Optimal Experimental Design for Staggered Rollouts

Ruoxuan Xiong
Emory University

Joint work with Susan Athey, Mohsen Bayati, Guido Imbens

https://arxiv.org/abs/1911.03764
Two motivating problems
Example 1: Marketplace experiments

But running experiments in a marketplace environment is challenging

**Example:** A ridesharing platform tests a new safety feature for drivers

- A/B testing suffers from network or contamination effect
- This safety feature can have drivers develop good driving habits and have a persistent impact on accident rates

“Today, Microsoft and several other leading companies—including Amazon, Booking.com, Facebook, and Google—each conduct more than 10,000 online controlled experiments annually, with many tests engaging millions of users.”
Example 2: Interventions for managing a contagious disease

• Experimentation for studying effects of non-pharmaceutical interventions

• Treatment and control groups may interact with one another

• Interventions may have carryover effects
Challenges in experimental design
Two challenges in experimental design

• **Challenge 1**: Interference and network effects

• **Challenge 2**: Carryover effects (instantaneous and lagged effects)

• Increasing drawn attention in operations research, statistics, and biostatistics communities to study the design of experiments to
  • Most efficiently learn the effect of a treatment (direct impact on experimental cost)
  • Address either one or both challenges

• This paper aims to address both challenges
Potential solutions for the first challenge

• **Challenge 1**: Interference and network effects
  
  • **Solution 1**: Run experiments at the individual level and account for the interference in the treatment effect estimation (e.g., Johari, Li, Liskovich, Weintraub 2021, Basse and Feller 2018)
    
    - **Pro**: Sample size is large
    - **Con**: Assumptions on the interference structure

  • **Solution 2**: Run experiments at the cluster level (e.g., Bojinov, Simchi-Levi, Zhao 2020, Candogan, Chen, Niazadeh 2021)
    
    - **Pro**: No interference between units
    - **Con**: Sample size tends to be small (statistical power is low)

• **This paper**: Solution 2
Potential solutions for the second challenge

- **Challenge 2**: Carryover effects
  - **Solution 1**: Run switchback experiments (e.g., Bojinov, Simchi-Levi, Zhao 2020)
    - Pro: Flexible
    - Con: Some practical constraints to switch back
  - **Solution 2**: Run experiments with staggered rollouts of treatment/stepped wedge design (e.g., Hussey and Hughes 2007, Hemming et al. 2015, Li, Turner and Preisser 2018)
    - Optimal design is a challenging problem!

- **This paper**: Solution 2
Panel data experiments, model and assumptions
Panel data experiments

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

- Benefits of panel data experiments
  1. Capable to estimate carryover effects
  2. Sample size is increased

<table>
<thead>
<tr>
<th></th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NY</td>
</tr>
<tr>
<td></td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>LA</td>
</tr>
<tr>
<td>-1: control</td>
<td>-1</td>
</tr>
<tr>
<td>1: treated</td>
<td>1</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

Optimal Design of Spatiotemporal Experiments
Our objective

- Optimally choose the treatment times for each unit in anticipation of most precisely estimating instantaneous and lagged effects

- Reduce the sample size requirement and directly minimize the experimental cost!

<table>
<thead>
<tr>
<th>Time</th>
<th>NY</th>
<th>SF</th>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td></td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td></td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td></td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

-1: control
1: treated

3/2/22
This is a challenging problem...

• We need to make two interrelated decisions
  
  1. Statistical decision: Which treatment effect estimation and inference approaches to use?

  2. Optimization decision: How to select treatment times to maximize precision?

    • Depends on the statistical decision
Tradeoffs in these two decisions

1. One can choose a simple estimator to simplify the optimization problem
   - E.g., Difference-in-means estimators (e.g., Bojinov, Simchi-Levi, Zhao, 2020)
   - The estimator can be inefficient

2. One can choose a powerful estimator, but the optimization problem is notoriously difficult to solve
   - E.g., Some sophisticated and efficient machine learning estimators (e.g., Athey, Bayati, Doudchenko, Imbens, Khosravi 2021)

