SUPPLEMENTARY APPENDIX FOR "CAUSAL INFERENCE WITH SPATIO-TEMPORAL DATA"

Georgia Papadogeorgou Kosuke Imai Jason Lyall Fan Li

Table of Contents

A Notation **2**

B Theoretical Proofs **2**

B.1 Regularity conditions . . . 2

B.2 Proofs: The propensity score as a balancing score . . . 5

B.3 Proofs: Asymptotic normality based on the true propensity score . . . 6

B.4 Proofs: Asymptotic normality based on the estimated propensity score . . . 13

B.5 Asymptotics for an increasing number of independent regions . . . 33

C The Hájek Estimator **37**

D Sensitivity analysis **38**

D.1 For the IPW estimator . . . 38

D.2 For the Hájek estimator . . . 39

E Additional Simulation Results on the Iraq-based scenario **40**

E.1 Asymptotic Variance and Bound, and Estimated Variance Bound . . . 40

E.2 Coverage of the Confidence Intervals for the IPW and Hájek Estimators . . . 41

E.3 Uncertainty Estimates . . . 44

E.4 Covariate Balance . . . 46

F Additional simulations on a square geometry **46**

F.1 The Simulation Design . . . 46

F.2 Simulation Results . . . 48

G Additional Empirical Results **55**

G.1 Visualization . . . 55

G.2 Empirical Results . . . 58

G.3 Single time point adaptive interventions . . . 58

A Notation

Table A.1: Notation.

Paths	$\overline{\boldsymbol{W}}_t$	Treatments over the time periods $1, \ldots, t$
	$\overline{\boldsymbol{w}}_t$	Realized treatment assignments for time periods $1, \ldots, t$
	$\overline{\mathcal{Y}}_t$	Collection of all potential outcomes for time periods $1, \ldots, t$
	$\overline{\boldsymbol{Y}}_t$	Observed outcomes for time periods $1, \ldots, t$
Intervention	M	The number of time periods over which we intervene
	h	Poisson point process intensity defining the stochastic intervention
Estimands	N_t, N	Expected number of outcome-active locations during time period t for an intervention over M time periods, and their average over time
	τ_t^M, τ^M	Expected change in the number of outcome-active locations comparing two interventions for time period t and their average over time
Estimators	$\widehat{Y}_t^M$	Estimated continuous surface the integral of which is used for calculating $\widehat{N}_t$
	$\widehat{N}_t$, $\widehat{N}$	Estimated expected number of points during time period t for an intervention taking place over the preceding M time periods, and their average over time
	$\widehat{\tau}_t^M$, $\widehat{\tau}^M$	Estimated expected change in the number of outcome-active locations for time period t comparing two interventions, and their average over time
Arguments	B	The set over which the number of outcome-active locations are counted

B Theoretical Proofs

B.1 Regularity conditions

For $\epsilon > 0$, we use $\mathcal{N}_\epsilon(A)$ to denote the $\epsilon-$neighborhood of a set A: $\mathcal{N}_\epsilon(A) = \{\omega \in \Omega : \text{there exists } a \in A \text{ with dist}(\omega, a) < \epsilon\}$. Also, we use ∂B to denote the boundary of B, formally defined as the set of points for which an open ball of any size centered at them includes points both in and outside B, i.e., $\partial B = \{s \in \Omega \text{ such that, for every } \epsilon > 0, \text{ there exist } s_1, s_2 \in \mathcal{N}_\epsilon(s) \text{ for which } s_1 \in B \text{ and } s_2 \notin B\}$.

Regularity conditions for asymptotic results when using the true or estimated propensity score The following assumption includes regularity conditions which are used to show asymptotic normality of the estimator based on the true or estimated propensity score:

Assumption A.1. *The following three conditions hold.*

(a) There exists $\delta_Y > 0$ *such that* $|S_{Y_t(\overline{\boldsymbol{w}}_t)}| < \delta_Y$ *for all* $t \in \mathcal{T}$ *and* $\overline{\boldsymbol{w}}_t \in \mathcal{W}^T$.

(b) Let $v_t = Var\left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} N_B(Y_t) \mid \overline{H}^*_{t-M}\right]$ *for* $t \geq M$. *Then, there exists* $v \in \mathbb{R}^+$ *such that* $(T-M+1)^{-1}\sum_{t=M}^{T} v_t \xrightarrow{p} v$ *as* $T \to \infty$.

(c) There exists $\delta_B > 0$ *and* $Q^* \in (1/2, 1)$ *such that*

$$P\left(\sum_{t=M}^{T} I\Big(\exists s \in S_{Y_t} \cap \mathcal{N}_{\delta_B}(\partial B)\Big) > T^{1-Q^*}\right) \to 0, \text{ as } T \to \infty.$$

Assumption A.1(a) states that there is an upper limit on the number of outcome-active locations at any time period and under any treatment path. In our application, it is reasonable to assume that the number of insurgent attacks occurring during any day is bounded. In Assumption A.1(b), $\overline{H}^*_t$ represents the expanded history preceding W_{t+1}, including previous treatments, all potential outcomes, and all potential confounders. Given the assumptions of bounded relative positivity and bounded number of outcome-active locations, Assumption A.1(b) is a weak condition, as it states that the average of bounded quantities converges. Lastly, Assumption A.1(c) states that the probability that we observe more than T^{1-Q^*} time periods with outcome-active locations within a δ_B-neighborhood of B's boundary goes to zero as the number of observed time periods increases. Since the size of the boundary's neighborhood can be arbitrarily small, this assumption is also reasonable. Informally, Assumption A.1(c) would be violated in our study if insurgent attacks occurred at the *boundary* of region B more often than during $\sqrt{T}$ time periods. As long as the regions B are decided upon substantive interest, we would expect this assumption to be satisfied. Alternatively, regions B can be defined by avoiding setting the region's boundary at observed outcome-active locations.

