Policy Evaluation for Temporal and/or Spatial Dependent Experiments

Shikai Luo[a*], Ying Yang[b*], Chengchun Shi[c*], Fang Yao[d],

ridesharing applications, for instance, most AB test experiment durations do not exceed 20 days (Shi et al., 2023), and the size of treatment effects typically ranges between 0.5% and 2% (Tang et al., 2019).

The primary objective of this paper is to develop a robust statistical framework for analyzing the causal connections between the policies implemented by these companies and their corresponding outcomes, even in the presence of the aforementioned challenges. Our four major contributions can be summarized as follows. Firstly, we address the challenges by introducing linear and neural network-based Varying Coefficient Decision Process (VCDP) models. These models accommodate dynamic treatment effects over time and/or space, even in the presence of non-stationarity, random effects, interference, and spatial spillovers. These models account for market features as mediators to incorporate historical policy carryover effects. Furthermore, by assuming network interference and employing mean field approximation (as detailed in Section 3.2), we effectively operate an "effective treatment" (Manski, 2013) or "exposure mapping" (Aronow and Samii, 2017) in the spatio-temporal system. Our approach extends beyond the switchback design to any dynamic treatment allocation setup.

Secondly, we develop estimation methods for our VCDPs. For linear VCDPs, we propose a two-step process involving the calculation of least squares estimates and kernel smoothing to refine the estimates. Kernel smoothing leverages neighboring observations across time and/or space, enhancing estimation efficiency and overcoming the challenge of weak signals and small sample sizes. Additionally, we decompose average treatment effects (ATEs) into Direct Effects (DE) and Indirect Effects (IE).

et al., 2020). Our VCDPs are closely related to the second and third types of models, but they focus on interference across time *and* space. Most aforementioned works studied the interference effect across time *or* space and were motivated by research questions in environmental and epidemiological studies. It remains unknown about their generalization to ride-sharing markets. Fourthly, recent models capture the interference effect via congestion or price effects in a marketplace (Munro et al., 2021; Wager and Xu, 2021; Johari et al., 2022). These solutions rely on an assumption of Markovanity or stationarity and are design-dependent. In contrast, our approach accommodates non-stationarity and is capable of managing non-Markovianity in scenarios where outcome errors exhibit time-correlated patterns.

Our proposal is closely related to a growing literature on off-policy evaluation (OPE) methods in sequential decision making (see Uehara et al., 2022, for a review). In the literature, augmented inverse propensity score weighting methods (see e.g., Zhang et al., 2013; Luedtke and Van Der Laan, 2016; Jiang and Li, 2016; Thomas and Brunskill, 2016) have been proposed for valid OPE. Nonetheless, these methods suffer from the curse of horizon (Liu et al., 2018) in that the variance of the resulting estimator grows exponentially fast with respect to m, leading to inefficient estimates in the large m setting. Efficient model-free OPE methods have been proposed by Kallus and Uehara (2020, 2022); Liao et al. (2020, 2021); Luckett et al. (2020); Shi et al. (2021, 2022b) under the Markov decision process (MDP, see e.g., Puterman, 2014) model assumption. Recently, Hu and Wager (2021) proposed a model-free OPE method in partially observed MDPs (POMDPs) that avoids the curse of horizon. Our proposal is model-based and is ultimately different from most existing model-free OPE methods that did not consider the random effects, spatial interference effects, and the decomposition into DE and IE. In addition, little has been done on OPE for spatio-temporal dependent experiments.

Finally, our paper is related to a line of works on quantitative approaches to ride-sharing platforms. In particular, Bimpikis et al. (2019) proposed supply-and-demand models and investigated the impact of the demand pattern on the platform's prices and profits. Castillo et al. (2017) studied how the surging prices can prevent wild goose chase (e.g., drivers pick up distant customers) and conducted regression analysis to verify the nonmonotonicity of supply on pickup times. However, estimation and inference of target policy's treatment effect have not been considered in these papers. Cohen et al. (2022) employed the difference in differences methods to estimate the treatment effects of different types of compensation on the engagement of riders who experienced a frustration. Their analysis is limited to staggered designs. Garg and Nazerzadeh (2022) studied the theoretical properties of driver-side payment mechanisms and compared additive surge against multiplicative surge numerically. However, they did not consider the spatial spillover effects of these policies. Our paper complements the existing literature by developing a general framework to efficiently infer a target policy's direct and indirect effects based on data collected from spatio-temporal dependent experiments and analyzing the advantage of switchback designs in the presence of spatio-temporal random effects.

1.2. Paper outline

The rest of the paper is organized as follows. In Section 2, we introduce a potential outcome framework for problem formulation, propose two novel temporal VCDP models under temporal dependent experiments, and develop estimation and testing procedures for both DE and IE. In Section 3, we further propose two spatio-temporal VCDP models under spatio-temporal dependent experiments and develop the associated estimation and testing procedures. In Section 4, we systematically investigate the theoretical properties of estimation and testing procedures (e.g., consistency and power) developed in Sections 2 and 3. We also illustrate the benefits of employing the switchback design in theory. In Section 5, we use numerical simulations to examine the finite sample performance of our estimation and testing procedures. Furthermore, we numerically explore the benefits of the switchback design. In Section 6, we apply the proposed procedures to evaluating different policies in Didi Chuxing.

2. Policy evaluation for temporal dependent experiments

In this section, we present the proposed methodology for policy evaluation in temporal dependent experiments for one experimental region.

2.1. A potential outcome framework

We use the potential outcome framework to present our model in non-stationary environments. We divide each day into m equally spaced nonoverlapping intervals. At each time interval, the platform can

implement either the new or old policy. We use A_τ to denote the policy implemented at the τth interval for any integer $\tau \geq 1$. Let S_τ be some state variables measured at the $(\tau - 1)$-th interval in a given day. All the states share the same support, which is assumed to be a compact subset of $\mathbb{R}^d$, where d denotes the dimension of the state. Let $Y_\tau \in \mathbb{R}$ be the outcome of interest measured at time τ.

Firstly, we define the average treatment effect (ATE) as the difference between the new and old policies. Let $\bar{a}_\tau = (a_1, \ldots, a_\tau)^\top \in \{0, 1\}^\tau$ denote a treatment history vector up to time τ, where 1 and 0 denote the new policy and the old one, respectively. We define $S^*_\tau(\bar{a}_{\tau-1})$ and $Y^*_\tau(\bar{a}_\tau)$ as the counterfactual state and the counterfactual outcome, respectively. Then ATE can be defined as follows.

DEFINITION 1. *ATE is the difference between two value functions given by*

$$ATE = \sum_{\tau=1}^{m} \mathbb{E}\{Y^*_\tau(\mathbf{1}_\tau) - Y^*_\tau(\mathbf{0}_\tau)\},$$

where $\mathbf{1}_\tau$ and $\mathbf{0}_\tau$ denote vectors of 1s and 0s of length τ, respectively.

Secondly, we can decompose ATE as the sum of direct effects (DE) and indirect effects (IE). Let R_τ denote the conditional mean function of the outcome given the data history,

$$\mathbb{E}\{Y^*_\tau(\bar{a}_\tau) | S^*_\tau(\bar{a}_{\tau-1}), Y^*_{\tau-1}(\bar{a}_{\tau-1}), S^*_{\tau-1}(\bar{a}_{\tau-2}), Y^*_{\tau-2}(\bar{a}_{\tau-2}), \ldots, S_1\} = R_\tau(a_\tau, S^*_\tau(\bar{a}_{\tau-1}), a_{\tau-1}, S^*_\tau(\bar{a}_{\tau-2}), \ldots, S_1).$$

LEMMA 1. *Under CA, SRA and PA, we have*

$$R_\tau(a_\tau, s_\tau, \ldots, s_1) = \mathbb{E}(Y_\tau | A_\tau = a_\tau, S_\tau = s_\tau, \ldots, S_1 = s_1), \quad (4)$$
$$\mathbb{E}\{R_\tau(a, S^*_\tau(\bar{a}_{\tau-1}), \ldots, S_1)\} = \mathbb{E}[\mathbb{E}[R_\tau(a, S_\tau, \ldots, S_1) | \{A_j = a_j\}_{1\le j<\tau}, \{S_j, Y_j\}_{1\le j<\tau}]]. \quad (5)$$

Lemma 1 implies that the causal estimand can be represented as a function of the observed data.

2.2. TVCDP model

We introduce two TVCDP regression models to model $Y_{i,\tau}$ and the conditional distribution of $S_{i,\tau}$ given the data history, forming the basis of our estimation and testing procedures. Suppose that the experiment is conducted over n days. Let $(S_{i,\tau}, A_{i,\tau}, Y_{i,\tau})$ be the state-policy-outcome triplet measured at the τth time interval of the ith day for $i = 1, \ldots, n$ and $\tau = 1, \ldots, m$. The proposed TVCDP model is composed of the following set of additive noise models,

$$\begin{aligned} Y_{i,\tau} &= f_{1,\tau}(S_{i,\tau}, A_{i,\tau}) + e_{i,\tau}, \\ S_{i,\tau+1} &= f_{2,\tau}(S_{i,\tau}, A_{i,\tau}) + \varepsilon_{i,\tau S}, \end{aligned} \quad (6)$$

where $f_{1,\tau}(\cdot)$ and $f_{2,\tau}(\cdot)$ are the regression functions.

We would like to highlight several key points. Firstly, in addition to defining the standard outcome regression model $f_{1,\tau}$ as described in equation (6), it is crucial to specify how past actions influence future states. This is accomplished through the inclusion of $f_{2,\tau}$, which plays a pivotal role in quantifying temporal interference effects.

Secondly, we introduce a specific assumption related to the error structure. This assumption is fundamental as it allows us to incorporate temporal random effects effectively.

ASSUMPTION 1. *(i) The outcome noise $e_{i,\tau} = \eta_{i,\tau} + \varepsilon_{i,\tau}$ is a combination of two mutually independent stochastic processes: day-specific temporal variation $\eta_{i,\tau}$ and measurement error $\varepsilon_{i,\tau}$. (ii) The processes $\{\eta_{i,\tau}\}_{i,\tau}$ are identical realizations of a zero-mean stochastic process with covariance function $\{\Sigma_\eta(\tau_1, \tau_2)\}_{\tau_1,\tau_2}$. Additionally, all components of $\Sigma_\eta(t_1, t_2)$ have bounded and continuous second derivatives with respect to t_1 and t_2. (iii) The measurement errors $\{\varepsilon_{i,\tau}\}_{i,\tau}$ and $\{\varepsilon_{i,\tau S}\}_{i,\tau}$ are independent over time. They have zero mean values and exhibit $Var(\varepsilon_{i,\tau}) = \sigma^2_{\varepsilon,\tau}$ and $Cov(\varepsilon_{i,\tau S}) = \Sigma_{\varepsilon,\tau S}$.*

It's important to note that the day-specific random effects are present only in the outcome regression model. However, our approach can be extended to scenarios where these random effects also exist in the state regression model. We provide a detailed discussion of this extension in Section 7. Additionally, it's worth mentioning that both the conditional mean and covariance functions, namely $f_{1,\tau}$, $f_{2,\tau}$, $\sigma^2_{\varepsilon,\tau}$, and $\Sigma_{\varepsilon,\tau S}$, are time-dependent. This captures the nonstationarity inherent in the data generating process.

Our TVCDP models (6) have strong connections with the MDP model that is commonly used in reinforcement learning. Specifically, models (6) reduce to non-stationary (or time-varying) MDP models (Kallus and Uehara, 2022) when there are no day-specific random effects in $\{e_{i,\tau}\}_{i,\tau}$. However, the proposed time varying models are no longer MDPs due to the existence of the day-specific random effects. In particular, $Y_{i,\tau}$ in (6) is dependent upon past responses given $Z_{i,\tau} = (1, S^\top_{i,\tau}, A_{i,\tau})^\top$, leading to the violation of the conditional independence assumption. In addition, the market features at each time serve as mediators that mediate the effects of past actions on the current outcome.

Next, we consider two specific function approximations for f_1 and f_2 and derive their related IE and DE as follows.

MODEL 1. *Linear temporal varying coefficient decision process (L-TVCDP) assumes*

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

where $\theta(\tau) = (\beta_0(\tau), \beta(\tau)^\top, \gamma(\tau))^\top$ is a $(d+2)\times 1$ vector of time-varying coefficients, $\Theta(\tau) = [\phi_0(\tau)\ \ \Phi(\tau)\ \ \Gamma(\tau)]$ is a $d \times (d+2)$ coefficient matrix and $Z_{i,\tau} = (1, S^\top_{i,\tau}, A_{i,\tau})^\top$.

Model 1 shares a close connection with the linear quadratic Gaussian model (LQG), well studied in the fields of RL and control theory (see, for example, Lale et al., 2021). To be more specific, Model 1 can be seen as a simplified, one-dimensional observation variant of LQG under certain conditions. This happens when the outcome regression model doesn't incorporate $A_{i,\tau}$ and the autocorrelated noise $\eta_{i,\tau}$. However,

there's a crucial distinction between LQG and our proposed model. In LQG, the state variables are hidden and must be deduced from the observed $Y_{i,\tau}$ values. This contrasts with similar models used in literature for estimating dynamic treatment effects (Lewis and Syrgkanis, 2020).

When $f\eta_{i,\tau}g_{i,\tau}$ become the fixed effects and satisfy $\eta_{i,\tau} = \eta_i$ for any i and τ, the outcome regression model of L-TVCDP includes both the day-specific fixed effects $f\eta_i g_i$ and the time-specific fixed effects $f\beta_0(\tau)g_\tau$. It is similar to the two-way fixed effects model in the panel data literature (De Chaisemartin and d'Haultfoeuille, 2020; Wooldridge, 2021; Arkhangelsky et al., 2021; Imai and Kim, 2021). Furthermore, we derive the closed-form expressions for DE and IE under L-TVCDP, whose proof can be found in Section S.3 of the supplementary document.

PROPOSITION 1. *Under the L-TVCDP model, we have* $DE = \sum_{\tau=1}^{m} \gamma(\tau)$ *and*

$$IE = \sum_{\tau=2}^{m} \beta(\tau)^\top \left\{ \sum_{k=1}^{\tau-1} \left(\Phi(\tau-1)\Phi(\tau-2)\dots\Phi(k+1)\right)\Gamma(k) \right\}, \tag{7}$$

where by convention, the product $\Phi(\tau-1)\Phi(\tau-2)\dots\Phi(k+1) = 1$ *when* $\tau - 1 < k+1$.

