CAUSAL INFERENCE AND EXPERIMENTAL DESIGN IN TWO-SIDED MARKETPLACES

University of North Carolina at Chapel Hill

Hongtu Zhu

All works were done when I worked at DiDi. Joint works with Jieping Ye, Fang Yao, Sikai Luo, Chengchun Shi, Ying Yang, Ting Li, Zhaohua Lu, and Yi Li among others.

Declaration of
Financial Interests or Relationships

Speaker Name: Hongtu Zhu

I have the following financial interest or relationship to disclose with regard to the subject matter of this presentation:

Company Name: DiDi Chuxing

Type of Relationship: Chief Scientist and Consultant

4.2018-3.2022

CONTENTS

Part I

Two-sided Marketplace

What is a Two-Sided Market?

Rochet –Tirole
2006

Two-sided markets are roughly defined as markets where one or several platforms enable interactions between end-users, and try to get the two (or multiple) sides "on board" by appropriately pricing each side.

Alvin E. Roth
Nobel Memorial Prize in Economic
@ SIGKDD 2018, 08/2018

"— In many markets, you care who you are dealing with, and prices don't do all the work

— (In some matching markets, we don't even let prices do any of the work...) "

UNC Biostatistics

Examples of Two-Sided Market

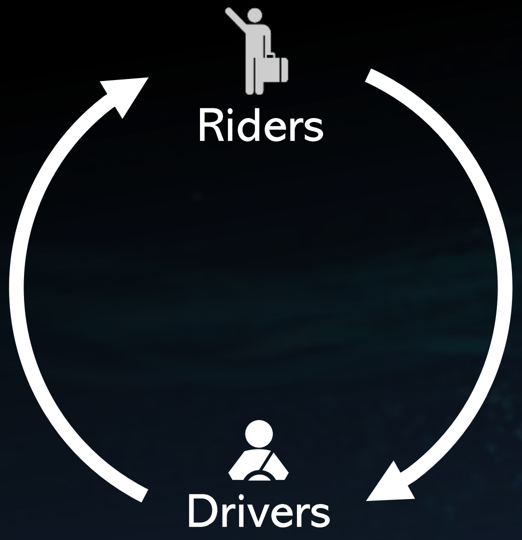
Networked Market	airbnb	ebay	amazon mechanical turk beta	DiDi Uber
Side 1	Hosts	Retailers	Organizations	Drivers
Side 2	Travelers	Consumers	Developers	Passengers
Platform Providers	Airbnb	eBay	Amazon	Ridesharing Platform

UNC Biostatistics

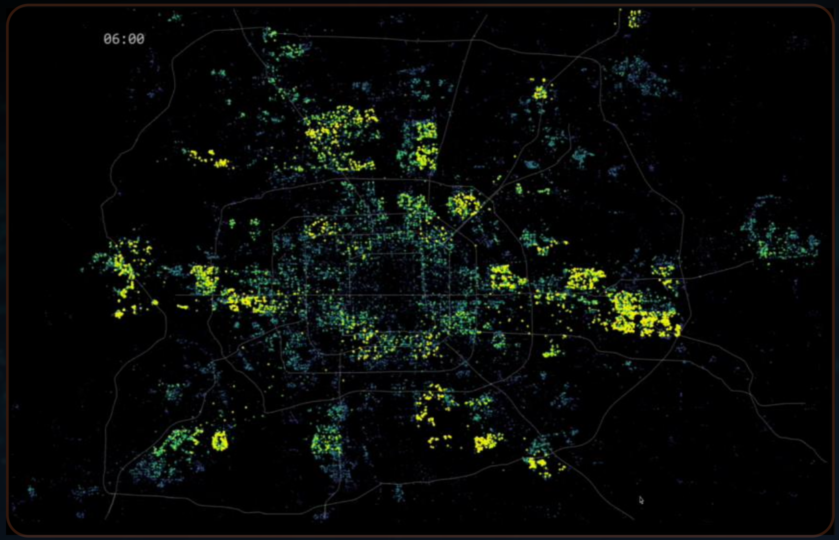
Ride-sharing Platform is a Complex Ecosystem

Spatio-temporal | Nonlinear | Interactive | Uncertainty | Causal

Two-sided Platform



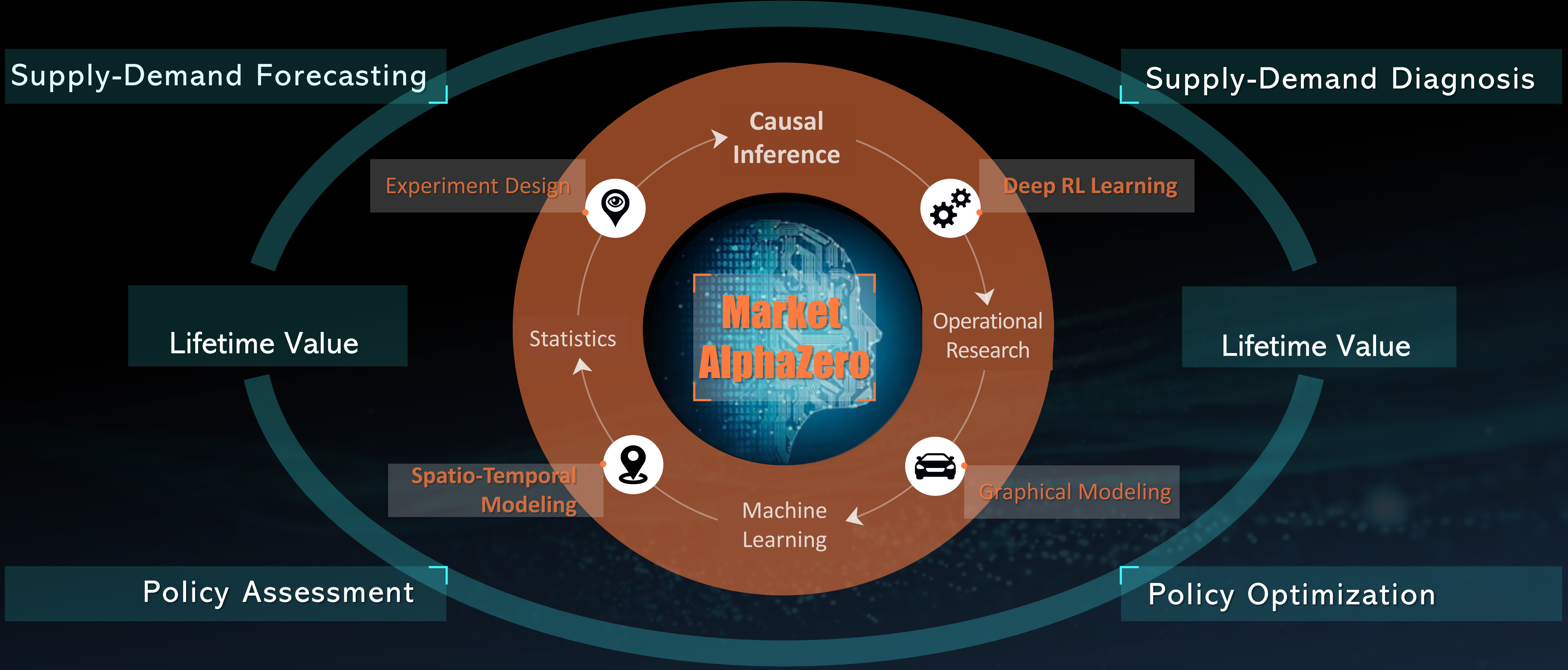
Complex Spatio-temporal System



UNC Biostatistics

Leverage Supply-Demand Network Effect

How to evaluate and improve the operational efficiency of ride-sharing platform?



UNC Biostatistics

Policy Evaluation

A/B Testing

Comparison btw new & old policies in spatio-temporal system

- How to design the experiments (or spatio-temporal units)?
- How to measure the treatment effects?

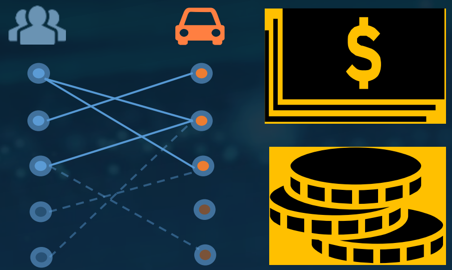
Challenges

Interference

Oversupply Regions

Large variation of key metrics

Interaction of Multiple Policies



UNC Biostatistics

Part II

Causal Inference and Experimental Design

Policy Evaluation for Temporal and /or Spatial Dependent Experiments

S Luo, Y Yang, C Shi, F Yao, J Ye, H Zhu. Policy Evaluation for Temporal and/or Spatial Dependent Experiments *JRSSB*, in press.

Policy evaluation

A/B Testing

Comparison btw new & old policies in spatio-temporal system

- Evaluating treatment effects
- Improve key platform metrics
- Exploring order dispatch policies and customer recommendation initiatives
- Leading to a more efficient and user-friendly transportation system

The Goal

Improve the service quality

Drivers

- Reduce empty driving

Riders

- Intelligent travel guidance
- Less queueing time

Platform

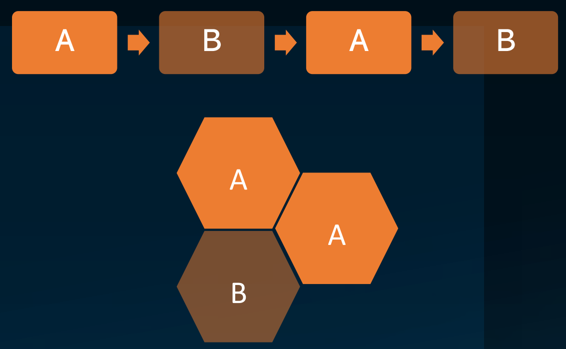
- Recognize the market
- Better dispatching and scheduling

Datasets

Data Description

Datasets collected from Didi Chuxing

- A time-dependent A/B experiment from 2021.12.1 to 2021.12.23， each day was divided 24 time intervals
- A new order dispatching policy aimed to increase the number of fulfilled ride requests and boost drivers’ total revenue
- A spatio-temporal dependent A/B experiment from 2020.2.19 to 2020.3.12, each day is divided into 48 time intervals
- A time-dependent A/A experiment from 2021.7.13-2021.9.17



Collected Data

Outcomes of Interests

- Drivers’ total income
- Answer Rate
- Completion Rate

Demand and Supply

Supply

Demand

- Drivers’ total online time
- Number of call orders

Treatment Effect Evaluation

Challenges

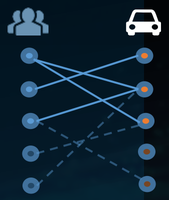
Statistical Challenges under the switchback design

