Evaluating Dynamic Conditional Quantile Treatment Effects with Applications in Ridesharing

Ting Li*[1], Chengchun Shi*[2], Zhaohua Lu[3], Yi Li[3], and Hongtu Zhu[4] *

[1]*Shanghai University of Finance and Economics*

[2]*London School of Economics and Political Science*

[3]*DiDi Chuxing*

[4]*University of North Carolina at Chapel Hill*

Abstract

Many modern tech companies, such as Google, Uber, and Didi, utilize online experiments (also known as A/B testing) to evaluate new policies against existing ones. While most studies concentrate on average treatment effects, situations with skewed and heavy-tailed outcome distributions may benefit from alternative criteria, such as quantiles. However, assessing dynamic quantile treatment effects (QTE) remains a challenge, particularly when dealing with data from ride-sourcing platforms that involve sequential decision-making across time and space. In this paper, we establish a formal framework to calculate QTE conditional on characteristics independent of the treatment. Under specific model assumptions, we demonstrate that the dynamic conditional QTE (CQTE) equals the sum of individual CQTEs across time, even though the conditional quantile of cumulative rewards may not necessarily equate to the sum of conditional quantiles of individual rewards. This crucial insight significantly streamlines the estimation and inference processes for our target causal estimand. We then introduce two varying coefficient decision process (VCDP) models and devise an innovative method to test the dynamic CQTE. Moreover, we expand our approach to accommodate data from spatiotemporal dependent experiments and examine both conditional quantile direct and indirect effects. To showcase the practical utility of our method, we apply it to three real-world datasets from a ride-sourcing platform. Theoretical findings and comprehensive simulation studies further substantiate our proposal.

Keywords: A/B testing, policy evaluation, quantile treatment effect, ridesourcing platform, spatialtemporal experiments, varying coefficient models.

*The first two authors contribute equally to this paper. Address for correspondence: Hongtu Zhu, Ph.D., E-mail: htzhu@email.unc.edu. This work was partially finished when Dr Zhu worked at Didi Chuxing. Li's research is partially supported by the National Science Foundation of China 12101388, CCF-DiDi GAIA Collaborative Research Funds for Young Scholars and Program for Innovative Research Team of Shanghai University of Finance and Economics. Shi's research is partially supported by an EPSRC grant EP/W014971/1. The content is solely the responsibility of the authors and does not necessarily represent the official views of any other funding agency.

1 Introduction

Online experiments, often referred to as A/B testing in computer science literature, are widely utilized by technology companies (e.g., Google, Netflix, Microsoft) to assess the effectiveness of new products or policies in comparison to existing ones. These companies have developed in-house A/B testing platforms for evaluating treatment effects and providing valuable experimental insights. Take ridesourcing platforms like Uber, Lyft, and Didi as examples. These platforms operate within intricate spatiotemporal ecosystems, dynamically matching passengers with drivers (see, for instance, Wang and Yang, 2019; Qin et al., 2020). They implement online experiments to explore various order dispatch policies and customer recommendation initiatives. These products hold the potential to enhance passenger engagement and satisfaction, diminish pickup waiting times, and boost driver earnings, ultimately leading to a more efficient and user-friendly transportation system.

In this study, we address the fundamental question of how to evaluate the difference between the *quantile* return of a new product (treatment) and that of an existing one. Although the average treatment effect (ATE) is widely used in the literature to quantify the difference between two policies (Imbens and Rubin, 2015; Kong et al., 2022),it only considers the average effect and does not account for variability around the expectation. In applications with skewed and heavy-tailed outcome distributions, decision-makers are more interested in the quantile treatment effect (QTE), which offers a more comprehensive characterization of distributional effects beyond the mean and is robust to heavy-tailed errors (see e.g., Abadie et al., 2002; Chernozhukov and Hansen, 2006). For example, in ridesourcing platforms, policymakers may want to determine which policy more effectively raises the lower tail of driver income. Furthermore, developing valid inferential tools for QTE can reveal how treatment effects differ by quantile and provide valuable information about the entire distribution.

Addressing the problem mentioned earlier presents two significant challenges. The first challenge involves efficiently inferring the dynamic QTE (quantile treatment effect), which is defined as the difference between the quantiles of cumulative outcomes under the new and old policies, in long horizon settings with weak signals. In contrast to single-stage decision-making, policy makers for ridesourcing platforms assign treatments sequentially over time and across various locations. Existing estimators, such as those based on (augmented) inverse probability weighting (see e.g., Wang et al., 2018, Section 4), are subject to the curse of horizon, as described by (Liu et al., 2018). This means their variances increase exponentially with respect to the horizon (i.e., the number of decision stages). Such approaches are inadequate in our context, where the horizon typically spans 24 or 48 stages and most policies improve key metrics by only 0.5% to 2% (Qin et al., 2022). Furthermore, unlike the average cumulative outcome, which can be broken down into the sum of individual outcome expectations, the quantile of cumulative outcomes generally does not equal the sum of individual quantiles. This makes estimating our causal effect extremely challenging. Existing efficient evaluation methods designed for mean return, such as those proposed by Kallus and Uehara (2022) cannot be easily adapted to our situation.

The second challenge arises from handling the interference effect caused by temporal and spatial proximities in spatiotemporal dependent experiments. This interference effect results in a treatment applied at one location influencing not only its own outcome, but also the outcomes at other locations. The current treatment is likely to affect both present and future outcomes. Neglecting these effects would produce a biased QTE estimator. As far as we are aware, there is no existing test capable of concurrently addressing both challenges.

1.1 Related work

A/B testing has been extensively researched in the literature, as evidenced by the works of Yang et al. (2017) and Zhou et al. (2020), among other references. In contrast to most existing A/B testing methods that focus on the Average Treatment Effect (ATE), Quantile Treatment Effects (QTE) have received less attention. Among the few available studies, Liu et al. (2019) proposed a scalable method to test QTE and construct associated confidence intervals. Moreover, Wang and Zhang (2021) developed a nonparametric method to estimate QTEs at a continuous range of quantile locations, including point-wise confidence intervals. More broadly, the estimation and inference of (conditional) QTEs have been considered in the causal inference literature, as seen in the works of Chernozhukov and Hansen (2006), Firpo (2007), and Blanco et al. (2020). However, these methods predominantly address single-stage decision-making. To the best of our knowledge, this paper represents the first attempt to explore QTE in temporally and/or spatially dependent experiments.

Our paper is closely related to the rapidly expanding body of literature on off-policy evaluation in sequential decision-making. 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). In recent years, several papers have explored policy evaluation beyond averages (Wang et al., 2018; Kallus et al., 2019; Qi et al., 2022). These works propose using (augmented) inverse probability weighted estimators to evaluate specific robust metrics under a given target policy. As noted previously, these methods are subject to the curse of horizon and become less effective in long-horizon settings. Most notably, policy evaluation in spatiotemporal dependent experiments remains unexplored in the aforementioned studies.

Analysis of temporal and spatial interference has received much attention recently. For example, Tchetgen and VanderWeele (2012), Hudgens and Halloran (2008) and Liu and Hudgens (2014) considered partial interference, where interference is possible between individuals within the same group (or through social interactions) but not between groups. This implies that potential outcomes are influenced not only by the treatment of the subject under consideration but also by the treatments of other subjects within the same group. Our focus differs from these studies as we delve into the realm of interference that extends across both time and space. However, similar to the above literature, we employ the partial interference assumption to handle the spatial interference as well. Recent proposals have investigated causal inference with temporal or spatial interference, including studies by Savje et al. (2021) and Hu et al. (2022), among others. However, these methods primarily focus on the average effect. Furthermore, our paper is closely related to the literature on distributional reinforcement learning (see e.g., Zhou et al., 2020). Despite this connection, these studies primarily concentrate on the policy learning problem, and the uncertainty quantification of a target policy's quantile value remains unexplored.

Lastly, our paper is connected to a line of research on quantitative analysis of ridesharing across various fields such as economics, operations research, statistics, and computer science (see e.g., Shi et al., 2022; Zhao et al., 2022). Nevertheless, quantile policy evaluation has not been examined in these papers.

1.2 Contributions

Our proposal offers three valuable contributions to existing literature. First, we present a framework for deducing dynamic conditional Quantile Treatment Effects (QTE), defined as dynamic QTE dependent on market features, irrespective of treatment history. While unconditional QTE may be of interest, as previously noted, it assumes a highly complex form in long horizon settings and is extremely challenging to identify when the signal is weak. In contrast, we demonstrate that under certain modeling assumptions, the proposed dynamic conditional QTE (CQTE) is equal to the sum of individual CQTE at each spatiotemporal unit, even though the conditional quantile of cumulative rewards does not necessarily equate to the sum of conditional quantiles of individual rewards. This finding significantly streamlines the estimation and inference processes for our causal estimand, making our proposal easily implementable in practice. Additionally, the estimated CQTE can exhibit a smaller variance compared to that of the unconditional counterpart.

Second, we introduce an innovative framework to test dynamic CQTE while accounting for the interference effect. We propose two Varying Coefficient Decision Process (VCDP) models, enabling the application of classical quantile regression (Koenker and Hallock, 2001) for parameter estimation and subsequent inference. We then develop a two-step method for estimating CQTE, along with a bootstrap-assisted procedure for testing CQTE. We further extend our proposal to analyze spatiotemporally dependent data and to test Conditional Quantile Direct Effects (CQDE) and Conditional Quantile Indirect Effects (CQIE).

Third, we thoroughly examine the theoretical and finite sample properties of our methods. Theoretically, we prove the consistency of our proposed test procedure, allowing the horizon to diverge with the sample size. Notably, classical weak convergence theorems (Van Der Vaart and Wellner, 1996) necessitate a fixed horizon and are not directly applicable. Empirically, we apply our proposed method to real datasets obtained from a leading ridesourcing platform to assess the dynamic quantile treatment effects of new policies.

1.3 Organization of the paper

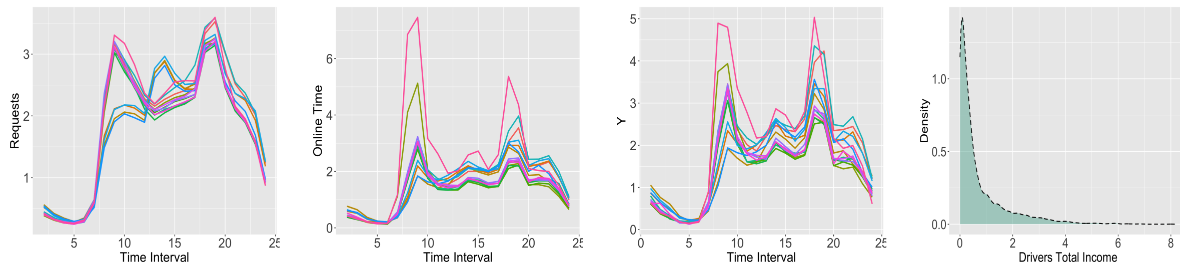
Section 2 describes data from online experiments. Section 3 covers temporally dependent experiments, and estimation and inference procedures. Section 4 extends the proposal to spatiotemporally dependent data. Section 5 decomposes CQTE into CQDE and CQIE. Section 6 evaluates ridesourcing dispatching and repositioning policies, and Section 7 assesses the finite sample performance using real-data-based simulations. Additional simulation studies, theoretical properties and their proofs are presented in the supplementary material.