MODEL 2. *Neural networks temporal varying decision process (NN-TVCDP) assumes*

$$\begin{aligned} Y_{i,\tau} &= g_0(\tau, S_{i,\tau}) \cdot \mathbb{I}(A_{i,\tau} = 0) + g_1(\tau, S_{i,\tau}) \cdot \mathbb{I}(A_{i,\tau} = 1) + e_{i,\tau}, \\ S_{i,\tau+1} &= G_0(\tau, S_{i,\tau}) \cdot \mathbb{I}(A_{i,\tau} = 0) + G_1(\tau, S_{i,\tau}) \cdot \mathbb{I}(A_{i,\tau} = 1) + \varepsilon_{i,\tau S}, \end{aligned}$$

where $\mathbb{I}(\cdot)$ *denotes the indicator function of an event and* $g_0(\cdot,\cdot)$, $g_1(\cdot,\cdot)$, $G_0(\cdot,\cdot)$, *and* $G_1(\cdot,\cdot)$ *are parametrized via some (deep) neural networks.*

Under NN-TVCDP, DE and IE are, respectively, given by

$$\mathrm{DE} = \sum_{\tau=1}^{m} \mathbb{E}\left\{ g_1\left(\tau, S_\tau^0\right) - g_0\left(\tau, S_\tau^0\right) \right\} \quad \text{and} \quad \mathrm{IE} = \sum_{\tau=1}^{m} \mathbb{E}\left\{ g_1\left(\tau, S_\tau^1\right) - g_1\left(\tau, S_\tau^0\right) \right\}, \tag{8}$$

where S_τ^0 and S_τ^1 are defined recursively by $S_\tau^0 = G_0(\tau - 1, S_{\tau-1}^0)$ and $S_\tau^1 = G_1(\tau - 1, S_{\tau-1}^1)$.

2.3. Estimation and testing procedures for DE in the L-TVCDP model

We describe our estimation and testing procedures for DE in the L-TVCDP model and present their pseudocode in Algorithm 1 as follows.

Algorithm 1 Inference of DE in the L-TVCDP model

1: Compute the OLS estimator $\widehat{\boldsymbol{\theta}}$ according to (9).
2: Employ kernel smoothing to compute a refined estimator $\widetilde{\boldsymbol{\theta}}$ according to (10) and calculate the estimate $\widehat{\mathrm{DE}}$ by (11).
3: Estimate the variance of $\widehat{\boldsymbol{\theta}}$ as follows:
4: (3.1). Estimate the conditional variance of $\boldsymbol{Y}_i$ given $fZ_{i,\tau}g_\tau$ using (12);
5: (3.2). Estimate the variance of $\widehat{\boldsymbol{\theta}}$ by the sandwich estimator (13).
6: Estimate the variance of $\widetilde{\boldsymbol{\theta}}$ by $\widetilde{\boldsymbol{V}}_\theta = \boldsymbol{\Omega}\widehat{\boldsymbol{V}}_\theta\boldsymbol{\Omega}^\top$ and compute the standard error of $\widehat{\mathrm{DE}}$, denoted by $\widehat{se}(\widehat{\mathrm{DE}})$.
7: Reject H_0^{DE} if $\widehat{\mathrm{DE}}/\widehat{se}(\widehat{\mathrm{DE}})$ exceeds the upper αth quantile of a standard normal distribution.

Step 1 of Algorithm 1 is to obtain an initial estimator of $\theta(\tau)$ by computing its ordinary least squares (OLS) estimator, defined as

$$\widehat{\theta}(\tau) = (\sum_{i=1}^{n} Z_{i,\tau} Z_{i,\tau}^\top)^{-1} (\sum_{i=1}^{n} Z_{i,\tau} Y_{i,\tau}) \quad \text{for} \;\; 1 \le \tau \le m. \tag{9}$$

Step 2 of Algorithm 1 is to employ kernel smoothing to refine the initial estimator. Specifically, for a given kernel function $K(\cdot)$, we introduce the refined estimator

$$\widetilde{\theta}(\tau) = (\widetilde{\beta}_0(\tau), \widetilde{\beta}(\tau)^\top, \widetilde{\gamma}(\tau))^\top = \sum_{\tau=1}^{m} \omega_{\tau,h}(t)\widehat{\theta}(\tau), \tag{10}$$

for any t 2 $[0, m]$ and a bandwidth parameter h, where $\omega_{\tau,h}(t) = K((t-\tau)/(mh))/\sum_{j=1}^{m} K((t-j)/(mh))$ is the weight function. Our DE estimator is given by

$$\widehat{\mathrm{DE}} = \sum_{\tau=1}^{m} \widetilde{\gamma}(\tau). \tag{11}$$

We will show in Section 4 that as $\min(n, m)$! 1, $\widehat{\mathrm{DE}}$ is asymptotically normal. To derive a Wald test for (2), it remains to estimate its variance $\mathrm{Var}(\widehat{\mathrm{DE}})$.

There are two major advantages of using the smoothing step here. First, it allows us to estimate the time-varying coefficient curve $\theta(t)$ without restricting t to the class of integers. Second, the smoothed estimator has smaller variance, leading to a more powerful test statistics. To elaborate, according to model (6) for L-TVCDP, the variation of the OLS estimator comes from two sources, the day-specific random effect and the measurement error. The use of smoothing removes the random fluctuations due to the measurement error. See Theorem 1 in Section 4 for a formal statement. This smoothing technique has been widely applied in the analysis of varying-coefficient models (see e.g., Zhu et al., 2014).

Step 3 of Algorithm 1 is to estimate the covariance matrix of the initial estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}^{\top}(1), \ldots, \widehat{\theta}^{\top}(m))^{\top}$. We first estimate the residual $e_{i,\tau}$ by $\widehat{e}_{i,\tau} = Y_{i,\tau} - Z_{i,\tau}^{\top}\widetilde{\theta}(\tau)$. It allows us to estimate the day-specific random effect via smoothing, i.e., $\widehat{\eta}_i(t) = \sum_{j=1}^{m} \omega_{j,h}(t)\widehat{e}_{i,\tau}$. Second, the measurement error can be estimated by $\widehat{\varepsilon}_{i,\tau} = \widehat{e}_{i,\tau} - \widehat{\eta}_{i,\tau}$ for any i and τ, where $\widehat{\eta}_{i,\tau} = \widehat{\eta}_i(\tau)$. Third, we estimate the conditional covariance matrix of $\boldsymbol{Y}_i = (Y_{i,1}, \ldots, Y_{i,m})^{\top}$ given $\mathsf{f}Z_{i,\tau}\mathsf{g}_{\tau}$ based on these estimated residuals. Under model (6) for L-TVCDP, the covariance between Y_{i,τ_1} and Y_{i,τ_2} conditional on $\mathsf{f}Z_{i,\tau}\mathsf{g}_{\tau}$ is given by $\Sigma_y(\tau_1, \tau_2) = \sigma^2_{\varepsilon,\tau_1}\mathbb{I}(\tau_1 = \tau_2) + \Sigma_{\eta}(\tau_1, \tau_2)$, which can be consistently estimated by

$$\widehat{\Sigma}_y(\tau_1, \tau_2) \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}^2_{i,\tau_1}\mathbb{I}(\tau_1 = \tau_2) + \frac{1}{n}\sum_{i=1}^{n} \widehat{\eta}_{i,\tau_1}\widehat{\eta}_{i,\tau_2}. \tag{12}$$

This allows us to estimate $\mathrm{Var}(\boldsymbol{Y}_i|\mathsf{f}Z_{i,\tau}\mathsf{g}_{\tau})$ by $\widehat{\boldsymbol{\Sigma}} = \mathsf{f}\widehat{\Sigma}_y(\tau_1, \tau_2)\mathsf{g}_{\tau_1,\tau_2}$. Finally, the covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be consistently estimated by the sandwich estimator,

$$\widehat{\boldsymbol{V}}_{\theta} = (\sum_{i=1}^{n} \boldsymbol{Z}_i^{\top}\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^{n} \boldsymbol{Z}_i^{\top}\widehat{\boldsymbol{\Sigma}}\boldsymbol{Z}_i)(\sum_{i=1}^{n} \boldsymbol{Z}_i^{\top}\boldsymbol{Z}_i)^{-1}, \tag{13}$$

where $\boldsymbol{Z}_i$ is a block-diagonal matrix computed by aligning $Z_{i,1}^{\top}$, …, $Z_{i,m}^{\top}$ along its diagonal.

Step 4 of Algorithm 1 is to estimate the covariance matrix of the refined estimator $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}^{\top}(1), \ldots, \widetilde{\theta}^{\top}(m))^{\top}$. A key observation is that each $\widetilde{\theta}(\tau)$ is essentially a weighted average of $\mathsf{f}\widehat{\theta}(\tau)\mathsf{g}_{\tau}$. Writing in matrix form, we have $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\Omega}\widehat{\boldsymbol{\theta}}$, where $\boldsymbol{\Omega}$ is a block-diagonal matrix computed by aligning $\omega_{1,h}(\tau)\boldsymbol{J}_p$, …, $\omega_{m,h}(\tau)\boldsymbol{J}_p$ along its diagonal and $\boldsymbol{J}_p$ is a $p \times p$ matrix of ones. As such, we estimate the covariance matrix of $\widetilde{\boldsymbol{\theta}}$ by $\widetilde{\boldsymbol{V}}_{\theta} = \boldsymbol{\Omega}\widehat{\boldsymbol{V}}_{\theta}\boldsymbol{\Omega}^{\top}$. This in turn yields a consistent estimator for the variance of $\widehat{\mathrm{DE}}$, as $\widehat{\mathrm{DE}}$ is a linear combination of $\widetilde{\boldsymbol{\theta}}$.

Step 5 of Algorithm 1 is to construct a Wald-type test statistic based on $\widehat{\mathrm{DE}}$ and its standard error $\widehat{se}(\widehat{\mathrm{DE}})$. We reject the null hypothesis in (2) if $\widehat{\mathrm{DE}}/\widehat{se}(\widehat{\mathrm{DE}})$ exceeds the upper αth quantile of a standard normal distribution. Size and power properties of the proposed test are investigated in Section 4.

2.4. Estimation and testing procedures for IE in the L-TVCDP model

We describe our estimation and testing procedures for IE in the L-TVCDP model and present their pseudocode in Algorithm 2 as follows.

Algorithm 2 Inference of IE in the L-TVCDP model

1: Compute the OLS estimator

$$\widehat{\boldsymbol{\Theta}} = \mathsf{f}\widehat{\Theta}(1), \ldots, \widehat{\Theta}(m-1)\mathsf{g}^{\top} = \mathsf{f}\sum_{i=1}^{n} \boldsymbol{Z}_{i,(-m)}\boldsymbol{Z}_{i,(-m)}^{\top}\mathsf{g}^{-1}\mathsf{f}\sum_{i=1}^{n} \boldsymbol{Z}_{i,(-m)}\boldsymbol{S}_{i,(-1)}^{\top}\mathsf{g},$$

where $\boldsymbol{S}_{i,(-1)}$ and $\boldsymbol{Z}_{i,(-m)}$ are block-diagonal matrices computed by aligning $S_{i,2}^{\top}$, …, $S_{i,m}^{\top}$ and $Z_{i,1}^{\top}$, …, $Z_{i,m-1}^{\top}$ along their diagonals, respectively.

2: Compute the refined estimator $\widetilde{\mathbf{\Theta}} = (\widetilde{\Theta}(1), \ldots, \widetilde{\Theta}(m-1))^\top = \mathbf{\Omega}\widehat{\mathbf{\Theta}}$.
3: Construct the plug-in estimator $\widehat{\mathrm{IE}}$ according to (14).
4: Compute the estimated residual $\widehat{\varepsilon}_{i,\tau S} = S_{i,\tau+1} - Z_{i,\tau}\widetilde{\Theta}(\tau)$ for any i and τ.
5: **for** $b = 1, \ldots, B$ **do**
Generate i.i.d. standard normal random variables $(\xi_i^b)_{i=1}^n$;
Generate pseudo outcomes $(\widehat{S}_{i,\tau}^b)_{i,\tau}$ and $(\widehat{Y}_{i,\tau}^b)_{i,\tau}$ according to (15);
Repeat Steps 1-2 in Algorithm 1 and Steps 1-3 in Algorithm 2 to compute $\widehat{\mathrm{IE}}^b$.
6: **end for**
7: Reject H_0^{IE} if $\widehat{\mathrm{IE}}$ exceeds the upper αth empirical quantile of $\widehat{\mathrm{IE}}^b - \widehat{\mathrm{IE}}_b$.

Steps 1-3 of Algorithm 2 are to compute a consistent estimator $\widehat{\mathrm{IE}}$ for IE. Specifically, in Step 1 of Algorithm 2, we apply OLS regression to derive an initial estimator $\widehat{\mathbf{\Theta}}$ for $\mathbf{\Theta} = (\Theta(1), \ldots, \Theta(m-1))^\top$. In Step 2 of Algorithm 2, we employ kernel smoothing to compute a refined estimator $\widetilde{\mathbf{\Theta}} = \mathbf{\Omega}\widehat{\mathbf{\Theta}}$ to improve its statistical efficiency, as in Algorithm 1. In Step 3 of Algorithm 2, we plug in $\widetilde{\mathbf{\Theta}}$ and $\widetilde{\boldsymbol{\theta}}$ for $\mathbf{\Theta}$ and $\boldsymbol{\theta}$ in model 1, leading to

$$\widehat{\mathrm{IE}} = \sum_{\tau=2}^{m} \widetilde{\beta}(\tau)^\top \left\{ \sum_{k=1}^{\tau-1} \left(\widetilde{\Phi}(\tau-1)\widetilde{\Phi}(\tau-2)\ldots\widetilde{\Phi}(k+1) \right) \widetilde{\Gamma}(k) \right\}, \tag{14}$$

where $\widetilde{\beta}(\tau)$, $\widetilde{\Phi}(\tau)$ and $\widetilde{\Gamma}(\tau)$ are the corresponding estimators for $\beta(\tau)$, $\Phi(\tau)$ and $\Gamma(\tau)$, respectively.