Regularity conditions for asymptotic results when using the estimated propensity score Next, we formalize the regularity conditions on the propensity score model. These conditions are used for establishing the asymptotic normality of the estimator based on the estimated propensity score.

Assumption A.2. *Assume that the parametric form of the propensity score indexed by* $\boldsymbol{\gamma}$, $f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma})$, *is correctly specified and differentiable with respect to* $\boldsymbol{\gamma}$, *and let* $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big) = \frac{\partial}{\partial \boldsymbol{\gamma}} \log f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma})$ *be twice continuously differentiable score functions. Let* $\boldsymbol{\gamma}_0$ *denote the true values of the parameters, where* $\boldsymbol{\gamma}_0$ *is in an open subset of the Euclidean space. Denote*

$\mathcal{F}_t = \overline{H}^*_{t-M+1} = \{\overline{\boldsymbol{W}}_{t-M+1}, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}$, *as in the proof of Theorem 1. We assume that the following conditions hold:*

1. *(a)* $E_{\boldsymbol{\gamma}_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\Big] < \infty$,

   (b) There exists a positive definite matrix V_{ps} *such that*

$$\frac{1}{T}\sum_{t=1}^{T} E_{\boldsymbol{\gamma}_0}\Big(\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)^\top \mid \mathcal{F}_{t-1}\Big) \xrightarrow{p} V_{ps}$$

   (c) $\dfrac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\| > \epsilon\sqrt{T}\Big) \mid \mathcal{F}_{t-1}\Big] \xrightarrow{p} 0$, *for all* $\epsilon > 0$,

2. *For all* k, j, *if we denote the* k^{th} *element of the* $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$ *vector by* $\boldsymbol{\psi}_k\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$ *and* $P_{kjt} = \frac{\partial}{\partial \gamma_j}\boldsymbol{\psi}_k\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)\big|_{\boldsymbol{\gamma}_0}$, *then* $E_{\boldsymbol{\gamma}_0}\left[|P_{kjt}|\right] < \infty$ *and there exists* $0 < r_{kj} \leq 2$ *such that* $\sum_{t=1}^{T} \frac{1}{t^{r_{kj}}} E_{\boldsymbol{\gamma}_0}\left(|P_{kjt} - E_{\theta_0}(P_{kjt} \mid \mathcal{F}_{t-1})|^{r_{kj}} \mid \mathcal{F}_{t-1}\right) \xrightarrow{p} 0$

3. *There exists an integrable function* $\ddot{\psi}(w_t, \overline{h}_{t-1})$ *such that* $\ddot{\psi}(w_t, \overline{h}_{t-1})$ *dominates the second partial derivatives of* $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$ *in a neighborhood of* $\boldsymbol{\gamma}_0$ *for all* $(w_t, \overline{h}_{t-1})$.

Assumption A.3. *Suppose that* $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$ *are the score functions of a propensity score model that satisfies Assumption A.2 with true parameters* $\boldsymbol{\gamma}_0$, *and*

$$s(\overline{h}_{t-1}, w_t, y_t; \boldsymbol{\gamma}) = \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})}\right] N_B(y_t) - N_{Bt}(F_h^M).$$

Then, the following conditions hold.

1. *There exists* $u \in \mathbb{R}^K$ *such that*

$$\frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\big[s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}_0)\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big) \mid \mathcal{F}_{t-1}\big] \xrightarrow{p} u,$$

2. *If* $P_{jt} = \dfrac{\partial}{\partial \boldsymbol{\gamma}_j} s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma})\Big|_{\boldsymbol{\gamma}_0}$, *where* $\boldsymbol{\gamma}_j$ *is the* j^{th} *entry of* $\boldsymbol{\gamma}$, *then there exists* $r_j \in (0, 2]$ *such that*

$$\sum_{t=1}^{T} \frac{1}{t^{r_j}} E_{\boldsymbol{\gamma}_0}\left(|P_{jt} - E_{\boldsymbol{\gamma}_0}(P_{jt} \mid \mathcal{F}_{t-1})|^{r_j} \mid \mathcal{F}_{t-1}\right) \xrightarrow{p} 0.$$

Remark A.1. *Given the previous assumptions, Assumption A.3 is quite weak. We look at the two parts separately:*

1. *For the k^{th} entry, we can write:*

$$\frac{1}{T-M+1}\sum_{t=M}^{T}\left|E_{\theta_0}\left[s(\overline{H}_{t-1},W_t,Y_t;\boldsymbol{\gamma}_0)\boldsymbol{\psi}_k\left(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\right)\mid\mathcal{F}_{t-1}\right]\right|\leq$$

$$\leq\frac{1}{T-M+1}\sum_{t=M}^{T}\sqrt{E_{\theta_0}\left[s(\overline{H}_{t-1},W_t,Y_t;\boldsymbol{\gamma}_0)^2\mid\mathcal{F}_{t-1}\right]}\sqrt{E_{\theta_0}\left[\boldsymbol{\psi}_k\left(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\right)^2\mid\mathcal{F}_{t-1}\right]}$$

(Cauchy-Schwarz)

$$\leq\frac{1}{2(T-M+1)}\sum_{t=M}^{T}\left(E_{\theta_0}\left[s(\overline{H}_{t-1},W_t,Y_t;\boldsymbol{\gamma}_0)^2\mid\mathcal{F}_{t-1}\right]+E_{\theta_0}\left[\boldsymbol{\psi}_k\left(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\right)^2\mid\mathcal{F}_{t-1}\right]\right)$$

$(2ab\leq a^2+b^2)$

$$\xrightarrow{p}\frac{1}{2}\left(v+[V_{ps}]_{kk}\right).$$

The proof that the first part converges to v will be shown in Equation (A.5), and the second part is based on Assumption A.2, where $[V_{ps}]_{kk}$ denotes the k^{th} diagonal entry of V_{ps}. Since the expression is already bounded at the limit, the assumption that it converges is reasonable. Furthermore, we have that $|u_k|\leq\frac{1}{2}(v+[V_{ps}]_{kk})$, where u_k is the k^{th} entry of u.