• The decision maker needs to carefully balance these two decisions!
Data model

For all \((i, t) \in \{1, \ldots, N\} \times \{1, \ldots, T\}\),

\[ Y_{it} = \alpha_i + \beta_t + X_i^T \cdot \theta_t + U_i^T \cdot V_t + \tau_0 \cdot z_{it} + \tau_1 \cdot z_{i,t-1} + \cdots + \tau_\ell \cdot z_{i,t-\ell} + \varepsilon_{it} \]

**Key variables/parameters:**

- \(Y_{it}\): potential outcome
- \(z_{it}\): treatment indicator \(z_{it} \in \{-1,1\}\) (decision variable)
- \(\tau_0\): instantaneous treatment effect
- \(\tau_j\): lagged \(j\)-th period treatment effect for \(j \geq 1\)

**Auxiliary variables/parameters:**

- \(\alpha_i\): unknown unit effect; \(\beta_t\): unknown time effect
- \(X_i\): observed covariates; \(\theta_t\): unknown time-varying coefficients
- \(U_i, V_t\): latent covariates and coefficients
- \(\varepsilon_{it}\): iid observational noise
Irreversible Treatment Adoption Pattern

• Treatment is irreversible: $z_{it} \leq z_{i(t+1)}$

Solution for the removable case is also provided in our paper
  • A less challenging problem!
Two types of panel data experiments

• Fixed-sample-size experiments
  • $N$ and $T$ are set pre-experimentation

• Sequential experiments
  • $N$ is fixed and $T$ varies (we can early stop the experiment)
  • More flexible and cost-effective!
Fixed-sample-size experiments

• Estimation and inference of treatment effects are less challenge

• **Key challenge**: Optimization of treatment times

• **Our contribution**: Provide provably near-optimal analytical solutions
Sequential experiments

• Treatment effect estimation can be biased due to the well-known **peeking challenge** (Johari, Koomen, Pekelis, Walsh 2017)!

• Optimizing treatment times pre-experimentation is generally infeasible!

• **Our contribution:** Propose the Precision-Guided Adaptive Experiment (PGAE) algorithm for adaptive treatment design and post-experimentation inference
  • Leverage ideas from Bayesian statistics, dynamic programming, and sample splitting
Fixed-sample-size experiments
Estimation

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

- We use generalized least squares (GLS) to estimate \( \tau_0, \tau_1, \ldots, \tau_\ell \)

\[
\arg \min_{\tau, \alpha, \beta, \theta} \sum_t \tilde{e}_t^\top \cdot W^{-1} \cdot \tilde{e}_t
\]

where \( \tilde{e}_t = \tilde{y}_t - \alpha - \beta_t \cdot \tilde{1} - X \cdot \theta_t - \tau_0 \cdot \tilde{z}_t - \cdots - \tau_\ell \cdot \tilde{z}_{t-\ell} \)

- GLS does not explicitly estimate \( U_i \) and \( V_t \)
- Optimal \( W \) is proportional to \( \text{Var}(\tilde{e}_t) \)
- Two nice properties of GLS:
  - Best Linear Unbiased Estimator
  - \( \text{Prec}(\hat{\tau}_0, \ldots, \hat{\tau}_\ell) \) as a quadratic function of \( Z \)
    - \( \text{Prec}(\hat{\tau}_0, \ldots, \hat{\tau}_\ell) = \text{Var} (\hat{\tau}_0, \ldots, \hat{\tau}_\ell)^{-1} \)
Optimization problem for treatment decisions

• Trace (T)-optimal design (Pukelsheim 2006)

\[
\max_{\{z_{it}\}} \text{trace}(\text{Prec}(\hat{\tau}_0, \ldots, \hat{\tau}_\ell))
\]

s. t. \quad z_{it} \leq z_{i(t+1)}

\[z_{it} \in \{-1, +1\}\]