Step 4 of Algorithm 2 is to compute the estimated residuals $\widehat{E}_{i,\tau} = S_{i,\tau+1} - Z_{i,\tau}\widetilde{\Theta}(\tau)$ for all i and τ, which are used to generate pseudo outcomes in the subsequent bootstrap step.

Step 5 of Algorithm 2 is to use bootstrap to simulate the distribution of $\widehat{\mathrm{IE}}$ under the null hypothesis. The key idea is to compute the bootstrap samples for $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\mathbf{\Theta}}$ and use the plug-in principle to construct the bootstrap samples for $\widehat{\mathrm{IE}}$. A key observation is that $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\mathbf{\Theta}}$ depend linearly on the random errors, so the wild bootstrap method (Wu et al., 1986) is applicable. We begin by generating i.i.d. standard normal random variables $(\xi_i)_{i=1}^n$. We next generate pseudo-outcomes given by

$$\widehat{S}_{i,\tau+1} = \widetilde{\Theta}(\tau)\widehat{Z}_{i,\tau} + \xi_i\widehat{\varepsilon}_{i,\tau S} \text{ and } \widehat{Y}_{i,\tau} = \widehat{Z}_{i,\tau}^\top\widetilde{\theta}(\tau) + \xi_i\widehat{e}_{i,\tau}, \tag{15}$$

where $\widehat{Z}_{i,\tau}$ is a version of $Z_{i,\tau}$ with $S_{i,\tau}$ replaced by $\widehat{S}_{i,\tau}$. Furthermore, we apply Steps 1-2 of Algorithm 1 and Steps 1-3 of Algorithm 2 to compute the bootstrap version of $\widehat{\mathrm{IE}}$ based on these pseudo outcomes in (15). The above procedures are repeatedly applied to simulate a sequence of bootstrap estimators $(\widehat{\mathrm{IE}}^b)_{b=1}^B$ based on which the decision region can be derived.

2.5. Estimation procedure in NN-TVCDP model

We first introduce how to estimate the regression functions g_0, g_1, G_0 and G_1. Take g_0 as an instance, we consider minimizing the following empirical objective function

$$\sum_{i=1}^{n}\sum_{\tau=1}^{m}(1 - A_{i,\tau})\left(Y_{i,\tau} - g_0(\tau, S_{i,\tau})\right)^2.$$

Instead of separately estimating $g_0(\tau, \bullet)$ for each τ, we treat τ as part of the features and jointly estimate $(g_0(\tau, \bullet))_\tau$ by solving the above optimization. It allows us to borrow information across different time points to improve the estimation accuracy.

Next, we introduce the estimation procedures for DE and IE. We impose a parametric model (e.g., Gaussian) for the density function $f_{\varepsilon_{\tau S}}$ of the measurement error $\varepsilon_{i,\tau S}$ and summarize the steps below.

1. Use neural networks to estimate g_0, g_1, G_0 and G_1 by solving their corresponding least square objective functions. Denote the corresponding estimators by $\widehat{g}_0$, $\widehat{g}_1$, $\widehat{G}_0$, and $\widehat{G}_1$, respectively.

2. Compute the residual $\widehat{\varepsilon}_{i,\tau S} = S_{i,\tau+1} - \left\{\widehat{G}_0(\tau, S_{i,\tau}) \cdot \mathbb{I}(A_{i,\tau} = 0) + \widehat{G}_1(\tau, S_{i,\tau}) \cdot \mathbb{I}(A_{i,\tau} = 1)\right\}$ and use $\widehat{\varepsilon}_{i,\tau S}$ to compute the density function estimator $\widehat{f}_{\varepsilon_{\tau S}}$.

3. Use Monte Carlo to estimate the distributions of the potential states $S^*_{i,\tau}(\mathbf{1}_{\tau-1})$ and $S^*_{i,\tau}(\mathbf{0}_{\tau-1})$ conditional on $S_{i,1}$. Specifically, for $\tau = 1, \ldots, m$, $i = 1, \ldots, n$, and $k = 1, \ldots, M$, we use $\widehat{f}_{\varepsilon_{\tau S}}$ to generate error residuals $\{\widehat{\varepsilon}_{i,\tau S,k}\}_{k=1}^{M}$, where M denotes the number of Monte Carlo replications. Next, we set $\widehat{S}^1_{i,1,k} = \widehat{S}^0_{i,1,k} = S_{i,1}$ for any i and k, and sequentially construct Monte Carlo samples $\{\widehat{S}^1_{i,\tau,k}\}_{k=1}^{M}, \{\widehat{S}^0_{i,\tau,k}\}_{k=1}^{M}$ by setting $\widehat{S}^1_{i,\tau+1,k} = \widehat{G}_1(\tau, \widehat{S}^1_{i,\tau,k}) + \widehat{\varepsilon}_{i,\tau S,k}$ and $\widehat{S}^0_{i,\tau+1,k} = \widehat{G}_0(\tau, \widehat{S}^0_{i,\tau,k}) + \widehat{\varepsilon}_{i,\tau S,k}$.

4. Based on (8), we estimate DE and IE by using

$$\widehat{DE} = \frac{1}{nM}\sum_{i=1}^{n}\sum_{k=1}^{M}\sum_{\tau=1}^{m}\left\{\widehat{g}_1(\tau, \widehat{S}^0_{i,k,\tau}) - \widehat{g}_0(\tau, \widehat{S}^0_{i,k,\tau})\right\} \quad \text{and}$$

$$\widehat{IE} = \frac{1}{nM}\sum_{i=1}^{n}\sum_{k=1}^{M}\sum_{\tau=2}^{m}\left\{\widehat{g}_1(\tau, \widehat{S}^1_{i,k,\tau}) - \widehat{g}_1(\tau, \widehat{S}^0_{i,k,\tau})\right\}.$$

3. Policy evaluation for spatio-temporal dependent experiments

In this section, we present the proposed methodology for policy evaluation in spatio-temproal dependent experiments by extending our proposal in temporal dependent experiments. We highlight several key differences between the spatio-temporal dependent experiment and the temporal dependent one.

3.1. A potential outcome framework

Firstly, we introduce the spatio-temporal dependent experiments as follows. Specifically, a city is split into r non-overlapping regions. Each region receives a sequence of policies over time and different regions may receive different policies at the same time. In our application, we employ the spatio-temporal dependent alternation design to randomize these policies. In each region, we independently randomize the initial policy (either A or B) and then apply the temporal alternation design. As discussed in the introduction, one major challenge for policy evaluation is that the spatial proximities will induce spatio-temporal interference among locations across time. In the example of ride-sharing platforms, for many call orders, their pickup locations and destinations belong to different regions. Therefore, applying an order dispatch policy at one region will change the distribution of drivers of its neighbouring areas as well, so the order dispatch policy at one location could influence outcomes of those neighbouring areas, inducing interference among spatial units.

Secondly, to quantify the spatio-temporal interference, we allow the potential outcome of each region to depend on polices applied to its neighbouring areas as well. Specifically, for the ιth region, let $\bar{a}_{\tau,\iota} = (\bar{a}_{1,\iota}, \ldots, \bar{a}_{\tau,\iota})^\top$ denote its treatment history up to time τ and $\mathcal{N}_\iota$ denote the neighbouring regions of ι. Let $\bar{a}_{\tau,[1:r]} = (\bar{a}_{\tau,1}, \ldots, \bar{a}_{\tau,r})^\top$ denote the treatment history associated with all regions. Similarly, let $S^*_{\tau,\iota}(\bar{a}_{\tau-1,[1:r]})$ and $Y^*_{\tau,[1:r]}(\bar{a}_{\tau,[1:r]})$ denote the potential state and outcome associated with the ιth region, respectively. Let $S^*_{\tau,[1:r]}(\bar{a}_{\tau-1,[1:r]})$ denote the set of potential states at time τ.

Similarly, we introduce CA and SRA in the spatio-temporal case as follows.

- **CA**. $S^*_{\tau+1,\iota}(\bar{A}_{\tau,[1:r]}) = S_{\tau+1,\iota}$ and $Y^*_{\tau,\iota}(\bar{A}_{\tau,[1:r]}) = Y_{\tau,\iota}$ for any $\tau \geq 1$ and $1 \leq \iota \leq r$, where $\bar{A}_{\tau,[1:r]}$ denotes the set of observed treatment history up to time τ.

- **SRA**. $A_{\tau,[1:r]}$, the set of observed policies at time τ, is conditionally independent of all potential variables given $S_{\tau,[1:r]}$ and $\{(S_{j,[1:r]}, A_{j,[1:r]}, Y_{j,[1:r]})\}_{j<\tau}$.

SRA automatically holds under the spatio-temporal alternation design, in which the policy assignment mechanism is conditionally independent of the data given the policies assigned at the initial time point.

Thirdly, we are interested in the overall treatment effects. Define ATE as the difference between the new and old policies aggregated over different regions.

Definition 2. *ATE is defined as the difference between two value functions given by*

$$ATE_{st} = \sum_{\iota=1}^{r}\sum_{\tau=1}^{m}\mathbb{E}\{Y^*_{\tau,\iota}(\mathbf{1}_{\tau,[1:r]}) - Y^*_{\tau,\iota}(\mathbf{0}_{\tau,[1:r]})\}.$$

Let $R_{\tau,\iota}$ denote the conditional mean function of $Y^*_{\tau,\iota}(\bar{a}_{\tau,[1:r]})$ given the past policies and potential states. Similarly, we can decompose ATE as the sum of DE and IE, which are, respectively, given by

$$\mathrm{DE}_{st} = \sum_{\iota=1}^{r}\sum_{\tau=1}^{m} \mathbb{E}\{R_{\tau,\iota}(\mathbf{1}_{\tau,[1:r]}, S^*_{\tau,\iota}(\mathbf{0}_{\tau-1,[1:r]}), \mathbf{0}_{\tau-1,[1:r]}, \ldots, S_1) - R_{\tau,\iota}(\mathbf{0}_{\tau,[1:r]}, S^*_{\tau,\iota}(\mathbf{0}_{\tau-1,[1:r]}), \mathbf{0}_{\tau-1,[1:r]}, \ldots, S_1)\},$$

$$\mathrm{IE}_{st} = \sum_{\iota=1}^{r}\sum_{\tau=1}^{m} \mathbb{E}\{R_{\tau,\iota}(\mathbf{1}_{\tau,[1:r]}, S^*_{\tau,\iota}(\mathbf{1}_{\tau-1,[1:r]}), \mathbf{1}_{\tau-1,[1:r]}, \ldots, S_1) - R_{\tau,\iota}(\mathbf{1}_{\tau,[1:r]}, S^*_{\tau,\iota}(\mathbf{0}_{\tau-1,[1:r]}), \mathbf{0}_{\tau-1,[1:r]}, \ldots, S_1)\}.$$

We aim to test the following hypotheses:

$$H_0^{DE}: \mathrm{DE}_{st} \le 0 \quad v.s \quad H_1^{DE}: \mathrm{DE}_{st} > 0, \tag{16}$$

$$H_0^{IE}: \mathrm{IE}_{st} \le 0 \quad v.s \quad H_1^{IE}: \mathrm{IE}_{st} > 0. \tag{17}$$

3.2. Spatio-temporal VCDP models

We introduce the spatio-temporal VCDP (STVCDP) models to model $Y_{\tau,\iota}$ and $S_{\tau,\iota}$, respectively. Suppose that the experiment is conducted across r regions over n days. Let $(S_{i,\tau,\iota}, A_{i,\tau,\iota}, Y_{i,\tau,\iota})$ denote the state-policy-outcome triplet measured from the ιth region at the τth time interval of the ith day for $i = 1, \ldots, n$, $\tau = 1, \ldots, m$, and $\iota = 1, \ldots, r$. The STVCDP model is given as follows,

$$\begin{aligned} Y_{i,\tau,\iota} &= f_{1,\tau,\iota}(S_{i,\tau,\iota}, A_{i,\tau,\iota}, \bar{A}_{i,\tau,\mathcal{N}_\iota}) + e_{i,\tau,\iota}, \\ S_{i,\tau+1,\iota} &= f_{2,\tau,\iota}(S_{i,\tau,\iota}, A_{i,\tau,\iota}, \bar{A}_{i,\tau,\mathcal{N}_\iota}) + \epsilon_{i,\tau,\iota}, \end{aligned}$$

where $\bar{A}_{i,\tau,\mathcal{N}_\iota}$ denotes the average of $\{A_{i,\tau,k}\}_{k\in\mathcal{N}_\iota}$, and $\{e_{i,\tau,\iota}, \epsilon_{i,\tau,\iota}\}$ are the random noises. In parallel to Assumption 1, we impose the following noise assumption for the STVCDP model.

Assumption 2. *(i) The outcome noise $e_{i,\tau,\iota} = \eta^I_{i,\tau,\iota} + \eta^{II}_{i,\tau,\iota} + \eta^{III}_{i,\tau,\iota} + \varepsilon_{i,\tau,\iota}$ can be decomposed into four mutually independent processes: $\{\eta^I_{i,\tau,\iota}\}$, $\{\eta^{II}_{i,\tau,\iota}\}$, $\{\eta^{III}_{i,\tau,\iota}\}$, and $\{\varepsilon_{i,\tau,\iota}\}$. (ii) The $\{\eta^I_{i,\tau,\iota}\}$, $\{\eta^{II}_{i,\tau,\iota}\}$ and $\{\eta^{III}_{i,\tau,\iota}\}$ are i.i.d. copies of some zero-mean random processes with covariance functions $\Sigma_{\eta^I}(\tau_1,\iota_1,\tau_2,\iota_2)$, $\Sigma_{\eta^{II}}(\tau_1,\iota_1,\tau_2,\iota_2)\mathbb{I}(\iota_1=\iota_2)$, and $\Sigma_{\eta^{III}}(\tau_1,\iota_1,\iota_2)\mathbb{I}(\tau_1=\tau_2)$, respectively. These covariance functions have bounded and continuously differentiable second-order derivatives. (iii) The measurement errors $\{\varepsilon_{i,\tau,\iota}\}_{i,\tau,\iota}$ and the state noises $\{\epsilon_{i,\tau,\iota}\}_{i,\tau,\iota}$ are independent over different location/time combinations, have zero means, and satisfy $\mathrm{Var}(\varepsilon_{i,\tau,\iota}) = \sigma^2_\varepsilon(\tau,\iota)$ and $\mathrm{Cov}(\epsilon_{i,\tau,\iota}) = \Sigma_{\epsilon,\tau,\iota}$.*

We make three remarks. Firstly, as per the STVCDP model, the outcome in the ιth region is influenced solely by the current actions $A_{i,\tau,\iota}$ and those from its neighboring areas. This assumption is often valid in various applications, such as ride-sharing platforms. For instance, the policy in one location may impact other locations only through its effect on the distribution of drivers. Within each time unit, a driver can travel at most from one location to its neighboring ones. Consequently, outcomes in one location are independent of policies applied to non-adjacent locations.