Data Generating Process

- Non-stationary
- Complex spatio-temporal patterns

Spatio-temporal random effects

- Supply and demand
- Violation of conditional independence



Interference

- Over time
- Over Space

Goal

Analyze the causal relationship between Platform policies and outcomes

Dynamic treatment effects

- Capture the random and interference over time and space

Temporal carryover effects

- Model market features as mediators

Spatial spillover effects

- Employ mean field approximation

Related Work

Causal Inference under Interference

- Four major types of models for the interference processes
  - Assume specific structure models to restrict the interference process (Lee, 2007)
  - The partial interference assumption (Sobel, 2006; Halloran and Hudgens, 2016; Pollmann,2020)
  - The local or network-based interference assumption (Bakshy et al., 2014; Aronow et al. 2020)
  - Capture the interference effect via congestion or price effects (Johari et al., 2022)

✓ Most aforementioned works studied the interference effect across time or space
✓ They were motivated by research questions in environmental and epidemiological studies
✓ It remains unknown about their generalization to ride-sharing markets

Off-policy Evaluation

- Augmented inverse propensity score weighting methods for valid OPE (Zhang et al., 2013; Jiang and Li, 2016)
- Efficient model-free OPE under the Markov decision process model assumption (Kallus and Uehara, 2020; Liao et al., 2021, 2022)

✓ The AIPW methods suffer from the curse of horizon
✓ The MDP model assumption excludes the existence of random effects and is typically violated in our application

ATE in Temporal dependent Experiments

Average Treatment Effect (ATE)

$$ATE = \sum_{t=1}^{m} E\{Y_t^*(1_t) - Y_t^*(0_t)\}$$

- Each day is divided into m intervals
- A_t: the policy implemented at tth interval
- S_t: state variables measured at tth interval

- Y_t: the outcome of interest measured at time t
- $\bar{a}_t = (a_1, \dots, a_t)^\top \in \{0,1\}^t$, the treatment history up to t
- $S_t^*\ (\bar{a}_{t-1})$ and $Y_t^*\ (\bar{a}_t)$ as the counterfactual state and outcome

Decomposition

- Conditional mean of the outcome given the data history

$$E\{Y_t^*(\bar{a}_t)|S_t^*(\bar{a}_{t-1},), Y_{t-1}^*(\bar{a}_{t-1}), \dots, S_1\} = R_t(a_t, S_t^*(\bar{a}_{t-1}), a_{t-1}, S_{t-1}^*(\bar{a}_{t-2}), \dots, S_1)$$

$$ATE = \sum_{t=1}^{m} E\{R_t(1, S_t^*(1_{t-1}), 1, S_{t-1}^*(1_{t-2}), \dots, S_1) - R_t(0_t, S_t^*(0_{t-1}), 0_{t-1}, S_{t-1}^*(0_{t-2}, \dots, S_1)\}$$

$$= \sum_{t=1}^{m} E\{R_t(1, S_t^*(0_{t-1}), 0, S_{t-1}^*(0_{t-2}), \dots, S_1) - R_t(0_t, S_t^*(0_{t-1}), 0_{t-1}, S_{t-1}^*(0_{t-2}, \dots, S_1)\}$$ → **DE**

$$+ \sum_{t=1}^{m} E\{R_t(1, S_t^*(1_{t-1}), 1, S_{t-1}^*(1_{t-2}), \dots, S_1) - R_t(1, S_t^*(0_{t-1}), 0, S_{t-1}^*(0_{t-2}), \dots, S_1)\}$$ → **IE**

Identification

Problem of interest

- Estimate ATE for the spatio-temporal dependent switchback experiments
- Test the following hypotheses

$$H_0^{DE}: \mathrm{DE} \le 0, \quad v.s. \quad H_1^{DE}: \mathrm{DE} > 0$$

$$H_0^{IE}: \mathrm{IE} \le 0, \quad v.s. \quad H_1^{IE}: \mathrm{IE} > 0$$

Estimable from the data

Lemma 1: Under consistency assumption, the sequential randomization assumption and the positivity assumption, the causal estimand can be represented as a function of the observed data such that

$$R_t(a_t, s_t, \dots, s_1) = E(Y_t | A_t = a_t, S_t = s_t, \dots, S_1 = s_1),$$

$$E\{R_t(a_t, S_t, \dots, S_1)\} = E\left[E\left(R_t(a_t, S_t, \dots, S_1) \middle| \{A_j = a_j\}_{1 \le j \le t}, \{S_j, Y_j\}_{1 \le j \le t}\right)\right],$$

TVCDP Models

Temporal VCDP

$$Y_{i,t} = f_{1,t}(Z_{i,t}) + e_{i,t} \qquad S_{i,t+1} = f_{2,t}(Z_{i,t}) + \varepsilon_{i,tS}$$

- Current State-Action pair $Z_{i,t} = (S_{i,t}^{\top}, A_{i,t})^{\top}$

Example: L-TVCDP

$$Y_{i,t} = \beta_0(t) + S_{i,t}^{\top}\beta(t) + A_{i,t}\gamma(t) + e_{i,t} = Z_{i,t}^{\top}\theta(t) + e_{i,t}$$

$$S_{i,t+1} = \phi_0(t) + \Phi(t)S_{i,t} + A_{i,t}\Gamma(t) + +\varepsilon_{i,tS} = \Theta(t)Z_{i,t} + \varepsilon_{i,tS}$$

$$\boldsymbol{DE} = \sum_{t=1}^{m} \gamma(t,\tau)$$

$$\boldsymbol{IE} = \sum_{t}^{m} \beta(t,\tau)^{\top} \left\{ \sum_{k=1}^{t-1} [\prod_{l=k+1}^{t-1} \Phi(l)]\Gamma(k) \right\}$$

Example: NN-TVCDP

$$Y_{i,t} = g_0(t, S_{i,t})I(A_{i,t} = 0) + g_1(t, S_{i,t})I(A_{i,t} = 1) + e_{i,t}$$

$$S_{i,t+1} = G_0(t, S_{i,t})I(A_{i,t} = 0) + G_1(t, S_{i,t})I(A_{i,t} = 1) + \varepsilon_{i,tS}$$

- g_0, g_1, G_0 and G_1 are parametrized via some (deep) neural networks

$$\boldsymbol{DE} = \sum_{t=1}^{m} E\{g_1(t, S_t^0) - g_0(t, S_t^0)\}$$

$$\mathbf{IE} = \sum_{t=1}^{m} E\{g_1(t, S_t^1) - g_0(t, S_t^0)\}$$

$$S_t^0 = G_0(t-1, S_{t-1}^0), S_t^1 = G_0(t-1, S_{t-1}^1)$$

Estimation and Testing in L-TVCDP

Inference of DE

$$
\begin{aligned}
Y_{i,t} &= \beta_0(t) + S_{i,t}^{\top}\beta(t) + A_{i,t}\gamma(t) + e_{i,t} = Z_{i,t}^{\top}\theta(t) + e_{i,t} \\
S_{i,t+1} &= \phi_0(t) + \Phi(t)S_{i,t} + A_{i,t}\Gamma(t) + +\varepsilon_{i,tS} = \Theta(t)Z_{i,t} + +\varepsilon_{i,tS}
\end{aligned}
$$

$$
\boldsymbol{DE} = \sum_{t=1}^{m} \boldsymbol{\gamma}(\boldsymbol{t}, \boldsymbol{\tau})
$$

Estimation and Testing Algorithm

- Compute the OLS estimator $\hat{\theta}$
- Employ kernel smoothing to compute a refined estimator $\tilde{\theta} = \Omega\hat{\theta}$ and obtain $\widehat{DE}$
- Estimate the variance of $\hat{\theta}$ by the sandwich estimator
- Estimate the variance of $\tilde{\theta}$ by $\tilde{V}_{\theta} = \Omega\hat{V}_{\theta}\Omega^{\top}$ and compute $\widehat{se}(\widehat{DE})$
- Reject H_0^{DE} if $\widehat{DE}/\widehat{se}(\widehat{DE})$ exceeds the upper αth quantile of $N(0,1)$

Estimation and Testing in L-TVCDP

Inference of IE

$$
\begin{aligned}
Y_{i,t} &= \beta_0(t) + S_{i,t}^\top \beta(t) + A_{i,t}\gamma(t) + e_{i,t} = Z_{i,t}^\top \theta(t) + e_{i,t} \\
S_{i,t+1} &= \phi_0(t) + \Phi(t) S_{i,t} + A_{i,t}\Gamma(t) + +\varepsilon_{i,tS} = \Theta(t) Z_{i,t} + \varepsilon_{i,tS}
\end{aligned}
$$

$$
\boxed{\boldsymbol{IE} = \sum_{t}^{m} \boldsymbol{\beta}(\boldsymbol{t},\boldsymbol{\tau})^\top \left\{ \sum_{k=1}^{t-1} \left[\prod_{l=k+1}^{t-1} \boldsymbol{\Phi}(\boldsymbol{l}) \right] \boldsymbol{\Gamma}(\boldsymbol{k}) \right\}}
$$

Estimation and Bootstrap-based Testing Algorithm

- Compute the OLS estimator $\widehat{\Theta} = \left(\widehat{\Theta}(1), \dots, \widehat{\Theta}(m-1)\right)^\top$
- Compute the refined estimator $\widetilde{\Theta} = \Omega\widehat{\Theta}$and obtain $\widehat{IE}$
- Compute the estimated residual $\hat{\varepsilon}_{i,tS} = S_{i,t+1} - \widetilde{\Theta}(t) Z_{i,t}$
- For $b = 1, \dots B$

  Generate i.i.d. standard normal variable $\{\xi_i^b\}_{i=1}^n$

  Generate pseudo outcomes

  Use the pseudo outcomes to compute $\widehat{IE}^\mathrm{b}$
- Reject H_0^{IE} if $\widehat{IE}$ exceeds the upper αth quantile of $\{\widehat{IE}^\mathrm{b} - \widehat{IE}\}_b$

$$
\boxed{
\begin{aligned}
\hat{S}_{i,t+1} &= \widetilde{\Theta}(t) \tilde{Z}_{i,t} + \xi_i \hat{\varepsilon}_{i,tS} \\
\hat{Y}_{i,t} &= \hat{Z}_{i,t}^\top \tilde{\theta}(t) + \xi_i \hat{e}_{i,t}
\end{aligned}
}
$$