2 Data Description

The purpose of this paper is to analyze three real datasets collected from Didi Chuxing, one of the world's leading ride-sharing companies. One dataset was collected during a time-dependent A/B experiment conducted in a city from December 10, 2021 to December 23, 2021. The goal of this experiment was to evaluate the performance of a newly designed order dispatching policy, which aimed to increase the number of fulfilled ride requests and boost drivers' total revenue. To protect privacy, we will not disclose the city name and the specific policy used. During the experiment, each day was divided into 24 equally spaced non-overlapping time intervals. The new policy (B) and the old policy (A) were alternated and assigned to these intervals every day. On each day, we randomly selected the treatment sequence from AB…AB and BA…BA with equal probability. Therefore, we switched between A and B within and between days, ensuring that each policy was used with equal probability at each time interval, meeting the positivity assumption. For more details, see Section 3.1. It is worth noting that such an alternating-time-interval design, also known as the switchback design, is commonly used in industries to reduce the variance of treatment effect estimators (Hu and Wager, 2022; Shi et al., 2022; Xiong et al., 2023). Further information can be found in the article by Lyft on experimentation in a ride-sharing marketplace (Chamandy, 2016).

The second dataset comes from a spatiotemporal-dependent experiment conducted in another city between February 19, 2020, and March 13, 2020. Each day is divided into 48 non-overlapping, equal time intervals, and the city is partitioned into 12 distinct, non-overlapping regions. On the first day of the experiment, the initial policy in each region is independently set to either the new or old policy with a 50% probability. The temporal alternation design for time-dependent experiments is then applied in each region.

In addition to the two datasets from A/B experiments mentioned earlier, we also analyze



(a) (b) (c) (d)

Figure 1: Scaled drivers' total income (a), request (b) and drivers' online time (c) in the temporal dependent A/B experiment and the estimated density of drivers' total income (d) in the spatial temporal dependent A/B experiment.

a third dataset collected from an A/A experiment. In this case, the two policies being compared are identical, and the treatment effect is zero. The experiment took place in a specific city from July 13, 2021 to September 17, 2021. This analysis serves as a sanity check to examine the size property of the proposed test. We expect that our test will not reject the null hypothesis when applied to this dataset, as the true effect is zero.

The ridesharing system dynamically connects passengers and drivers in real-time. All three datasets include the number of call orders and the total online time of drivers for each time interval. These metrics represent the supply and demand in this two-sided market. The platform's outcomes include the drivers' total income, the answer rate (the number of call orders responded to), and the completion rate (the number of call orders completed) for each time interval. In our study, we are interested in determining whether the new policy improves drivers' total income at various quantile levels.

The datasets exhibit four distinct characteristics. First, the horizon duration is typically much longer (e.g., 24 or 48) than the experiment duration, while the treatment effect is usually weak (e.g., 0.5%-2%). Second, both demand and supply are spatiotemporal networks that interact across time and location, as observed in panels (a) and (b) of Figure 1, which display the number of call orders and drivers' online time. Third, the outcome of interest follows a non-normal and heavy-tailed distribution, illustrated in panels (c) and (d) of Figure 1. Finally, there are interference effects over time and space, demonstrated in Figure 2, with temporal interference effects occurring when past actions impact future outcomes.

We focus on answering three key questions in these datasets:

(Q1) How can we quantify treatment effects across various quantile levels for the time-

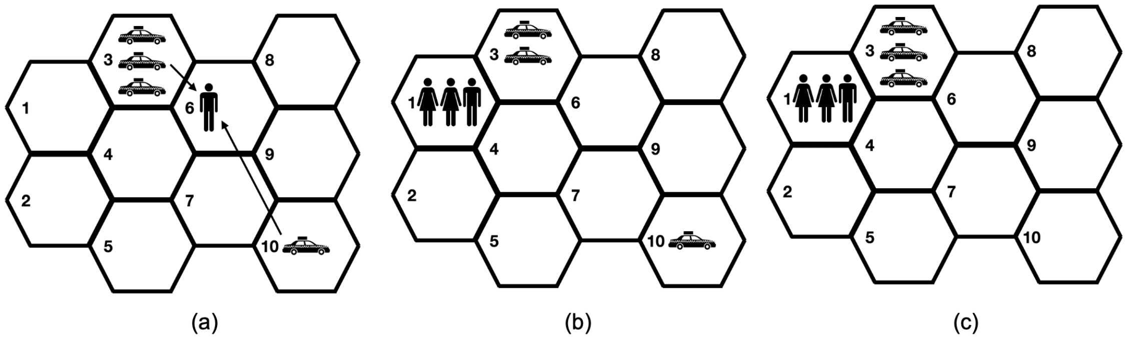


Figure 2: This example illustrates the temporal interference effect in ridesharing, where assigning different drivers to pick up a passenger significantly impacts future ride requests. (a) A city with 10 regions has a passenger in region 6 needing a ride, with three drivers in region 3 and one in region 10. Two actions are possible: assigning a driver from region 3 or region 10. (b) Assigning a driver from region 3 might result in an unmatched future request due to the driver in region 10 being too far from region 1. (c) Assigning the driver in region 10 could lead to all future ride requests being matched, preserving all three drivers in region 3.

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 spatiatemporal dependent experiment data?

(Q3) How to determine whether or not to replace the old policy with the new one? These questions drive the methodological development outlined in Sections 3 and 4.

3 Testing CQTE in temporal dependent experiments

In this section, we explicitly state the test hypotheses for our first research question (Q1) and explore the primary challenge encountered in experiments exhibiting temporal dependence. Subsequently, we detail the key technical assumptions that enable the cumulative quantile treatment effect (CQTE) to be equivalent to the sum of individual CQTEs. Finally, to address the third research question (Q3), we present the proposed estimation and testing strategies for our investigation.

3.1 CQTE, test hypotheses and assumptions

We consider the temporal alternation design with a sequence of treatments over time. Specifically, we divide each day into m non-overlapping intervals. The platform can implement either one of the two policies at each time interval. For any $t \geq 1$, let A_t denote the policy implemented at the tth time interval where $A_t = 1$ represents exposure to the new policy and $A_t = 0$ represents exposure to the old policy. Let S_t and Y_t denote the state (e.g., the supply and demand) and the outcome at time t, respectively.

To formulate our problem, we adopt a potential outcome framework (Rubin, 2005). Specifically, we define $\bar{a}_t = (a_1, \ldots, a_t)^\top \in \{0,1\}^t$ as the treatment history up to time t. We also define $S_t^*(\bar{a}_{t-1})$ and $Y_t^*(\bar{a}_t)$ as the counterfactual state and counterfactual outcome, respectively, that would have occurred had the platform followed the treatment history $\bar{a}_t$. Our primary interest lies in quantifying the difference between the τth quantile of the cumulative outcomes under the new policy and that under the old policy, denoted as the quantile treatment effect (QTE):

$$\mathrm{QTE} = Q_\tau\Big(\sum_{t=1}^{m} Y_t^*(\mathbf{1}_t)\Big) - Q_\tau\Big(\sum_{t=1}^{m} Y_t^*(\mathbf{0}_t)\Big),$$

where $\mathbf{1}_t$ and $\mathbf{0}_t$ are vectors of 1s and 0s of length t, respectively, and $Q_\tau(\cdot)$ denotes the quantile function at the τth level.

However, learning such an unconditional dynamic QTE from our experimental dataset is highly challenging. Remember that in our A/B experiment, the old and new policies are assigned alternately over the m time intervals. Nevertheless, the target policy we aim to evaluate corresponds to the global policy, which allocates the new or old policy globally throughout each day. This leads to an off-policy setting where the target policy differs from the behavior policy that generates the data. Existing off-policy quantile evaluation methods based on inverse probability weighting, such as those presented by Wang et al. (2018), are inefficient in our setting with a moderately large m. Off-policy evaluation (OPE) methods, including Shi et al. (2020), Liao et al. (2021), and Kallus and Uehara (2022), are semiparametrically efficient[1] in long-horizon settings. Despite this, these methods primarily focus on the mean return, making it difficult to adapt them for quantile evaluation due to the nonlinear quantile function Q_τ. To illustrate, note that there is no guarantee that $Q_\tau(\sum_{t=1}^m Y_t^*(\mathbf{1}_t)) - Q_\tau(\sum_{t=1}^m Y_t^*(\mathbf{0}_t)) = \sum_{t=1}^m \Big[Q_\tau(Y_t^*(\mathbf{1}_t)) - Q_\tau(Y_t^*(\mathbf{0}_t))\Big]$ in general. This

[1]Shi et al. (2020) and Liao et al. (2021) proposed to use the direct method based on linear sieves or kernels. However, the resulting estimators are semiparametrically efficient as well.

observation motivates us to seek an alternative definition for QTE.

Second, let $\mathcal{E}_t$ represent the set of features (e.g., extreme weather events) that have an impact on the outcomes up to time t, but are not influenced by the treatment history. This means that for any treatment history $\bar{a}_t$, the potential outcome of these features remains the same, i.e., $\mathcal{E}_t^*(\bar{a}_t) = \mathcal{E}_t$. By definition, S_1 is an element of $\mathcal{E}_t$, which ensures that $\mathcal{E}_t$ is non-empty for any t. We introduce CQTE as follows:

$$\text{CQTE}_\tau = Q_\tau(\sum_{t=1}^{m} Y_t^*(\mathbf{1}_t)|\mathcal{E}_m) - Q_\tau(\sum_{t=1}^{m} Y_t^*(\mathbf{0}_t)|\mathcal{E}_m). \tag{1}$$

The CQTE is a reasonable measure because the set of conditioning variables remains consistent under both new and old policies. When $m = 1$, this definition reduces to the one used in single-stage decision-making, as discussed in previous literature, such as Chernozhukov and Hansen (2006).

Conditioning has several benefits. Firstly, it offers a more convenient way to estimate the dynamic Quantile Treatment Effect (QTE) by aggregating individual QTEs over time. This approach simplifies the estimation process and reduces computational complexity. Secondly, conditioning can also help to reduce the variance of the resulting QTE estimator by removing the need to account for variability in the relevant characteristics. By conditioning on certain variables, researchers can effectively control for confounding factors and produce more accurate estimates of treatment effects.

Third, we introduce the concept of Summed Conditional Quantile Treatment Effects (SCQTE), which represents the sum of individual Conditional Quantile Treatment Effects (CQTE) over time. The SCQTE is defined as follows:

$$\text{SCQTE}_\tau = \sum_{t=1}^{m} Q_\tau(Y_t^*(\mathbf{1}_t)|\mathcal{E}_t) - \sum_{t=1}^{m} Q_\tau(Y_t^*(\mathbf{0}_t)|\mathcal{E}_t).$$

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. Although the quantile function is not additive, we demonstrate in the following proposition that SCQTE is equal to CQTE under specific modeling assumptions. Even in scenarios where the underlying assumptions may not strictly hold, focusing on SCQTE still offers valuable insights. It describes the average of the marginal quantile treatment effect across all time intervals, which assesses how the distribution of outcomes changes in response to the treatment at each time point.