Secondly, in our spatial interference model, we adopt the mean field approximation. Under this approach, the outcome $Y_{\tau,\iota}$ and next state $S_{\tau+1,\iota}$ in a given region depend on the treatments of neighboring regions $\{A_{\tau,k}\}_{k\in\mathcal{N}_\iota}$ only through their average $\bar{A}_{\tau,\mathcal{N}_\iota}$. The mean field approximation is a commonly used technique in multi-agent reinforcement learning for policy learning and evaluation. It's worth noting that studies, such as Shi et al. (2022a), have shown that the average effect $\bar{A}_{\tau,\mathcal{N}_\iota}$ effectively summarizes the impact of $\{A_{\tau,k}\}_{k\in\mathcal{N}_\iota}$. This approach aligns with assumptions frequently made in the causal inference literature dealing with spatial interference (Sobel, 2006; Hudgens and Halloran, 2008; Zigler et al., 2012; Perez-Heydrich et al., 2014; Sobel and Lindquist, 2014; Liu et al., 2016; Sävje et al., 2021).

Thirdly, besides the average effect, alternative low-dimensional summary statistics of $\{A_{ij} : j \in \mathcal{N}_\iota\}$ can be considered, such as $\sum_{j\in\mathcal{N}_\iota}\theta_{\iota j}A_{ij}$ and $\theta_\iota \mathbb{I}_{\{\sum_{j\in\mathcal{N}_\iota}A_{ij}>0\}}$ (Hu et al., 2022). The resulting estimation and inference procedures can be similarly derived.

Similar to model (6), we allow general function approximation for f_1 and f_2. To save space, we focus on linear STVCDP models (L-STVCDP) in the rest of this section. Meanwhile, the proposed estimation procedure can be extended to handle neural network STVCDP models, as in Section 2.4. The proposed L-STVCDP model is given as follows,

$$\begin{aligned} Y_{i,\tau,\iota} &= \beta_0(\tau,\iota) + S^\top_{i,\tau,\iota}\beta(\tau,\iota) + A_{i,\tau,\iota}\gamma_1(\tau,\iota) + \bar{A}_{i,\tau,\mathcal{N}_\iota}\gamma_2(\tau,\iota) + e_{i,\tau,\iota}, \\ S_{i,\tau+1,\iota} &= \phi_0(\tau,\iota) + \Phi(\tau,\iota)S_{i,\tau,\iota} + A_{i,\tau,\iota}\Gamma_1(\tau,\iota) + \bar{A}_{i,\tau,\mathcal{N}_\iota}\Gamma_2(\tau,\iota) + \epsilon_{i,\tau,\iota}, \end{aligned} \tag{18}$$

where $Z_{i,\tau,\iota} = (1, S_{i,\tau,\iota}^\top, A_{i,\tau,\iota}, \bar{A}_{i,\tau,\mathcal{N}_\iota})^\top$.

Similar to (7), we can show that DE_{st} and IE_{st} are equal to the following,

$$\begin{aligned} \mathrm{DE}_{st} &= \textstyle\sum_{\iota=1}^{r}\sum_{\tau=1}^{m} \boldsymbol{f}\gamma_1(\tau,\iota) + \gamma_2(\tau,\iota)\boldsymbol{g}, \\ \mathrm{IE}_{st} &= \textstyle\sum_{\iota=1}^{r}\sum_{\tau=1}^{m} \beta(\tau,\iota)^\top \left[\sum_{k=1}^{\tau-1} \left(\Phi(\tau-1,\iota)\ldots\Phi(k+1,\iota)\right) \boldsymbol{f}\Gamma_1(k,\iota) + \Gamma_2(k,\iota)\boldsymbol{g}\right], \end{aligned} \tag{19}$$

where the product $\Phi(\tau-1,\iota)\ldots\Phi(k+1,\iota) = 1$ when $\tau - 1 < k+1$. These two identities form the basis of our test procedure.

3.3. Estimation and testing procedures for DE and IE

We first describe our estimation and testing procedures for DE under the spatio-temporal alternation design and present the pseudocode in Algorithms S.1 of Section S.1 of the supplementary document to save space.

Step 1 of Algorithm S.1 is to independently apply Steps 1 and 2 of Algorithm 1 detailed in Section 2.3 to the data subset $\boldsymbol{f}(Z_{i,\tau,\iota}, Y_{i,\tau,\iota})\boldsymbol{g}_{i,\tau}$ for each region ι in order to compute a smoothed estimator $\widetilde{\boldsymbol{\theta}}_{st}^0(\iota) = \boldsymbol{f}\widetilde{\theta}_{st}^0(1,\iota)^\top, \ldots, \widetilde{\theta}_{st}^0(m,\iota)^\top\boldsymbol{g}^\top$ for $\boldsymbol{f}\theta(1,\iota)^\top, \ldots, \theta(m,\iota)^\top\boldsymbol{g}^\top$.

Step 2 of Algorithm S.1 is to employ kernel smoothing again to spatially smooth each component of $\widetilde{\boldsymbol{\theta}}_{st}^0(\iota)$ across all ι 2 $\boldsymbol{f}1, \ldots, r\boldsymbol{g}$. Specifically, we compute $\widetilde{\boldsymbol{\theta}}_{st}(\iota) = \boldsymbol{f}\widetilde{\theta}_{st}(1,\iota)^\top, \ldots, \widetilde{\theta}_{st}(m,\iota)^\top\boldsymbol{g}^\top$ as the resulting refined estimator, given by $\widetilde{\theta}_{st}(\tau,\iota) = \sum_{\ell=1}^{r} \kappa_{\ell,h_{st}}(\iota)\widetilde{\theta}_{st}^0(\tau,\ell)$, where $\kappa_{\ell,h_{st}}(\cdot)$ defined in (S.2) is a normalized kernel function with bandwidth parameter h_{st}.

We remark that we employ kernel smoothing twice in order to estimate the varying coefficients. In the first step, we temporally smooth the least square estimator to compute $\widetilde{\boldsymbol{\theta}}_{st}^0(\iota)$. In the second step, we further spatially smooth $\widetilde{\boldsymbol{\theta}}_{st}^0(\iota)$ to compute $\widetilde{\boldsymbol{\theta}}_{st}(\iota)$. Therefore, the estimator $\widetilde{\boldsymbol{\theta}}_{st}(\iota)$ has smaller variance than $\widetilde{\boldsymbol{\theta}}_{st}^0(\iota)$, since we borrow information across neighboring regions to improve the estimation efficiency. To elaborate this point, the random effect in (18) can be decomposed into three parts: $\eta_{i,\tau,\iota}^I + \eta_{i,\tau,\iota}^{II} + \eta_{i,\tau,\iota}^{III}$. Temporally smoothing the varying coefficient estimator removes the random fluctuations caused by $\eta_{i,\tau,\iota}^{III}$ and the measurement error. Spatially smoothing the estimator further removes the random fluctuations caused by $\eta_{i,\tau,\iota}^{II}$. This in turn implies that the proposed test under the spatio-temporal design is more powerful than the one developed in Section 2 under the temporal design. Such an observation is consistent with our numerical findings in Section 5.2.

Steps 3 and 4 of Algorithm S.1 are to estimate the covariance matrix of $(\widetilde{\boldsymbol{\theta}}_{st}(1), \ldots, \widetilde{\boldsymbol{\theta}}_{st}(r))^\top$, denoted by $\widetilde{\boldsymbol{V}}_{\theta,st}$. These two steps are very similar to Steps 3 and 4 of Algorithm 1. Specifically, we first estimate the measurement errors and random effects based on the estimated varying coefficients. We next use the sandwich formula to compute the estimated covariance matrix for the initial least-square estimator. Then the estimated covariance matrix for $\widetilde{\boldsymbol{\theta}}_{st}^0(\iota)$ can be derived accordingly. We use $\widetilde{\boldsymbol{V}}_{\theta,st}$ to denote the corresponding covariance matrix estimator.

Step 5 of Algorithm S.1 is to compute the Wald-type test statistic and its standard error estimator. Specifically, let $\widetilde{\gamma}_1(\tau,\iota)$ and $\widetilde{\gamma}_2(\tau,\iota)$ be the last two elements of $\widetilde{\theta}_{st}(\tau,\iota)$, we have $\widehat{\mathrm{DE}}_{st} = \sum_{\iota=1}^{r}\sum_{\tau=1}^{m} \boldsymbol{f}\widetilde{\gamma}_1(\tau,\iota) + \widetilde{\gamma}_2(\tau,\iota)\boldsymbol{g}$. We will show in Theorem 6 that $\widehat{\mathrm{DE}}_{st}$ is asymptotically normal. In addition, its standard error $\widehat{se}(\widehat{\mathrm{DE}}_{st})$ can be derived based on $\widetilde{\boldsymbol{V}}_{\theta,st}$. This yields our Wald-type test statistic $T_{st} = \widehat{\mathrm{DE}}_{st}/\widehat{se}(\widehat{\mathrm{DE}}_{st})$. We reject the null hypothesis if T_{st} exceeds the upper αth quantile of a standard normal distribution.

We next describe our estimation and testing procedures for IE. The method is very similar to the one discussed in Section 2.4. We sketch an outline of the algorithm to save space. Details are presented in S.2 of Section S.1 of the supplementary document. Specifically, we first plug in the set of smoothed estimators $\boldsymbol{f}\widetilde{\Theta}_{st}(\tau,\iota)\boldsymbol{g}_{\tau,\iota}$ and $\boldsymbol{f}\widetilde{\theta}_{st}(\tau,\iota)\boldsymbol{g}_{\tau,\iota}$ for $\boldsymbol{f}\Theta(\tau,\iota)\boldsymbol{g}_{\tau,\iota}$ and $\boldsymbol{f}\theta(\tau,\iota)\boldsymbol{g}_{\tau,\iota}$ to compute $\widehat{\mathrm{IE}}_{st}$, the plug-in estimator of IE_{st}. We next estimate the measurement errors and random effects and then apply the parametric bootstrap method to compute the bootstrap statistics $\boldsymbol{f}\widehat{\mathrm{IE}}_{st}^b\boldsymbol{g}_b$. Finally, we reject H_0^{IE} if $\widehat{\mathrm{IE}}_{st}$ exceeds the upper αth empirical quantile of $\boldsymbol{f}\widehat{\mathrm{IE}}_{st}^b - \widehat{\mathrm{IE}}\boldsymbol{g}_b$.

To conclude this section, we remark that in Sections 2 and 3, we focus on testing one-sided hypotheses for the direct and indirect effects. However, the proposed method can be easily extended to test two-sided hypotheses as well.

4. Theoretical Analysis

In this section, we systematically investigate the asymptotic properties of the proposed estimators and test statistics in L-TVCDP and derive the convergence rates of our causal estimands in NN-TVCDP. We also explore the benefits of employing the swtichback design and study the theoretical properties of our estimator in the spatio-temporal dependent experiments.

Firstly, we impose the following regularity assumptions for the temporal dependent experiments using L-TVCDP.

ASSUMPTION 3. *The kernel function $K(\cdot)$ is a symmetric probability density function on $[-1,1]$ and is Lipschitz continuous.*

ASSUMPTION 4. *The covariate $\boldsymbol{Z}_i$s are i.i.d.; for $1 \le \tau \le m$, $\mathbb{E}(Z_{i,\tau}^\top Z_{i,\tau})$ 2 $\mathbb{M}^{p\times p}$ is invertible; all components of $\theta(t)$ have bounded and continuous second derivatives with respect to t.*

ASSUMPTION 5. *There exists $0 < q < 1$ such that the absolute values of eigenvalues of $\Phi(\tau)$ are smaller than q, and there exist some constants M_Γ and M_β such that $\|\Gamma(\tau)\|_\infty \le M_\Gamma$ and $\|\beta(\tau)\|_\infty \le M_\beta$. $\{\beta(\tau)\}_{2\le\tau\le m}$, $\{\Phi(l)\}_{2\le l\le m-1}$, and $\{\Gamma(k)\}_{1\le k\le m-1}$ must not be all zero. $\Theta(\tau)$ has a continuous second-order partial derivative.*

Assumption 3 is mild as the kernel $K(\cdot)$ is user-specified. Assumption 4 has been commonly used in the literature on varying coefficient models (see e.g., Zhu et al., 2014). Assumption 5 ensures that the time series is stationary, since $\Phi(\tau)$ is the autoregressive coefficient. It is commonly imposed in the literature on time series analysis (Shumway and Stoffer, 2010).

Before presenting the theoretical properties of the proposed method for L-TVCDP, we introduce some notation. For $1 \le \tau_1, \tau_2 \le m$, define $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\Sigma}_\eta$ to be the $m \times m$ matrices $\{\Sigma_y(\tau_1,\tau_2)\}_{\tau_1,\tau_2}$ and $\{\Sigma_\eta(\tau_1,\tau_2)\}_{\tau_1,\tau_2}$, respectively. We define

$$\boldsymbol{V}_{\widehat{\theta}} = (\mathbb{E}\boldsymbol{Z}_i^\top \boldsymbol{Z}_i)^{-1}\mathbb{E}(\boldsymbol{Z}_i^\top \boldsymbol{\Sigma}_y \boldsymbol{Z}_i)(\mathbb{E}\boldsymbol{Z}_i^\top \boldsymbol{Z}_i)^{-1} \quad \text{and} \quad \boldsymbol{V}_{\widetilde{\theta}} = (\mathbb{E}\boldsymbol{Z}_i^\top \boldsymbol{Z}_i)^{-1}\mathbb{E}(\boldsymbol{Z}_i^\top \boldsymbol{\Sigma}_\eta \boldsymbol{Z}_i)(\mathbb{E}\boldsymbol{Z}_i^\top \boldsymbol{Z}_i)^{-1}$$

as the asymptotic covariance matrices of $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$, respectively. Let $\boldsymbol{V}_{\widehat{\theta}}(\tau,\tau)$ and $\boldsymbol{V}_{\widetilde{\theta}}(\tau,\tau)$ denote the submatrices of $\boldsymbol{V}_{\widehat{\theta}}$ and $\boldsymbol{V}_{\widetilde{\theta}}$ that correspond to the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$, respectively. We first compare the mean squared error (MSE) of the OLS estimator $\widehat{\theta}(\tau)$ against that of the smoothed estimator $\widetilde{\theta}(\tau)$ based on L-TVCDP.