Estimation in NN-TVCDP

Inference of DE

$$Y_{i,t} = g_0(t, S_{i,t})I(A_{i,t} = 0) + g_1(t, S_{i,t})I(A_{i,t} = 1) + e_{i,t}$$

$$S_{i,t+1} = G_0(t, S_{i,t})I(A_{i,t} = 0) + G_1(t, S_{i,t})I(A_{i,t} = 1) + \varepsilon_{i,tS}$$

Estimation and Testing Algorithm

- Use neural networks to obtain $\hat{g}_0, \hat{g}_1, \hat{G}_0$ and $\hat{G}_1$
- Employ the residual $\hat{\varepsilon}_{i,tS}$ and compute the density function estimator $\hat{f}_{\varepsilon_{tS}}$
- Use Monter Carlo $(\mathrm{k} = 1, \ldots, M)$ to estimate the distribution of $S^*_{i,t}(1_{t-1})$ and $S^*_{i,t}(0_{t-1})$ conditional on $S_{i,1}$
- Obtain the estimator

$$\boldsymbol{DE} = \frac{\mathbf{1}}{\mathbf{nM}}\sum_{i=1}^{n}\sum_{k=1}^{M}\sum_{t=1}^{m}\hat{g}_1(t, \hat{S}^0_{i,k,t}) - \hat{g}_0(t, \hat{S}^0_{i,k,t})\}$$

$$\boldsymbol{IE} = \frac{\mathbf{1}}{\mathbf{nM}}\sum_{i=1}^{n}\sum_{k=1}^{M}\sum_{t=1}^{m}\hat{g}_1(t, \hat{S}^1_{i,k,t}) - \hat{g}_0(t, \hat{S}^0_{i,k,t})\}$$

ATE in Spatio-Temporal dependent Experiments

Average Treatment Effect (ATE)

$$ATE = \sum_{\iota=1}^{r}\sum_{t=1}^{m} E\{Y_{t,\iota}^{*}(1_{t,[1:r]}) - Y_{t,\iota}^{*}(0_{t,[1:r]})\}$$

- $\bar{a}_{t,\iota} = (a_{1,\iota}, \dots, a_{t,\iota})^{\top} \in \{0,1\}^{t}$, the treatment history up to t for the ιth region
- $\bar{a}_{t,\iota} = (\bar{a}_{t,\iota}, \dots, \bar{a}_{t,r})^{\top}$ denote the treatment history associated with all regions
- $S_{t,\iota}^{*}\ (\bar{a}_{t-1,[1:r]})$ and $Y_{t,\iota}^{*}\ (\bar{a}_{t-1,[1:r]})$ as the counterfactual state and outcome for the ιth region

Decomposition

$$DE_{st} = \sum_{\iota}^{r}\sum_{t=1}^{m} E\{R_{t,\iota}(1_{t,[1:r]}, S_t^{*}(0_{t-1,[1:r]}), 0_{t-1,[1:r]}, S_1) - R_{t,\iota}(0_{t,[1:r]}, S_t^{*}(0_{t-1,[1:r]}), 0_{t-1,[1:r]}, \dots, S_1\}$$

$$DE_{st} = \sum_{\iota=1}^{r}\sum_{t=1}^{m} E\{R_{t,\iota}(1_{t,[1:r]}, S_t^{*}(1_{t-1,t-1}), 1_{t-1,[1:r]}, S_1) - R_t(1_{t,[1:r]}, S_t^{*}(0_{t-1,\iota}), 0_{t-1,[1:r]}, \dots, S_1))\}$$

$$H_0: DE_{\tau st} \le 0 \quad v.s. \quad DE_{\tau st} > 0$$

$$H_0: IE_{\tau st} \le 0 \quad v.s. \quad IE_{\tau st} > 0$$

Estimation and Testing in L-STVCDP

Model

$$Y_{i,t,\iota} = \beta_0(t,\iota) + S_{i,t,\iota}^{\top}\beta(t,\iota) + A_{i,t,\iota}\gamma_1(t,\iota) + \bar{A}_{i,t,N_\iota}\gamma_2(t,\iota) + e_{i,t,\iota}$$

$$S_{i,t+1,\iota} = \phi_0(t,\iota) + \Phi(t,\iota)S_{i,t,\iota} + A_{i,t,\iota}\Gamma_1(t,\iota) + \bar{A}_{i,t,N_\iota}\Gamma_2(t,\iota) + \varepsilon_{i,t,\iota}$$

$$\mathrm{DE}_{st} = \sum_{\iota=1}^{r}\sum_{\tau=1}^{m}\{\gamma_1(\tau,\iota) + \gamma_2(\tau,\iota)\}$$

Wald statistic

$$\mathrm{IE}_{st} = \sum_{\iota=1}^{r}\sum_{\tau=1}^{m}\beta(\tau,\iota)^{\top}\left\{\sum_{k=1}^{\tau-1}\left(\prod_{j=k+1}^{\tau-1}\Phi(j,\iota)\right)(\Gamma_1(k,\iota) + \Gamma_2(k,\iota))\right\}$$

Gaussian appx.
Multiplier bootstrap

Theoretical Analysis

Validity of test for DE

Theorem 1: Under suitable conditions, if the bandwidth $h = o\left(n^{-\frac{1}{4}}\right)$, and $mh \to 0$ $m \gg \sqrt{n}$, and as $n \to \infty$, then under H_0^{DE},

$$\mathrm{P}\left(\frac{\widehat{DE}}{\widehat{se}\left(\widehat{DE}\right)} > z\_\alpha\right) = \alpha + o(1),$$

It approaches to 1 under under H_0^{IE}.

Validity of test for IE

Theorem 2: Under suitable conditions, if the bandwidth $h = o\left(n^{-\frac{1}{4}}\right), m \asymp n^c$ for some $0.5 < c < 1.5$, and $mh \to 0$ as $n \to \infty$,

$$\sup_z \left|P\left(\widehat{IE} - IE \leq z\right) - P\left(\widehat{IE}^b - \widehat{IE} \middle| Data \leq z\right)\right| \leq C\left(\sqrt{n}h^2 + \sqrt{n}m + n^{-1/8}\right)$$

with probability approaching 1, for some positive constant C.

Theoretical Analysis

Switchback and alternating day design

Theorem 3: Suppose $\Sigma_e(t_1, t_2)$ is nonnegative for any t_1, t_2, then under L-TVCDP, as as $n \to \infty$

$$nMSE\left(\widehat{DE}_{sb}\right) \leq nMSE\left(\widehat{DE}_{ad}\right) + o(1).$$

If $\Sigma_e(t_1, t_2) = c\rho^{|t_1 - t_2|}$, then

$$\frac{MSE\left(\widehat{DE}_{sb}\right)}{MSE\left(\widehat{DE}_{ad}\right)} = \frac{(1-\rho)^2}{(1+\rho)^2} + o(1).$$

- The larger the ρ, the smaller the variance ratio
- When $\rho = 0.5$, MSE of DE under the switch back design is approximately 9 times smaller than that under the alternating-day design.

Real Data Based Simulation

Temporal alternative design-setting

- Simulation experiments are conducted based on two real dataset collected from the A/A experiment
- Obtain the other estimates by setting $\gamma(t,\tau) = \Gamma(t) = 0$ and obtain the estimated error processes

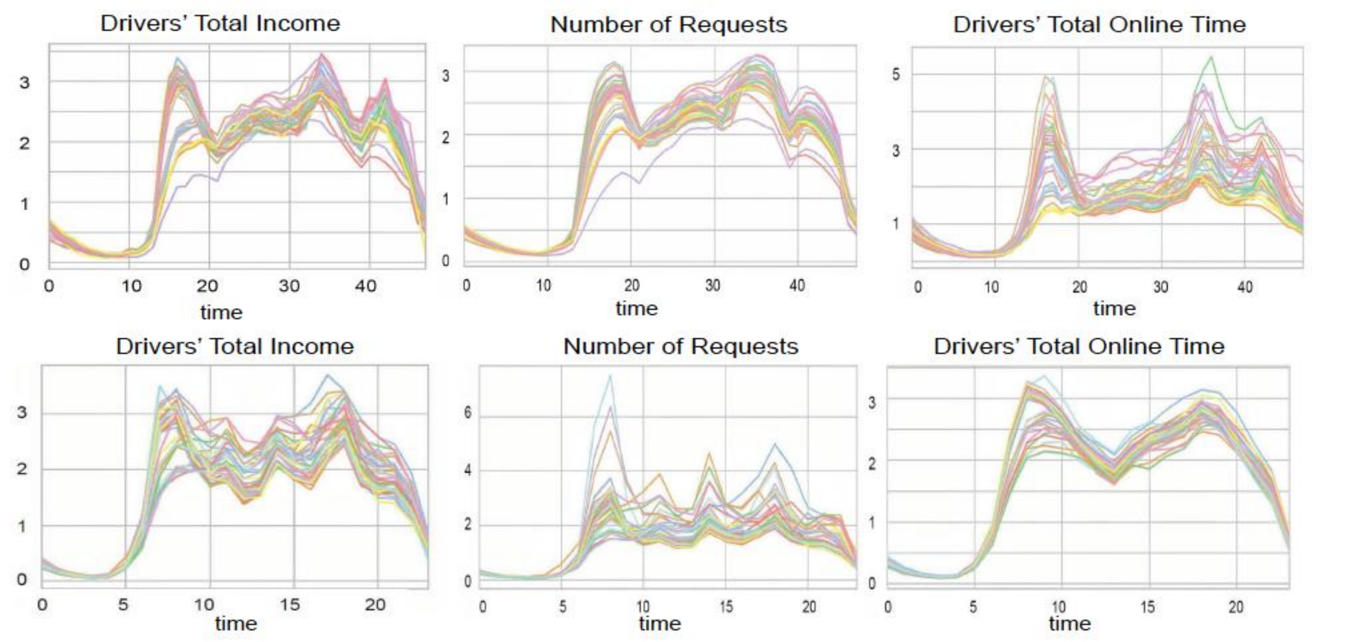


Figure 1: Scaled business metrics from City A (the first row) and City B (the second row) across 40 days, including drivers' total income, the numbers of requests and drivers' total online time.