• Other objective functions: e.g., determinant (D)-optimal design
  - No analytical solutions in general
  - Numerical solutions for D-optimal design in our paper
Lemma 1 (Separable quadratic representation, instantaneous effect only). Suppose $X_i$ and $u_i$ are centered (i.e., $\sum_i X_i = 0$ and $\sum_i u_i = 0$) and orthogonal to each other. Under mild regularity conditions, $\text{trace}(\text{Prec}(\hat{\tau}_0, \ldots, \hat{\tau}_\ell))$ takes the form of

$$\text{trace}(\text{Prec}(\hat{\tau}_0, \ldots, \hat{\tau}_\ell)) = \frac{N}{\sigma^2} \left\{ T - \left[ (\bar{\omega}^T \cdot P_1 \cdot \bar{\omega} + 2 \cdot \bar{b}^T \cdot \bar{\omega}) + \left( \sum_k \bar{\omega}^{x_k}^T \cdot P_1 \cdot \bar{\omega}^{x_k} \right) + \left( \frac{1}{N} \cdot \bar{z}^T \cdot M_u \cdot \bar{z} \right) \right] \right\}$$

where $\bar{z} \in \{-1, +1\}^{NT}$, $\bar{\omega}$ and $\bar{b} \in [-1, +1]^T$ with $\omega_t = \frac{1}{N} \sum_i z_{it}$ and $b_t = \frac{T+1-2t}{T}$, $\bar{\omega}^{x_k} \in R^T$ with $\omega_t^{x_k} = \frac{1}{N} \sum_i X_{ik} \cdot z_{it}$, $P_1 = I_T - \frac{1}{T} \cdot 1_1^T \cdot 1_1^T$, and $M_u = P_1 \otimes U (I_{d_u} + U^T \cdot U)^{-1} U^T$.

- $\text{max} \ \text{trace}(\text{Prec}(\hat{\tau}_0, \ldots, \hat{\tau}_\ell))$ is equivalent to minimizing the sum of three quadratic functions $f_1(\bar{z}) + f_X(\bar{z}) + f_U(\bar{z})$
- With lagged effects, the decomposition is similar but the expressions of $f_1(\bar{z})$, $f_X(\bar{z})$, and $f_U(\bar{z})$ are more complicated
Optimal solution

• If there is a treatment design \( \hat{Z} \) that simultaneously minimizes \( f_1(\hat{Z}), f_X(\hat{Z}), \) and \( f_U(\hat{Z}), \) then \( \hat{Z} \) maximizes \( \text{trace}(\text{Prec}(\hat{\tau}_0, \ldots, \hat{\tau}_\ell)) \)

• Let us find the optimal solution of \( f_1(\hat{Z}), f_X(\hat{Z}), \) and \( f_U(\hat{Z}) \)
Optimal solution of $f_1(\vec{z})$

**Theorem 1** (Part A: minimize $f_1(\vec{z})$). Suppose the assumptions in Lemma 1 hold. Any treatment design that satisfies $\omega_t = \omega_{\ell,t}^*$ minimizes $f_1(\vec{z})$. If $\ell = 0$, then $\omega_{\ell,t}^* = \frac{2t-1-T}{T}$. If $\ell = 1$, then $\omega_{\ell,t}^* = -1 + \frac{2(t-1)}{T-1}$. We provide the closed-form expression of $\omega_{\ell,t}^*$ for general $\ell$ in our paper.
Optimal solution of $f_X(\vec{z})$

**Theorem 1** (Part B: minimize $f_X(\vec{z})$). Suppose the assumptions in Lemma 1 hold. Any treatment design that satisfies
\[
\frac{1}{N} \sum_{i} X_i \cdot z_{it}
\]
to be the same for all $t$ minimizes $f_X(\vec{z})$.

- If $X_i \in \{x_1, x_2, \cdots, x_G\}$ and $G$ is finite,
  - Stratum ($O_g = \{i: X_i = x_g\}$): group of units with observed covariate value
  - Stratification minimizes $f_X(\vec{z})$
    - Stratification: Treatment design whose $\sum_{t \in O_g} z_{it}$ is the same across all strata $O_g$
Theorem 1 (Part C: minimize $f_U(\bar{z})$). Suppose the assumptions in Lemma 1 hold. Any treatment design that satisfies
\[ \frac{1}{N} \sum_i U_i \cdot z_{it} \] to be the same for all $t$ minimizes $f_U(\bar{z})$.