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 $U \overset{d}{\sim}$ Unif$(0,1)$, which is independent of $\{\mathcal{E}_t\}_t$. Furthermore, assume that $\phi_t(\mathcal{E}_t, \mathbf{1}_t, \tau)$ and $\phi_t(\mathcal{E}_t, \mathbf{0}_t, \tau)$ are non-decreasing functions of τ for any $\mathcal{E}_t$. Under these conditions, we find that $CQTE_\tau = SCQTE_\tau$.*

Proposition 1 establishes the equivalence between CQTE and SCQTE and serves as a fundamental building block for our proposal. It allows us to focus on SCQTE, which is a simplified version of CQTE. This simplification 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. For more details, see Sections 3.2 and 3.3. Moreover, the proposed model in Proposition 1 is related to the structural quantile model in the quantile regression literature (Chernozhukov and Hansen, 2005, 2006). These models assume that, conditional on the covariate $X = x$, the potential outcome $Y^*(a) = q(a, x, U)$ for $a = 0, 1$ and $U \sim U(0,1)$, where $q(d, x, \tau)$ is strictly increasing in τ. The uniformly distributed variable U serves as a rank variable that characterizes the heterogeneity of the outcome across different quantile levels. More discussion on the rank invariance assumption can be found in Section E of the supplementary material. Under the non-decreasing constraint, the τth conditional quantile of $Y^*(a)|X = x$ can be shown to equal $q(a, x, \tau)$. The discussion on this assumption can be found in Section F. Proposition 1 motivates us to focus on testing the following hypotheses for each quantile level τ:

$$H_0 : \text{CQTE}_\tau \leq 0 \quad \text{versus} \quad H_1 : \text{CQTE}_\tau > 0. \tag{2}$$

These hypotheses test whether the treatment effect at the τth quantile is non-negative or positive, respectively.

In this study, we utilize the consistency assumption (CA), sequential ignorability assumption (SRA), and positivity assumption (PA) to identify the causal estimand. Similar assumptions are frequently used in the dynamic treatment regime literature for learning optimal dynamic treatment policies (Gill and Robins, 2001). The consistency assumption (CA) states that the potential state and outcome, given the observed data history, should align with the actual observed state and outcome. The sequential ignorability assumption (SRA) demands that the action be conditionally independent of all potential variables, given the past data history. In our application, the SRA is inherently satisfied as the policy is assigned according to the alternating-time-interval design, independent of data history. The positivity assumption (PA) necessitates that the probability of $\{A_t = 1\}$, given the current state, must be strictly confined between zero and one for any $t \geq 1$[2]. Under the alternating-time-interval design, the PA can be easily satisfied if the treatment in the initial time interval is randomized. It is essential to note that the CA, SRA, and PA enable the consistent estimation of the potential outcome distribution using the observed data.

3.2 VCDP models

Suppose that the experiment is conducted over n consecutive days. Let $(S_{i,j}, A_{i,j}, Y_{i,j})$ be the state-treatment-outcome triplet measured at the jth time interval of the ith day for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. We assume that these triplets are independent across different days, but may be dependent within each day over time.

We begin by introducing two varying coefficient decision process models, one for the outcome and the other for the state. The first model characterizes the conditional quantile of the outcome and is given by

$$Y_{i,t} = \beta_0(t, U_i) + S_{i,t}^\top \beta(t, U_i) + A_{i,t}\gamma(t, U_i) = Z_{i,t}^\top \theta(t, U_i), \tag{3}$$

where $Z_{i,t} = (1, S_{i,t}^\top, A_{i,t})^\top \in \mathbb{R}^{d+2}$, $\theta(t, U_i) = (\beta_0(t, U_i), \beta(t, U_i)^\top, \gamma(t, U_i))^\top \in \mathbb{R}^{d+2}$ is a vector of time-varying coefficients, and $U_i \sim U(0, 1)$ is the rank variable. Model (3) extends the idea of using rank variables to represent unobserved heterogeneity across different quantiles in a single-stage study to sequential decision making.

[2]Our positivity assumption is weaker than the standard positivity assumption which requires the probability of $\{A_t = 1\}$ given the observed data history to be strictly confined between zero and one. Such a relaxation is facilitated by the Markov assumption, which allows us to model the current outcome and future state solely based on the current state-action pair, as elaborated in Equations (7) and (8)

The second model characterizes the conditional mean of the observed state variables,

$$S_{i,t+1} = \phi_0(t) + \Phi(t)S_{i,t} + A_{i,t}\Gamma(t) + E_i(t+1) = \Theta(t)Z_{i,t} + E_i(t+1), \quad (4)$$

where $\phi_0(t)$ and $\Gamma(t)$ are d-dimensional vectors, $\Phi(t)$ is a $d \times d$ matrix of autoregressive coefficients, and $\Theta(t) = [\phi_0(t) \quad \Phi(t) \quad \Gamma(t)]$ is a $d \times (d+2)$ coefficient matrix. The term $E_i(t+1)$ is a random error term whose conditional mean given $Z_{i,t}$ equals zero. In addition, $\{E_i(t)\}_t$ are independent over time. It can be checked by testing the autocorrelation of the residuals, such as the well-known Ljung-Box test. Therefore, the conditional expectation of $S_{i,t+1}$ given $Z_{i,t}$ is: $\mathbb{E}(S_{i,t+1}|Z_{i,t}) = \Theta(t)Z_{i,t}$.

It is worth noting that models (3) and (4) belong to the class of varying-coefficient regression models. The existing literature on this topic mainly focuses on estimating the relationships between scalar predictors and scalar responses (Sherwood and Wang, 2016), or between scalar predictors and functional responses (Zhang et al., 2022), or between longitudinal predictors and responses (Wang et al., 2009). However, to the best of our knowledge, none of these works have utilized varying-coefficient regression models for policy evaluation in sequential decision making.

While we assume the residual process $\{E_i(t)\}$ is independent over time, it is worthwhile to note that the state process $\{S_{i,t}\}$ is correlated and dependent. Specifically, this state process adheres to a martingale sequence that satisfies the Markov property. In the absence of this independence assumption, the Markov assumption would be violated. It is well-known that efficient policy evaluation is extremely challenging in non-Markovian environments (see e.g., Uehara et al., 2022). Moreover, models (3) and (4) are valid when the potential outcomes satisfy similar assumptions in the quantile varying coefficient models. Please refer to equations (S2) and (S3) in the supplementary material for more details.

If the residual $E_i(t+1)$ is independent of the treatment history and U_i, we can define $\mathcal{E}_t$ as $\{S_1, E(1), \cdots, E(t)\}$. Under this condition, the assumptions in Proposition 1 are satisfied. Hence, under the proposed VCDP models, CQTE is equivalent to SCQTE. Next, we introduce the following function:

$$\begin{aligned}\phi_t(\mathcal{E}_t, \bar{a}_t, U) &= \beta_0(t,U) + a_t\gamma(t,U) + \beta(t,U)^\top\Big(\sum_{k=1}^{t-1}\Big\{\prod_{l=k+1}^{t-1}\Phi(l)[\phi_0(k)+a_k\Gamma(k)]\Big\}\Big) \\ &\quad + \prod_{l=1}^{t-1}\Phi(l)S_1 + \sum_{k=2}^{t}[\prod_{l=k}^{t-1}\Phi(l)E(k)],\end{aligned}$$

The subsequent proposition offers a closed-form formula for CQTE.

Proposition 2. *Assuming that CA and Equations (S2) and (S3) in the supplementary material hold, U is independent of $\{\mathcal{E}_t\}_t$, and $\phi_t(\mathcal{E}_t, \mathbf{1}_t, \tau)$ and $\phi_t(\mathcal{E}_t, \mathbf{0}_t, \tau)$ are non-decreasing in τ for any $\mathcal{E}_t$, then we have*

$$CQTE_\tau = SCQTE_\tau = \sum_{t=1}^{m}\gamma(t,\tau) + \sum_{t=2}^{m}\beta(t,\tau)^\top\left\{\sum_{k=1}^{t-1}\left[\prod_{l=k+1}^{t-1}\Phi(l)\right]\Gamma(k)\right\}, \tag{5}$$

where the product $\prod_{l=k+1}^{t-1}\Phi(l) = 1$ when $t-1 < k+1$.

Proposition 2 enables us to estimate CQTE through SCQTE under certain assumptions. Among these, the monotonicity assumption can be satisfied under various conditions. For example, it holds when $\beta_0(t,\tau)$, $\gamma(t,\tau)$, and all elements in $\beta(t,\tau)$ are strictly increasing in τ, and $\phi_0(t)$, all elements in $\Phi(t)$, and $\Gamma(t)$ are positive. Additionally, when $\Gamma(t) = 0$ and $\phi_0(t) = 0$ for any t, it suffices to require $\gamma(t,\tau)$ and $\beta_0(t,\tau)$ to be strictly increasing in τ.

To evaluate policy value, we need to estimate the model parameters β, γ, Φ, and Γ. Notice that under the conditions of Proposition 2, we have that:

$$Y_{i,t} = \beta_0(t,\tau) + S_{i,t}^\top\beta(t,\tau) + A_{i,t}\gamma(t,\tau) + e_i(t,\tau) = Z_{i,t}^\top\theta(t,\tau) + e_i(t,\tau), \tag{6}$$

where $e_i(t,\tau)$ is the error term, defined as $Z_{i,t}^\top[\theta(t,U) - \theta(t,\tau)]$, and its conditional τ-th quantile given $Z_{i,t}$ equals zero. Therefore, we can employ ordinary quantile regression to learn β and γ. Meanwhile, since the residuals $E_i(t)$s are independent over time, ordinary least-squares regression is applicable to the state regression model to estimate Φ and Γ. We detail our estimating procedure in the next section.

3.3 Estimation and inference procedures

In this subsection, we outline the procedures for estimating and testing CQTE based on the results in Proposition 2. We first estimate the regression coefficients in models (6) and (4). We then plug these estimates into (5) to estimate CQTE. Finally, we develop a bootstrap-assisted procedure to test CQTE.