$$\tilde{\gamma}(t,\tau) = \left(\frac{\delta}{100}\right) E(Y_t)$$

$$\tilde{\Gamma}(t,\tau) = \left(\frac{\delta}{100}\right) E(S_t)$$

Real Data Based Simulation

Temporal alternative design-results

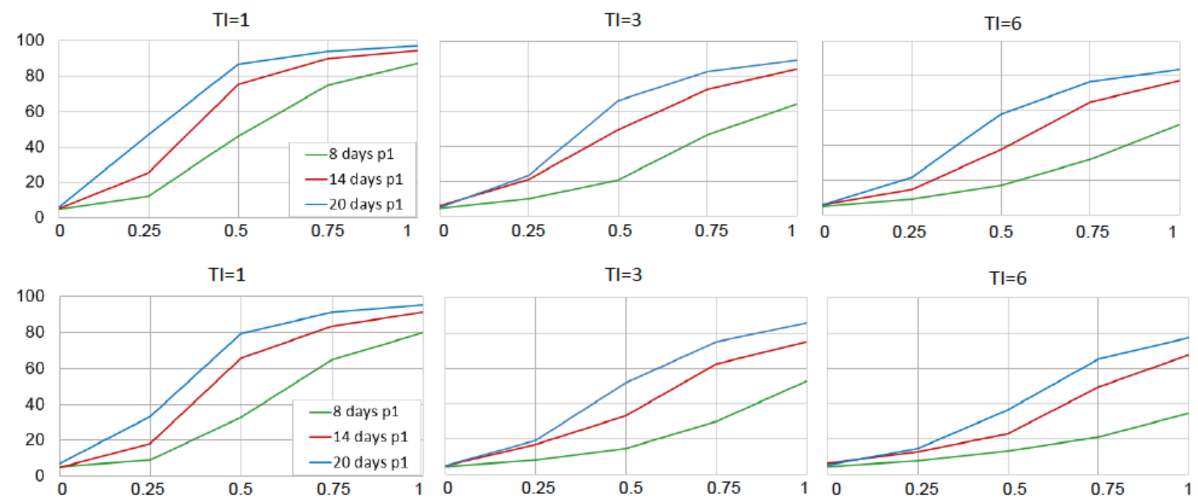


Figure 3: Simulation results for L-TVCDP: empirical rejection rates of the proposed test for IE under different combinations of (n, δ, TI). Synthetic data are simulated based on the real dataset from city A (the first row) and city B (the second row).

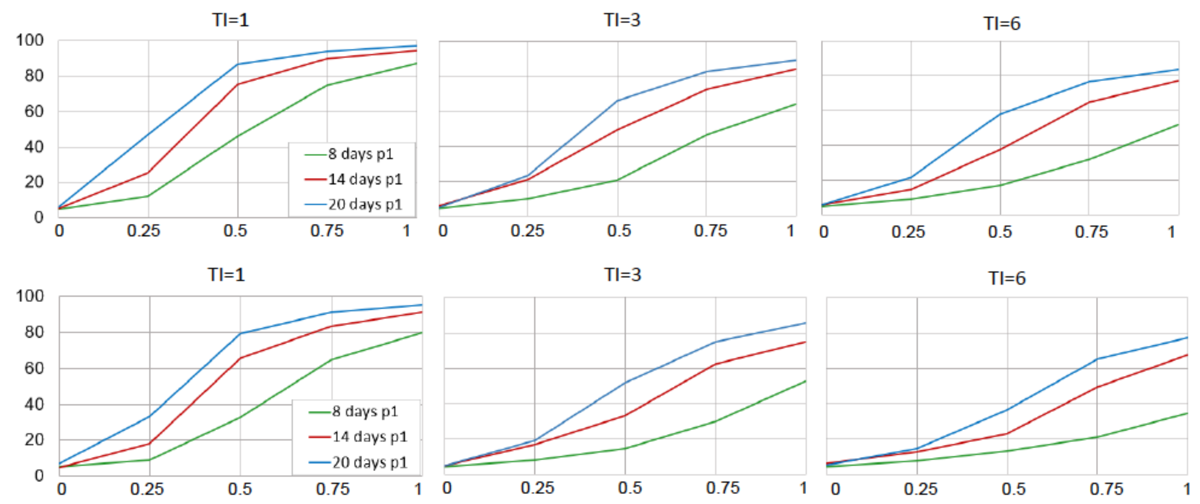


Figure 3: Simulation results for L-TVCDP: empirical rejection rates of the proposed test for IE under different combinations of (n, δ, TI). Synthetic data are simulated based on the real dataset from city A (the first row) and city B (the second row).

- The more frequently we switch back and forth between the two policies, the more powerful the resulting test

Real Data Analysis

Temporal Experiment

Table 1: One sided p-values of the proposed test for DE, when applied to eight datasets collected from the A/A or A/B experiment based on the temporal alternation design.

	AA			AB		
	DTI(%)	ART(%)	CRT(%)	DTI(%)	ART(%)	CRT(%)
S_1	0.527	0.435	0.442	0.000	0.000	0.003
S_2	0.232	0.126	0.209	0.000	0.763	0.661
S_3	0.378	0.379	0.567	0.700	0.637	0.839
S_4	0.348	0.507	0.292	0.198	0.000	0.133

Table 2: One sided p-values of the proposed test for IE, when applied to eight datasets collected from the A/A or A/B experiment based on the temporal alternation design. Drivers' total income is set to be the outcome of interest.

	S1		S2		S3		S4	
	AA	AB	AA	AB	AA	AB	AA	AB
p-value	0.334	0.001	0.341	0.003	0.254	0.589	0.427	0.168

- Four cities with policies $S_1, \dots S_4$
- policy S1 is proposed to reduce the answer time
- Both policy S2 and policy designed to reduce drivers' idle time ratio.
- S4 aims to balance drivers' downtime and their average pick-up distance.

- DTI: drivers' tatoal income
- ART: anwser rate
- CRT: completion rate

Real Data Analysis

Spatiotemporal Experiment

- The city is divided into 17 regions.
- Policies are implemented based on alternating 30-minute time intervals within each region.
- Outcome: drivers' total income
- State variable: the number of call orders

Table 3: One sided p-values of the proposed test, when applied to two datasets collected from the A/A or A/B experiment based on the spatio-temporal alternation design. Drivers' total income is set to be the outcome of interest.

	DE		IE	
	AA	AB	AA	AB
p-value	0.176	0.001	0.334	0.000

- The new policy significantly increases drivers' income.
- It fails to reject the null hypotheses for AA data

Evaluating Dynamic Conditional Quantile Treatment Effects

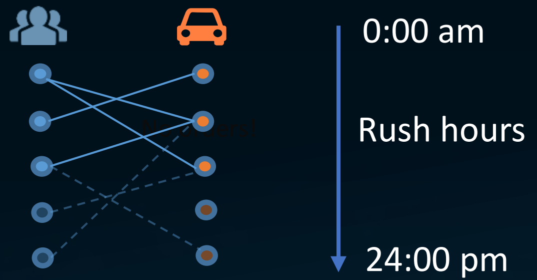
T Li, C Shi, Z Lu, Y Li, H Zhu. Evaluating Dynamic Conditional Quantile Treatment Effects with Applications in Ridesharing JASA, in revision.

Treatment Effect Evaluation

Additional Challenges

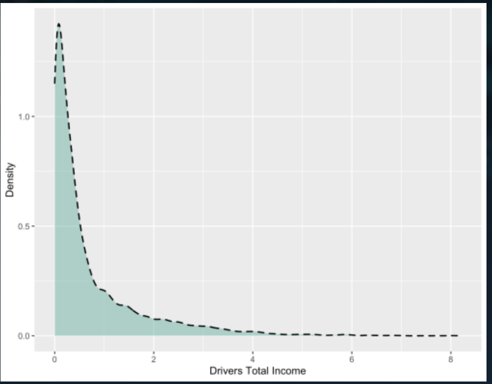
Spatio-temporal data

Non-stationary data generating process



Interference

Non-normal and heavy-tailed outcome



Datasets

Characteristics

- Long horizon, multi-stage decision making
- The treatment effect is usually weak
- Both supply and demand are spatiotemporal networks that interact across time and location
- The outcome of interest follows a non-normal and heavy-tailed distribution

Interested Questions

(Q1) How can we quantify treatment effects across various quantile levels for the time-dependent A/B experiment data in order to gain a comprehensive understanding of the new policy's effects within the city?

(Q2) How to evaluate the quantile treatment effects for the above spatio-temporal dependent experiment data?

(Q3) How to determine whether or not to replace the old policy with the new one?

Related Work

A/B testing

- Most existing A/B testing methods that focus on the Average Treatment Effect.
- Liu et al. (2019) proposed a scalable method to test QTE and construct associated confidence intervals.
- Wang and Zhang (2021) developed a nonparametric method to estimate QTEs at a continuous range of quantile locations.
- Chernozhukov and Hansen (2006), Fripo (2007) and Blanco et al. (2020) considered the estimation of (conditional) QTEs.

These methods address single-stage decision-making.

Off-policy Evaluation

- The majority of existing studies primarily concentrate on inferring the expected return under a fixed target policy or a data-dependent estimated optimal policy (Zhang et al.; 2013, Shi et al. 2020; Kallus and Uehara, 2022).
- Wang et al.(2018), Qi et al. (2022) and Xu et al. (2022) proposed using inverse probability weighted estimators to evaluate specific robust metrics under a given target policy.

These methods are subject to the curse of horizon and become less effective in long-horizon settings

Policy evaluation in spatiotemporal dependent experiments remains unexplored

CQTE in Temporal dependent Experiments

Quantile Treatment Effect (QTE)

$$QTE_\tau = Q_\tau\left(\sum_{t=1}^{m} Y_t^*(1_t)\right) - Q_\tau\left(\sum_{t}^{m} Y_t^*(0_t)\right)$$

- A_t: the policy implemented at tth interval
- S_t: state variables measured at tth interval
- Y_t: the outcome of interest measured at time t
- $\bar{a}_t = (a_1, \dots, a_t)^\top \in \{0,1\}^t$, the treatment history up to t
- $S_t^* \ (\bar{a}_{t-1})$ and $Y_t^* \ (\bar{a}_t)$ as the counterfactual state and outcome

Challenges

- Existing off-policy quantile evaluation methods are inefficient in our setting with a moderately large m.
- It is difficult to adapt methods focusing on the mean return for quantile evaluation due to the nonlinear quantile function

Conditional QTE (CQTE)

$$CQTE_\tau = Q_\tau\left(\sum_{t=1}^{m} Y_t^*(1_t)\,|\mathcal{E}_m\right) - Q_\tau\left(\sum_{t=1}^{m} Y_t^*(0_t)\,|\mathcal{E}_m\right)$$

- $\mathcal{E}_t$: the set of features that have an impact on the outcomes up to time t, but are not influenced by the treatment history
- When $m = 1$, it reduces to single –stage decision making

Benefits

- It offers a more more convenient way to estimate the dynamic Quantile Treatment Effect
- It can help to reduce the variance of the resulting QTE estimator by removing the need to account for variability in the relevant characteristics

Summed CQTE

SCQTE the sum of individual Conditional Quantile Treatment Effects (CQTE) over time

$$SCQTE_\tau = \sum_{t=1}^{m} Q_\tau(Y_t^*(1_t)|\mathcal{E}_t) - \sum_{t=1}^{m} Q_\tau(Y_t^*(0_t)|\mathcal{E}_t)$$

- Compared to CQTE, SCQTE is easier to learn from observed data
- For example, one can fit a quantile regression model at each stage, estimate individual CQTE values, and then sum these estimators together.