PROPOSITION 2. *Suppose $\lambda_{\min}(\boldsymbol{V}_{\widehat{\theta}}(\tau,\tau))$ and $\lambda_{\min}(\boldsymbol{V}_{\widetilde{\theta}}(\tau,\tau))$ are uniformly bounded away from zero for any τ. Under Assumptions 3 and 4, we have*

$$\sum_{\tau=1}^{m} MSE(\widehat{\theta}(\tau)) \quad n^{-1}trace(\boldsymbol{V}_{\widehat{\theta}}),\ \sum_{\tau=1}^{m} MSE(\widehat{\theta}(\tau)) \quad n^{-1}trace(\boldsymbol{V}_{\widetilde{\theta}}) + O(mh^4 + m^{-1}).$$

Proposition 2 has an important implication. Both trace($\boldsymbol{V}_{\widehat{\theta}}$) and trace($\boldsymbol{V}_{\widetilde{\theta}}$) are of the order of magnitude $O(m)$. When m $\sqrt{n}$ or h^4 n^{-1}, the squared bias of $\widetilde{\boldsymbol{\theta}}$ may dominate its variance. Hence, the OLS estimator $\widehat{\boldsymbol{\theta}}$ may achieve a smaller MSE. When m $\sqrt{n}$ and $h^4 = O(n^{-1}m)$, the two MSEs are of the same order of magnitude and it remains unclear which one is smaller. When m $\sqrt{n}$ and $h^4 = o(n^{-1})$, the variance of $\widetilde{\boldsymbol{\theta}}$ dominates its squared bias. Moreover, $\boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_\eta$ is strictly positive definite, so is $\boldsymbol{V}_{\widehat{\theta}} - \boldsymbol{V}_{\widetilde{\theta}}$. As a result, $\widetilde{\boldsymbol{\theta}}$ achieves a smaller MSE. In our applications, m is moderately large and the condition m $\sqrt{n}$ is likely to be satisfied. With properly chosen bandwidth, we expected the smoothed estimator achieves a smaller MSE.

Secondly, we present the limiting distributions of $\widehat{\theta}(\tau)$ and $\widetilde{\theta}(\tau)$ and prove the validity of our test for DE based on L-TVCDP.

THEOREM 1. *Suppose $\lambda_{\min}(\boldsymbol{V}_{\widehat{\theta}}(\tau,\tau))$ and $\lambda_{\min}(\boldsymbol{V}_{\widetilde{\theta}}(\tau,\tau))$ are uniformly bounded away from zero for any τ. Under Assumptions 1, 3 and 4, for any $(d+2)$-dimensional vectors $\boldsymbol{a}_{n,1}$, $\boldsymbol{a}_{n,2}$, with unit ℓ_2 norm,*

(i) $\sqrt{n}\boldsymbol{a}_{n,1}^\top\{\widehat{\theta}(\tau) - \theta(\tau)\} \Big/ \sqrt{\boldsymbol{a}_{n,1}^\top \boldsymbol{V}_{\widehat{\theta}}(\tau,\tau)\boldsymbol{a}_{n,1}} \xrightarrow{d} N(0,1)$ *as $n \to \infty$ for any τ;*

(ii) Suppose $m \to \infty, h \to 0$, and $hm \to \infty$ as $n \to \infty$. Then $\sqrt{n}\boldsymbol{a}_{n,2}^\top\{\widetilde{\theta}(\tau) - \theta(\tau)\} \Big/ \sqrt{\boldsymbol{a}_{n,2}^\top \boldsymbol{V}_{\widetilde{\theta}}(\tau,\tau)\boldsymbol{a}_{n,2}} \xrightarrow{d} N(b_n, 1)$ *as $n \to \infty$ for any τ, where the bias $b_n = O(\sqrt{n}h^2 + \sqrt{n}m^{-1})$.*

(iii) Suppose $h = o(n^{-1/4})$, $m \gtrsim \sqrt{n}$ *and the sum of all elements in* $m^{-2}\boldsymbol{V}_{\widetilde{\gamma}}$ *is bounded away from zero where* $\boldsymbol{V}_{\widetilde{\gamma}}$ *denotes the submatrix of* $\boldsymbol{V}_{\widetilde{\theta}}$ *which corresponds to the asymptotic covariance matrix of* $\widetilde{\boldsymbol{\theta}}$*. Then for the hypotheses* (2)*, under* H_0^{DE} *,* $\mathbb{P}(\widehat{DE}/\widehat{se}(\widehat{DE}) > z_\alpha) = \alpha + o(1)$*; under* H_1^{DE}*,* $\mathbb{P}(\widehat{DE}/\widehat{se}(\widehat{DE}) > z_\alpha) \to 1$*, where* z_α *denotes the upper* α*th quantile of a standard normal distribution.*

Theorem 1 has several important implications. First, the bias of the smoothed estimator $\widetilde{\boldsymbol{\theta}}$ decays with m. In cases where m is fixed, the kernel smoothing step is not preferred as it will result in an asymptotically biased estimator. Second, each $\widetilde{\theta}(\tau)$ converges at a rate of $O_p(n^{-1/2})$ under the assumption that $\lambda_{\min}(\boldsymbol{V}_{\widetilde{\theta}}(\tau,\tau))$ is bounded away from zero. The rate $O_p(n^{-1/2}m^{-1/2})$ cannot be achieved despite that we have a total of nm observations, since the random errors $\{e_\tau g_\tau\}$ are not independent. We also remark that in the extreme case where $\{e_\tau g_\tau\}$ are independent, we can set $h \asymp (nm)^{-1/5}$ and $\widetilde{\theta}(\tau)$ attains the classical nonparametric convergence rate $O_p((nm)^{-2/5})$. Third, since $\boldsymbol{V}_{\widehat{\theta}} - \boldsymbol{V}_{\widetilde{\theta}}$ is strictly positive, this similarly implies that the smoothed estimator is more efficient when $b_n = o(1)$, or equivalently, $h = o(n^{-1/4})$ and $m \gtrsim \sqrt{n}$. Finally, in the proof of Theorem 1, we show that the covariance estimator $\widetilde{\boldsymbol{V}}_\theta$ is consistent. This together with asymptotic distribution of $\widetilde{\boldsymbol{\theta}}$ yields the the consistency of our test in (iii).

Thirdly, we present the validity of the proposed parametric bootstrap procedure for IE under the temporal alternation design based on L-TVCDP.

THEOREM 2. *Suppose that there is some constant* $0 < c_1 \le 1$ *such that* $c_1 \le \mathbb{E}\|\varepsilon_{\tau,S}\|^2$ *and* $\mathbb{E}e_\tau^2 \le c_1^{-1}$ *for all* $1 \le \tau \le m$*. Suppose that* $h = o(n^{-1/4})$, $m \asymp n^{c_2}$ *for some* $1/2 \le c_2 < 3/2$ *and* $mh \to 1$*. Then under the assumptions in Theorem 1 and Assumption 5, with probability approaching 1, we have*

$$\sup_z |\mathbb{P}(\widehat{IE} - IE \le z) - \mathbb{P}(\widehat{IE}^b - \widehat{IE} \le z | Data)| \le C(\sqrt{n}h^2 + \sqrt{n}m^{-1} + n^{-1/8}),$$

where C *is some positive constant.*

We have several remarks. The derivation of Theorem 2 is non-trivial when m diverges with n. Specifically, since $\widehat{\mathrm{IE}}$ is a very complicated function of the estimated varying coefficients (see Equation (14)), its limiting distribution is not well-defined. To prove Theorem 2, we derive a nonasymptotic error bound on the difference between the distribution of $\widehat{\mathrm{IE}}$ and that of the bootstrap statistics conditional on the data. As a result, it ensures that the type-I error can be well-controlled and the power approaches one. Please refer to the proof of Theorem 2 in the supplementary document for details. Finally, we require m to diverge with n at certain rate. In settings with a small or fixed m, one can apply the proposed bootstrap procedure to the unsmoothed estimator $\widehat{\boldsymbol{\theta}}$. The resulting test procedure remains valid regardless of whether m is fixed or not.

Fourthly, we illustrate the advantage of employing the switchback design in the presence of temporal random effects. As commented in the introduction, the switchback design assigns different treatments at adjacent time points $A_{i,1} = 1 - A_{i,2} = A_{i,3} = \ldots = A_{i,2t-1} = 1 - A_{i,2t}$, whereas the alternating-day design assigns fixed treatment $A_{i,1} = A_{i,2} = A_{i,3} = \ldots = A_{i,2t-1} = A_{i,2t}$ within each day for any i and t. In the switchback design, the random effects at adjacent time points can cancel with each other when estimating the causal effect, yielding a more efficient estimator. To elaborate this point, we compare the mean square errors of the proposed estimators under the switchback design against those under an alternating-day design where the new and old policies are daily switched back and forth. To simplify the analysis, we focus on the case where the state is one-dimensional and assume the treatment effect estimators are constructed based on the unsmoothed OLS estimators (see Section S.12.3 for details). Let $\mathrm{MSE}(\widehat{\mathrm{DE}}_{sb})$ and $\mathrm{MSE}(\widehat{\mathrm{DE}}_{ad})$ denote the mean squared errors of DE estimators under the switchback design and the alternating-day design, respectively.

THEOREM 3. *Suppose that the state is one-dimensional,* $\Sigma_\eta(\tau_1,\tau_2)$ *is nonnegative for any* τ_1 *and* τ_2 *and Assumptions 1 and 4 hold. When* $\{\Phi(\tau)g_\tau\}$ *and* $\{\Gamma(\tau)g_\tau\}$ *are of the same signs, respectively, i.e. for any* τ_1, τ_2*,* $\Phi(\tau_1)\Phi(\tau_2) \ge 0$ *and* $\Gamma(\tau_1)\Gamma(\tau_2) \ge 0$*, then as* $n \to 1$*, we have*

$$n\mathrm{MSE}(\widehat{DE}_{sb}) \le n\mathrm{MSE}(\widehat{DE}_{ad}) + o(1),$$

where the equality holds only when $\Sigma_\eta(j,k) = 0$ *for any* j,k *such that* $|j-k| = 1, 3, 5, \ldots$.

To ensure that DE achieves a much smaller MSE under the switchback design, we only require that the random effects are non-negatively correlated and that the correlation $\Sigma(j,k)$ is nonzero for some $j-k=1,3,5,\ldots$. These conditions are automatically satisfied when the random effects are positively correlated. We next provide a close-formed expression for the ratio of the two MSEs under an AR(1) noise structure and the constraint that $\Gamma(1)=\Gamma(2)=\cdots=\Gamma(m-1)=0$.

COROLLARY 1. *Suppose that for any $1 \le \tau_1,\tau_2 \le m$, $\Sigma_e(\tau_1,\tau_2)=c\rho^{|\tau_1-\tau_2|}$ for some constant $c>0$. Then under assumptions of Theorem 3, when $\Gamma(1)=\Gamma(2)=\cdots=\Gamma(m-1)=0$, we have as $n,m \to \infty$,*

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

It can be seen from Corollary 1 that the larger the ρ, the smaller the variance ratio. In particular, when $\rho=0.5$, MSE of DE under the switchback design is approximately 9 times smaller than that under the alternating-day design. We next consider IE.

THEOREM 4. *Suppose $m=2$. Under Assumptions 1 and 4, we have*

$$n\{MSE(\widehat{IE}_{ad})-MSE(\widehat{IE}_{sb})\}=o(1).$$

Theorem 4 suggests that the IE estimators under the two designs have comparable MSEs. This together with Theorem 3 underscores the superiority of the switchback design, particularly when $m=2$. However, as m exceeds 2, determining the closed-form expression for $\mathrm{MSE}(\widehat{\mathrm{IE}})$ becomes exceedingly complex, making it challenging to directly compare the two designs. Addressing this complexity and extending the comparison for cases where $m>2$ is a task we reserve for future research.

Fifth, we establish the convergence rates of the estimated DE and IE for NN-VCDP.

THEOREM 5. *Suppose that $f_{\varepsilon_{\tau S}}$ is Lipschitz, meaning that for any τ, there exists a constant $L_f>0$ such that $|f_{\varepsilon_{\tau S}}(x)-f_{\varepsilon_{\tau S}}(y)| \le L_f\|x-y\|_2$, where $\|\cdot\|_2$ represents the Frobenius norm. Additionally, assume that the NN-based learners satisfy $\mathbb{E}\{\widehat{G}_a(\tau,S_\tau)-G_a(\tau,S_\tau)\}^2 \le \Delta_1^2(n,m)$ and $\mathbb{E}\{\widehat{g}_a(\tau,S_\tau^{a_1})-g_a(\tau,S_\tau^{a_1})\}^2 \le \Delta_2^2(n,m)$, where $a \in \{0,1\}$ and $\Delta_1(n,m)$ and $\Delta_2(n,m)$ are specific functions. The density estimator should fulfill $\int_x |f_{\varepsilon_{\tau S}}(x)-\widehat{f}_{\varepsilon_{\tau S}}(x)|dx=O_p(\Delta_3(n,m))$ for some function Δ_3. Both g_a and $\widehat{g}_a$ must be uniformly bounded. Moreover, the ratio of the density function of the potential state S_τ^a to the density of the observed state S_τ must be bounded by $\overline{\omega}$ for any τ and a. Then, as $\min(n,m) \to \infty$, we obtain the following convergence results:*

$$\begin{aligned}\widehat{DE}-DE &= O_p\left(m\overline{\omega}\Delta_2(n,m)+m^2\Delta_1(n,m)+m^2L_f\overline{\omega}\Delta_3(n,m)+\frac{m}{n}\sqrt{\log(nm)}\right),\\ \widehat{IE}-IE &= O_p\left(m\overline{\omega}\Delta_2(n,m)+m^2\Delta_1(n,m)+m^2L_f\overline{\omega}\Delta_3(n,m)+\frac{m}{n}\sqrt{\log(nm)}\right).\end{aligned}$$

Since the convergence rates of NN-based learners have been widely studied in the literature (see e.g., Shen et al., 2019; Schmidt-Hieber, 2020; Shen et al., 2022; Yan and Yao, 2023), these results can be used to establish the convergence rates of $\widehat{G}_a$ and $\widehat{g}_a$.

