Design and Analysis of Panel Data Experiments

Ruoxuan Xiong

Emory University

Aug 17 @ AI/OR Workshop

Optimal experimental design for staggered rollouts with Susan Athey, Mohsen Bayati, Guido Imbens (XABI'19)

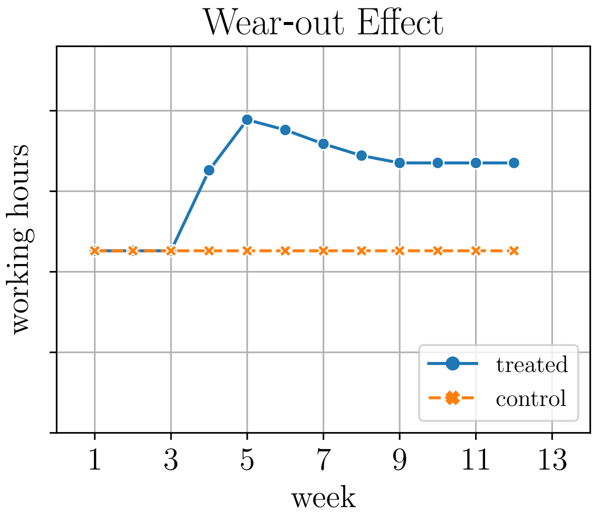
Bias-variance tradeoffs for designing simultaneous temporal experiments with Alex Chin, Sean Taylor, Susan Athey (XCTA'22)

EMORY

Example: Marketplace experiments

A ridesharing platform tests an app feature to improve driver experience

- Conventional A/B testing suffers from network or contamination effect
- The impact of the feature may fade over time



Two key challenges

1. Interference and network effects

2. Carryover effects (instantaneous and lagged effects)

Panel data experiments can help addressing **both** challenges

Panel data experiments

- Run an experiment on N units (e.g. cities) for T time periods (e.g. days, weeks)
  - Repeated treatment decisions for the same N units
  - Repeated observations on the same N units

	day 1	day 2	day 3	day 4
NY	Control	Control	Treatment	Treatment
SF	Treatment	Treatment	Treatment	Treatment
ATL	Control	Treatment	Treatment	Treatment
:				

- Three main advantages
  - Increase sample size
  - Allow for heterogeneity in unit means and common time shocks
  - Study the impact of sustaining treatment over multiple periods

Treatment allocation schemes

Switchback experiments

	day 1	day 2	day 3	day 4
NY	Treatment	Treatment	Control	Control
SF	Treatment	Control	Control	Treatment
ATL	Control	Treatment	Control	Treatment
:				

Pro: Flexible
Con: Some practical constraints to switch back

[Bojinov, Simchi-Levi, Zhao'20]

Irreversible treatment adoption experiments

	day 1	day 2	day 3	day 4
NY	Control	Control	Treatment	Treatment
SF	Treatment	Treatment	Treatment	Treatment
ATL	Control	Treatment	Treatment	Treatment
:				

Simultaneous or staggered adoption pattern
Optimal design is a challenging problem!

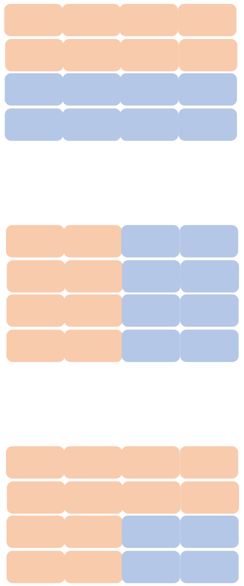
[Hussey and Hughes'07, Hemming et al.'15, Li, Turner and Preisser'18]

Decision making problem

- **Objective:** Most precisely estimate instantaneous and lagged effects
  - Can reduce sample size requirement and experimental cost!

- **Decision variables:** Optimally choose the treatment times for each unit
  - Integer programming problem

	day 1	day 2	day 3	day 4
NY	Control	Control	Treatment	Treatment
SF	Treatment	Treatment	Treatment	Treatment
ATL	Control	Treatment	Treatment	Treatment
:				



Z_{ff}?

Z_{ba}?

Z_{ffba}?