Proposition 1: Suppose that for any time point t, $Y_t^*(\bar{a}_t)$ follows the structural quantile model $Y_t^*(\bar{a}_t) = \phi_t(\mathcal{E}_t, \bar{a}_t, U)$ for a specific deterministic function ϕ_t and a uniformly distributed random variable $U \sim U(0,1)$,which is independent of $\{\mathcal{E}_t\}_{t=1}^m$. Furthermore, assume that $\phi_t(\mathcal{E}_t, 1_t, \tau)$ and $\phi_t(\mathcal{E}_t, 0_t, \tau)$ are strictly increasing functions of τ for any $\mathcal{E}_t$. Under these conditions, we find that

$$SCQTE_\tau = CQTE_\tau$$

Summed CQTE

Proposition 1 serves as a fundamental building block for our proposal.

- It greatly facilitates the estimation and inference procedures that follow, which rely on fitting a quantile regression model at each time point to learn the SCQTE.
- It is related to the structural quantile model in the quantile regression literature (Chernozhukov and Hansen, 2005, 2006)
- These models assume that conditional on $X = x$, the potential outcome

$$\boldsymbol{Y^*(a) = q(a, x, U), U \sim U(0,1)}$$

$\boldsymbol{q(a, x, \tau)}$ is strictly increasing in τ

- $\boldsymbol{U}$: a rank variable that characterizes the heterogeneity of the outcome across different quantile levels.
- Under the monotonicity constraint

$$\boldsymbol{Q_\tau(Y^*(a) | X = x) = q(a, x, \tau)}$$

Testing CQTE

Testing whether the treatment effect at the τth quantile is non-negative or positive

$$H_0: CQTE_\tau \leq 0 \quad v.s. \quad CQTE_\tau > 0$$

$$\Downarrow$$

$$H_0: SCQTE_\tau \leq 0 \quad v.s. \quad SCQTE_\tau > 0$$

Assumptions

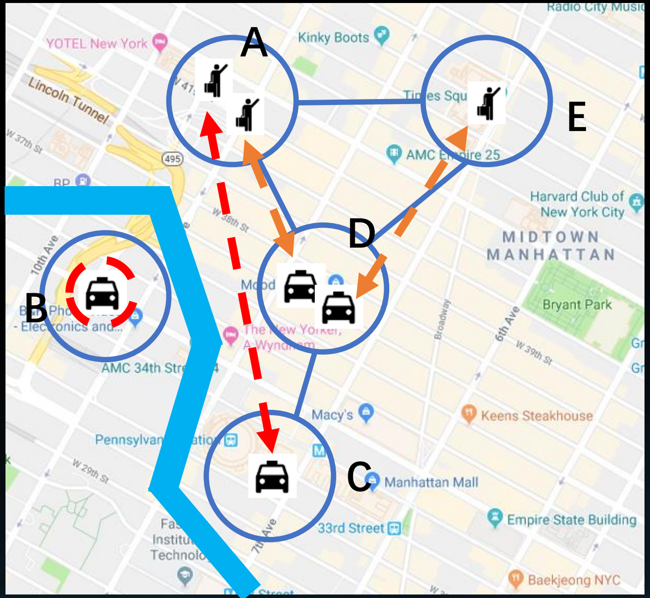
- **Consistency Assumption**
  the potential state and outcome, given the observed data history, should align with the actual observed variables
- **Sequential ignorability assumption**
  the action be conditionally independent of all potential variables, given the past data history
- **Positivity assumption**
  the probability of $\{A_t = 1\}$ given the observed data history, must be strictly between zero and one for any $t \geq 1$

VCDP Models

Temporal VCDP

$$Y_{i,t} = \beta_0(t,U_i) + S_{i,t}^{\top}\beta(t,U_i) + A_{i,t}\gamma(t,U_i) = Z_{i,t}^{\top}\theta(t,U_i)$$

$$S_{i,t+1} = \phi_0(t) + \Phi(t)S_{i,t} + A_{i,t}\Gamma(t) + E_i(t+1) = \Theta(t)Z_{i,t} + E_i(t+1)$$



- $\boldsymbol{U_i \sim U(0,1)}$ is the rank variable, which represents unobserved heterogeneity
- $E\left(E_i(t+1) \middle| S_{i,t}, A_{i,t}\right) = 0, E_i(t)$ are independent over time
- The temporal independence between $E_i(t+1)$ implies that the state vector satisfies the Markov property

$$\boldsymbol{CQTE_\tau = SQTE_\tau = \sum_{t=1}^{m}\gamma(t,\tau) + \sum_{t}^{m}\beta(t,\tau)^{\top}\left\{\sum_{k=1}^{t-1}\left[\prod_{l=k+1}^{t-1}\Phi(l)\right]\Gamma(k)\right\}}$$

Two –step Estimation

Step One

$$\hat{\theta}(t,\tau) = argmin \sum_i \rho_\tau \left(Y_{i,t} - Z_{i,t}^\top \theta(t,\tau) \right), \qquad t = 1, \ldots, m$$

$$\widehat{\Theta}^{(v)}(t) = argmin \sum_i \left(S_{i,t}^{(v)} - Z_{i,t}^\top \Theta^{(v)}(t) \right)^2, \qquad t = 1, \ldots, m-1, v = 1, \ldots, d$$

Reduce Variance

Step Two

$$\tilde{\theta}(t,\tau) = \sum_j^m \omega_{j,h} \hat{\theta}(j,\tau), \qquad t = 1, \ldots, m$$

$$\widehat{\Theta}^{(v)}(t) = \sum_j^m \omega_{j,h} \widehat{\Theta}^{(v)}(j), \qquad t = 1, \ldots, m-1, v = 1, \ldots, d$$

$$\widehat{CQTE}_\tau = \sum_{t=1}^m \tilde{\gamma}(t,\tau) + \sum_t^m \tilde{\beta}(t,\tau)^\top \left\{ \sum_{k=1}^{t-1} \left[\prod_{l=k+1}^{t-1} \tilde{\Phi}(l) \right] \tilde{\Gamma}(k) \right\}$$

Testing Procedure

Bootstrap Testing

- Step 1: Compute the estimators $\tilde{\theta}(t,\tau)$ and $\tilde{\Theta}(t)$
- Step 2: Estimate the residuals
- Step 3: for each b, generate i.i.d random variables by randomly sampling the residuals with replacement, then generate pseudo outcomes by the fitted value and the sampled residuals
- Step 4: compute the bootstrap estimates $\tilde{\theta}^b(t,\tau)$ and $\tilde{\Theta}^b(t)$ and the bootstrapped statistic $T_\tau^b = \widehat{CTQE}_\tau^b$
- Step 5: Repeat steps 3-4 B times, reject H_0 if the statistic T_τ exceeds the upper α quantile of $T_\tau^b - T_\tau$

Theorem 1: Under suitable conditions, if the bandwidth $h = o\left(n^{-\frac{1}{4}}\right), m \asymp n^c$ for some $0.5 < c < 1.5$, and $mh \to 0$ as $n \to \infty$,

$$\sup_{\epsilon} \sup_{z} \left|P(T_\tau - CQTE_\tau \leq z) - P\left(T_\tau^b - T_\tau | Data \leq z\right)\right| \leq C(\sqrt{n}h^2 + \sqrt{n}m + n^{-1/8})$$

with probability approaching 1, for some $\epsilon \in (0,1)$ and some positive constant C.

Extension to Spatiotemporal Experiment

CQTE and SCQTE

$$CQTE_{\tau st} = Q_\tau\left(\sum_{\iota=1}^{r}\sum_{t=1}^{m} Y_t^*(1_{t,[1:r]})\,|\mathcal{E}_{m,[1:r]}\right) - Q_\tau\left(\sum_{\iota=1}^{r}\sum_{t=1}^{m} Y_t^*(0_{t,[1:r]})\,|\mathcal{E}_{m,[1:r]}\right)$$

$$SCQTE_{\tau st} = \sum_{\iota=1}^{r}\sum_{t=1}^{m} Q_\tau\left(Y_t^*(1_{t,[1:r]})|\mathcal{E}_{m,[1:r]}\right) - \sum_{\iota=1}^{r}\sum_{t=1}^{m} Q_\tau\left(Y_t^*(0_{t,[1:r]})|\mathcal{E}_{m,[1:r]}\right)$$

- $\bar{a}_{t,\iota} = (a_{1,\iota}, \dots, a_{t,\iota})^\top \in \{0,1\}^t$, the treatment history up to t for the ιth region
- $S^*_{t,\iota}\ (\bar{a}_{t-1,[1:r]})$ and $Y^*_{t,\iota}\ (\bar{a}_{t-1,[1:r]})$ as the counterfactual state and outcome for the ιth region

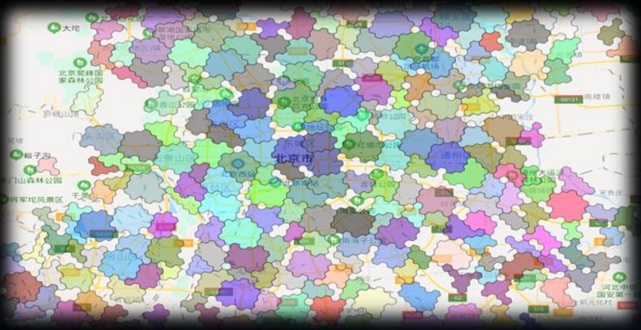
$$H_0: CQTE_{\tau st} \le 0 \quad v.s. \quad CQTE_{\tau st} > 0$$

Spatiotemporal Models

Extension

$$Y_{i,t,\iota} = \beta_0(t,\iota,U_i) + S_{i,t,\iota}^{\top}\beta(t,\iota,U_i) + A_{i,t,\iota}\gamma_1(t,\iota,U_i) + \bar{A}_{i,t,\mathcal{N}_\iota}\gamma_2(t,\iota,U_i)$$

$$S_{i,t+1,\iota} = \phi_0(t,\iota) + \Phi(t,\iota)S_{i,t,\iota} + A_{i,t,\iota}\Gamma_1(t,\iota) + \bar{A}_{i,t,\mathcal{N}_\iota}\Gamma_2(t,\iota) + E_i(t+1,\iota)$$