Finally, we impose the following regularity assumptions for the proposed tests in spatio-temporal dependent experiments based on L-STVCDP.

ASSUMPTION 6. *For any τ,ι, $\mathbb{E}(Z_{i,\tau,\iota}^\top Z_{i,\tau,\iota})$ is invertible; $\theta(\tau,\iota)$, $\Sigma_{\eta^I}(\tau_1,\tau_2,\iota_1,\iota_2)$, $\Sigma_{\eta^{II}}(\tau_1,\iota_1,\tau_2)$, and $\Sigma_{\eta^{III}}(\tau_1,\iota_1,\iota_2)$ have bounded and continuous second-order derivatives.*

ASSUMPTION 7. *There exists $q<1$ such that the absolute values of eigenvalues of $\Phi(\tau,\iota)$ are smaller than q. In addition, there exist M_Γ and $M_\beta<\infty$ such that $\|\Gamma_1(\tau,\iota)+\Gamma_2(\tau,\iota)\|_\infty \le M_\Gamma$ and $\|\beta(\tau,\iota)\|_\infty \le M_\beta$. $\Theta(\tau,\iota)$ has a bounded and continuous second-order derivative.*

With these assumptions, we present the asymptotic properties of our DE and IE estimators and their associated test statistics for the spatio-temporal dependent experiments based on L-STVCDP. Define

$$\boldsymbol{V}_{\widetilde{\theta}_{st}}(\tau_1,\iota_1,\tau_2,\iota_2)=\{\mathbb{E}Z_{i,\tau_1,\iota_1}Z_{i,\tau_1,\iota_1}^\top\}^{-1}\mathbb{E}\{Z_{i,\tau_2,\iota_2}Z_{i,\tau_1,\iota_1}^\top\Sigma_{\eta^I}(\tau_1,\iota_1,\tau_2,\iota_2)\}\{\mathbb{E}Z_{i,\tau_2,\iota_2}Z_{i,\tau_2,\iota_2}^\top\}^{-1}$$

as the asymptotic covariance between $\sqrt{n}\widetilde{\theta}_{st}(\tau_1,\iota_1)$ and $\sqrt{n}\widetilde{\theta}_{st}(\tau_2,\iota_2)$.

THEOREM 6. *Suppose* $\lambda_{\min}(\boldsymbol{V}_{\widetilde{\theta}_{st}})$ *is bounded away from zero. Under Assumptions 2, 3 and 6, for any set of* $(d+2)$*-dimensional vectors* $\{B_{\tau,\iota}\}_{\tau,\iota}$*, we have as* $n, m, r \to \infty$*,* $h, h_{st} \to 0$ *and* $mh, rh_{st} \to \infty$ *that*

(i) For any set of $(d+2)$*-dimensional vectors* $\{B_{\tau,\iota}\}_{\tau,\iota}$ *with* $\sum_{\tau_1,\tau_2,\iota_1,\iota_2} B_{\tau_1,\iota_1}^\top \boldsymbol{V}_{\widetilde{\theta}_{st}}(\tau_1,\iota_1,\tau_2,\iota_2) B_{\tau_2,\iota_2} \geq c\sum_{\tau,\iota}\|B_{\tau,\iota}\|_2^2$ *for some constant* $c>0$*, we have*

$$\sqrt{\bar{n}}\sum_{\tau,\iota}[B_{\tau,\iota}^\top \widetilde{\theta}_{st}(\tau,\iota) - \theta_{st}(\tau,\iota)] \Big/ \sqrt{\sum_{\tau_1,\tau_2,\iota_1,\iota_2} B_{\tau_1,\iota_1}^\top \boldsymbol{V}_{\widetilde{\theta}_{st}}(\tau_1,\iota_1,\tau_2,\iota_2) B_{\tau_2,\iota_2}} \xrightarrow{d} N(b_{n,st}, 1),$$

where the bias $b_{n,st} = O(\sqrt{\bar{n}}h^2 + \sqrt{\bar{n}}h_{st}^2 + \sqrt{\bar{n}}m^{-1} + \sqrt{\bar{n}}r^{-1})$*.*

(ii) Suppose $h, h_{st} = o(n^{-1/4})$ *and* $m, r \gtrsim \sqrt{\bar{n}}$*. Then for the hypotheses* (16)*,* $\mathbb{P}(\widehat{DE}_{st}/\widehat{se}(\widehat{DE}_{st}) > z_\alpha) = \alpha + o(1)$ *under* H_0^{DE} *and* $\mathbb{P}(\widehat{DE}_{st}/\widehat{se}(\widehat{DE}_{st}) > z_\alpha) \to 1$ *under* H_1^{DE}*.*

THEOREM 7. *Suppose that there are some constants* $0 < c_1 \leq 1$ *such that* $c_1 \leq \mathbb{E}\varepsilon_{\tau,\iota,S}^2, \mathbb{E}e_{\tau,\iota}^2 \leq c_1^{-1}$ *for all* $1 \leq \tau \leq m$*,* $1 \leq \iota \leq r$*, and that* $h, h_{st} = o(n^{-1/4})$*,* $m, r \gtrsim \sqrt{\bar{n}}$ *and* $mr \asymp n^{c_2}$ *for some constant* $c_2 < 3/2$*. Then under Assumptions of Theorem 6 and Assumption 7, with probability approaching 1,*

$$\sup_z |\mathbb{P}(\widehat{IE}_{st} - IE_{st} \leq z) - \mathbb{P}(\widehat{IE}_{st}^b - \widehat{IE}_{st} \leq z | Data)| \leq C(\sqrt{\bar{n}}h^2 + \sqrt{\bar{n}}h_{st}^2 + \sqrt{\bar{n}}m^{-1} + \sqrt{\bar{n}}r^{-1} + n^{-1/8}), \quad (20)$$

where C *is some positive constant.*

Theorem 6 establishes the limiting distribution of the proposed DE estimator for the spatio-temporal dependent experiments. Similar to Proposition 2, we can show that the smoothed estimator is more efficient when $m, r \gtrsim \sqrt{\bar{n}}$ and $h^4, h_{st}^4 = o(n^{-1})$. In addition, Theorem 7 allows both m and r to be either fixed, or diverge with n, and is thus applicable to a wide range of applications.

5. Real data based simulations

5.1. Temporal alternation design

In this section, we conduct Monte Carlo simulations to examine the finite sample properties of the proposed test statistics based on L-TVCDP and L-STVCDP models. To generate data under the temporal alternation design, we design two simulation environments based on two real datasets obtained from Didi Chuxing. The first dataset is collected from a given city A from Dec. 5th, 2018 to Jan. 13th, 2019. Thirty-minutes is defined as one time unit. The second dataset is from another city B, from May 17th, 2019 to June 25th, 2019. One-hour is defined as one time unit. Both contain data for 40 days. Due to privacy, we only present scaled metrics in this paper. Figure 1 depicts the trend of some business metrics over time across 40 different days. These metrics include drivers' total income, the number of requests and drivers' total online time. Among them, the first quantity is our outcome of interest and the last two are considered as the state variables to characterize the demand and supply networks. As expected, these quantities show a similar pattern, achieving the largest values at peak time.

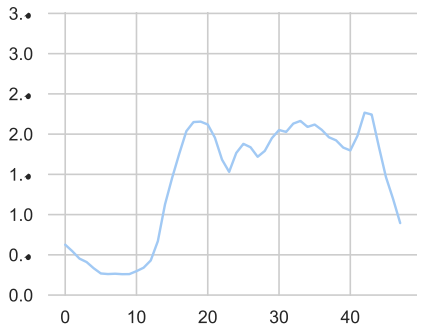

Fig. 1. Scaled business metrics from City A (the first row) and City B (the second row) across 40 days, including drivers' total income, the numbers of requests and drivers' total online time.

We next discuss how to generate synthetic data based on the real datasets. The main idea is to fit the proposed L-TVCDP models to the real dataset and apply the parametric bootstrap to simulate the data. Let $\widetilde{\beta}_0(\tau)$, $\widetilde{\beta}(\tau)$, $\widetilde{\phi}_0(\tau)$, and $\widetilde{\Phi}(\tau)$ denote the smoothed estimators for $\beta_0(\tau)$, $\beta(\tau)$, $\phi_0(\tau)$ and $\Phi(\tau)$, respectively. We set $\widetilde{\gamma}(\tau)$ and $\widetilde{\Gamma}(\tau)$ to $(\delta/100) \times (\sum_{i,\tau} Y_{i,\tau}/nm)$ and $(\delta/100) \times (\sum_{i,\tau} S_{i,\tau}/nm)$, respectively. As such, the parameter δ controls the degree of the treatment effects. Specifically, the null holds if $\delta = 0$ and the alternative holds if $\delta > 0$. It corresponds to the increase relative to the outcome (state). We next generate the policies according to the temporal alternation design and simulate the responses and states based on the fitted model. Let TI denote the time span we implement each policy under the alternation design. For instance, if TI $= 3$, then we first implements one policy for three hours, then switch to the other for another three hours and then switch back and forth between the two policies. We consider three choices of n *2* *f*$8, 14, 20$*g*, fives choices of δ *2* *f*$0, 0.25, 0.5, 0.75, 1$*g* and three choices of TI *2* *f*$1, 3, 6$*g*. This corresponds to a total of 45 cases. The bandwidth is set $h = Cn^{-1/3}$, where C is selected by the 5-fold cross validation method.

In Figure 2, we depict the empirical rejection probabilities of the proposed test for DE, aggregated over 400 simulations, for all combinations. It can be seen that our test controls the type-I error and its power increases as δ increases. In addition, the empirical rejection rates decreases as TI increases. This phenomenon suggests that the more frequently we switch back and forth between the two policies, the more powerful the resulting test. It is due to the positive correlation between adjacent observations. To elaborate, consider the extreme case where we switch policies at each time. The policies assigned at any two adjacent time points are different. As such, the random effect cancels with each other, yielding an efficient estimator. We conduct some additional simulations using the numbers of answered requests and finished requests of cities A and B as responses (see Figure S.2 in the supplement). Results are very similar and are reported in Figures S.3–S.4 in the supplementary document. See also Tables S.1–S.2 in the supplementary document.

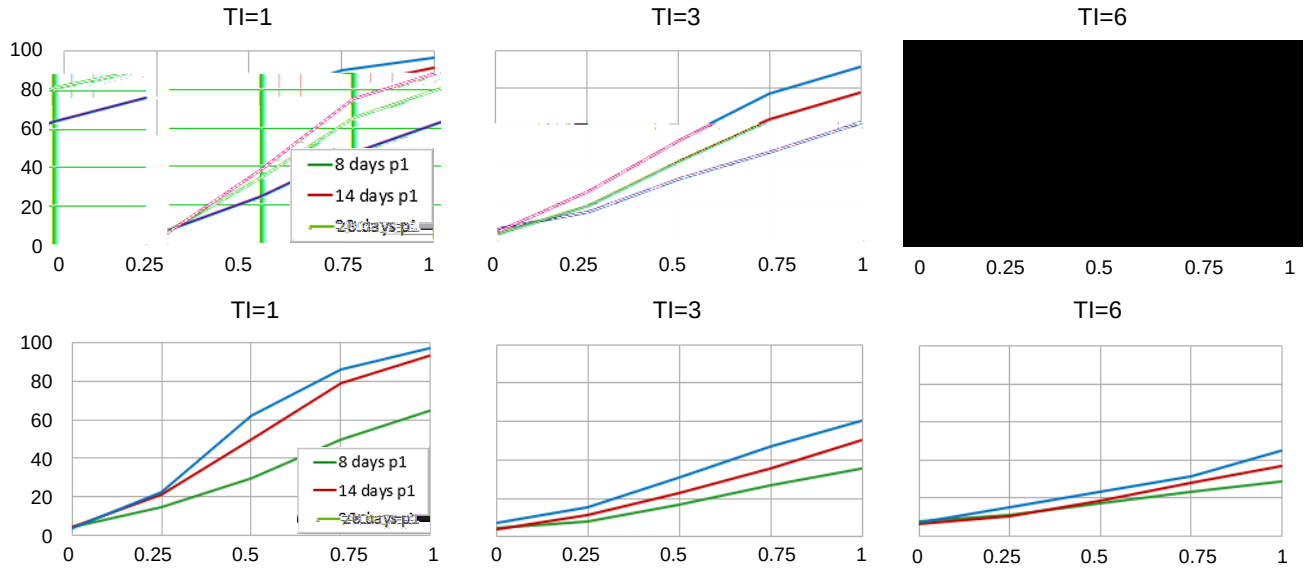

Fig. 2. Simulation results for L-TVCDP: empirical rejection rates (expressed as percentages) of the proposed test for DE under different combinations of (n, δ, TI) and types of outcomes. Synthetic data are simulated based on the real dataset from city A (the first row) and city B (the second row).

To infer IE, we set the outcome to drivers' total online income. The empirical rejection probabilities of the proposed test for IE are reported in Figure 3. Results are aggregated over 400 simulations. Similarly, the proposed test is consistent. Its power increases with the sample size and δ. In addition, its power under $\mathrm{TI} = 1$ is much larger than those under $\mathrm{TI} = 3$ or 6. This suggests that we shall switch back and forth between the two policies as frequently as possible to maximize the power property of the test (see also Tables S.3–S.4 in Supplementary document).