Fixed-sample-size experiments

- N and T are set pre-experiment

- Consider a flexible outcome model
  - Nonstationarity and heterogeneity

$$Y_{it}(Z) = \alpha_i + \beta_t + X_i^\top \cdot \theta_t + U_i^\top \cdot V_t + \tau_0 \cdot Z_{it} + \tau_1 \cdot Z_{i,t-1} + \cdots + \tau_\ell \cdot Z_{i,t-\ell} + \varepsilon_{it}$$

Two-way fixed effects (**unknown**)

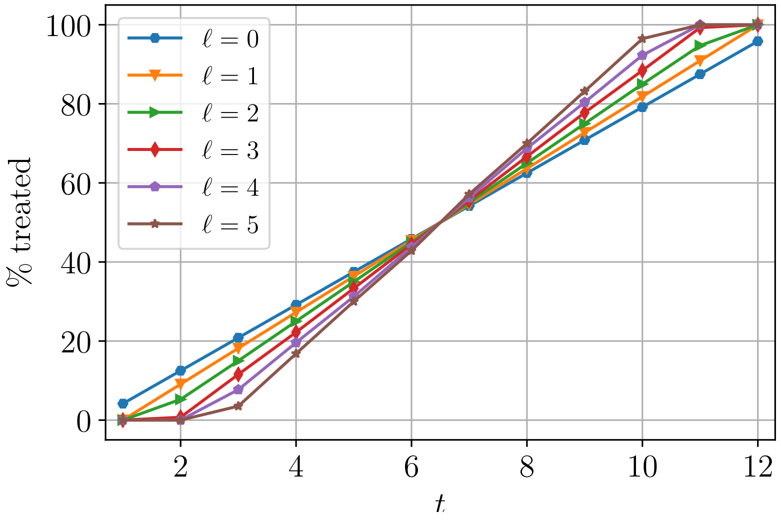
Observed and latent covariates with **unknown** time-varying coefficients

Binary treatment variables with **unknown** treatment effect parameters

- Formulate an integer program to solve Z that maximizes the estimation precision of instantaneous effect $\hat{\tau}_0$ and lagged effects $\hat{\tau}_1, \dots, \hat{\tau}_\ell$, post-experiment

Optimality conditions and choosing a design

- Provide the optimal conditions for the integer program [XABI'19]
  - Optimal treated fraction at every time period
  - Covariate conditions
- Develop an algorithm to choose a design with provably guarantee to the optimal integer solution (within $1 + O(1/N^2)$) [XABI'19]



Optimal treated fraction for a 12-period experiment

Algorithm 2: Choose a treatment design for each stratum g

```
Inputs: |O_g|, [ω*_{ℓ,t}]_{t∈[T]}
for t = 1, ..., T do
    if frac(|O_g|·(1+ω*_{ℓ,t}) / 2T) < 1/2 or frac(|O_g|·(1+ω*_{ℓ,t}) / 2T) = 1/2 with (1+ω*_{ℓ,t}) / 2T < 1/2 then
        N_{g,t} ← ⌊|O_g|·(1+ω*_{ℓ,t}) / 2T⌋ ;
    else
        N_{g,t} ← ⌈|O_g|·(1+ω*_{ℓ,t}) / 2T⌉ ;
    end
end
f(·) ← a random function that shuffles {1, 2, ..., |O_g|};
Z_g ← [−1]^{|O_g|×T} ;
for i = 1, ..., |O_g| do
    for t = 1, ..., T do
        if f(i) ≤ N_{g,t} then
            z_{g,it} ← 1 ;
        else
            z_{g,it} = −1 ;
        end
    end
end
return Z_g;
```

Sequential experiments

- N is fixed and T varies (we can early stop the experiment)

- More flexible and cost-effective!

