}$.

- **Error bound for** $\hat\theta_{i,T:T+h}$**:** Under the regularity conditions

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

- **Asymptotic normality**: $\delta_{NT}\left(\hat\theta_{i,T:T+h} - \theta_{i,T:T+h}\right)/\sigma_{i,T,h} \xrightarrow{d} \mathcal{N}(0,1)$, $\sigma^2_{i,T,h} = \sigma^{2,\text{est}}_{i,T,h} + \sigma^{2,\text{for}}_{i,T,h}$.
  - $\sigma^{2,\text{est}}_{i,T,h}$: Uncertainty due to missing entries
  - $\sigma^{2,\text{for}}_{i,T,h}$: Uncertainty due to fitting the VAR(1) model
- Can compute the HAC estimator of variance $\hat\sigma^2_{i,T,h}$ and construct $100(1-\alpha)\,\%$ forecast confidence interval.

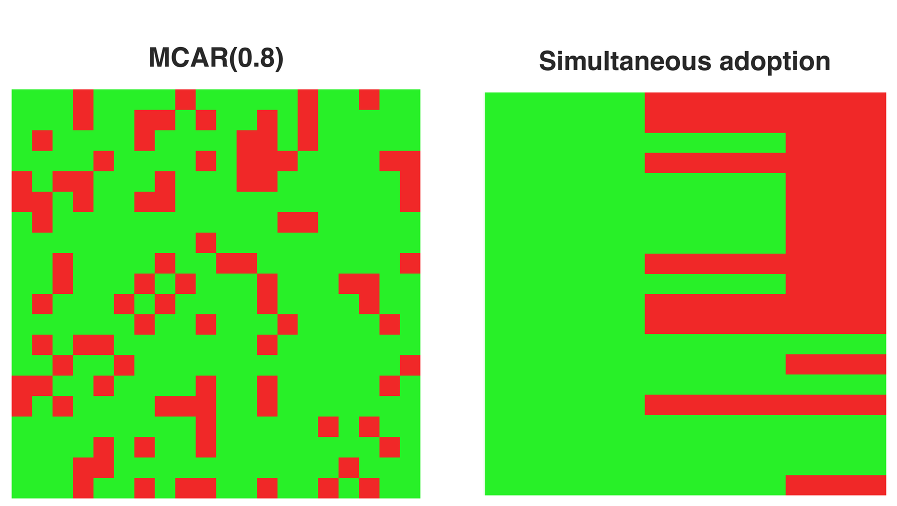
☑ Results hold for several observation patterns, can be generalized to more general linear processes

- Missing completely at random (MCAR), $W_{i,t} \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(p)$
- Staggered adoption design (often appears in *synthetic controls*)