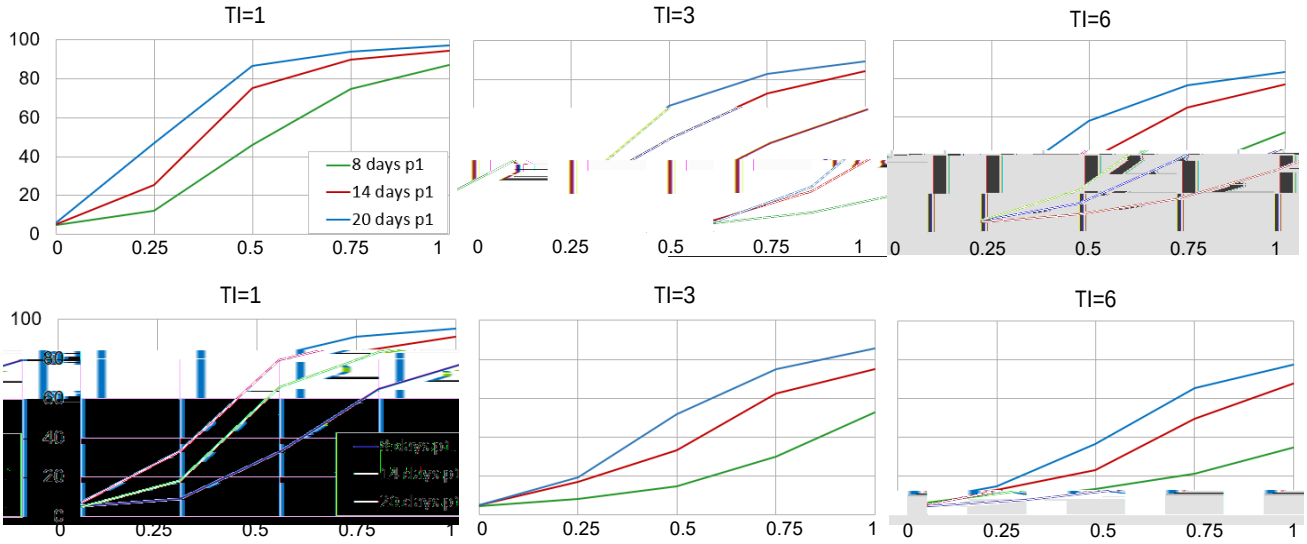

Fig. 3. Simulation results for L-TVCDP: empirical rejection rates (expressed as percentages) of the proposed test for IE under different combinations of (n, δ, TI). Synthetic data are simulated based on the real dataset from city A (the first row) and city B (the second row).

5.2. Spatio-temporal alternation design

To generate data under the spatio-temporal alternation design, we create a simulation environment based on the real dataset from city A. We divide the city into 10 non-overlapping regions. We plot these variables associated with 3 particular regions, over the first 10 days in Figure 4. It can be seen that although the daily trends differ across regions, the state and the response are highly correlated.

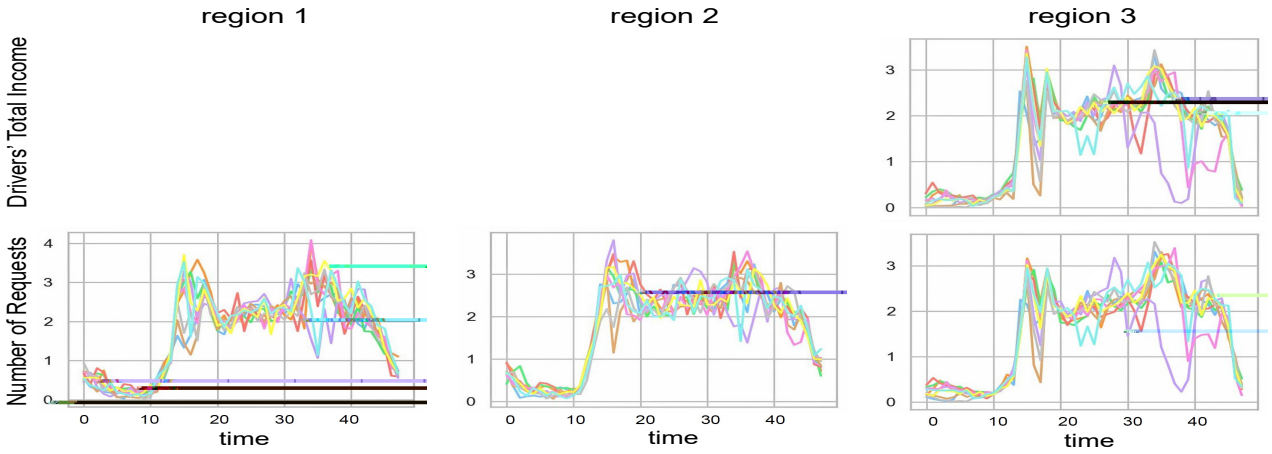

Fig. 4. Number of call requests and drivers' total income across different regions and days. The values are scaled for privacy concerns.

We fit the proposed models in (18) to the real dataset to estimate the varying coefficients and the variances of the random errors. Then we manually set the treatment effects $\widehat{\gamma}(\tau,\iota)$ and $\widehat{\Gamma}(\tau,\iota)$ to $(\delta_1/100) \times (\sum_{i=1}^{n}\sum_{\tau=1}^{m} Y_{i,\tau,\iota}/nm)$ and $(\delta_2/100) \times (\sum_{i=1}^{n}\sum_{\tau=1}^{m} S_{i,\tau,\iota}/nm)$ for some constants δ_1 and $\delta_2 > 0$. We consider both the temporal and spatio-temporal alternation designs, and simulate the data via parametric bootstrap.

We also consider three choices of n *2* *f*$8, 14, 20$*g*, three choices of TI *2* *f*$1, 3, 6$*g* and three choices of δ_1, δ_2 *2* *f*$0, 0.5, 1$*g*. This yields a total of 81 combinations under each design. The rejection probabilities of the proposed tests for DE and IE tests are reported in Figures 5 and 6 (see also Tables S.5 and S.6 in the supplementary document). It can be seen that the type I error rates of the proposed test are close to the nominal level under both designs. More importantly, the power under spatio-temporal alternation design is higher than that of temporal alternation design in all cases. The reason is twofold. First, under the spatio-temporal design, we independently randomize the initial policy for each region, and adjacent regions may receive different policies. Observations across adjacent areas are likely to be positively correlated. As such, the variance of the estimated treatment effects will be smaller than that under the temporal design where all regions receive the same policy at each time. Second, we employ kernel smoothing twice when computing $\widehat{\mathrm{DE}}_{st}$ and $\widehat{\mathrm{IE}}_{st}$, as discussed in Section 3. This results in a more efficient estimator. In addition, compared with the results in Tables S.1 and S.3, it can be seen that the test that focuses on the entire city has better power property than the one that considers a particular region in general. Finally, the power decreases with TI and increases with n, δ_1 and δ_2.

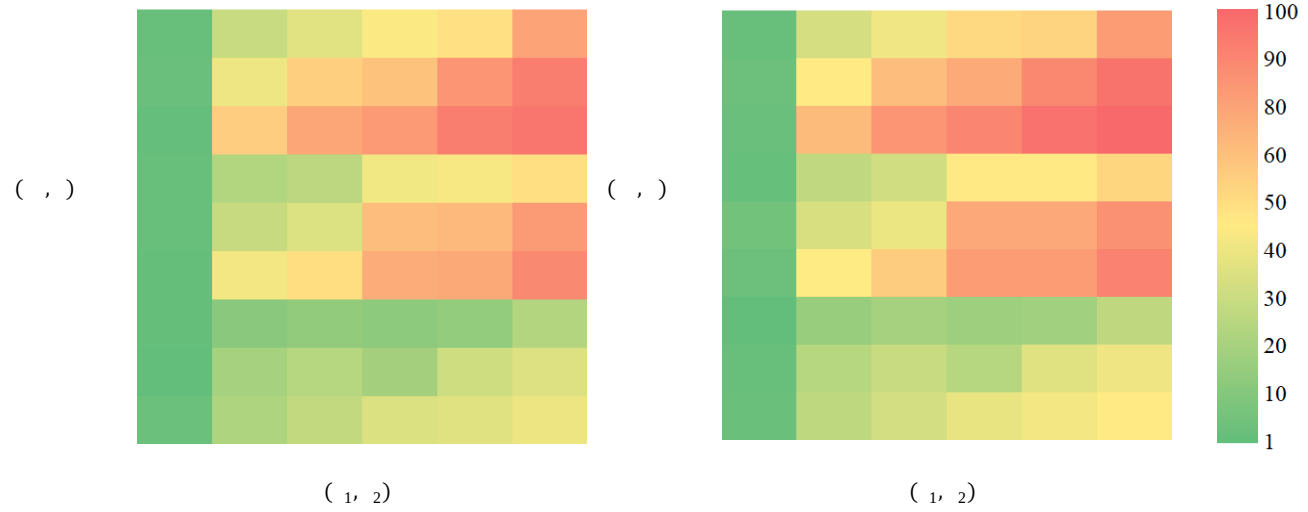

Fig. 5. Simulation results for L-STVCDP: the empirical rejection probabilities of the proposed test test for DE

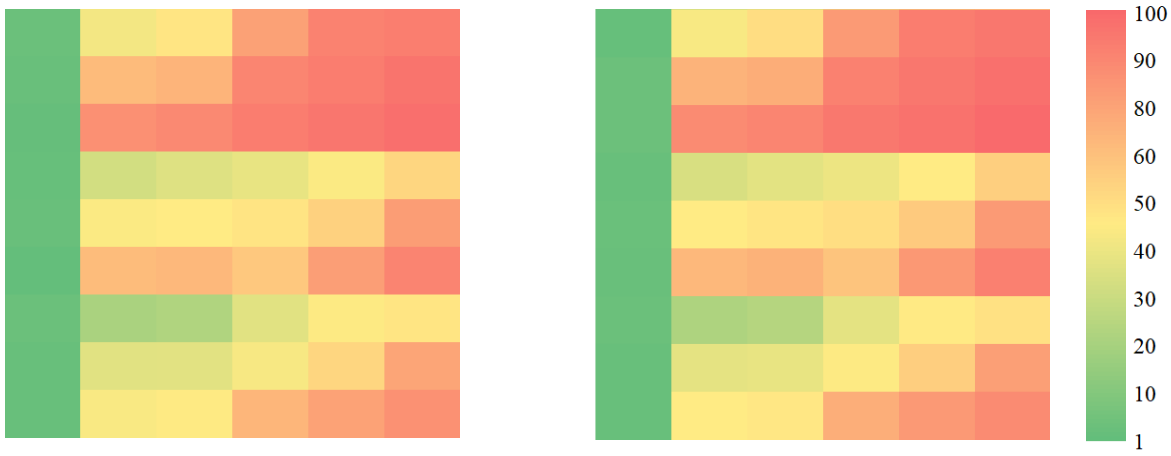

Fig. 6. Simulation results for L-STVCDP: the empirical rejection probabilities of the proposed test test for IE under the temporal alternation design (left panel) and the spatio-temporal alternation design (right panel).

6. Real data analysis

In this section, we apply the proposed tests based on L-TVCDP and L-STVCDP to a number of real datasets from Didi Chuxing to examine the treatment effects of some newly developed order dispatch and vehicle reposition policies. Due to privacy, we do not publicize the names of these policies.

We first consider four data sets collected from four online experiments under the temporal alternation design. All the experiments last for 14 days. Policies are executed based on alternating half-hourly time intervals. We denote the cities, in which these experiments take place, as C_1, C_2, C_3, and C_4 and their corresponding policies as S_1, S_2, S_3, and S_4, respectively. For each policy, we are interested in its effect on three key business metrics, including drivers' total income, the answer rate, and the completion rate. Similar to Section 5.1, we use the number of call orders and drivers' total online time to construct the time-varying state variables.

All the new policies are compared with some baseline policies in order to evaluate whether they improve some business outcomes. Specifically, in city C_1, policy S_1 is proposed to reduce the answer time (the time period between the time when an order is requested and the time when the order is responded by the driver). This in turn meets more call orders requests. Both policy S_2 in city C_2 and policy S_3 in city C_3 are designed to guide drivers to regions with more orders in order to reduce drivers' idle time ratio. Policies S_2 and S_3 are designed to assign more drivers to areas with more orders. This in turn reduces drivers' downtime and increase their income. Policy S_4 aims to balance drivers' downtime and their average pick-up distance.

We also apply our test to another four datasets collected from four A/A experiments which compare the standard policy against itself. These A/A experiments are conducted two weeks before the A/B experiments. Each lasts for 14 days and thirty-minutes is defined as one time unit. We remark that the A/A experiment is employed as a sanity check for the validity of the proposed test. We expect our test will not reject the null when applied to these datasets, since the sole standard policy is used.

We fit the proposed L-TVCDP models to each of the eight datasets. In Figures 7 and 8, we plot the predicted outcomes against the observed values and plot the corresponding residuals over time for policy S_1. Results for policies S_2–S_4 are represented in Figure S.5 in the supplementary article. It can be seen that the predicted outcomes are very close to the observed values, suggesting that the proposed model fits the data well. P-values of the proposed tests are reported in Tables 1 and 2. As expected, the proposed test does not reject the null hypothesis when applied to all datasets from A/A experiments. When applied to the data from A/B experiments, it can be seen that the new policy S_1 directly improves the answer rate and the completion rate, while increasing drivers' total income in city C_1. It also significantly increases drivers' income in the long run. Policy S_2 has significant direct and indirect effects on drivers' income as expected. Policy S_4 significantly increases the immediate answer rate, while improving the overall passenger satisfaction. However, policy S_3 is not significantly better than the standard policy.

We further apply the proposed test to two real datasets collected from an A/A and A/B 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, with DTI, ART and CRT corresponding to drivers' total income, the answer rate and the completion rate, respectively.

	AA			AB		
	DTI(%)	ART(%)	CRT(%)	DTI(%)	ART(%)	

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

under the spatio-temporal alternation design, conducted in city C_5. This city is partitioned into 17 regions. Within each region, more than 90% orders are answered by drivers in the same region. Similar to the temporal alternation design, both experiments last for 14 days and 30-minutes is set as one time unit. We take the number of requests as the state variables and drivers' total income as the outcome, as in Section 5.2. In Figures 9 and 10, we plot the fitted drivers' total income and the fitted number of requests against their observed values, and plot the corresponding residuals over time. We only present results associated with 2 regions in the city for space economy. The fitted values and residuals associated with other regions are similar and we do not present them to save space. It can be seen that the proposed models fit these datasets well. In addition, we report the p-values of the proposed test in Table 3. It can be seen that the new policy significantly increases drivers' income. When applied to the dataset from the A/A experiment, it fails to reject either null hypothesis.