Let $S_{i,t+1}^{(\nu)}$, $\phi_0^{(\nu)}(t)$, and $\Gamma^{(\nu)}(t)$ denote the ν-th entries of $S_{i,t+1}$, $\phi_0(t)$, and $\Gamma(t)$, respectively. Let $\Phi^{(\nu)}(t)$ and $\Theta^{(\nu)}(t)$ denote the ν-th rows of $\Phi(t)$ and $\Theta(t)$, respectively. It follows from (4) that:

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

We propose a two-step procedure to estimate $\theta(t,\tau)$ and $\Theta(t)$. In the first step, we minimize the following functions:

$$\widehat{\theta}(t,\tau) = \arg\min \sum_i \rho_\tau(Y_{i,t} - Z_{i,t}^{\top}\theta(t,\tau)), \quad \text{for } t = 1, \ldots, m, \tag{7}$$

$$\widehat{\Theta}^{(\nu)}(t) = \arg\min \sum_i [S_{i,t+1}^{(\nu)} - Z_{i,t}^{\top}\Theta^{(\nu)}(t)]^2, \quad \text{for } \nu = 1, \ldots, d; \quad t = 1, \ldots, m-1. \tag{8}$$

These one-step estimates can be computed easily but suffer from large variances as they rely solely on observations at time t. In contrast, the true coefficient $\theta(t,\tau)$ and $\Theta(t)$ are expected to possess smoothness across time t. To fully use the information across time, we employ the kernel smoothing technique to reduce the variances of these initial estimators, achieve smoothness, and identify weak signals (Zhu et al., 2014). Specifically, for a given kernel function $K(\cdot)$, the second-step estimators $\widetilde{\theta}(t,\tau)$ and $\widetilde{\Theta}_\tau^{(\nu)}(t)$ are defined as:

$$\widetilde{\theta}(t,\tau) = \sum_{j=1}^{m} \omega_{j,h}(t)\widehat{\theta}(j,\tau), \quad \text{for } t = 1, \ldots, m, \tag{9}$$

$$\widetilde{\Theta}^{(\nu)}(t) = \sum_{j=1}^{m} \omega_{j,h}(t)\widehat{\Theta}^{(\nu)}(j), \quad \text{for } \nu = 1, \ldots, d, \quad t = 1, \ldots, m, \tag{10}$$

where $\omega_{j,h}(t) = K((j-t)/(mh))/\sum_{k=1}^{m} K((k-t)/(mh))$ is the weight function and h denotes the kernel bandwidth. Among various kernel functions, we adopt the widely used Gaussian kernel $K(t) = \exp(-t^2)$. Essentially, the estimates $\widetilde{\Theta}^{(\nu)}(t)$ and $\widetilde{\theta}(t,\tau)$ are calculated as weighted averages of the one-step estimates $\{\widehat{\theta}(j,\tau), j = 1,\ldots,m\}$, with the weights being modulated by the kernel function. This methodology facilitates a more stable and continuous estimation, which is especially valuable in the presence of weak signals or when assessing parameters that are expected to be consistent over time. Given $\widetilde{\theta}(t,\tau)$ and $\widetilde{\Theta}_{\tau}^{(\nu)}(t)$, we can compute the following CQTE estimator:

$$\widehat{\mathrm{CQTE}}_{\tau} = \sum_{t=1}^{m} \widetilde{\gamma}(t,\tau) + \sum_{t=2}^{m} \widetilde{\beta}(t,\tau)^{\top} \left\{ \sum_{k=1}^{t-1} \left[\prod_{l=k+1}^{t-1} \widetilde{\Phi}(l) \right] \widetilde{\Gamma}(k) \right\}. \tag{11}$$

To test (2), we use the test statistic T_{τ}, which is set to $\widehat{\mathrm{CQTE}}_{\tau}$. Under the null hypothesis, T_{τ} is expected to be negative or close to zero. Therefore, we reject the null hypothesis for a large value of $\widehat{\mathrm{CQTE}}_{\tau}$. However, deriving the limiting distribution of T_{τ} for large m is complicated due to the complex dependence of $\widehat{\mathrm{CQTE}}_{\tau}$ on the estimated model parameters. To address this issue, we use the bootstrap method to simulate the distribution of $\widehat{\mathrm{CQTE}}_{\tau}$ under the null hypothesis. Specifically, we modify the bootstrap method proposed by Horowitz and Krishnamurthy (2018) and adapt it to our setting as follows. Horowitz and Krishnamurthy (2018) proposed to resample the estimated residuals to infer the conditional quantile function in a nonparametric quantile regression model. In our case, to handle the dependence over time, we resample the entire error process (see Step 3 below for details).

The bootstrap method for $\widehat{\mathrm{CQTE}}_{\tau}$ is implemented as follows:

- Step 1. Compute the estimators $\widetilde{\theta}(t,\tau)$ and $\widetilde{\Theta}(t)$ in (9) and (10).
- Step 2. Estimate the residuals by $\widehat{e}_i(t,\tau) = Y_{i,t} - Z_{i,t}^{\top}\widetilde{\theta}(t,\tau)$ for $t = 1,\ldots,m$ and $\widehat{E}_i(t+1) = S_{i,t+1} - \widetilde{\Theta}(t)Z_{i,t}$ for $t = 1,\ldots,m-1$.
- Step 3. For $i = 1, ..., n$, define the random vectors $\widehat{e}_i(\tau) = (\widehat{e}_i(1,\tau),\cdots,\widehat{e}_i(m,\tau))$ and $\widehat{E}_i = (\widehat{E}_i(2),\cdots,\widehat{E}_i(m))$. For each bootstrap iteration indexed by b, we sample the entire residual process with replacement. Specifically, we generate a bootstrap sample consisting of n error processes $\{e_1^b(\tau),\ldots,e_n^b(\tau)\}$, by resampling n random vectors from the original set $\{\widehat{e}_1(\tau),\ldots,\widehat{e}_n(\tau)\}$ with replacement. Similarly, we construct another bootstrap sample of n error processes $\{E_1^b,\ldots,E_n^b\}$, by resampling n random vectors from $\{\widehat{E}_1,\ldots,\widehat{E}_n\}$ with replacement. Next, we generate pseudo outcomes

$\{\widehat{S}_{i,t}^b\}_{i,t}$ and $\{\widehat{Y}_{i,t}^b\}_{i,t}$ as follows,

$$\widehat{S}_{i,t+1}^b = \widetilde{\Theta}(t)\widehat{Z}_{i,t}^b + E_i^b(t+1) \text{ and } \widehat{Y}_{i,t}^b = \widehat{Z}_{i,t}^{b\top}\widetilde{\theta}(t,\tau) + e_i^b(t,\tau). \tag{12}$$

- Step 4. For each b, compute the bootstrap estimates $\widetilde{\theta}^b(t,\tau)$ and $\widetilde{\Theta}^b(t)$ according to equations (7)-(10) using the pseudo outcomes $\{(\widehat{S}_{i,t}^b, \widehat{Y}_{i,t}^b) : i, t\}$.
- Step 5. For each b, compute the bootstrapped statistic $T_\tau^b = \widehat{\mathrm{CQTE}}_\tau^b$.
- Step 6. Repeat Steps 3-5 B times. Given a significance level α, reject H_0 (see (2)) if the statistic T_τ exceeds the upper αth empirical quantile of $\{T_\tau^b - T_\tau\}_{b=1}^B$.

In the supplementary material, we present Theorem S1, which rigorously establishes the consistency of the aforementioned bootstrap method. It's worth noting that the bootstrap consistency theory elaborated in Horowitz and Krishnamurthy (2018) isn't readily applicable to our context, where m can increase along with the sample size.

In practice, we may select the bandwidth h by using K-fold cross-validation. To satisfy the $o(n^{-1/4})$ order condition, following a similar spirit to Seo and Linton (2007) and Porter and Yu (2015), we choose h among $H_n = [C_1(\log n)n^{-1/2}, C_2(\log n)n^{-1/2}]$ for some positive constants C_1 and C_2. For each potential bandwidth parameter, the cross-validation criterion is computed as the sum of the prediction errors after leaving out one of the K folds; smaller errors are preferred. Specifically, we randomly split the full sample into K subsets, denoted by $\{\mathcal{D}_j : j = 1, \ldots, K\}$. For $j = 1, \ldots, K$ and $i \in \mathcal{D}_j$, let $\widehat{S}_{i,t+1}^{(-\mathcal{D}_j)}(h)$ and $\widehat{Y}_{i,t}^{(-\mathcal{D}_j)}(h)$ represent the estimated mean for $S_{i,t+1}$ and the estimated quantile for $Y_{i,t}$, respectively. These estimates are computed using bandwidth h based on the dataset that excludes the observations in $\mathcal{D}_j$. We then select the optimal bandwidths h_θ for $\widetilde{\theta}$ and h_Θ for $\widetilde{\Theta}$ by minimizing the following objective functions:

$$\mathrm{CV}_\theta(h) = \sum_{j=1}^{K}\sum_{i\in\mathcal{D}_j}\sum_{t=1}^{m}\rho_\tau(Y_{i,t} - \widehat{Y}_{i,t}^{(-\mathcal{D}_j)}(h)), \quad \mathrm{CV}_\Theta(h) = \sum_{j=1}^{K}\sum_{i\in\mathcal{D}_j}\sum_{t=1}^{m-1}(S_{i,t+1} - \widehat{S}_{i,t+1}^{(-\mathcal{D}_j)}(h))^2.$$

We set $K = 2$ in our simulation studies and real data analysis.

4 Extension to spatiotemporal dependent experiments

In this section, we aim to address (Q2) and expand upon the method proposed in Section 3 to analyze data from spatiotemporal dependent experiments involving multiple non-overlapping regions receiving distinct treatments in a sequential manner over time. Let r represent the number of these non-overlapping regions. As previously discussed, these experiments are not only subject to temporal interference effects but also exhibit spatial interference, whereby the policy implemented in one location may influence the outcomes in other locations.

4.1 Test hypotheses

For the ι-th region, we use $\bar{a}_{t,\iota} = (\bar{a}_{1,\iota}, \ldots, \bar{a}_{t,\iota})^\top$ to denote its treatment history up to time t. Let $\bar{a}_{t,[1:r]} = (\bar{a}_{t,1}, \ldots, \bar{a}_{t,r})^\top$ represent the treatment history across all regions. Similarly, define $S^*_{t,\iota}(\bar{a}_{t-1,[1:r]})$ and $Y^*_{t,\iota}(\bar{a}_{t,[1:r]})$ as the potential observation and outcome for the ι-th region, respectively. The set of potential observations at time t is denoted as $S^*_{t,[1:r]}(\bar{a}_{t-1,[1:r]})$.

In the spatiotemporal context, our focus is on the cumulative quantile treatment effects, aggregated over all regions. We define CQTE and SCQTE at the τ-th quantile level as

$$\text{CQTE}_{\tau st} = Q_\tau\Big(\sum_{\iota=1}^{r}\sum_{t=1}^{m} Y^*_{t,\iota}(\mathbf{1}_{t,[1:r]})|\mathcal{E}_{m,[1:r]}\Big) - Q_\tau\Big(\sum_{\iota=1}^{r}\sum_{t=1}^{m} Y^*_{t,\iota}(\mathbf{0}_{t,[1:r]})|\mathcal{E}_{m,[1:r]}\Big),$$

$$\text{SCQTE}_{\tau st} = \sum_{\iota=1}^{r}\sum_{t=1}^{m} Q_\tau\Big(Y^*_{t,\iota}(\mathbf{1}_{t,[1:r]})|\mathcal{E}_{t,[1:r]}\Big) - \sum_{\iota=1}^{r}\sum_{t=1}^{m} Q_\tau\Big(Y^*_{t,\iota}(\mathbf{0}_{t,[1:r]})|\mathcal{E}_{t,[1:r]}\Big),$$

respectively, where $\mathcal{E}_{t,[1:r]}$ denotes the set of characteristics independent of the treatment history up to time t across all regions. For a given quantile level τ, our goal is to test whether a new policy outperforms the old one as follows:

$$H_0 : \text{CQTE}_{\tau st} \le 0 \quad \text{versus} \quad H_1 : \text{CQTE}_{\tau st} > 0. \tag{13}$$

Compared to the testing problem in (2), (13) focuses on global treatment effects aggregated over time and regions. We assume the consistency assumption holds. Similar to Section 3, under the spatial alternating-time-interval design, one can show that the sequential ignorability assumption and the positivity assumption are automatically satisfied, ensuring that CQTE is identifiable from the observed data.