2. *This assumption limits how much the derivative of $s(\overline{h}_{t-1},w_t,y_t;\boldsymbol{\gamma})$ can vary around its conditional expectation. As we will see in Lemma A.2, this derivative can be re-written as a sum that involves three terms: the number of outcome active locations, the inverse probability ratios, and the score functions. The first two of these terms are bounded, and Assumption A.2 already controls how variable the score functions can be. Thus, this assumption is also reasonable.*

B.2 Proofs: The propensity score as a balancing score

Proof of Proposition 1. Note that $f(W_t=w\mid e_t(w),\overline{H}_{t-1})=f(W_t=w\mid\overline{H}_{t-1})=e_t(w)$ since $e_t(w)$ is a function of $\overline{H}_{t-1}$. Therefore, it suffices to show that $f(W_t=w\mid e_t(w))=e_t(w)$:

$$f(W_t=w\mid e_t(w))=E[f(W_t=w\mid\overline{H}_{t-1})\mid e_t(w)]=E[e_t(w)\mid e_t(w)]=e_t(w). \tag{A.1}$$

□

Proof of Proposition 2.

$$f\left(W_t=w\mid\overline{\boldsymbol{W}}_{t-1},\overline{\mathcal{Y}}_T,\overline{\mathcal{X}}_T\right)$$

$$
\begin{aligned}
&= f\big(W_t = w \mid \overline{H}_{t-1}, \overline{\boldsymbol{W}}_{t-1}, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\big) && \text{(Since } \overline{H}_{t-1} \subset \{\overline{\boldsymbol{W}}_{t-1}, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}\text{)} \\
&= f\big(W_t = w \mid \overline{H}_{t-1}\big) && \text{(From Assumption 1)} \\
&= e_t(w) \\
&= f(W_t = w \mid e_t(w)) && \text{(From (A.1))}
\end{aligned}
$$

□

B.3 Proofs: Asymptotic normality based on the true propensity score

Proof of Theorem 1. Note that the collection of variables temporally precedent to treatment at time period t is the expanded history $\overline{H}^*_{t-1}$, defined in Assumption A.1. The expanded history $\overline{H}^*_{t-1}$ is a filtration generated by the collection of potential confounders $\overline{\mathcal{X}}_T$, the collection of potential outcomes $\overline{\mathcal{Y}}_T$, and the previous treatments, and satisfies $\overline{H}^*_{t-1} \subset \overline{H}^*_t$.

Let $err_t = \widehat{N}_{Bt}(F_h) - N_{Bt}(F_h)$ be the estimation error for time period t and lag M. We will decompose err_t in two components, one corresponding to the error due to the treatment assignment (A_{1t}), and the other corresponding to the error due to spatial smoothing (A_{2t}). Since the bandwidth parameter of the kernel depends on T, we write K_{b_T} instead of K_b. Specifically,

$$
\begin{aligned}
err_t &= \left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)}\right] \int_B \sum_{s \in S_{Y_t}} K_{b_T}(\omega, s)\mathrm{d}\omega - N_{Bt}(F_h^M) \\
&= \underbrace{\left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)}\right] N_B(Y_t) - N_{Bt}(F_h^M)}_{A_{1t}} + \\
&\qquad \underbrace{\left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)}\right] \left[\int_B \sum_{s \in S_{Y_t}} K_{b_T}(\omega, s)\mathrm{d}\omega - N_B(Y_t)\right]}_{A_{2t}}.
\end{aligned}
\tag{A.2}
$$

We show that

1. $\sqrt{T}\big(\frac{1}{T-M+1}\sum_{t=M}^{T} A_{1t}\big)$ is asymptotically normal, and
2. $\sqrt{T}\big(\frac{1}{T-M+1}\sum_{t=M}^{T} A_{2t}\big)$ converges to zero in probability.

Asymptotic normality of the first error.

We use the central limit theorem for martingale difference series (Theorem 4.16 of van der Vaart (2010)) to establish the asymptotic normality of $(T-M+1)^{-1}\sum_{t=M}^{T} A_{1t}$.

Claim. A_{1t} is a martingale difference series with respect to the filtration $\mathcal{F}_t = \overline{H}^*_{t-M+1}$.