- If $U_i \in \{u_1, u_2, \ldots, u_G\}$ and $G$ is finite,
  - Stratum ($O_g = \{i: u_i = u_g\}$): group of units with latent covariate value
  - Stratification minimizes $f_X(\bar{z})$
    - Stratification: Treatment design whose $\sum_{t \in O_g} z_{it}$ is the same across all strata $O_g$
Combining $f_1(\mathbf{z})$, $f_X(\mathbf{z})$, and $f_U(\mathbf{z})$ together

**Theorem 1** (Part A: minimize $f_1(\mathbf{z})$). Suppose the assumptions in Lemma 1 hold. Any treatment design that satisfies

$$\omega_t = \omega^*_t, \text{ and } \frac{1}{N} \sum_i \left[ \begin{array}{c} X_i \\ U_i \end{array} \right] \cdot z_{it}$$

to be the same for all $t$ maximizes $\text{trace}(\text{Prec}(\hat{\tau}_0, ..., \hat{\tau}_\ell))$.

- If $(X_i, U_i) \in \{(x_1, u_1), (x_2, u_2), \ldots, (x_G, u_G)\}$ and $G$ is finite,
  - Stratum $O_g = \{i: X_i = x_g, U_i = u_g\}$: group of units with observed and latent covariate value
  - Stratification maximizes $\text{trace}(\text{Prec}(\hat{\tau}_0, ..., \hat{\tau}_\ell))$
    - Stratification: Treatment design that satisfies $\sum_{i \in O_g} z_{it} = \omega^*_{i,t}$ for all $O_g$

- In practice, we do not know the value of $U_i$
  - Historical control data has information about $U_i$
    - Estimate $U_i$ on historical control data and use clustering algorithms to “stratify”
    - We propose a data-driven local search algorithm to improve the treatment design based on historical control data

\[
\begin{array}{c|c|c}
\text{ } & -1 & 1 \\
-1 & -1 & 1 \\
-1 & 1 & 1 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{ } & -1 & -1 \\
-1 & 1 & 1 \\
-1 & 1 & 1 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
\text{ } & -1 & 1 \\
-1 & 1 & 1 \\
\end{array}
\]
Empirical application

• MarketScan medical claims databases
  • Inpatient and outpatient claim records from early 2007 to mid 2017
  • Primary diagnosis is influenza according to ICD-9-CM diagnosis codes
  • 21,277 inpatient admissions and 9,678,572 outpatient records for influenza

• Study effect of interventions (e.g., face cover, social distancing, vaccine) on flu rate
  • Aggregate at the Metropolitan Statistical Area (MSA) level and month
  • Focus on the flu peak season (October to April)

• Other applications in our paper: Medical home visits, grocery expenditure, Lending Club loans
Empirical application: Comparison

- Design matrices

\[
\begin{array}{cccc}
-1 & -1 & 1 & 1 \\
-1 & -1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
\end{array}
\]

- Design matrices

\[
\begin{array}{cccc}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
\end{array}
\]

- Design matrices

\[
\begin{array}{cccc}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
\end{array}
\]

- Design matrices

\[
\begin{array}{cccc}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
\end{array}
\]

\(Z_{ff}\)
50% control 50%
treated at every
time period

\(Z_{ba}\)
First half control
second half all
treated

\(Z_{ffba}\)
First half all
control second
half half-treated

\(Z_{opt}\)
Nonlinear staggered design: \(\omega^*_t\)
Results for flu data

- Apply a synthetic treatment to the observed data, denoted by $Y_{it}(0)$, based on a specific $Z$ assuming $\ell = 2$

$$Y_{it} = Y_{it}(0) + \tau_0 z_{it} + \tau_1 z_{i,t-1} + \tau_2 z_{i,t-2}$$

- Estimate $\tau_0$, $\tau_1$, and $\tau_2$ from $Y_{it}$ using GLS, and compare $\sum_{j=0}^{2}(\hat{\tau}_j - \tau_j)^2$ from the data generated by various $Z$

  - $Z_{opt}$ requires fewer than 50% units to achieve the same estimation error as $Z_{ff}, Z_{ba}, Z_{ffba}$
  - Save experimental cost by at least 50%!