4.2 Spatiotemporal VCDP models

Suppose that the experiment last for n days, and each day is divided into m time intervals. For $i = 1, \ldots, n$, $t = 1, \ldots, m$, and $\iota = 1, \ldots, r$, let $(S_{i,t,\iota}, A_{i,t,\iota}, Y_{i,t,\iota})$ represent the state-treatment-outcome triplet measured from the ιth region at the t-th time interval of the i-th day. For each ι, $\mathcal{N}_\iota$ denotes the neighbouring regions of ι. To model the quantiles of $Y_{t,\iota}$ and $S_{t,\iota}$, we extend the two VCDP models in Section 3 to two spatialtemporal VCDP (STVCDP) models in this section.

The first STVCDP model describes the quantile structure of the outcome,

$$\begin{aligned} Y_{i,t,\iota} &= \beta_0(t,\iota,U_i) + S_{i,t,\iota}^\top \beta(t,\iota,U_i) + A_{i,t,\iota}\gamma_1(t,\iota,U_i) + \bar{A}_{i,t,\mathcal{N}_\iota}\gamma_2(t,\iota,U_i) \\ &= Z_{i,t,\iota}^\top \theta(t,\iota,U_i), \end{aligned} \tag{14}$$

where $\bar{A}_{i,t,\mathcal{N}_\iota}$ denotes the average of $\{A_{i,t,k}\}_{k\in\mathcal{N}_\iota}$, $Z_{i,t,\iota} = (1, S_{i,t,\iota}^\top, A_{i,t,\iota}, \bar{A}_{i,t,\mathcal{N}_\iota})^\top$, and $\theta(t,\iota,U_i) = (\beta_0(t,\iota,U_i), \beta(t,\iota,U_i)^\top, \gamma_1(t,\iota,U_i), \gamma_2(t,\iota,U_i))^\top$. Model (14) is based on two key assumptions. Firstly, it is assumed that the effect of treatments in other regions on the conditional quantile of $Y_{i,t,\iota}$ is limited to those of its neighboring regions, as long as each experimental region is large enough. This is because drivers can only travel between neighboring regions in one time unit, meaning that treatments in non-neighboring regions are not expected to impact $Y_{i,t,\iota}$. Secondly, it is assumed that the influence of treatments in neighboring regions on the conditional quantile of $Y_{i,t,\iota}$ is only through the mean of the treatments. This is a common mean-field assumption used to model spillover effects (e.g., Hudgens and Halloran, 2008; Shi et al., 2022). Following a similar spirit to Shi et al. (2022), the mean-field assumption can be tested using observed data by investigating the conditional independence between $Y_{i,t,\iota}$ and treatments from the neighboring regions given $\{S_{i,t,\iota}, A_{i,t,\iota}, \bar{A}_{i,t,\mathcal{N}_\iota}\}$. Our proposed method is readily extensible to contexts where the effects of treatments in neighboring regions are influenced solely through their quantiles. Specifically, the term $\bar{A}_{i,t,\mathcal{N}\iota}$ in Equations (14) and (15) below can be replaced with the quantiles of the treatments $\{A_{i,t,k}\}_{k\in\mathcal{N}\iota}$. The theoretical results remain largely unchanged and can be established in a similar manner, as they are derived on the framework of conditional quantile regression. In particular, variations in some of the predictors do not impact the theoretical validity of our approach.

The second STVCDP model models the conditional distribution of the next state given the current state-action pair as follows:

$$\begin{aligned} S_{i,t+1,\iota} &= \phi_0(t,\iota) + \Phi(t,\iota)S_{i,t,\iota} + A_{i,t,\iota}\Gamma_1(t,\iota) + \bar{A}_{i,t,\mathcal{N}_\iota}\Gamma_2(t,\iota) + E_i(t+1,\iota) \\ &= \Theta(t,\iota)Z_{i,t,\iota} + E_i(t+1,\iota), \end{aligned} \tag{15}$$

where $\Theta(t,\iota) = [\phi_0(t,\iota), \Phi(t,\iota), \Gamma_1(t,\iota), \Gamma_2(t,\iota)] \in \mathbb{R}^{d\times(d+3)}$ and $\Phi(t,\iota)$ is a $d\times d$ matrix of autoregressive coefficients. The conditional mean of each entry in the error process $E_i(t,\iota)$ given $Z_{i,t,\iota}$ is zero. The error process is required to be independent over time, although it may be dependent across different locations. The varying coefficients are required to be smooth over the entire spatial domain, which will help to reduce the variances of the model estimators and improve the accuracy of the CQTE estimator. The models (14) and (15) hold under the assumption that the potential outcomes satisfy the quantile varying coefficient models, as described in the supplementary material (models (S4) and (S5)).

The following proposition provides a closed-form expression for $\text{CQTE}_{\tau st}$ and proves that $\text{CQTE}_{\tau st} = \text{SCQTE}_{\tau st}$. Let

$$\begin{aligned} \phi_{t,\iota}(\mathcal{E}_{t,\iota}, \bar{a}_{t,[1:r]}, U) &= \beta_0(t,\iota,U) + a_{t,\iota}\gamma_1(t,\iota,U) + \bar{a}_{t,N_\iota}\gamma_2(t,\iota,U) \\ &+ \beta(t,\iota,U)^\top \Big(\sum_{k=1}^{t-1} \Big\{ \prod_{l=k+1}^{t-1} \Phi(l,\iota)[\phi_0(k,\iota) + a_{k,\iota}\Gamma_1(k,\iota) + a_{k,N_\iota}\Gamma_2(k,\iota)] \Big\} \Big) \\ &+ \prod_{l=1}^{t-1} \Phi(l,\iota)S_{1,\iota} + \sum_{k=2}^{t} [\prod_{l=k}^{t-1} \Phi(l,\iota)E(k,\iota)], \end{aligned}$$

where $\mathcal{E}_{t,\iota} = \{S_{1,\iota}, E(2,\iota), \ldots, E(t,\iota)\}$ and the product $\prod_{l=k+1}^{t-1}\Phi(j,\iota) = 1$ when $t-1 < k+1$.

Proposition 3. *Suppose that CA and the conditions in equations (S4) and (S5) of the supplementary material hold, and that U is independent of the collection of error processes $\{\mathcal{E}_{t,\iota}\}_{t,\iota}$. Furthermore, assume that the functions $\phi_{t,\iota}(\mathcal{E}_{t,\iota}, \mathbf{1}_{t,[1:r]}, \tau)$ and $\phi_{t,\iota}(\mathcal{E}_{t,\iota}, \mathbf{0}_{t,[1:r]}, \tau)$ are non-decreasing in τ for any $\mathcal{E}_{t,\iota}$. Then, we have*

$$\begin{aligned} CQTE_{\tau st} = SCQTE_{\tau st} &= \sum_{\iota=1}^{r}\sum_{t=1}^{m}\{\gamma_1(t,\iota,\tau) + \gamma_2(t,\iota,\tau)\} \\ &+ \sum_{\iota=1}^{r}\sum_{t=2}^{m}\beta(t,\iota,\tau)^\top \left\{ \sum_{k=1}^{t-1} \left[\prod_{j=k+1}^{t-1} \Phi(j,\iota) \right] [\Gamma_1(k,\iota) + \Gamma_2(k,\iota)] \right\}. \end{aligned}$$

Proposition 3 provides a foundation for constructing a plug-in estimator for $\mathrm{CQTE}_{\tau st}$. This forms the basis of the proposed inference procedure, which we discuss in more detail in the next section. Additionally, from models (14) and (15), we can obtain the expression

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

where $e_i(t,\iota,\tau)$ is the residual term, defined as $Z_{i,t,\iota}^\top[\theta(t,\iota,U)-\theta(t,\iota,\tau)]$, and its conditional τ-th quantile given $Z_{i,t,\iota}$ is equal to zero. It is worth mentioning that these models can be further extended to incorporate the effects of states from neighboring regions on the immediate outcome by including another mean-field term $\Phi_2(t,\iota)\bar{S}_{i,t,\mathcal{N}_\iota}$, where $\bar{S}_{i,t,\mathcal{N}_\iota} = \sum_{\iota' \in \mathcal{N}_\iota} S_{i,t,\iota'}/\mathcal{N}_\iota$. In this case, the closed-form expression for $\mathrm{CQTE}_{\tau st}$ can be similarly derived.

4.3 Estimation and inference procedures

In this subsection, we outline the estimation and testing procedures for $\mathrm{CQTE}_{\tau st}$.

Firstly, we calculate raw estimators of the unknown coefficients in the two STVCDP models. For each region ι, we employ standard quantile regression and linear regression as shown in (7) and (8) to the data subsets $\{(Z_{i,t,\iota}, Y_{i,t,\iota})\}_{i,t}$ and $\{(Z_{i,t,\iota}, S_{i,t+1,\iota})\}_{i,t}$ to obtain the initial estimators $\widehat{\theta}(t,\iota,\tau)$ and $\widehat{\Theta}(t,\iota)$ for $\theta(t,\iota,\tau)$ and $\Theta(t,\iota)$, respectively. Next, we apply kernel smoothing techniques as illustrated in (9) and (10) to refine these initial estimators over time. We denote the resulting estimators as $\widetilde{\theta}^0(t,\iota,\tau)$ and $\widetilde{\Theta}^0(t,\iota)$.

Secondly, we further refine these raw estimators by employing kernel smoothing to borrow information across space. Specifically, we define $\widetilde{\theta}(t,\iota,\tau) = \sum_{\ell=1}^r \kappa_{\ell,h}(\iota)\widetilde{\theta}^0(t,\ell,\tau)$ and $\widetilde{\Theta}^{(\nu)}(t,\iota) = \sum_{\ell=1}^r \kappa_{\ell,h_{st}}(\iota)\widetilde{\Theta}^{0(\nu)}(t,\ell)$, where $\widetilde{\Theta}^{0(\nu)}(t,\iota)$ is the νth column of $\widetilde{\Theta}^0(t,\iota)$ and $\kappa_{\ell,h_{st}}(\iota)$ is a normalized kernel function with bandwidth parameter h_{st}. The kernel function $\kappa_{\ell,h_{st}}(\iota)$ is given by

$$\kappa_{\ell,h_{st}}(\iota) = \frac{K((u_\iota - u_\ell)/h_{st})K((v_\iota - v_\ell)/h_{st})}{\sum_{j=1}^r K((u_\iota - u_j)/h_{st})K((v_\iota - v_j)/h_{st})},$$

where (u_ι, v_ι) represents the longitude and latitude of region ι. Consequently, regions with smaller spatial distances contribute more significantly.

Thirdly, we estimate $\mathrm{CQTE}_{\tau st}$ by substituting the refined estimators $\widetilde{\theta}_{\tau st}(t,\iota)$ and $\widetilde{\Theta}_{st}(t,\iota)$ and use the resulting estimator $\widehat{\mathrm{CQTE}}_{\tau st}$ as the test statistic $T_{\tau st}$. Finally, we introduce a bootstrap method to test (13). During each iteration, we resample the estimated error processes to obtain the bootstrap estimates $\widetilde{\theta}^b_{\tau st}(t,\iota)$ and $\widetilde{\Theta}^b_{st}(t,\iota)$, and the bootstrapped statistic $T^b_{\tau st} = \widehat{\mathrm{CQTE}}^b_{\tau st}$. We reject H_0 in (13) if $T_{\tau st}$ exceeds the upper αth empirical quantile of $\{T^b_{\tau st} - T_{\tau st}\}_{b=1}^B$. As this approach is highly similar to the one presented in Section 3.3, we omit further details for brevity.