To prove this, we show that $E(|A_{1t}|) < \infty$ and $E(A_{1t} \mid \mathcal{F}_{t-1}) = E(A_{1t} \mid \overline{H}^*_{t-M}) = 0$. For the first part, Assumptions 2 and Assumption A.1(a) imply that A_{1t} is bounded and hence $E[|A_{1t}|] < \infty$:

$$|A_{1t}| \;\le\; \left| \prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} N_B(Y_t) \right| + \left| N_{Bt}(F_h^M) \right| \;\le\; \delta_W^M \delta_Y + \delta_Y \tag{A.3}$$

For the second part, it suffices to show

$$E\left\{ \left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} \right] N_B(Y_t) \;\middle|\; \overline{H}^*_{t-M} \right\} \;=\; N_{Bt}(F_h^M),$$

where the expectation is taken with respect to the assignment of treatments $\boldsymbol{W}_{(t-M+1):t}$.

$$\begin{aligned}
& E\left\{ \left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} \right] N_B(Y_t) \;\middle|\; \overline{H}^*_{t-M} \right\} \\
&= \int \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j)} \right] N_B\Big(Y_t\big(\overline{\boldsymbol{W}}_{t-M}, \underbrace{w_{t-M+1}, \ldots, w_t}_{\boldsymbol{w}_{(t-M+1):t}} \big) \Big) \times \\
&\qquad f(w_{t-M+1} \mid \overline{H}^*_{t-M}) f(w_{t-M+2} \mid \overline{H}^*_{t-M}, W_{t-M+1}) \cdots \times \\
&\qquad f(w_t \mid \overline{H}^*_{t-M}, \boldsymbol{W}_{(t-M+1):(t-1)}) \, \mathrm{d}\boldsymbol{w}_{(t-M+1):t} \\
&= \int \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j)} \right] N_B\Big(Y_t\big(\overline{\boldsymbol{W}}_{t-M}, w_{t-M+1}, \ldots, w_t \big) \Big) \times \\
&\qquad f(w_{t-M+1} \mid \overline{H}^*_{t-M}) f(w_{t-M+2} \mid \overline{H}^*_{t-M+1}) \cdots f(w_t \mid \overline{H}^*_{t-1}) \, \mathrm{d}\boldsymbol{w}_{(t-M+1):t} \\
&\qquad\qquad \text{(because } \overline{H}^*_{t'+1} = \overline{H}^*_{t'} \cup \{W_{t'+1}\}) \\
&= \int N_B\Big(Y_t\big(\overline{\boldsymbol{W}}_{t-M}, w_{t-M+1}, \ldots, w_t \big) \Big) \left[\prod_{j=t-M+1}^{t} f_h(w_j) \right] \mathrm{d}\boldsymbol{w}_{(t-M+1):t} \qquad \text{(By Assumption 1)} \\
&= N_{Bt}(F_h^M).
\end{aligned} \tag{A.4}$$

This proves that A_{1t} is a martingale difference series with respect to filtration $\mathcal{F}_{t-1}$.

Claim. $(T-M+1)^{-1} \sum_{t=M}^{T} E\{A_{1t}^2 I(|A_{1t}| > \epsilon\sqrt{T-M+1}) \mid \mathcal{F}_{t-1}\} \xrightarrow{p} 0$ for every $\epsilon > 0$.

Let $\epsilon > 0$. Note that A_{1t} is bounded by $\delta_Y(\delta_W^M + 1)$ (see Equation (A.3)). Choose T_0 as

$$\begin{aligned}
T_0 \;&=\; \operatorname*{argmin}_{t \in \mathbb{N}^+} \{ \epsilon\sqrt{t-M+1} > \delta_Y(\delta_W^M + 1) \} \\
&=\; \operatorname*{argmin}_{t \in \mathbb{N}^+} \left\{ t > M - 1 + \left[\frac{\delta_Y(\delta_W^M + 1)}{\epsilon} \right]^2 \right\}
\end{aligned}$$

$$= \left\lceil M - 1 + \left\lceil \frac{\delta_Y(\delta_W^M + 1)}{\epsilon} \right\rceil^2 \right\rceil.$$

Then, for $T > T_0$, we have that $\epsilon\sqrt{T - M + 1} > \epsilon\sqrt{T_0 - M + 1} > \delta_Y(\delta_W^M + 1)$ which leads to $I(|A_{1t}| > \epsilon\sqrt{T + M + 1}) = 0$ and $E(A_{1t}^2 I(|A_{1t}| > \epsilon\sqrt{T - M + 1}) \mid \mathcal{F}_{t-1}) = 0$. This proves the claim.

We now combine the above claims to establish the asymptotic normality of the first error. Since A_{1t} has mean zero, $E(A_{1t}^2 \mid \mathcal{F}_{t-1}) = \text{Var}(A_{1t} \mid \mathcal{F}_{t-1})$, and since $N_{Bt}(F_h)$ is fixed,

$$\text{Var}(A_{1t} \mid \mathcal{F}_{t-1}) = \text{Var}\left(\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)} N_B(Y_t) \mid \overline{H}_{t-M}^* \right)$$

which yields

$$\frac{1}{T - M + 1} \sum_{t=M}^{T} E(A_{1t}^2 \mid \mathcal{F}_{t-1}) \xrightarrow{p} v, \tag{A.5}$$

from Assumption A.1(b). Combining these results, using that $\sqrt{T}/\sqrt{T - M + 1} \to 1$ and Theorem 4.16 of van der Vaart (2010), we have the desired result,

$$\sqrt{T}\left(\frac{1}{T - M + 1} \sum_{t=M}^{T} A_{1t} \right) \xrightarrow{d} N(0, v).$$

Convergence to zero of the second error.