Leveraging latent dynamics $\Longrightarrow$ Accurate forecast

Simulation setting

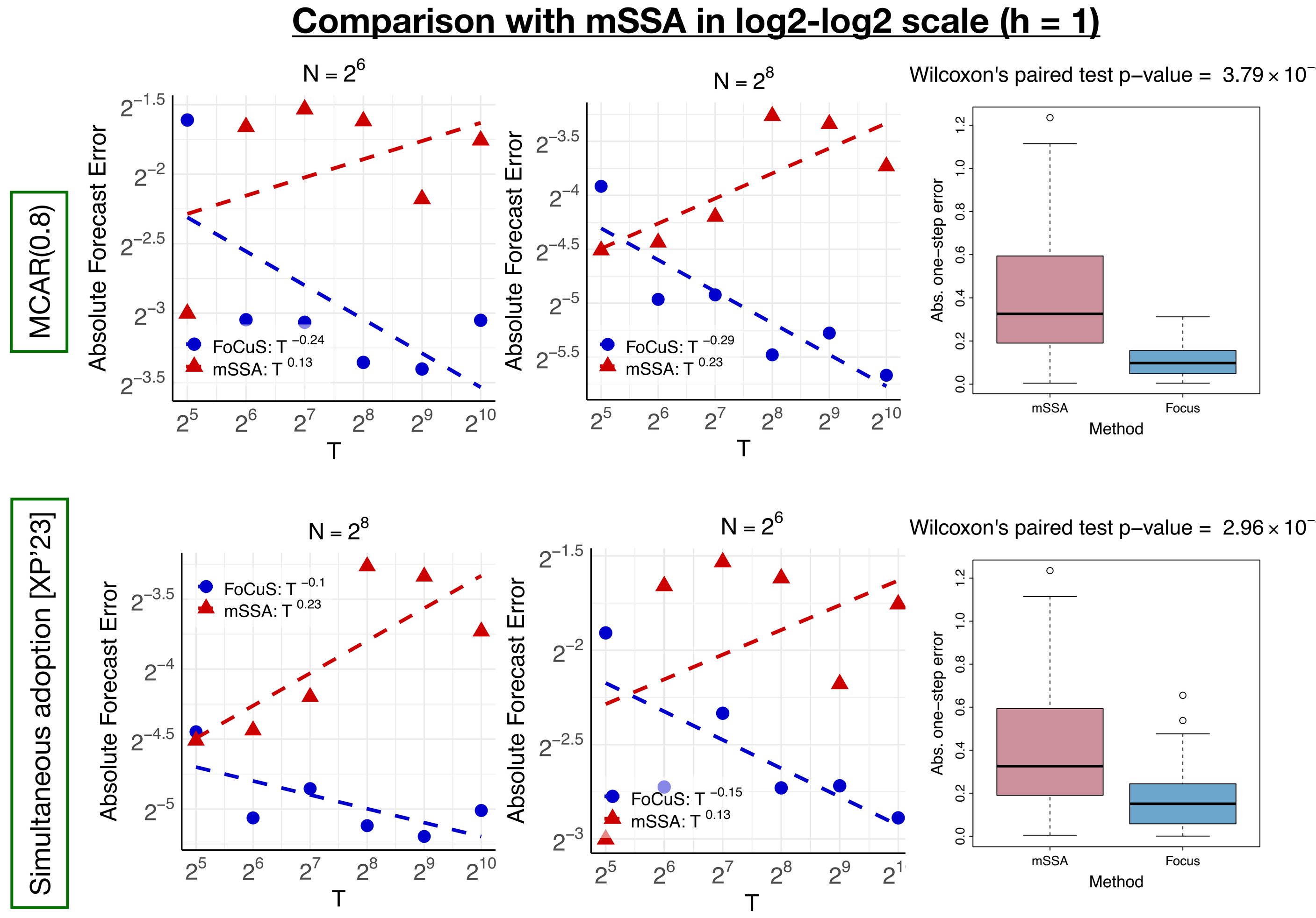
- Benchmark: mSSA and SyN-BEATS.
  — Does not capture temporal correlation of factors.
- Generative model: One-factor model, i.i.d loadings and noise.
- Factors: Quadratic $(100t^2/T^2)$ + AR(1) factors with coefficient 0.5



- Use spline smoothing + FOCUS on estimated factors.
- **Performance metric**: Median of absolute forecast error over 50 rep.

$$|\hat\theta_{i,T:T+h} - \theta_{i,T:T+h}|$$

Comparison with mSSA in log2-log2 scale (h = 1)



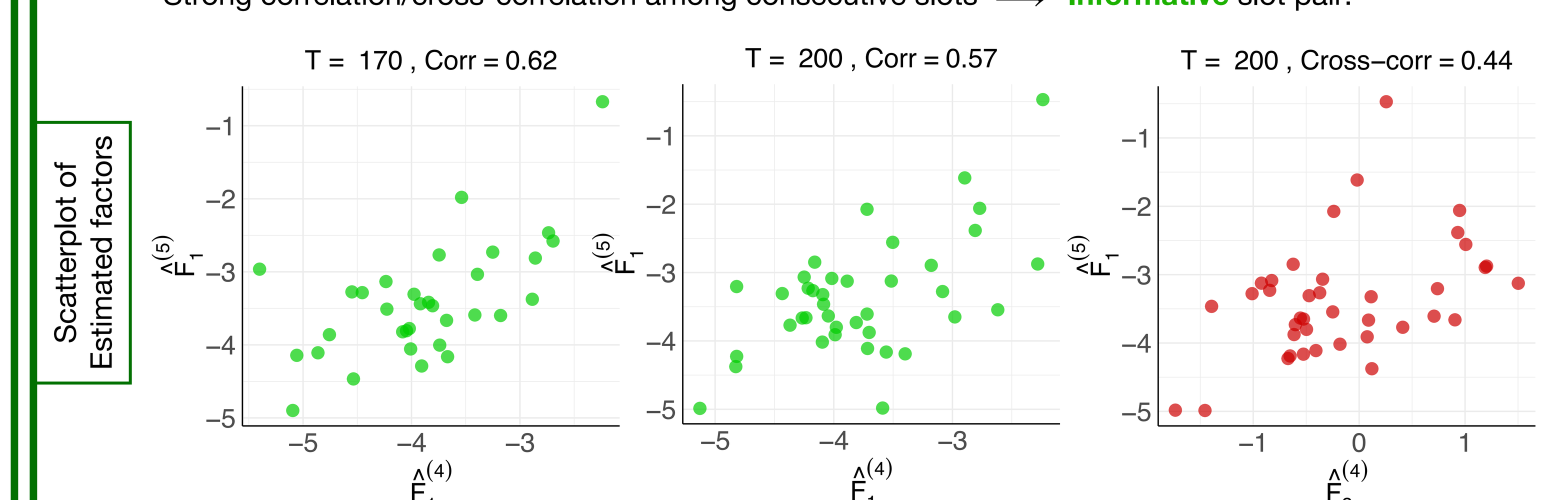
Comparison with SyN-BEATS (T = 32)