Similar to the bandwidth selection method used in temporal-dependent experiments, we employ the K-fold cross-validation to simultaneously optimize the two bandwidths for both kernel smoothing procedures. Detailed formulations can be found in Section C of the supplementary material.

5 Direct and indirect effects

Recall that Proposition 2 provides the closed-form expression of CQTE_τ, which is

$$\sum_{t=1}^m \gamma(t,\tau) + \sum_{t=2}^m \beta(t,\tau)^\top \left\{ \sum_{k=1}^{t-1} \left[\prod_{l=k+1}^{t-1} \Phi(l) \right] \Gamma(k) \right\}.$$

Consequently, we can divide the quantile treatment effect into two components. Specifically, the first term $\sum_{t=1}^m \gamma(t,\tau)$ of CQTE_τ represents the direct effect of the treatment on the immediate outcome, expressed as

$$\mathrm{CQDE}_\tau = Q_\tau\Big(\sum_{t=1}^m Y_t^*(\mathbf{1}_t)|\mathcal{E}_m \Big) - Q_\tau\Big(\sum_{t=1}^m Y_t^*(0,\mathbf{1}_{t-1})|\mathcal{E}_m \Big).$$

Observe that for each t, the two potential outcomes $Y_t^*(\mathbf{1}_t)$ and $Y_t^*(0,\mathbf{1}_{t-1})$ differ in the treatment received at time t, but they share the same treatment history. The second term $\sum_{t=2}^m \beta(t,\tau)^\top \left\{ \sum_{k=1}^{t-1} \left[\prod_{l=k+1}^{t-1} \Phi(l) \right] \Gamma(k) \right\}$ quantifies the carryover effects of past treatments on the current outcome, defined as

$$\mathrm{CQIE}_\tau = Q_\tau\Big(\sum_{t=1}^m Y_t^*(0,\mathbf{1}_{t-1})|\mathcal{E}_m \Big) - Q_\tau\Big(\sum_{t=1}^m Y_t^*(\mathbf{0}_t)|\mathcal{E}_m \Big).$$

Similar decompositions have been considered in Li and Wager (2022) and Shi et al. (2022).

The corresponding testing hypotheses are given by

$$H_0^D : \text{CQDE}_\tau \leq 0 \quad \text{versus} \quad H_1^D : \text{CQDE}_\tau > 0, \tag{16}$$

$$H_0^I : \text{CQIE}_\tau \leq 0 \quad \text{versus} \quad H_1^I : \text{CQIE}_\tau > 0. \tag{17}$$

Testing these hypotheses not only enables us to determine whether the new policy is significantly better than the old one or not, but also helps us understand how the new (or the old) policy outperforms the other.

To test (16) and (17), we use the two-step estimators in (9) and (10) to construct the plug-in estimators $\widehat{\text{CQDE}}_\tau$ and $\widehat{\text{CQIE}}_\tau$ for CQDE and CQIE, respectively. Next, we employ the bootstrap method in Section 3.3 to approximate the limiting distributions of $\widehat{\text{CQDE}}_\tau$ and $\widehat{\text{CQIE}}_\tau$ under the null hypotheses. We note that although $\widehat{\text{CQDE}}_\tau$ has a tractable limiting distribution and is asymptotically normal, estimating its asymptotic variance without using bootstrap remains challenging.

Finally, we can similarly define the direct effect and indirect effect in the spatiotemporal design as follows,

$$\begin{aligned}
\text{CQDE}_{\tau st} &= \sum_{\iota=1}^{r}\sum_{t=1}^{m}\{\gamma_1(t,\iota,\tau)+\gamma_2(t,\iota,\tau)\},\\
\text{CQIE}_{\tau st} &= \sum_{\iota=1}^{r}\sum_{t=2}^{m}\beta(t,\iota,\tau)^\top\left[\sum_{k=1}^{t-1}\left(\prod_{j=k+1}^{t-1}\Phi(j,\iota)\right)\{\Gamma_1(k,\iota)+\Gamma_2(k,\iota)\}\right].
\end{aligned}$$

The estimation and inference procedures can be derived similarly.

6 Real Data Analysis

To address (Q1)-(Q3), we apply the proposed test procedures to the three real datasets obtained from Didi Chuxing introduced in Section 2.

Firstly, we examine the dataset from a temporally dependent A/B experiment conducted from Dec 10, 2021 to Dec 23, 2021. As detailed in Section 2, two order dispatch policies are tested in alternating one-hour time intervals. The new policy, in comparison to the old one, is designed to fulfill more call orders and elevate drivers' total income. As for the choices of the observation variables, our recommended approach is to carefully select state variables that effectively capture the demand and supply dynamics of the ridesharing platform while exerting a substantial influence on the outcome of interest. To bolster the selection process, we advocate subjecting potential state variables to the Ljung-Box test. State variables that pass this test tend to be more fitting; their residuals reflect a temporal independence congruent with our modeling prerequisites. We set drivers' total income as the outcome, and the observation variables include the number of call orders and drivers' total online time. To address question (Q1), we apply model (3) to elucidate the correlation structure between supply and demand and model (4) to elucidate the temporal interference effects. For question (Q3), we utilize the testing procedure described in Section 3.3 for these temporally dependent experiments. As a means to validate the proposed test, we also apply our procedure to the A/A dataset outlined in Section 2, where a single order dispatch strategy is employed. We anticipate that our test will not reject the null hypothesis when applied to this dataset.

In Figure 3, we display the estimated residuals of the outcome over time for $\tau \in \{0.1, 0.5, 0.9\}$ of the A/B experiment. As can be seen from Figure 3, some residuals are significantly larger than others, suggesting that the outcome likely originates from heavy-tailed distributions. This reinforces the use of quantile treatment effects for policy evaluation. We further investigate the correlations of $E_i(t)$ in real-world data scenarios. For the temporally A/B experiment, involving observation variables such as the number of call orders and drivers' total online time, we apply the Ljung-Box test to assess the correlations of the two residual processes. The resulting p-values for the residuals of the number of call orders and drivers' total online time are 0.083 and 0.162, respectively, indicating no/weak autocorrelations over time. Table 1 presents the p-values of the proposed test for CQTE_τ, CQDE_τ, and CQIE_τ, respectively. Furthermore, Figure 4 illustrates the estimated treatment effects and the $p-$values across various quantiles for the A/B experiment. As expected, the proposed test does not reject the null hypothesis at any quantile level when applied to the A/A experiment. However, when applied to the A/B 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. For comparison, we also report the p-values for testing the average direct and indirect effects in Table 1. These p-values

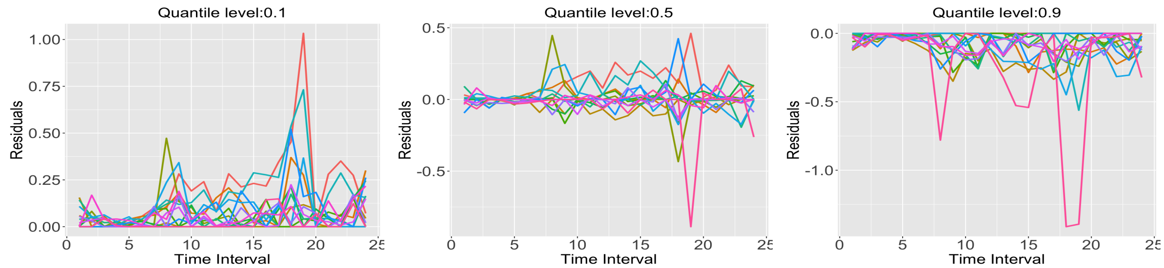


Figure 3: Estimated residuals of drivers' total income at quantile levels 0.1, 0.5, and 0.9 in the temporal dependent A/B experiment.

Table 1: p-values of the proposed test for CQDE_τ and CQIE_τ, as well as p-values of direct effect and indirect effect for average effects for both datasets from the A/A experiment and A/B experiment, utilizing the time-alternation design.

	pvalues for AA		pvalues for AB	
τ	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	0.530	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
average effect	0.800	0.220	0.046	0.956

are calculated by replacing the quantile function in the proposed test procedure with the mean function. Similar to the proposed test, in the A/A experiment, both the direct effect and the indirect effect are not significant. For the A/B experiment, the direct effect is significant at 5% significance level and the indirect effect is not significant. To the contrary, the proposed quantile-based test suggests that the direct effect is significant only at higher quantile levels (when $\tau \geq 0.2$). This highlights the strengths of our test. Namely, it enables us to evaluate treatment effects across different quantiles, thereby providing a richer and more comprehensive understanding than a sole focus on the average effect would allow.

Secondly, we analyze the dataset from the spatiotemporal dependent experiment as described in Section 2. Recall that in this experiment, the city is divided into 12 regions. Policies are implemented based on alternating 30-minute time intervals within each region. We concentrate on a data subset collected from 7 am to midnight each day, as there are relatively few order requests from midnight to 7 am. The drivers' total income and the

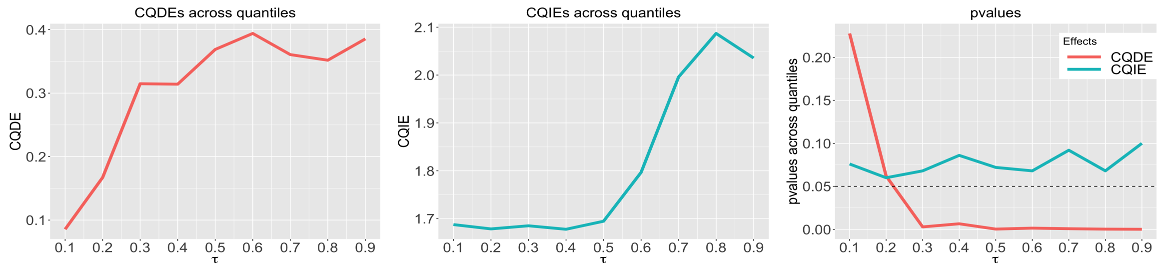


Figure 4: Estimates of CQDE_τ and CQIE_τ and their p-values across quantile levels for the A/B experiment under the temporal design.

number of call orders are designated as the outcome and state variable, respectively. We fit the spatiotemporal VCDP models (14) and (15) to address (Q2), and apply the testing procedure from Section 4.3 to address (Q3) for this spatiotemporal dependent experiment. Our aim is to determine whether the new policy has significant treatment effects on drivers' total income across various quantile levels.

Before we fit the models, we conduct the conditional independence test the conditional independence between $Y_{i,t,\iota}$ and treatments from the neighboring regions given $\{S_{i,t,\iota}, A_{i,t,\iota}, \bar{A}_{i,t,\mathcal{N}_\iota}\}$ in each region, leading to 12 p-values. The minimal p-value of the 12 regions is 0.144, indicating that the mean-filed assumption holds for this dataset. For each quantile level, we implement the proposed estimation and testing procedures on the data. The p-values are generated through the bootstrap procedure outlined in Section 4, utilizing 500 bootstrap samples. The estimation and testing results for $\text{CQDE}_{\tau st}$ and $\text{CQIE}_{\tau st}$ are summarized in Table 2 and Figure 5. The treatment effects are significant at most quantile levels, and both the estimated direct and indirect effects are positive across all quantiles. Generally, these effects escalate with the quantile level. However, the new policy doesn't seem to boost the lower quantile of the outcome (e.g., $\tau = 0.1$). These results underline the heterogeneous effects of the new policy across different quantile levels. Furthermore, we calculate the p-values for testing the average direct and indirect effects and report them in Table 2. While both average effects are found to be statistically significant, our proposed quantile-based test reveals that the direct effect is significant only when $\tau \geq 0.1$. Additionally, while the mean of the estimated quantile effects—aggregated across various quantile levels—closely approximates the ATE, these effects exhibit variability around ATE across different quantiles. Once again, these findings underscore the merits of our proposed quantile-based approach, which facilitates a more nuanced and comprehensive understanding

Table 2: p-values and estimators of $\text{CQDE}_{\tau st}$ and $\text{CQIE}_{\tau st}$ for the spatiotemporal data, as well as p-values of direct effect and indirect effect for average effects.