The second error compares the integral of the kernel-smoothed outcome surface over the region of interest B with the actual number of points within the set B. We show that as T goes to infinity, and since the bandwidth of the kernel converges to 0, the error due to kernel smoothing also goes to zero. Specifically, we will show that

$$\sqrt{T}\left(\frac{1}{T - M + 1} \sum_{t=M}^{T} A_{2t} \right) \xrightarrow{p} 0.$$

Let $c_t = \prod_{j=t-M+1}^{t} f_h(W_j)/e_j(W_j)$, and write

$$\left| \frac{1}{T - M + 1} \sum_{t=M}^{T} A_{2t} \right| = \left| \frac{1}{T - M + 1} \sum_{t=M}^{T} c_t \left[\int_B \sum_{s \in S_{Y_t}} K_{b_T}(\omega; s) \mathrm{d}\omega - N_B(Y_t) \right] \right|.$$

Then,

$$\int_B \sum_{s \in S_{Y_t}} K_{b_T}(\omega; s) \mathrm{d}\omega - N_B(Y_t)$$

$$
\begin{aligned}
&= \sum_{s \in S_{Y_t} \cap B} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega + \sum_{s \in S_{Y_t} \cap B^c} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega - N_B(Y_t) \\
&= \sum_{s \in S_{Y_t} \cap B} \left[1 - \int_{B^c} K_{b_T}(\omega; s) \mathrm{d}\omega \right] + \sum_{s \in S_{Y_t} \cap B^c} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega - N_B(Y_t) \\
&= \sum_{s \in S_{Y_t} \cap B^c} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega - \sum_{s \in S_{Y_t} \cap B} \int_{B^c} K_{b_T}(\omega; s) \mathrm{d}\omega.
\end{aligned}
$$

This shows that the error from smoothing the outcome surface at time t comes from (1) the kernel weight from points outside of B that falls within B, and (2) the kernel weight from points inside B that falls outside B. Using this, we write:

$$
\left| \frac{1}{T-M+1} \sum_{t=M}^{T} A_{2t} \right| = \left| \frac{1}{T-M+1} \sum_{t=M}^{T} c_t \left[\sum_{s \in S_{Y_t} \cap B^c} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega - \sum_{s \in S_{Y_t} \cap B} \int_{B^c} K_{b_T}(\omega; s) \mathrm{d}\omega \right] \right|.
$$

Take $\epsilon > 0$, and $Q \in (1/2, Q^*)$ where Q^* is the one in Assumption A.1(c). Then, we will show that $P(T^Q \{ | \frac{1}{T-M+1} \sum_{t=M}^{T} A_{2t} | \} > \epsilon) \to 0$ as $T \to \infty$, which implies that the second error converges to zero faster than $\sqrt{T}$ (since $Q > 1/2$).

$$
\begin{aligned}
&P\left(T^Q \left\{ \left| \frac{1}{T-M+1} \sum_{t=M}^{T} A_{2t} \right| \right\} > \epsilon \right) \\
&= P\left(\left| \frac{1}{T-M+1} \sum_{t=M}^{T} c_t \left[\sum_{s \in S_{Y_t} \cap B^c} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega - \sum_{s \in S_{Y_t} \cap B} \int_{B^c} K_{b_T}(\omega; s) \mathrm{d}\omega \right] \right| > \frac{\epsilon}{T^Q} \right) \\
&\leq P\left(\frac{1}{T-M+1} \sum_{t=M}^{T} c_t \sum_{s \in S_{Y_t} \cap B^c} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega > \frac{\epsilon}{2T^Q} \right) + \\
&\qquad P\left(\frac{1}{T-M+1} \sum_{t=M}^{T} c_t \sum_{s \in S_{Y_t} \cap B} \int_{B^c} K_{b_T}(\omega; s) \mathrm{d}\omega > \frac{\epsilon}{2T^Q} \right),
\end{aligned}
$$

where the last equation holds because $|A - B| > \epsilon$ implies that at least one of $|A|, |B| > \epsilon/2$. Also, since all quantities are positive, we can drop the absolute value. Then, since $c_t \leq \delta_W^M$ from Assumption 2,

$$
\begin{aligned}
&P\left(T^Q \left\{ \left| \frac{1}{T-M+1} \sum_{t=M}^{T} A_{2t} \right| \right\} > \epsilon \right) \\
&\leq P\left(\frac{1}{T-M+1} \sum_{t=M}^{T} \sum_{s \in S_{Y_t} \cap B^c} \int_B K_{b_T}(\omega; s) \mathrm{d}\omega > \frac{\epsilon}{2T^Q \delta_W^M} \right) +
\end{aligned}
$$

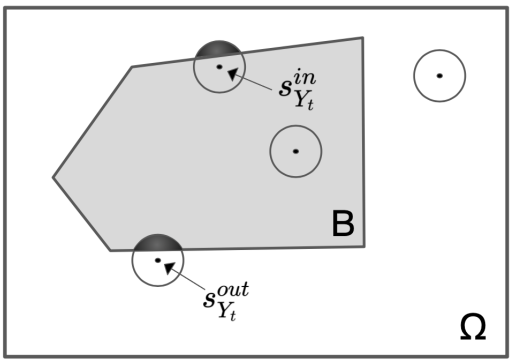


Figure A.1: Kernel-smoothed outcome surface, and points $s^{in}_{Y_t}, s^{out}_{Y_t}$ as the points closest to the boundary of B that lie within and outside B respectively. The amount of kernel weight falling within B from points outside of B is necessarily less or equal to the kernel weight from $s^{out}_{Y_t}$ (shaded), and similarly for $s^{in}_{Y_t}$.

$$P\left(\frac{1}{T-M+1}\sum_{t=M}^{T}\sum_{s\in S_{Y_t}\cap B}\int_{B^{\mathsf{c}}} K_{b_T}(\omega;s)\mathrm{d}\omega > \frac{\epsilon}{2T^Q\delta_W^M}\right).$$