- Key challenges
  - Experiment termination rule
    - What is an appropriate rule and how to implement the rule
  - Peeking challenge [Johari, Koomen, Pekelis, Walsh'17]
    - Treatment effect estimation based by experiment termination rule
  - Infeasibility to optimize treatment times pre-experiment
    - Optimal solution depends on T

An initial stab at the problem

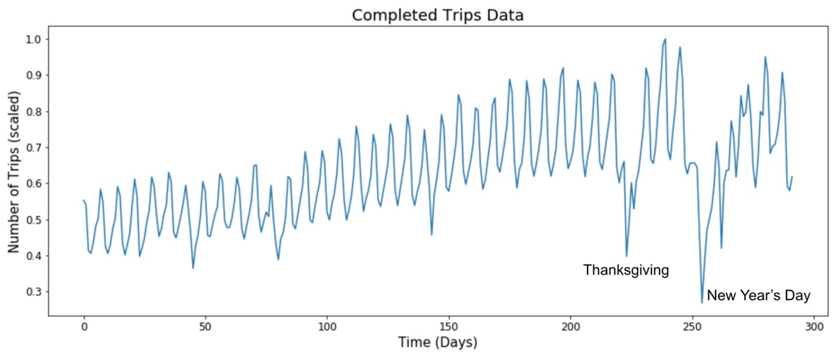
- We propose the Precision-Guided Adaptive Experimentation (P-GAE) algorithm [XABI'19]
  - Leverage ideas from dynamic programming, empirical Bayes, and sample splitting
  - Key features
    - Adaptive treatment decisions
    - Precision-based experiment termination rule
    - Valid statistical inference post-experiment
  - We provide theoretical guarantees for P-GAE
    - Asymptotic: consistency, normality, and efficiency

Micro-level perspective

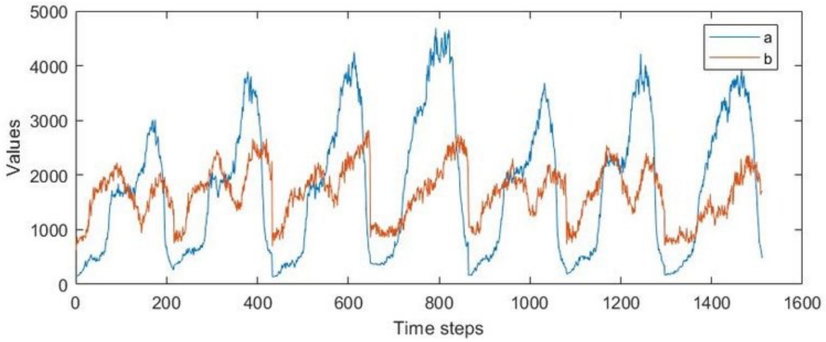
- Micro-level data: raw data are **events**, like rider checking price, outcome is whether rider requested a ride
  - Large sample size, but analysis is more challenging [Johari, Li, Liskovich, Weintraub'21]

- Additional considerations when analyzing event data

Irregular event density



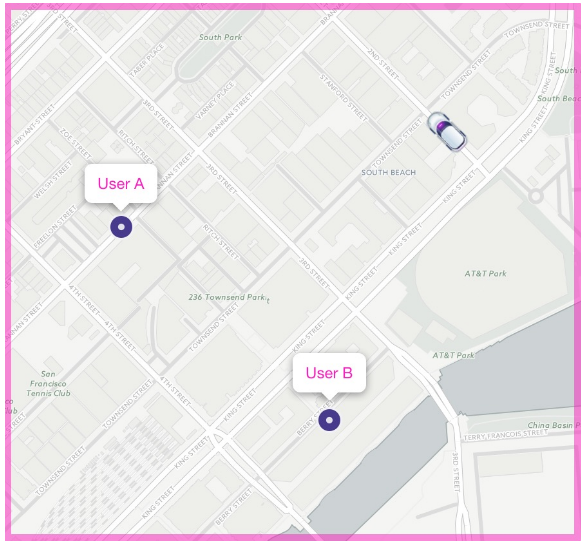
Correlation in event outcomes



Additional considerations

- Additional considerations when analyzing event data

Spillover effects



Simultaneous experiments

11:00 am 11:30 am 12:00 pm 12:30 pm 01:00 pm

Treatment 1	Treatment 1	Control 1	Control 1
Treatment 2	Control 2	Control 2	Treatment 2
Control 3	Treatment 3	Control 3	Treatment 3
Control 4	Treatment 4	Treatment 4	Control 4

EMORY

Error analysis and design of experiments

- Analyze the mean-squared error (MSE) in treatment effect estimation [XCTA'22]
  - Bias affected by event density, spillover effects, simultaneous experiments
  - Variance affected by event density, correlation in event outcomes

- Study how partition time and space (irregularly) to minimize MSE [XCTA'22]

Thank you!