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
average effect	0.001	0.040	5.525	16.496

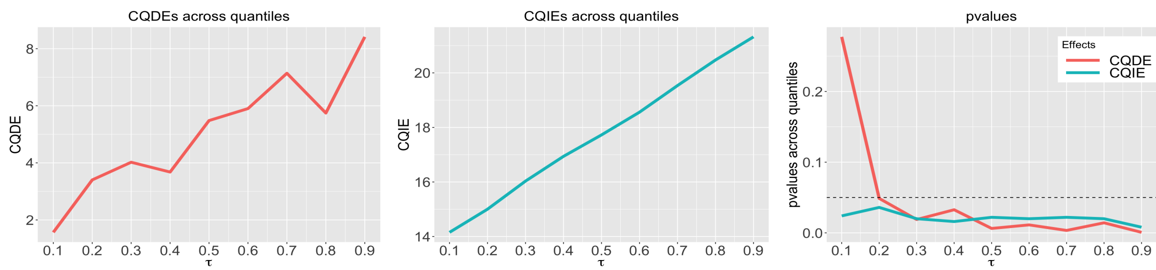


Figure 5: Estimates of $\text{CQDE}_{\tau st}$ and $\text{CQIE}_{\tau st}$ and pvalues for the spatiotemporal data across quantiles.

of treatment effects across various quantile levels.

Finally, we display the scaled outcomes, and residuals for the representative region 5 over time, with $\tau \in \{0.1, 0.5, 0.9\}$, in Figure S4. It is evident that there may be several outliers in the data. This observation further supports the use of quantiles as the evaluation metric. Similar patterns are observed for other regions as well.

7 Real data based simulations

In this section, we evaluate the finite sample performance of the proposed estimation and testing procedures through simulations. Simulation experiments are conducted based on the real dataset collected from the A/A experiment described in Section 2. Recall that one hour is defined as a time unit, and drivers' total income within each time unit is set as the outcome of interest. The observation variables correspond to the number of call orders and drivers' total online time. These variables characterize the demand and supply of the ridesharing platform and have a substantial impact on the outcome.

Example 1. In this example, we investigate the finite sample performance of the proposed test CQTE, CQDE and CQIE, respectively. For a given quantile level τ, we fit the proposed VCDP models (3) and (4) to the data by setting $\gamma(t,\tau) = \Gamma(t) = 0$, since the two policies being compared are essentially the same. This enables us to obtain the estimated model parameters $\widetilde{\beta}_{0\tau}(t)$, $\widetilde{\beta}_{\tau}(t)$, $\widetilde{\phi}_0(t)$, and $\widetilde{\Phi}(t)$ and the estimated error processes $\widetilde{e}_i(t,\tau)$ and $\widetilde{E}_i(t)$ for $1 \leq t \leq 24$ and $1 \leq i \leq 68$. To simulate data, we set $\widetilde{\gamma}(t,\tau) = \delta Q_\tau(Y_t)$ and $\widetilde{\Gamma}(t) = \delta E(S_t)$ for some constant $\delta \geq 0$, where $Q_\tau(Y_t)$ and $E(S_t)$ denote the (elementwise) empirical τ-th quantile of $\{Y_{it}\}_i$ and empirical mean of $\{S_{it}\}_i$, respectively. The constant δ controls the strength of CQTE, no CQTE exists if $\delta = 0$, and the new policy is better if $\delta > 0$.

We employ the bootstrap method for data generation. Specifically, in each simulation run, we randomly sample n initial observations and n error processes with replacement. Then, we generate n days of data according to the proposed VCDP models:

$$
\begin{aligned}
\widetilde{Y}_{i,t} &= \widetilde{\beta}_0(t,\tau) + \widetilde{S}_{i,t}^\top \widetilde{\beta}(t,\tau) + A_{i,t}\widetilde{\gamma}(t,\tau) + \widetilde{e}_i(t,\tau), \\
\widetilde{S}_{i,t+1} &= \widetilde{\phi}_0(t) + \widetilde{\Phi}(t)\widetilde{S}_{i,t} + A_{i,t}\widetilde{\Gamma}(t) + \widetilde{E}_i(t+1),
\end{aligned}
$$

based on these samples and the estimated model parameters. The treatments $A_{i,t}$ are generated according to the temporal alternation design. Specifically, we first implement one policy for TI time units, then switch to the other policy for another TI time units, and alternate between the two policies. We consider a wide range of simulation settings by setting $\tau \in \{0.2, 0.5, 0.8\}$, $n \in \{20, 40\}$, $\text{TI} \in \{1, 3\}$, and $\delta \in \{0, 0.01, 0.025, 0.05, 0.075, 0.1\}$. For each scenario, we generate 500 simulation runs to compute the empirical type-I error rate and power. The significance level is fixed at 5% throughout the simulation.

Figure 6 presents the empirical rejection rates of the proposed test for CQTE (refer also to Table S1 in the supplementary material). The type-I error is around the nominal level in all cases. The empirical power generally increases with the sample size and approaches 1 as the signal strength δ increases to 0.1. Furthermore, the empirical power increases with the quantile level τ, which is expected since $\widetilde{\gamma}(\tau,t)$ is set to be proportional to $Q_\tau(Y_t)$, whose values increase with the quantile level. These results validate our theoretical assertions.

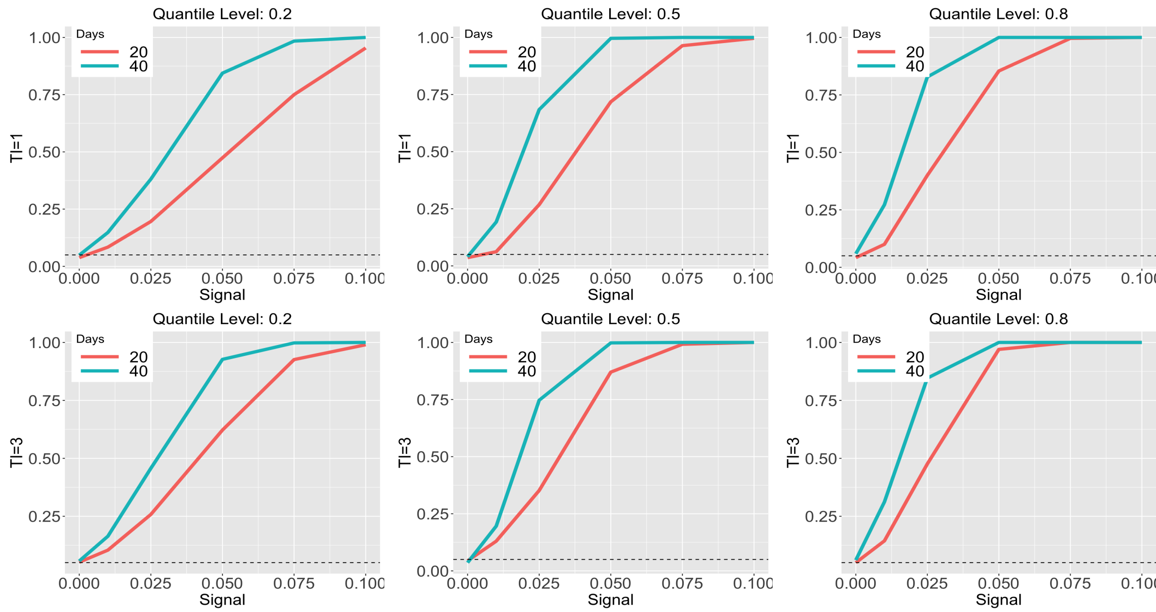


Figure 6: Empirical rejection rates of the proposed test for CQTE_τ. TI equals 1 for the top panels and 3 for the bottom panels. The quantile level $\tau = 0.2$, 0.5 and 0.8, from left to right plots.

We also report the empirical rejection rates of the proposed test for CQDE and CQIE in Figures S1 and S2 of the supplementary material, respectively. The results are very similar to those of CQTE. It is worth noting that the power for CQDE is generally larger than that of CQTE, whereas the power for CQIE is generally smaller than that for CQTE. This is because the test statistics of CQIE have larger variances than those of CQDE.

Example 2. In this example, following a suggestion from one of the reviewers, we compare the proposed CQTE test with two baseline methods. The first method, denoted as "NoInterference", ignores temporal interference and treats each time interval as independent. The second method is designed to test the average treatment effect (ATE), which is commonly used in ridesharing platforms. We conduct the ATE test by replacing the quantile function in the proposed test procedure with the mean function.

The data settings are similar to those in Example 1, with the exception that we scale both the outcome and observation variables by their standard deviation. Additionally, we fix $\widetilde{\gamma}(t,\tau) = 0$ and generate data from the following heterogeneous VCDP model,

$$\begin{aligned}
\widetilde{Y}_{i,t} &= \widetilde{\beta}_0(t,\tau) + \widetilde{S}_{i,t}^\top \widetilde{\beta}(t,\tau) + A_{i,t}\widetilde{\gamma}(t,\tau) + 0.2(\widetilde{S}_{i,t}^{(1)})^2 \widetilde{e}_i(t,\tau), \\
\widetilde{S}_{i,t+1} &= \widetilde{\phi}_0(t) + \widetilde{\Phi}(t)\widetilde{S}_{i,t} + A_{i,t}\widetilde{\Gamma}(t) + 0.2(\widetilde{S}_{i,t}^{(1)})^2 \widetilde{E}_i(t+1),
\end{aligned}$$

where $S_{i,t}^{(1)}$ represents the first observational covariate. Consequently, the error processes at each time point are heteroscedastic.

Figure 7 displays the empirical rejection rates of the proposed CQTE test and the NoInterference method for CQTE, as well as the test for ATE when $n = 40$. The results for $n = 20$ can be found in the supplementary material and exhibit a similar pattern to the $n = 40$ results. Notably, the NoInterference method fails to detect the treatment effect in all scenarios, as expected, since the treatment has no direct effect but exhibits an indirect effect on the outcome. In contrast, both the proposed CQTE test and the ATE test not only maintain Type-I error control but also effectively identify causal effects. It's important to note that heteroscedastic errors can impact the estimation of $\widetilde{\beta}(t, \tau)$ and subsequently affect the indirect effect. Since $\widetilde{\beta}(t, \tau)$ behaves differently across various quantile levels, the magnitude of this influence varies accordingly. Specifically, for the lower quantile level $\tau = 0.2$, the ATE test exhibits higher statistical power compared to the proposed method. However, for quantile levels $\tau = 0.5$ and $\tau = 0.8$, the CQTE test demonstrates superior power when contrasted with the ATE test. This difference in power is primarily due to the influence of heteroscedastic errors on estimation and tests based on linear regression. In contrast, the proposed test, which relies on quantile regression, proves to be robust even in the presence of such heteroscedastic errors. These results highlight the value of the proposed test in detecting treatment effects when compared to baseline methods.