- $\bar{A}_{i,t,\mathcal{N}_\iota}$ denotes the average of the treatments of its neighboring regions

$$CQTE_{\tau st} = \sum_{\iota=1}^{r}\sum_{\tau=1}^{m}\{\gamma_1(\tau,\iota)+\gamma_2(\tau,\iota)\} + \sum_{\iota=1}^{r}\sum_{\tau=1}^{m}\beta(\tau,\iota)^{\top}\left\{\sum_{k=1}^{\tau-1}\left(\prod_{j=k+1}^{\tau-1}\Phi(j,\iota)\right)(\Gamma_1(k,\iota)+\Gamma_2(k,\iota))\right\}$$

Direct and Indirect Effects

CQDE and CQIE

- CQDE: direct effect of the treatment at time t.

$$CQDE_\tau = Q_\tau\left(\sum_t^m Y_t^*(1_t)\,|\mathcal{E}_m\right) - Q_\tau\left(\sum_t^m Y_t^*(0, 1_{t-1})\,|\mathcal{E}_m\right) \Rightarrow CQDE_\tau = \sum_{t=1}^m \gamma(t,\tau)$$

- CQIE: carryover effects of past treatments on the current outcome

$$CQDE_\tau = Q_\tau\left(\sum_t^m Y_t^*(0, 1_{t-1})\,|\mathcal{E}_m\right) - Q_\tau\left(\sum_t^m Y_t^*(0_t)\,|\mathcal{E}_m\right) \Rightarrow CQIE_\tau = \sum_t^m \beta(t,\tau)^\top\left\{\sum_{k=1}^{t-1}\left[\prod_{l=k+1}^{t-1}\Phi(l)\right]\Gamma(k)\right\}$$

$$H_0: CQDE_\tau \le 0 \quad v.s. \quad CQDE_\tau > 0$$

$$H_0: CQTE_\tau \le 0 \quad v.s. \quad CQTE_\tau > 0$$

Simulation Results

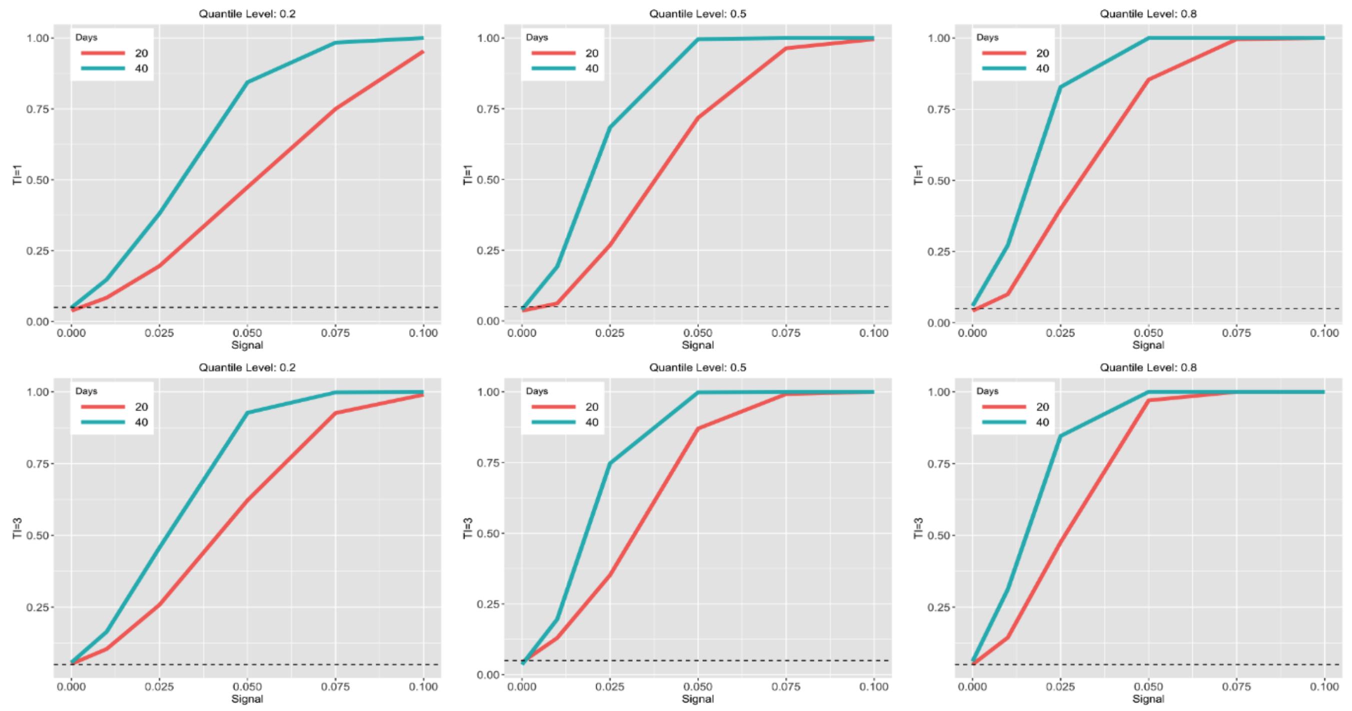


Figure 7: Empirical rejection rates of the proposed test for CQTE_τ. TI equals 1 for the top panels and 3 for the bottom panels. The quantile level $\tau = 0.2$, 0.5 and 0.8, from left to right plots.

- Outcome of Interest: drivers‘ total income
- State variable : the number of call orders and drivers' total online time
- Obtain the other estimates by setting $\gamma(t,\tau) = \Gamma(t) = 0$ and obtain the estimated error processes

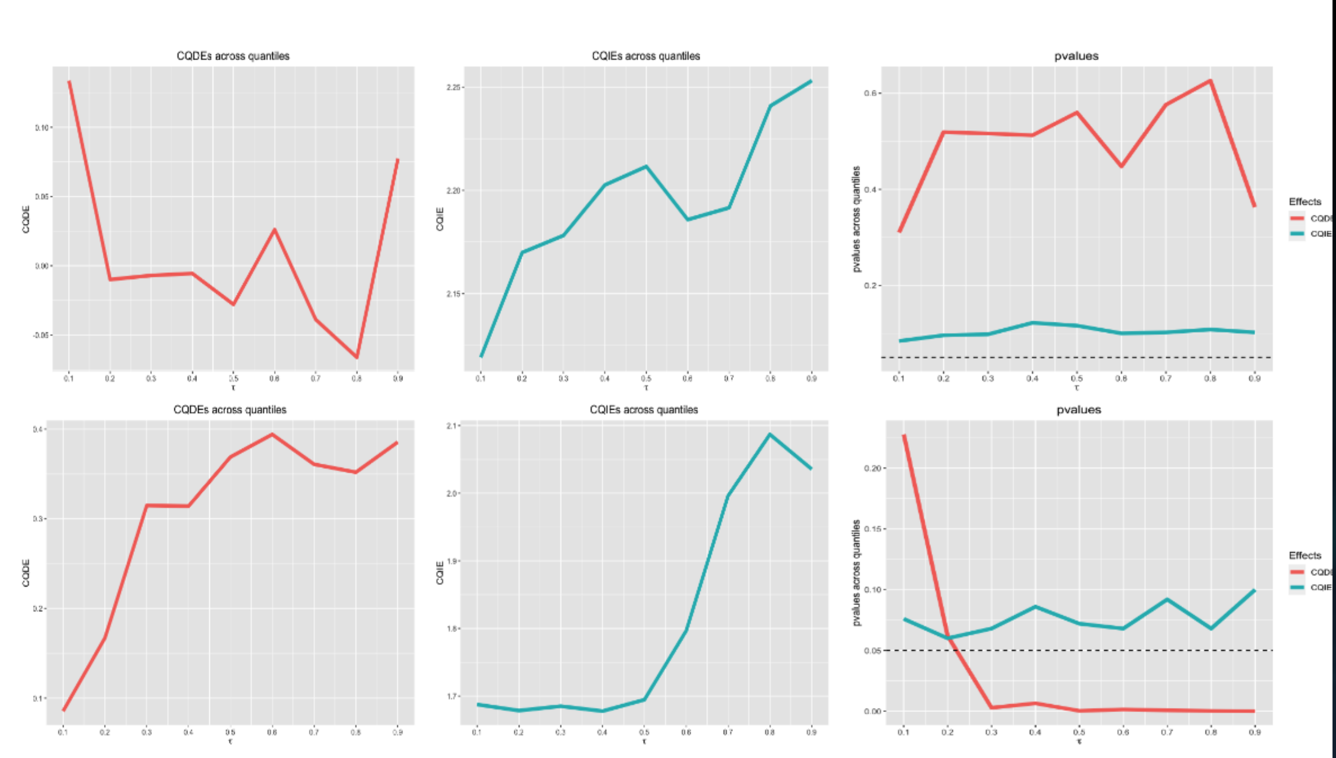
$$\tilde{\gamma}(t,\tau) = \delta Q_\tau(Y_t)$$
$$\tilde{\Gamma}(t,\tau) = \delta E(S_t)$$

Real Data Analysis

Temporal Experiment Results

τ	pvalues for AA		pvalues for AB	
	CQDE_τ	CQIE_τ	CQDE_τ	CQIE_τ
0.1	0.286	0.084	0.208	0.076
0.2	0.522	0.096	0.080	0.060
0.3	0.53	0.098	0.002	0.068
0.4	0.568	0.122	0.010	0.086
0.5	0.536	0.116	2e-4	0.072
0.6	0.464	0.100	0.002	0.068
0.7	0.548	0.102	7e-4	0.092
0.8	0.606	0.108	2e-4	0.068
0.9	0.322	0.102	7e-5	0.100

- The new policy is designed to fulfill more call orders and elevate drivers' total income.



- The proposed test does not reject the null hypothesis at any quantile level when applied to the A/A experiment.
- The new policy demonstrates significant quantile direct effects on the business outcome at most quantile levels .
- In contrast, the indirect effects are not significant.

Real Data Analysis

Spatiotemporal Experiment Results

τ	$\text{pvalue}_{\text{CQDE}_{\tau st}}$	$\text{pvalue}_{\text{CQIE}_{\tau st}}$	$\widehat{\text{CQDE}}_{\tau st}$	$\widehat{\text{CQIE}}_{\tau st}$
0.1	0.290	0.024	1.566	14.153
0.2	0.072	0.036	3.403	15.002
0.3	0.026	0.020	4.022	16.032
0.4	0.032	0.016	3.678	16.939
0.5	0.010	0.022	5.482	17.725
0.6	0.004	0.020	5.902	18.559
0.7	0.004	0.022	7.139	19.535
0.8	0.006	0.014	5.746	20.473
0.9	7e-4	0.008	8.414	21.320

- The city is divided into 12 regions.
- Policies are implemented based on alternating 30-minute time intervals within each region.
- Outcome: drivers’ total income
- State variable: the number of call orders

- The treatment effects are significant at most quantile levels
- Both the estimated direct and indirect effects are positive across all quantiles.
- The new policy doesn't seem to boost the lower quantile of the outcome.
- It reveals the heterogeneous effects of the new policy across different quantile levels.