![Graph showing $\sum_j(\hat{\tau}_j - \tau_j)^2$ vs. $N$ for different designs]

$T = 7$
Empirical application: Comparison

• Design matrices

- Design matrices
  - $\mathbf{Z}_{\text{opt}}$: Nonlinear staggered design with $\omega_t^*$
  - $\mathbf{Z}_{\text{opt,linear}}$: Linear staggered design (optimal when $\ell = 0$)
  - $\mathbf{Z}_{\text{opt,stratified}}$: Nonlinear staggered design with $\omega_t^*$ and $k$-group stratification using historical data
Results for flu data

- Non-linear staggered design with stratification $Z_{\text{opt, stratified}}$ requires fewer than 80% units to achieve the same estimation error as $Z_{\text{opt}}$ and $Z_{\text{opt, linear}}$.
- Save experimental cost by more than 20% by using historical data and stratification!
Sequential experiments
Two challenges in sequential experiments

• Challenge 1: Peeking challenge
  • Treatment effect may be overestimated by chance early in the experiment, leading to the premature termination of the experiment

• Challenge 2: Infeasibility to optimize treatment times pre-experimentation
  • \( \omega_{\ell,t}^* \) depends on \( T \) that is unknown pre-experimentation
    • Recall Theorem 1, \( \omega_{\ell,t}^* = \frac{2t-1-T}{T} \) for \( \ell = 0 \) and \( \omega_{\ell,t}^* = -1 + \frac{2(t-1)}{T-1} \) for \( \ell = 1 \)
Potential solutions for the first challenge

- **Challenge 1: Peeking challenge**
  - Solution 1: Correction for the estimated effect size (e.g., Liu and Hall, 1999)
  - Solution 2: A better design of stopping rules (e.g., Wei, 1977)
  - Solution 3: Construction of valid $p$-values (e.g., Johari et al, 2017)

- **Our Precision-Guided Adaptive Experimentation (PGAE) algorithm**
  - Associated with Solution 2: Precision-based stopping rules (Chow and Robbins 1965, Glynn and Whitt 1992, Singham and Schruben 2012)
  - Valid statistical inference using sample-splitting idea from machine learning
Potential solutions for the second challenge

- Challenge 2: Infeasibility to optimize treatment times pre-experimentation
  - Solution: Adaptive treatment designs (Bhat, Farias, Moallemi, Sinha 2019)

- Our Precision-Guided Adaptive Experimentation (PGAE) algorithm
  - Update our belief about $T$ via Bayesian statistics
  - Adaptive treatment decisions based on our belief about $T$ via dynamic programming
PGAE

- PGAE has five main components

- We provide theoretical guarantees for PGAE
  - Asymptotic properties of PGAE: consistency, asymptotic normality, and efficiency
  - Our approach to construct belief is justified

- Model: $Y_{it} = \alpha_i + \beta_t + \tau \cdot z_{it} + \varepsilon_{it}$
  - From Lemma 1, $\text{Prec}(\hat{\tau}) = \frac{N}{\sigma_{\varepsilon}^2} \{T - f_1(\tilde{z})\}$
  - Equivalent to learn $\sigma_{\varepsilon}^2$
Key components in PGAE

- Component 1: Partition units into static treatment units (STU) and adaptive treatment units (ATU)

Data

STU $S_{sta}$

ATU $S_{ada,1}$

ATU $S_{ada,2}$

Treatment design is set pre-experimentation
Update our belief about $T$

$S_{sta}$ is a small set

Treatment design is adaptively chosen
Key components in PGAE

• Component 2: Update our belief about $T$

1. Estimate $\sigma^2 = E[\varepsilon_i^2]$ and $\xi^2 = E[(\varepsilon_i^2 - \sigma^2)^2]$.

2. Approximate our belief about $\sigma^2$ by a normal distribution centered at $\hat{\sigma}^2$ and scaled by a function of $\hat{\xi}$, $\hat{\sigma}$, and $t$.