Use $s^{out}_{Y_t}$ to denote the point in S_{Y_t} that lies outside B and is the closest to B: $s^{out}_{Y_t} = \{s \in S_{Y_t} \cap B^{\mathsf{c}} : \text{dist}(s,B) = \min_{s'\in S_{Y_t}\cap B^{\mathsf{c}}} \text{dist}(s',B)\}$. Similarly, $s^{in}_{Y_t}$ is the point in $S_{Y_t} \cap B$ that is closest to B^{c}. These points are shown graphically in Figure A.1. Because there are at most δ_Y outcome-active locations, from the definition of $s^{in}_{Y_t}, s^{out}_{Y_t}$, and because kernels are defined to be decreasing in distance, we have that

$$\begin{aligned}
&P\left(T^Q\left\{\left|\frac{1}{T-M+1}\sum_{t=M}^{T}A_{2t}\right|\right\} > \epsilon\right) \\
&\quad\le P\left(\frac{1}{T-M+1}\sum_{t=M}^{T}\int_{B} K_{b_T}(\omega;s^{out}_{Y_t})\mathrm{d}\omega > \frac{\epsilon}{2T^Q\delta_W^M\delta_Y}\right) \\
&\qquad\quad + P\left(\frac{1}{T-M+1}\sum_{t=M}^{T}\int_{B^{\mathsf{c}}} K_{b_T}(\omega;s^{in}_{Y_t})\mathrm{d}\omega > \frac{\epsilon}{2T^Q\delta_W^M\delta_Y}\right) \\
&\quad= \underbrace{P\left(\sum_{t=M}^{T}\int_{B} K_{b_T}(\omega;s^{out}_{Y_t})\mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q\delta_W^M\delta_Y}\right)}_{B_1} \\
&\qquad\quad + \underbrace{P\left(\sum_{t=M}^{T}\int_{B^{\mathsf{c}}} K_{b_T}(\omega;s^{in}_{Y_t})\mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q\delta_W^M\delta_Y}\right)}_{B_2}.
\end{aligned}$$

We show that B_1, B_2 converge to zero separately. Take B_1:

$$\begin{aligned} B_1 \;=\; & P\left(\sum_{t=M}^{T} \int_B K_{b_T}(\omega; s_{Y_t}^{out}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} \;\middle|\; \sum_{t=M}^{T} I(s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)) > T^{1-Q^*} \right) \\ & \qquad \times \; P\left(\sum_{t=M}^{T} I(s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)) > T^{1-Q^*} \right) \\ & + P\left(\sum_{t=M}^{T} \int_B K_{b_T}(\omega; s_{Y_t}^{out}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} \;\middle|\; \sum_{t=M}^{T} I(s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)) \leq T^{1-Q^*} \right) \\ & \qquad \times \; P\left(\sum_{t=M}^{T} I(s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)) \leq T^{1-Q^*} \right) \end{aligned}$$

From Assumption (c) we have that

$$\begin{aligned} & P\left(\sum_{t=M}^{T} I(s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)) > T^{1-Q^*} \right) \\ & \qquad \leq \; P\left(\sum_{t=M}^{T} I\Big(\exists s \in S_{Y_t} \cap \mathcal{N}_{\delta_B}(\partial B) \Big) > T^{1-Q^*} \right) \to 0, \end{aligned}$$

and $\lim_{T\to\infty} B_1$ is equal to

$$\lim_{T\to\infty} P\left(\sum_{t=M}^{T} \int_B K_{b_T}(\omega; s_{Y_t}^{out}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} \;\middle|\; \sum_{t=M}^{T} I(s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)) \leq T^{1-Q^*} \right).$$

Studying the latter quantity, we have that

$$\begin{aligned} & P\left(\sum_{t=M}^{T} \int_B K_{b_T}(\omega; s_{Y_t}^{out}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} \middle| \sum_{t=M}^{T} I(s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)) \leq T^{1-Q^*} \right) \\ & \leq P\left(\sum_{\substack{t=M \\ s_{Y_t}^{out} \notin \mathcal{N}_{\delta_B}(\partial B)}}^{T} \int_B K_{b_T}(\omega; s_{Y_t}^{out}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} - T^{1-Q^*} \right) \\ & \leq P\left(\sum_{\substack{t=M \\ s_{Y_t}^{out} \notin \mathcal{N}_{\delta_B}(\partial B)}}^{T} \int_{\omega:\|\omega\|>\delta_B} K_{b_T}(\omega; \mathbf{0}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} - T^{1-Q^*} \right) \\ & \leq P\left((T-M+1) \int_{\omega:\|\omega\|>\delta_B} K_{b_T}(\omega; \mathbf{0}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} - T^{1-Q^*} \right) \\ & = I\left((T-M+1) \int_{\omega:\|\omega\|>\delta_B} K_{b_T}(\omega; \mathbf{0}) \mathrm{d}\omega > \frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} - T^{1-Q^*} \right) \end{aligned} \tag{A.6}$$

where the first inequality follows from the fact that at most T^{1-Q^*} time periods had $s_{Y_t}^{out}$ within δ_B of set's B boundary, and $\int_B K_{b_T}(\omega; s_{Y_t}^{out}) \leq 1$ for those time periods. The second inequality follows from the fact that during the remaining time periods $s_{Y_t}^{out}$ was further than δ_B from B and $\int_B K_{b_T}(\omega; s_{Y_t}^{out}) \leq \int_{\omega:\|\omega - s_{Y_t}^{out}\|>\delta_B} K_{b_T}(\omega; s_{Y_t}^{out}) = \int_{\omega:\|\omega\|>\delta_B} K_{b_T}(\omega; \mathbf{0})$. The third inequality follows from not excluding the time periods with $s_{Y_t}^{out} \in \mathcal{N}_{\delta_B}(\partial B)$. Finally, the last equality holds because there is no uncertainty in the statement so the probability turns to an indicator.