Table: 100 × median of the ratio of abs. forecast errors (FOCUS/SyN-BEATS). For all experiments, Wilcoxon paired test with the forecast errors yield p-value<< 0.05.

h	N = 32	N = 64
1	4.83	5.15
2	5.43	5.17
3	4.74	5.02
4	4.85	4.80
5	5.81	5.69

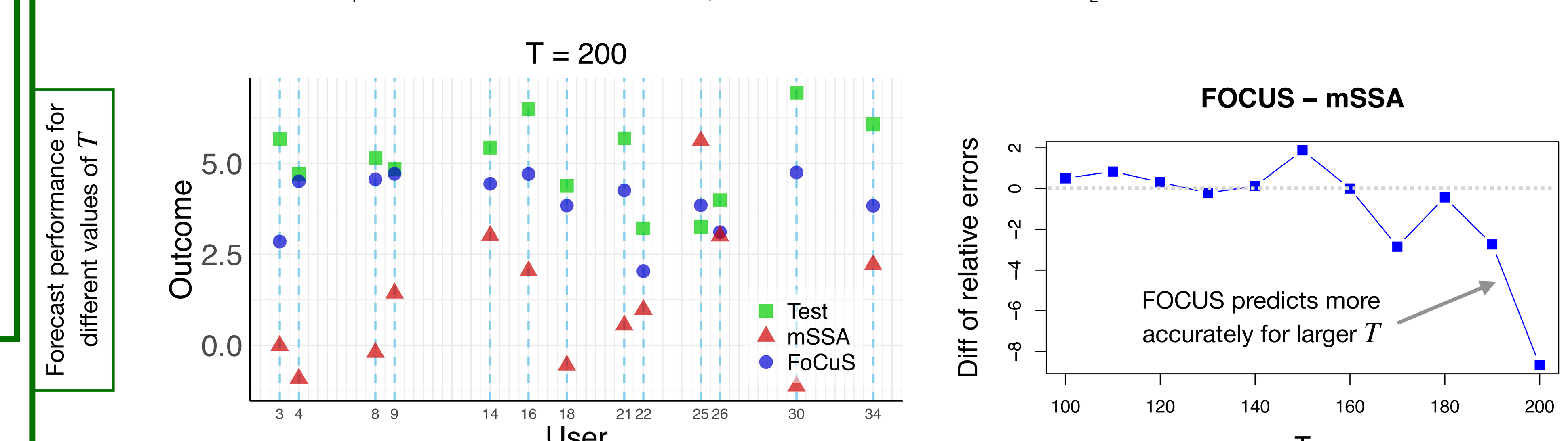
FOCUS more accurately forecasts the steps under nudge

- $Y = \log(1 + \text{jbsteps30})$, $W = I\{\text{available, nudged}\}$, $N = 36$, $T \in \{100, 110, \ldots, 200\}$.
- Strong correlation/cross-correlation among consecutive slots $\Longrightarrow$ Informative slot pair!



Informative pair

Late evening (4) and before sleep (5)



References

Bai, J., & Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, *116*(536), 1746-1763.

Jin, S., Miao, K., & Su, L. (2021). On factor models with random missing: Em estimation, inference, and cross validation. *Journal of Econometrics*, *222*(1), 745-777.

Cahan, E., Bai, J., & Ng, S. (2023). Factor-based imputation of missing values and covariances in panel data of large dimensions. *Journal of Econometrics*, *233*(1), 113-131.

Xiong, R., & Pelger, M. (2023). Large dimensional latent factor modeling with missing observations and applications to causal inference. *Journal of Econometrics*, *233*(1), 271-301.

Goldin, J., Nyarko, J., & Young, J. (2022). Forecasting algorithms for causal inference with panel data. *arXiv preprint arXiv:2208.03489*.

Agarwal, A., Alomar, A., & Shah, D. (2022). On multivariate singular spectrum analysis and its variants. *ACM SIGMETRICS Performance Evaluation Review*, *50*(1), 79-80.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, *120*(1), 387-422.