3. Map the belief about $\sigma^2$ to the belief about $T$ by Monte Carlo simulations, denoted by $P_t(\bar{T})$. 
Key components in PGAE

- Component 3: Treatment decisions for ATU based on our belief about \( T \) via dynamic programming

\[
\omega_{ada,t+1} = \arg \min_{\omega_{t+1} : \omega_{ada,t} \leq \omega_{t+1}} \mathbb{E}_{T \sim P_t}[f_1(\bar{z})] \\
\omega_{ada,s} = \frac{1}{N} \sum_i z_{is}, \text{ for } s \leq t \\
\omega_{t+1} = \frac{1}{N} \sum_i z_{i,t+1}
\]
Key components in PGAE

• Component 4: Experiment termination

1. Estimate $\sigma^2$, denoted by $\sigma^2_{ada,1}$
2. Use $\sigma^2_{ada,1}$ to estimate precision
3. If the estimated precision is larger than the threshold, terminate the experiment. Otherwise, keep running the experiment.
Key components in PGAE

• Component 5: Statistical inference post experimentation

Data → STU $S_{sta}$ → ATU $S_{ada,1}$ → ATU $S_{ada,2}$ → Estimate $\tau$ → Construct confidence interval, or statistical test

Estimate $\sigma_\epsilon^2$
Asymptotic properties of PGAE

Theorem 2 (PGAE). Suppose we use PGAE to the sequential experiment and our experiment is run for $T$ periods in total. For a finite $T$ and $p_{sta} = \frac{|S_{sta}|}{N} < 1$, as $N \to \infty$,

$$
\sqrt{NT} \begin{pmatrix}
\hat{\tau}_{all} \\
\hat{\sigma}_{\hat{\sigma}_{ada,1}}^2 \\
\hat{\sigma}_{\hat{\sigma}_{ada,2}}^2
\end{pmatrix} \xrightarrow{d} N \left( \begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix}, \begin{pmatrix}
\sigma_{\hat{\sigma}}^2 \cdot \left(1 - \frac{f_1(\tilde{z})}{T}\right)^{-1} & 0 \\
0 & 0.5(1 - p_{sta}) \cdot \left(\hat{\xi}_\hat{\sigma}^2 + \frac{1}{(T - 1)\sigma_{\hat{\sigma}}^2}\right)
\end{pmatrix} \right)
$$

- The stopping rule that depends on $\hat{\sigma}_{\hat{\sigma}_{ada,1}}^2$ is asymptotically independent of $\hat{\tau}_{all}$ and $\hat{\sigma}_{\hat{\sigma}_{ada,2}}^2$, and therefore independent of the statistical inference based on $\hat{\tau}_{all}$ and $\hat{\sigma}_{\hat{\sigma}_{ada,2}}$.

- $\hat{\tau}_{all}$ is as efficient as the treatment effect estimator from the fixed-sample-size experiment with treatment design $\tilde{z}$.

- $\hat{\sigma}_{\hat{\sigma}}^2$ is asymptotically normal, justifying the normal approximation of our belief about $\sigma_{\hat{\sigma}}^2$. 

Optimal Design of Spatiotemporal Experiments

3/2/22
Empirical application: Results for flu data

- Adaptive design from PGAE reduces errors by 20% compared to benchmark ($Z_{\text{opt}}$ with $T_{\text{max}}$) and oracle ($Z_{\text{opt}}$ with $T^*$)
  - Could potentially save experimental cost by 20%

![Graph showing comparison of design, benchmark, oracle, and adaptive methods over different $T_{\text{max}}$ values.](image_url)
Conclusion

• Panel data experiments
  • Fixed-sample-size experiments
    • Provide provably near-optimal analytical solutions
    • Could reduce experimental cost by more than 60%

• Sequential experiments
  • Propose the Precision-Guided Adaptive Experiment (PGAE) algorithm for adaptive treatment design and post-experimentation inference
    • Combines ideas from Bayesian statistics, dynamic programming, and sample splitting
  • Could further reduce experimental cost by more than 20%
Supplementary slides
Key components in PGAE

• Component 2: Update our belief about $T$

1. Estimate $\sigma_\xi^2 = E[\varepsilon_1^2]$ and $\xi_\xi = E\left[(\varepsilon_1^2 - \sigma_\xi^2)^2\right]$

2. Normal approximation of the belief about $\sigma_\xi^2$
   
   $\sigma_\xi^2 \sim N(\hat{\sigma}_\xi^2, (\hat{\sigma}_\xi^2 + \hat{\sigma}_\xi^2/(t - 1))/(\text{S}_8/9 \cdot t))$

3. Map the belief about $\sigma_\xi^2$ to the belief about $T$ by Monte Carlo simulations, denoted by $P_t(\hat{T})$
Thank you!