Since $b_T \to 0$ as $T \to 0$, there exists $T_1 \in \mathbb{N}$ such that $b_T < \delta_B$ and $\int_{\omega:\|\omega\|>\delta_B} K_{b_T}(\omega; \mathbf{0}) \mathrm{d}\omega = 0$ for all $T \geq T_1$. Also, since $\frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} - T^{1-Q^*} \to \infty$, there exists $T_2 \in \mathbb{N}$ such that $\frac{\epsilon(T-M+1)}{2T^Q \delta_W^M \delta_Y} - T^{1-Q^*} > 1$ for all $T \geq T_2$. Then, for all $T \geq T_0 = \max\{T_1, T_2\}$ we have that the quantity in Equation (A.6) is equal to 0, showing that $\lim_{T\to\infty} B_1 = 0$. Similarly, we can show that $\lim_{T\to\infty} B_2 = 0$.

Combining all of these results we have that

$$P\left(T^Q \left\{ \left| \frac{1}{T-M+1} \sum_{t=M}^{T} A_{2t} \right| \right\} > \epsilon \right) \to 0,$$

as $T \to \infty$, establishing that the second error converges to zero faster than $1/\sqrt{T}$. □

Proof of Lemma 1. Define $\Psi_t = \left[\widehat{N}_{Bt}(F_h^M)\right]^2 - v_t^*$. Then, Ψ_t is a martingale difference series with respect to $\mathcal{F}_t = \overline{H}_{t-M+1}$ since the following two hold: (1) $E(|\Psi_t|) < \infty$ since Ψ_t is bounded, and (2) $\mathrm{E}(\Psi_t \mid \mathcal{F}_{t-1}) = \mathrm{E}\left\{ \left[\widehat{N}_{Bt}(F_h^M)\right]^2 \mid \overline{H}_{t-M}^* \right\} - v_t^* = 0$. Also, since $\widehat{N}_{Bt}(F_h^M)$ is bounded we have that $\sum_{t=M}^{\infty} t^{-2} \mathrm{E}\left(\Psi_t^2\right) < \infty$. From Theorem 1 in Csörgö (1968) we have that

$$\frac{1}{T-M+1} \sum_{t=M}^{T} \Psi_t = \frac{1}{T-M+1} \sum_{t=M}^{T} \left[\widehat{N}_{Bt}(F_h^M)\right]^2 - \frac{1}{T-M+1} \sum_{t=M}^{T} v_t^* \xrightarrow{p} 0.$$

□

We use the results above to acquire asymptotic normality of the estimator for the causal effect, $\widehat{\tau}_B(F_{h_1}^M, F_{h_2}^M)$:

Theorem A.1. *Suppose that Assumptions 1 and 2 as well as the regularity conditions (Assumption A.1) hold. If the bandwidth $b_T \to 0$, then we have that*

$$\sqrt{T}(\widehat{\tau}_B(F_{h_1}^M, F_{h_2}^M) - \tau_B(F_{h_1}^M, F_{h_2}^M)) \xrightarrow{d} N(0, \eta),$$

as $T \to \infty$, for some $\eta > 0$. Finally, an upper bound of the asymptotic variance η can be consistently estimated by

$$\frac{1}{T-M+1} \sum_{t=M}^{T} \left[\widehat{\tau}_{Bt}(F_{h_1}^M, F_{h_2}^M)\right]^2 \xrightarrow{p} \eta^* \geq \eta.$$

Proof. In order to prove the asymptotic normality of $\widehat{\tau}_B(F_{h_1}^M, F_{h_2}^M)$ we will rely on results in the proof of Theorem 1 above. Take

$$\widehat{\tau}_{Bt}(F_{h_1}^M, F_{h_2}^M) - \tau_{Bt}(F_{h_1}^M, F_{h_2}^M) =$$
$$\left\{ \prod_{j=t-M+1}^{t} \frac{f_{h_2}(W_j)}{e_j(W_j)} - \prod_{j=t-M+1}^{t} \frac{f_{h_1}(W_j)}{e_j(W_j)} \right\} \int_B \sum_{s \in S_{Y_t}} K_{b_T}(\omega, s) \mathrm{d}\omega - \tau_{Bt}(F_{h_1}^M, F_{h_2}^M) =$$
$$\underbrace{\left\{ \prod_{j=t-M+1}^{t} \frac{f_{h_2}(W_j)}{e_j(W_j)} - \prod_{j=t-M+1}^{t} \frac{f_{h_1}(W_j)}{e_j(W_j)} \right\} N_B(Y_t) - \tau_{Bt}(F_{h_1}^M, F_{h_2}^M) +}_{C_{1t}}$$
$$\underbrace{\left[\prod_{j=t-M+1}^{t} \frac{f_{h_2}(W_j)}{e_j(W_j)} \right] \left[\int_B \sum_{s \in S_{Y_t}} K_{b_T}(\omega, s) \mathrm{d}\omega - N_B(Y_t) \right] -}_{C_{2t}^2}$$
$$\underbrace{\left[\prod_{j=t-M+1}^{t} \frac{f_{h_1}(W_j)}{e_j(W_j)} \right] \left[\int_B \sum_{s \in S_{Y_t}} K_{b_T}(\omega, s) \mathrm{d}\omega - N_B(Y_t) \right]}_{C_{2t}^1}$$

Following steps identical to showing $\sqrt{T}\left[(T-M+1)^{-1}\sum_{t=M}^{T} A_{2t}\right] \xrightarrow{p} 0$ in the proof of Theorem 1, we can equivalently show that $\sqrt{T}\left[(T-M+1)^{-1}\sum_{t=M}^{T} C_{2t}^1\right] \xrightarrow{p} 0$ and $\sqrt{T}\left[(T-M+1)^{-1}\sum_{t=M}^{T} C_{2t}^2\right] \xrightarrow{p} 0$.