Real Data Analysis

Spatiotemporal Experiment Results

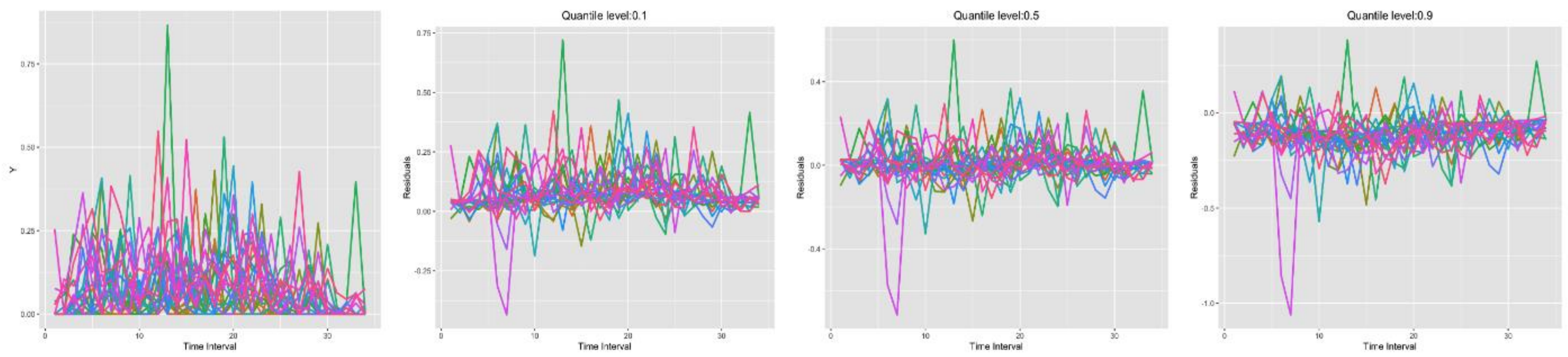


Figure 6: Scaled values of drivers' total income and their estimated residuals at quantile levels 0.1, 0.5, 0.9 in Regions 5.

- There may be several outliers in the data
- This observation further supports the use of quantiles as the evaluation metric.

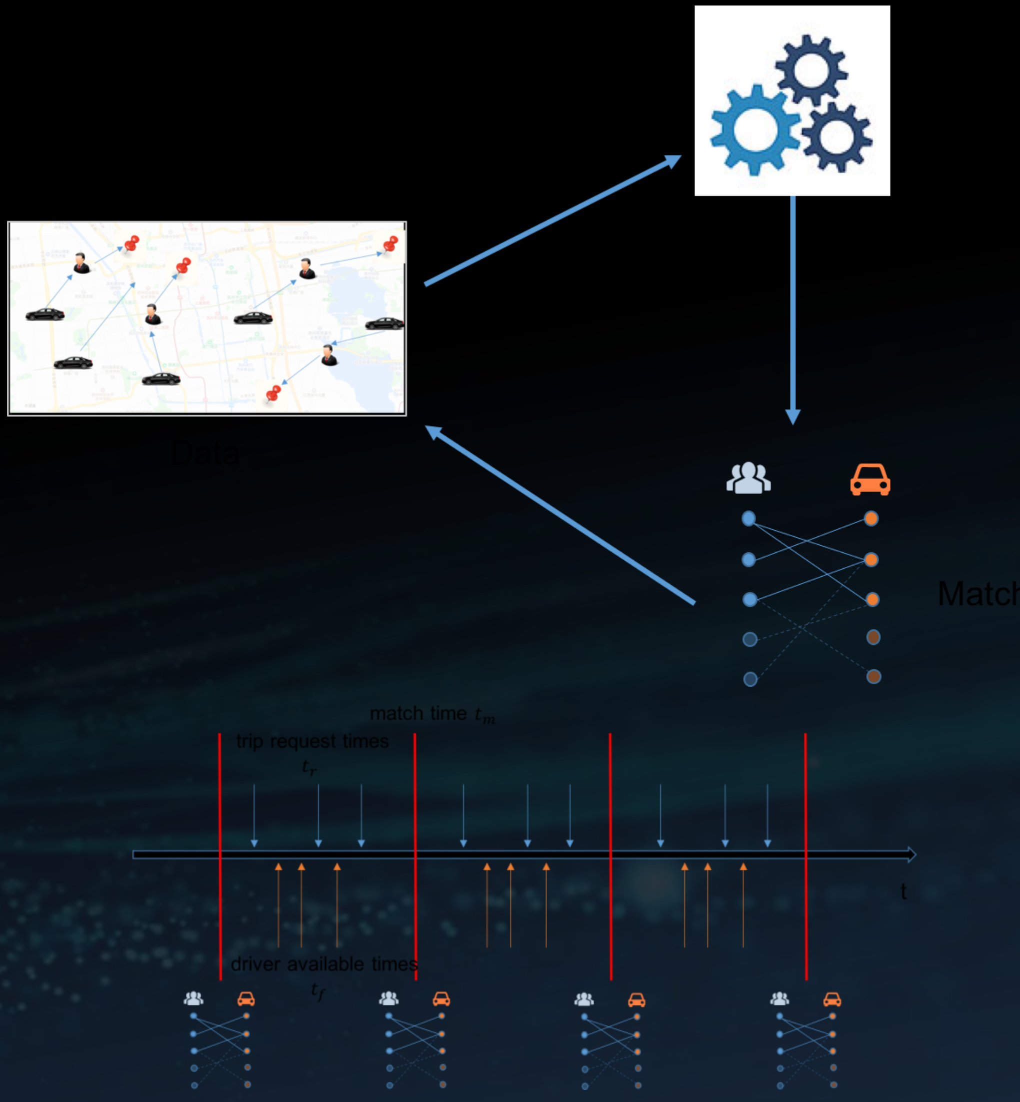
Optimal Dynamic Treatment Allocation for Efficient Policy Evaluation

T Li, C Shi,J Wang, F. Zhou, H Zhu. Optimal Dynamic Treatment Allocation for Efficient Policy Evaluation in Sequential Decision Making. NeurIPS 2023.

Dynamic Treatment Allocation

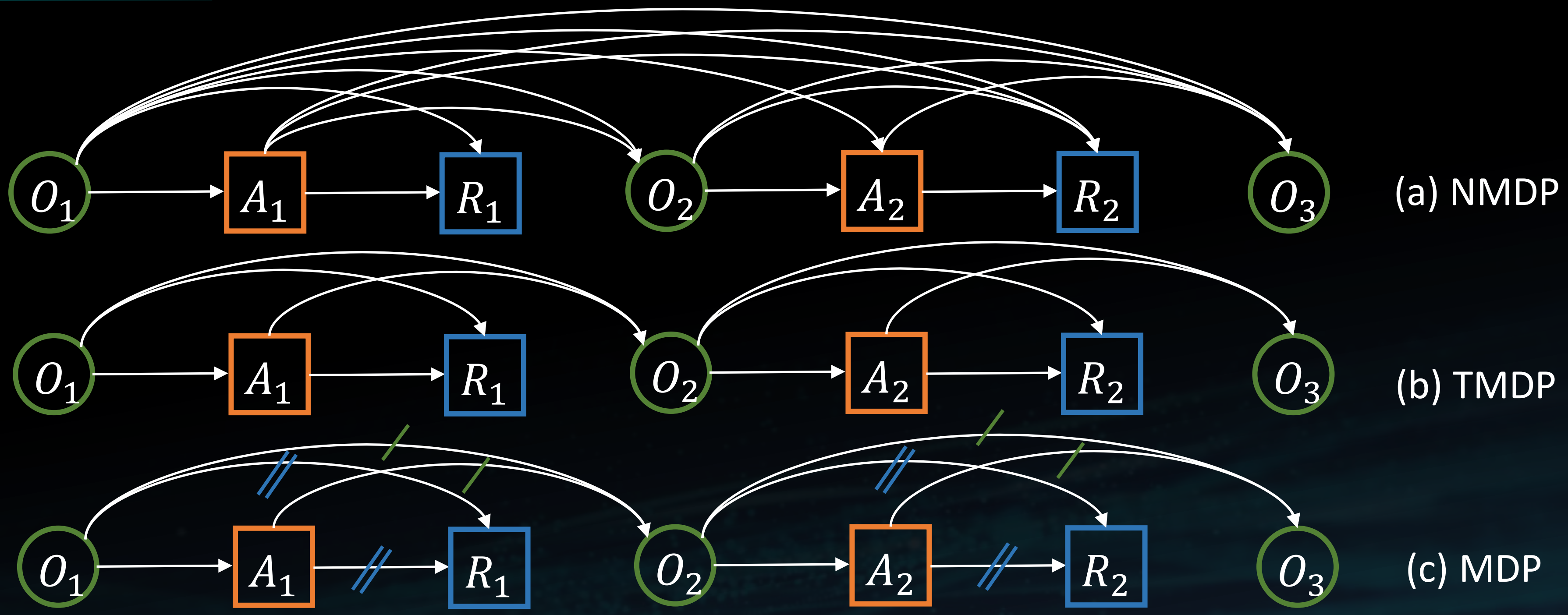
Motivation

- Prior to the full-scale deployment of any product, an accurate evaluation of its potential impact is crucial.
- Generation of the experimental dataset is critical, it can substantially influence the precision of the subsequent treatment effect estimator.
- A carefully designed experiment can significantly improve the accuracy of the treatment effect estimator and the statistical power
- Most existing designs fails to account for temporal carryover effects.



Dynamic Treatment Allocation

The Goal



- Design dynamic treatment allocation method in sequential decision making with carryover effects over time
- Study optimal designs that aim to maximize the amount of information to estimate treatment effects accurately

Related Work

Experimental designs

- There is an extensive body of literature on experimental design for clinical trials, with a multitude of optimal designs proposed.
  - D-optimality; D_A-optimality (Jones and Goos, 2009; Atkinson and Pedrosa, 2017)
  - A-optimality; A_A-optimality (Sverdlov and Rosenberger, 2013; Yin and Zhou, 2017)
  - Covariate-adaptive designs (Zhu and Hu, 2019)
  - Response-adaptive designs (Yu et al. 2022)
  - Covariate-adaptive response-adaptive designs (Zhang et al., 2007)

  ✓ These methods were designed for i.i.d. data and thus are not directly applicable to our settings.