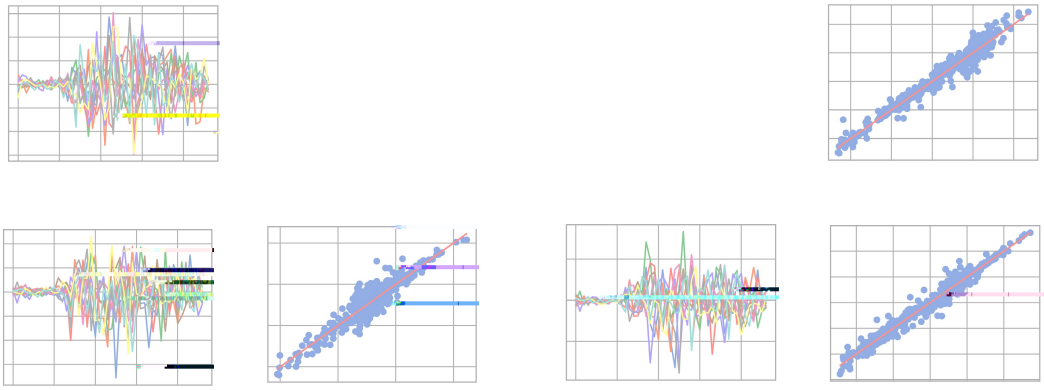

Fig. 9. Plots of the fitted drivers' income against the observed values, as well as the corresponding residuals. Data are collected from an A/A or A/B experiment under the spatio-temporal alternation design.

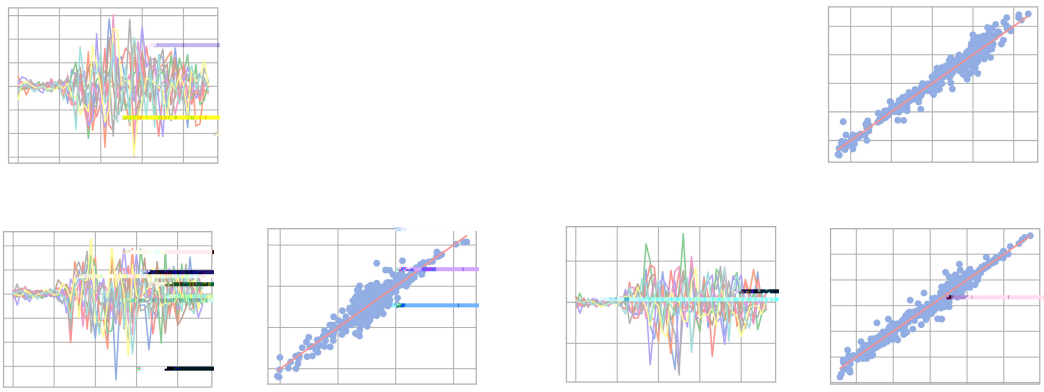

Fig. 10. Plots of the fitted number of orders against the observed values, as well as the corresponding residuals. Data are collected from an A/A or A/B experiment under the spatio-temporal alternation design.

7. Discussion

In this study, driven by the need for policy evaluation in technological companies, we thoroughly examine AB testing for temporal and/or spatial dependent experiments, particularly in scenarios characterized by weak signals, (spatio)-temporal random effects, and intricate interference structures. Our approach offers two key benefits. Firstly, it accommodates the switchback design, which can significantly enhance testing power. As explained earlier, by applying diverse treatments to neighboring time points, we can potentially offset the impact of random effects at these times, resulting in more efficient estimations of treatment effects. Secondly, we break down the ATE into its DE and IE components. We then advocate for testing these effects separately. This separation aids decision-makers in gaining a clearer understanding of how different policies function and in devising more effective strategies and designs. Further details can be found in Section S.12.4 of the supplementary document.

There are several intriguing avenues for future research. Firstly, considering Assumptions 1 and 2, it's worth exploring scenarios where errors in the state regression model are not necessarily independent over time. This can be achieved by incorporating random effects into the state regression model, allowing for correlated errors over time. However, this introduces dependencies between these random effects, which in turn affects the conditional independence of past and future features. Consequently, the Markov assumption is violated, and applying existing OPE methods and our proposal from Section 2 directly would result in biased policy value estimations. In Section S.12.1 of the supplementary document, we present two approaches to mitigate this endogeneity bias. Secondly, we can delve into situations involving a large number of state variables. However, in ride-sharing platforms, it's reasonable to assume that the dimension of state variables is fixed. This typically involves a two-dimensional market feature, encompassing the number of call orders and the number of available drivers. We outline potential extensions to high-dimensional settings in Section S.12.2 of the supplementary document. Thirdly, while the interference structure examined in this work is general, it remains relatively simple. It would be intriguing to explore more complex structural interferences across both space and time. Lastly, addressing statistical inference for deep neural networks remains an open challenge. This could represent a significant step toward incorporating deep learning into causal inference, offering promising directions for future research.

Acknowledgments

The first three authors, Drs. Luo, Yang, and Shi, contributed equally to this paper. This work was finished when Drs. Luo, Ye, and Zhu worked at Didi Chuxing. The authors would like to express their gratitude to the Editors, the Associate Editor and the two reviewers for their constructive and insightful comments.

Supplementary material

Supplementary data is available online at Journal of the Royal Statistical Society online.

Conflict of interests

None declared.

Funding

Yang's research is partially supported by China Postdoctoral Science Foundation (No. 2022TQ0360, 2022M723334) and the National Natural Science Foundation of China (No. 71988101). 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 funding agency.

Data availability

The code can be found on our GitHub page at https://github.com/BIG-S2/STVCM. We have included detailed instructions in the "Readme" file on how to reproduce Figures 1-10 and Tables 1-3. The real data used in our study is proprietary and cannot be shared publicly. However, we have provided simulated datasets that can yield similar results for broader accessibility and understanding.

References

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E. and Rus, D. (2017) On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, **114**, 462–467.

Arkhangelsky, D., Imbens, G. W., Lei, L. and Luo, X. (2021) Double-robust two-way-fixed-effects regression for panel data. *arXiv preprint arXiv:2107.13737*.

Aronow, P. M. and Samii, C. (2017) Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, **11**, 1912–1947.

Aronow, P. M., Samii, C. and Wang, Y. (2020) Design-based inference for spatial experiments with interference. *arXiv preprint arXiv:2010.13599*.

Bakshy, E., Eckles, D. and Bernstein, M. S. (2014) Designing and deploying online field experiments. In *Proceedings of the 23rd International Conference on World Wide Web*, 283–292.

Bimpikis, K., Candogan, O. and Saban, D. (2019) Spatial pricing in ride-sharing networks. *Operations Research*, **67**, 744–769.

Bojinov, I. and Shephard, N. (2019) Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*, **114**, 1665–1682.

Boruvka, A., Almirall, D., Witkiewitz, K. and Murphy, S. A. (2018) Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association*, **113**, 1112–1121.

Castillo, J. C., Knoepfle, D. and Weyl, G. (2017) Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, 241–242.

Cohen, M. C., Fiszer, M. D. and Kim, B. J. (2022) Frustration-based promotions: Field experiments in ride-sharing. *Management Science*, **68**, 2432–2464.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020) Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, **110**, 2964–96.

Garg, N. and Nazerzadeh, H. (2022) Driver surge pricing. *Management Science*, **68**, 3219–3235.

Hagiu, A. and Wright, J. (2019) The status of workers and platforms in the sharing economy. *Journal of Economics & Management Strategy*, **28**, 97–108.

Halloran, M. E. and Hudgens, M. G. (2016) Dependent happenings: A recent methodological review. *Current Epidemiology Reports*, **3**, 297–305.

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

Hu, Y. and Wager, S. (2021) Off-policy evaluation in partially observed markov decision processes under sequential ignorability. *arXiv preprint arXiv:2110.12343*.

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

Imai, K. and Kim, I. S. (2021) On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, **29**, 405–415.

Jiang, N. and Li, L. (2016) Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, 652–661. PMLR.

Johari, R., Li, H., Liskovich, I. and Weintraub, G. Y. (2022) Experimental design in two-sided platforms: An analysis of bias. *Management Science*, **68**, 7065–7791.

Kallus, N. and Uehara, M. (2020) Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *Journal of Machine Learning Research*, **21**.

— (2022) Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning. *Operations Research*, **70**, 3282–3302.

Lale, S., Azizzadenesheli, K., Hassibi, B. and Anandkumar, A. (2021) Adaptive control and regret minimization in linear quadratic gaussian (lqg) setting. In *2021 American Control Conference (ACC)*, 2517–2522. IEEE.

Larsen, N., Stallrich, J., Sengupta, S., Deng, A., Kohavi, R. and Stevens, N. T. (2023) Statistical challenges in online controlled experiments: A review of a/b testing methodology. *The American Statistician*, 1–15.

Lee, L. (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, **140**, 333–374.

Lewis, G. and Syrgkanis, V. (2020) Double/debiased machine learning for dynamic treatment effects via g-estimation. *arXiv preprint arXiv:2002.07285*.

Liao, P., Klasnja, P. and Murphy, S. (2021) Off-policy estimation of long-term average outcomes with

applications to mobile health. *Journal of the American Statistical Association*, **116**, 382–391.

Liao, P., Qi, Z., Klasnja, P. and Murphy, S. (2020) Batch policy learning in average reward markov decision processes. *arXiv preprint arXiv:2007.11771*.

Liu, L., Hudgens, M. G. and Becker-Dreps, S. (2016) On inverse probability-weighted estimators in the presence of interference. *Biometrika*, **103**, 829–842.

Liu, Q., Li, L., Tang, Z. and Zhou, D. (2018) Breaking the curse of horizon: Infinite-horizon off-policy estimation. In *Advances in Neural Information Processing Systems*, vol. 31.

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E. and Kosorok, M. R. (2020) Estimating dynamic treatment regimes in mobile health using v-learning. *Journal of the American Statistical Association*, **115**, 692–706.

Luedtke, A. R. and Van Der Laan, M. J. (2016) Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, **44**, 713–742.

Manski, C. F. (2013) Identification of treatment response with social interactions. *The Econometrics Journal*, **16**, S1–S23.

Munro, E., Wager, S. and Xu, K. (2021) Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*.

Perez-Heydrich, C., Hudgens, M. G., Halloran, M. E., Clemens, J. D., Ali, M. and Emch, M. E. (2014) Assessing effects of cholera vaccination in the presence of interference. *Biometrics*, **70**, 731–741.

Pollmann, M. (2020) Causal inference for spatial treatments. *arXiv preprint arXiv:2011.00373*.

Puelz, D., Basse, G., Feller, A. and Toulis, P. (2019) A graph-theoretic approach to randomization tests of causal effects under general interference. **arXiv**, 1910.10862v1.

Puterman, M. L. (2014) *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

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

Reich, B. J., Yang, S., Guan, Y., Giffin, A. B., Miller, M. J. and Rappold, A. (2020) A review of spatial causal inference methods for environmental and epidemiological applications. **arXiv**, 2007.02714v1.

Rubin, D. (1980) Discussion of "randomization analysis of experimental data in the fisher randomization test" by d. basu. *Journal of the American Statistical Association*, **75**, 591–593.

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

Schmidt-Hieber, J. (2020) Nonparametric regression using deep neural networks with relu activation function. *The Annals of Statistics*, **48**, 1875–1897.

Shen, Z., Yang, H. and Zhang, S. (2019) Deep network approximation characterized by number of neurons. *arXiv preprint arXiv:1906.05497*.

— (2022) Optimal approximation rate of relu networks in terms of width and depth. *Journal de Mathématiques Pures et Appliquées*, **157**, 101–135.

Shi, C., Wan, R., Chernozhukov, V. and Song, R. (2021) Deeply-debiased off-policy interval estimation. In *International conference on machine learning*, 9580–9591. PMLR.

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

Shi, C., Wang, X., Luo, S., Zhu, H., Ye, J. and Song, R. (2023) Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework. *Journal of the American Statistical Association*, **118**, 2059–2071.

Shi, C., Zhang, S., Lu, W. and Song, R. (2022b) Statistical inference of the value function for reinforcement learning in infinite-horizon settings. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **84**, 765–793.

Shumway, R. and Stoffer, D. (2010) *Time series analysis and its applications with R examples (3rd ed.)*. Springer.

Sobel, M. E. (2006) What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*, **101**, 1398–1407.

Sobel, M. E. and Lindquist, M. A. (2014) Causal inference for fmri time series data with systematic errors of measurement in a balanced on/off study of social evaluative threat. *Journal of the American Statistical Association*, **109**, 967–976.

Tang, X., Qin, Z., Zhang, F., Wang, Z., Xu, Z., Ma, Y., Zhu, H. and Ye, J. (2019) A deep value-network based approach for multi-driver order dispatching. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1780–1790.

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2012) On causal inference in the presence of interference. *Statistical Methods in Medical Research*, **21**, 55–75.

Thomas, P. and Brunskill, E. (2016) Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning*, 2139–2148. PMLR.

Uehara, M., Shi, C. and Kallus, N. (2022) A review of off-policy evaluation in reinforcement learning. *arXiv preprint arXiv:2212.06355*.

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

Verbitsky-Savitz, N. and Raudenbush, S. W. (2012) Causal inference under interference in spatial settings: A case study evaluating community policing program in chicago. *Epidemiologic Methods*, **1**, 107–130.

Wager, S. and Xu, K. (2021) Experimenting in equilibrium. *Management Science*, **67**, 6694–6715.

Wooldridge, J. M. (2021) Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

Wu, C.-F. J. et al. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, **14**, 1261–1295.

Yan, S. and Yao, F. (2023) Nonparametric regression for repeated measurements with deep neural networks. *arXiv preprint arXiv:2302.13908*.

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

Zhou, F., Luo, S., Qie, X., Ye, J. and Zhu, H. (2021) Graph-based equilibrium metrics for dynamic supply–demand systems with applications to ride-sourcing platforms. *Journal of the American Statistical Association*, **116**, 1688–1699.

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

Zigler, C. M., Dominici, F. and Wang, Y. (2012) Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics*, **13**, 289–302.