Additionally, we conduct another simulation study in Example S2 of the supplementary material where the causal effect exists but ATE=0. In this case, the test for ATE fails to capture the treatment effect, while the proposed method not only captures distributional treatment effects but also unveils distinct treatment effects at different quantile levels.

8 Discussion

As we did not impose any constraints, our proposed estimation procedure may not guarantee monotonicity with respect to the quantile location. In the existing literature, three prevalent approaches are employed to mitigate the issue of crossing quantile curves in quantile estimation methods, the post-processing approach that involves sorting or monotonically rearranging the original functions (e.g. Chernozhukov et al., 2010), the stepwise procedure

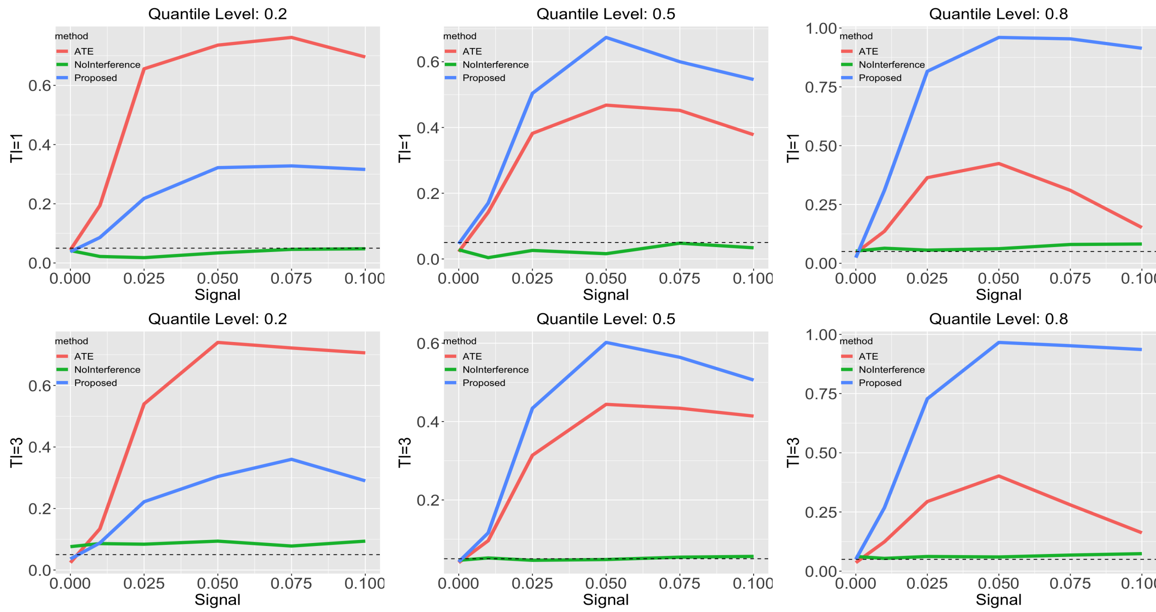


Figure 7: Empirical rejection rates of the proposed test and the NoInterference method for CQTE_τ and the test for ATE when $n = 40$. TI equals 1 for the top panels and 3 for the bottom panels. The quantile level $\tau = 0.2$, 0.5 and 0.8, from left to right plots.

that iteratively adds an extra set of non-crossing constraints to the quantile model (e.g. Andriyana and Gijbels, 2017), and the simultaneous estimation approach that estimates all quantiles concurrently while incorporating non-crossing constraints (e.g. Bondell et al., 2010). To address the crossing issue in our proposed method, we could consider adapting these existing strategies. For example, inspired by Bondell et al. (2010), we could impose a non-crossing restriction such that $z^\top\theta(t, \tau_j) \geq z^\top\theta(t, \tau_{j-1}), j = 2, \ldots, q$ for any z and any desired quantile levels $\tau_1 < \cdots < \tau_q$ when estimating the model parameters via (7). These extensions are worthy of further investigation.

References

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

Andriyana, Y. and I. Gijbels (2017). Quantile regression in heteroscedastic varying coefficient models. *AStA Advances in Statistical Analysis 101*(2), 151–176.

Blanco, G., X. Chen, C. A. Flores, and A. Flores-Lagunes (2020). Bounds on average and quantile treatment effects on duration outcomes under censoring, selection, and noncompliance. *Journal of Business & Economic Statistics 38*(4), 901–920.

Bondell, H. D., B. J. Reich, and H. Wang (2010). Noncrossing quantile regression curve estimation. *Biometrika 97*(4), 825–838.

Chamandy, N. (2016). Experimentation in a ridesharing marketplace. *URL: https://eng. lyft. com/experimentation-in-a-ridesharing-marketplace-b39db027a66e*.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics 132*(2), 491–525.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica 75*(1), 259–276.

Gill, R. D. and J. M. Robins (2001). Causal inference for complex longitudinal data: the continuous case. *The Annals of Statistics 29*(6), 1785–1811.

Horowitz, J. L. and A. Krishnamurthy (2018). A bootstrap method for constructing pointwise and uniform confidence bands for conditional quantile funtions. *Statistica Sinica 28*(4), 2609–2632.

Hu, Y., S. Li, and S. Wager (2022). Average direct and indirect causal effects under interference. *Biometrika 109*(4), 1165–1172.

Hu, Y. and S. Wager (2022). Switchback experiments under geometric mixing. *arXiv preprint arXiv:2209.00197*.

Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482), 832–842.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kallus, N., X. Mao, and M. Uehara (2019). Localized debiased machine learning: Efficient inference on quantile treatment effects and beyond. *arXiv preprint arXiv:1912.12945*.

Kallus, N. and M. Uehara (2022). Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning. *Operations Research 70*(6), 3035–3628.

Koenker, R. and K. F. Hallock (2001). Quantile regression. *Journal of Economic Perspectives 15*(4), 143–156.

Kong, D., S. Yang, and L. Wang (2022). Identifiability of causal effects with multiple causes and a binary outcome. *Biometrika 109*(1), 265–272.

Li, S. and S. Wager (2022). Network interference in micro-randomized trials. *arXiv preprint arXiv:2202.05356*.

Liao, P., P. Klasnja, and S. Murphy (2021). Off-policy estimation of long-term average outcomes with applications to mobile health. *Journal of the American Statistical Association 116*(533), 382–391.

Liu, L. and M. G. Hudgens (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association 109*(505), 288–301.

Liu, M., X. Sun, M. Varshney, and Y. Xu (2019). Large-scale online experimentation with quantile metrics. *arXiv preprint arXiv:1903.08762*.

Liu, Q., L. Li, Z. Tang, and D. Zhou (2018). Breaking the curse of horizon: infinite-horizon off-policy estimation. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 5361–5371.

Porter, J. and P. Yu (2015). Regression discontinuity designs with unknown discontinuity points: Testing and estimation. *Journal of Econometrics 189*(1), 132–147.

Qi, Z., J.-S. Pang, and Y. Liu (2022). On robustness of individualized decision rules. *Journal of the American Statistical Association (Accepted)*.

Qin, Z., X. Tang, Y. Jiao, F. Zhang, Z. Xu, H. Zhu, and J. Ye (2020). Ride-hailing order dispatching at didi via reinforcement learning. *Informs Journal on Applied Analytics 50*(5), 272–285.

Qin, Z. T., H. Zhu, and J. Ye (2022). Reinforcement learning for ridesharing: An extended survey. *Transportation Research Part C: Emerging Technologies 144*, 103852.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association 100*(469), 322–331.

Savje, F., P. M. Aronow, and M. G. Hudgens (2021). Average treatment effects in the presence of unknown interference. *The Annals of Statistics 49*(2), 673–701.

Seo, M. H. and O. Linton (2007). A smoothed least squares estimator for threshold regression models. *Journal of Econometrics 141*(2), 704–735.

Sherwood, B. and L. Wang (2016). Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics 44*(1), 288–317.

Shi, C., W. Lu, and R. Song (2020). Breaking the curse of nonregularity with subagging—inference of the mean outcome under optimal treatment regimes. *Journal of Machine Learning Research 21*(176), 1–67.

Shi, C., R. Wan, G. Song, S. Luo, R. Song, and H. Zhu (2022). A multi-agent reinforcement learning framework for off-policy evaluation in two-sided markets. *arXiv preprint arXiv:2202.10574*.

Shi, C., X. Wang, S. Luo, H. Zhu, J. Ye, and R. Song (2022). Dynamic causal effects evaluation in A/B testing with a reinforcement learning framework. *Journal of the American Statistical Association (Accepted)*.

Shi, C., S. Zhang, W. Lu, and R. Song (2020). Statistical inference of the value function for reinforcement learning in infinite horizon settings. *arXiv preprint arXiv:2001.04515*.

Tchetgen, E. J. T. and T. J. VanderWeele (2012). On causal inference in the presence of interference. *Statistical methods in medical research 21*(1), 55–75.

Uehara, M., H. Kiyohara, A. Bennett, V. Chernozhukov, N. Jiang, N. Kallus, C. Shi, and W. Sun (2022). Future-dependent value-based off-policy evaluation in pomdps. *arXiv preprint arXiv:2207.13081*.

Van Der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer.

Wang, H. and H. Yang (2019). Ridesourcing systems: A framework and review. *Transportation Research Part B: Methodological 129*, 122–155.

Wang, H. J., Z. Zhu, and J. Zhou (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics 37*(6B), 3841–3866.

Wang, L., Y. Zhou, R. Song, and B. Sherwood (2018). Quantile-optimal treatment regimes. *Journal of the American Statistical Association 113*(523), 1243–1254.

Wang, W. and X. Zhang (2021). Conq: Continuous quantile treatment effects for large-scale online controlled experiments. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 202–210.

Xiong, R., A. Chin, and S. Taylor (2023). Bias-variance tradeoffs for designing simultaneous temporal experiments. In *The KDD'23 Workshop on Causal Discovery, Prediction and Decision*, pp. 115–131. PMLR.

Yang, F., A. Ramdas, K. G. Jamieson, and M. J. Wainwright (2017). A framework for multi-A (rmed)/B (andit) testing with online FDR control. *Advances in Neural Information Processing Systems 30*.

Zhang, B., A. A. Tsiatis, E. B. Laber, and M. Davidian (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika 100*(3), 681–694.

Zhang, Z., X. Wang, L. Kong, and H. Zhu (2022). High-dimensional spatial quantile function-on-scalar regression. *Journal of the American Statistical Association 117*(539), 1563–1578.

Zhao, Z., X. Chen, X. Zhang, and Y. Zhou (2022). Dynamic car dispatching and pricing: Revenue and fairness for ridesharing platforms. *arXiv preprint arXiv:2207.06318*.

Zhou, F., J. Wang, and X. Feng (2020). Non-crossing quantile regression for distributional reinforcement learning. *Advances in Neural Information Processing Systems 33*, 15909–15919.

Zhou, Y., Y. Liu, P. Li, and F. Hu (2020). Cluster-adaptive network A/B testing: From randomization to estimation. *arXiv preprint arXiv:2008.08648*.

Zhu, H., J. Fan, and L. Kong (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association 109*(507), 1084–1098.