- Ugander et al. (2013), Li et al. (2019) and Leung (2022), among others studied experimental designs with spatial/network spillover effects.

- A few designs have been developed for modern technological companies (Nandy et al. 2021, Johari et al. 2022)

  ✓ These studies did not utilize NMDP or TMDP models for experimental designs.

Design in NMDPS

Efficiency Bound for ATE

$$ATE = T^{-1}\sum_{t=1}^{T}[E^1(R_t) - E^0(R_t)]$$

$$EB_1(\pi^b) = T^{-2}\sum_{a\in\{0,1\}}\sum_{t=1}^{T} E^{\pi^b}\left[\sigma_t(H_t,a)\prod_{k\le t}\frac{I(A_k = a)}{\pi_k^b(a|H_k)}\right] + T^{-2}Var(V_1^1(O_1) - V_1^0(O_1))$$

- π^b : the behavior policy that generated the experimental data
- O_t: time varying features
- R_t: reward at time t

- History dependent policy, $\{\pi_t\}_{t\ge 1}$ with $\pi_t(\cdot\,|H_t)$
- H_t : the observed data up to time t
- $V_t^a(h) = \sum_{k=t}^{T} E^a(R_k|H_t = h)$ is the value function
- $\sigma_t^2(H_t, a)$: conditional variance of the temporal difference given H_t

The proposed designs

$$\pi^{b*} = argmin\, EB_1(\pi^b)$$

✓ Our objective lies in the design of an optimal behavior policy so that the mean squared error of the subsequent ATE estimator is minimized.

Implementation and Evaluation in NMDPS

The proposed design

Theorem 1: In NMDP, $\boldsymbol{\pi}^{b*}$ satisfies (1) for any $a \in \{0,1\}$,

$$\pi_1^{b*}(a|O_1) = \frac{\sigma_*(O_1, a)}{\sigma_*(O_1, 0) + \sigma_*(O_1, 1)}, \qquad \sigma_*^2(O_1, a) = E^a[\left\{\sum_t R_t - E^a R_t\right\}^2 |O_1, A_1 = a]$$

(2) for any $a \in \{0,1\}$, $\pi_2^{b*}(A_1|H_2) = \pi_3^{b*}(A_1|H_3) = \cdots = \pi_T^{b*}(A_1|H_T) = 1$ almost surely, or equivalently $A_1 = A_2, \ldots, A_T$ under $\boldsymbol{\pi}^{b*}$

Treatment Allocation Algorithm for NMDPS

- The burn-in period m_0 for each global policy and the termination day n
- Run each global policy for m_0 days
- While $2m_0 < m \leq n$, using the collected data to estimate the unknown terms
- Assign $A_1^{(m)}$ according to the plugged in probability
- Set $A_2^{(m)} = \cdots = A_T^{(m)} = A_1^{(m)}$

Design in TMDPS

Efficiency Bound for ATE

$$EB_2(\pi^b) = T^{-2}\sum_{a\in\{0,1\}}\sum_{t=1}^{T} E^{\pi^b}\left[\frac{I(A_t = a)p_t^a(O_t)}{p_t^b(O_t, a)}\right]^2 + T^{-2}Var(V_1^1(O_1) - V_1^0(O_1))$$

- $p_t^1(\cdot)$: the probability density function of O_t under the new policy
- $p_t^0(\cdot)$: the probability density function of O_t under the control policy
- $p_t^b(\cdot,\cdot)$: the probability density function of (O_t, A_t) under the behavior policy

✓ In contrast to NMDPs, the marginal distribution function p_t^b in TMDPs cannot be represented in a closed-form as a function of π^b

✓ The dependence of p_t^b on π^b makes it exceptionally challenging to identify the optimal π^b that minimizes $EB_2(\pi^b)$

✓ We shift our focus to finding the optimal in-class behavior policy

The proposed designs

$$\pi^{b*} = argmin_{\pi^b\in\Pi^b} EB_1(\pi^b) \qquad \Pi^b = \{\pi^b : \pi_2^b(A_1|H_2) = \pi_3^b(A_1|H_3) = \cdots = \pi_T^b(A_1|H_T) = 1\}$$

✓ The optimal behavior policy π^{b*} belongs to Π^b in NMDPs

✓ This not the case in TMDPs without additional assumptions.

Implementation and Evaluation in TMDPS

The proposed design

Theorem 2: Under β-mixing condition, an asymptotically optimal in-class behavior policy $\boldsymbol{\pi}^{b*}$ satisfies (1) for any $a \in \{0,1\}$,

$$\pi_1^{b*}(a|O_1) = \frac{\sigma_{a*}}{\sigma_{1*} + \sigma_{a*}}, \qquad \sigma_{a*}^2 = E^a[\sigma_t^2(O_t, a)]$$

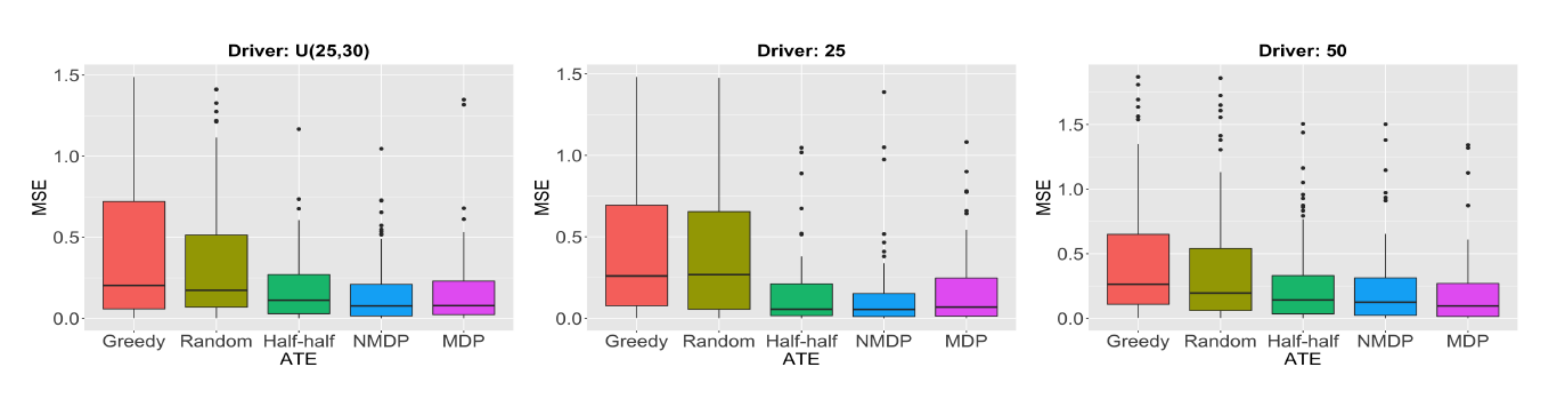
(2) for any $a \in \{0,1\}$, $\pi_2^{b*}(A_1|H_2) = \pi_3^{b*}(A_1|H_3) = \cdots = \pi_T^{b*}(A_1|H_T) = 1$ almost surely.

Additionally, suppose the proportionality condition holds such that $\sigma_t^2(o,1)/\ \sigma_t^2(o,0) = c$ for some constant c and any o, t. Then $\boldsymbol{\pi}^{b*}$ is the optimal one among all the candidate behavior policies.

- ✓ The estimation algorithm is similar to that in NMDP by adaptively assign the treatment by plugged-in estimators
- ✓ The design under MDP is similar to TMDP with the difference that σ_t^2 are not time varying given (O_t, a)

Experiment

I. Synthetic Dispatch

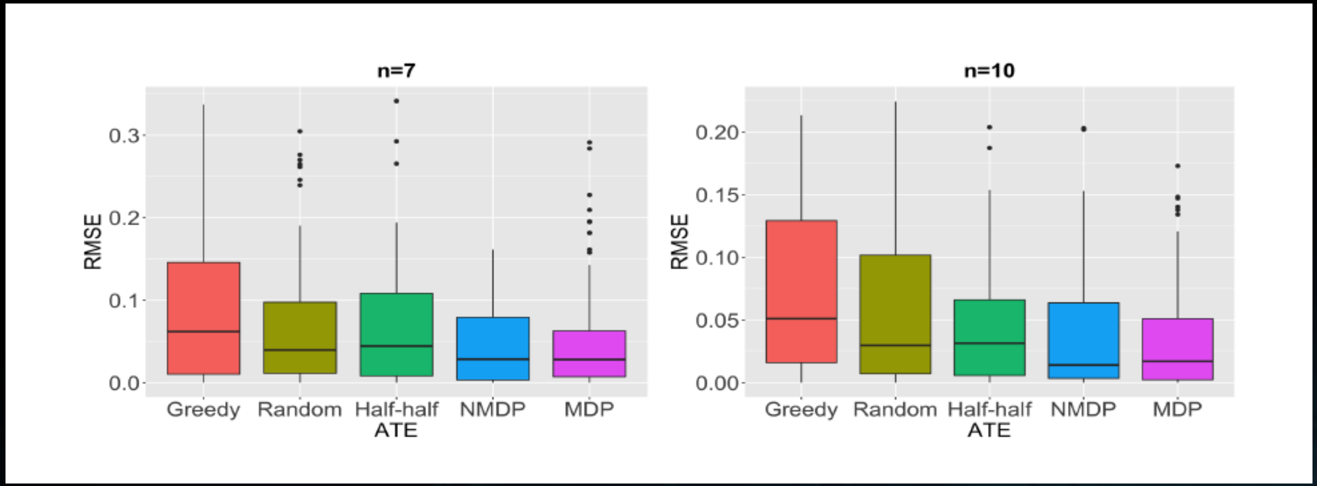


- Greedy: ϵ-greedy method
- Random: $P(A_{i,t} = 1) = 0.5$
- Half-half: treatment for the first $n/2$ days, the control for the remaining days
- NMDP: the proposed design under NMDP
- MDP: the proposed design nuder MDP

- Construct a small-scale synthetic dispatch environment.
- Simulate drivers and orders in a 9×9 spatial grid with a duration of 20 time steps each day

Experiment

I. Real-Data Dispatch



- A dispatch simulator based on a city-scale order-driver historical dataset from Didi Chuxing.
- Generate data based on the historical dataset.
- The distributions of drivers and orders are set to be identical to the distributions of historical data.
- The proposed method outperforms all its counterparts.

Thanks!