Therefore, all we need to show is that $\sqrt{T}\left[(T-M+1)^{-1}\sum_{t=M}^{T} C_{1t}\right] \xrightarrow{d} N(0,\eta)$. We will do so by showing again that C_{1t} is a martingale difference series with respect to the filtration $\mathcal{F}_{t-1}$:

1. Since $E(|A_{1t}|) < \infty$, from the triangular inequality we straightforwardly have that $E(|C_{1t}|) < \infty$.

2. Since $E(A_{1t}|\mathcal{F}_{t-1}) = 0$, we also have that $E(C_{1t}|\mathcal{F}_{t-1}) = 0$, from linearity of expectation.

Then, using the triangular inequality and Equation (A.3), we have that C_{1t} is bounded by $2\delta_Y(\delta_W^M + 1)$. Then, for $\epsilon > 0$, choosing $T_0 = \underset{t \in \mathbb{N}^+}{\operatorname{argmin}}\{\epsilon\sqrt{t-M+1} > 2\delta_Y(\delta_W^M + 1)\}$ satisfies that, for $T > T_0$, $E(C_{1t}^2 I(|C_{1t}| > \epsilon\sqrt{T-M+1})|\mathcal{F}_{t-1}) = 0$. Combining these results, we have that $\sqrt{T}\left[\widehat{\tau}_{Bt}(F_{h_1}^M, F_{h_2}^M) - \tau_{Bt}(F_{h_1}^M, F_{h_2}^M)\right] \to N(0,\eta)$.

To show $(T-M+1)^{-1}\sum_{t=M}^{T}\left\{\left[\widehat{\tau}_{Bt}(F_{h_1}^M, F_{h_2}^M)\right]^2 - E\left\{\left[\tau_{Bt}(F_{h_1}^M, F_{h_2}^M)\right]^2 | \overline{H}_{t-M}^*\right\}\right\} \xrightarrow{p} 0$, the proof follows exactly the same way as the proof of Lemma 1 and is omitted here.

□

B.4 Proofs: Asymptotic normality based on the estimated propensity score

We will prove the asymptotic normality of the proposed estimators when the propensity score is estimated using a correctly specified parametric model. We extend Theorem 4.16 of van der Vaart (2010) to

multivariate martingale difference series. To our knowledge, this result is new even though the related results exist in the continuous time setting (Küchler *et al.*, 1999; Crimaldi and Pratelli, 2005). Under some additional assumptions on the martingale series, we show that the solution to the empirical estimating equation is also asymptotically normal. This result will be crucial in establishing the asymptotic normality of the maximum likelihood estimator for the propensity score model parameters. Finally, we combine these results and apply them to our specific context.

Theorem A.2 (Central limit theorem for multivariate martingale difference series)**.** *Let* $X_t = (X_{1t}, X_{2t}, \ldots, X_{Kt})^\top$ *be a multivariate martingale difference series with respect to the filtration* $\mathcal{F}_t$ *in that* $E[X_t \mid \mathcal{F}_{t-1}] = 0$ *and* $E[\|X_t\|] < \infty$*, where* $\|X_t\| = \sqrt{X_t^\top X_t} = \sqrt{\sum_{k=1}^K X_{kt}^2}$*. Suppose that the following conditions hold.*

1. *There exists positive definite matrix* $V \in \mathbb{R}^{K\times K}$ *such that* $\frac{1}{T}\sum_{t=1}^T E\Big(X_t X_t^\top \mid \mathcal{F}_{t-1}\Big) \xrightarrow{p} V$*,*
2. $\frac{1}{T}\sum_{t=1}^T E\Big[\|X_t\|^2 I\Big(\|X_t\| > \epsilon\sqrt{T}\Big) \mid \mathcal{F}_{t-1}\Big] \xrightarrow{p} 0$*, for all* $\epsilon > 0$*.*

Then, we have,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T X_t \xrightarrow{d} N(0, V).$$

Proof. We will use the Cramer-Wold device. We show that for every $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_K) \in \mathbb{R}^K$, it holds that $\alpha^\top \frac{1}{\sqrt{T}}\sum_{t=1}^T X_t \xrightarrow{d} \alpha^\top N(0, V)$. If this is true, it is implied that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T X_t \xrightarrow{d} N(0, V).$$

Clearly, if α is the zero-vector, the result is trivial. So we focus on vectors α such that $\|\alpha\| \neq 0$.

Define $Y_t = \alpha^\top X_t$. First we show that Y_t is a martingale difference series with respect to $\mathcal{F}_t$:

$$E(|Y_t|) \leq \sum_{k=1}^K |\alpha_k| E(|X_{kt}|) \leq \sum_{k=1}^K |\alpha_k| E(\|X_t\|) < \infty, \quad \text{and}$$
$$E(Y_t \mid \mathcal{F}_{t-1}) = \sum_{k=1}^K \alpha_k E(X_{kt} \mid \mathcal{F}_{t-1}) = 0,$$

since X_t is a martingale difference series with respect to $\mathcal{F}_t$. So Y_t is also a martingale difference series with respect to $\mathcal{F}_t$. Next we will show that the conditions of Theorem 4.16 of van der Vaart (2010) hold

for Y_t. For Condition 2 we will use the fact that $\alpha\alpha^\top$ is a rank 1 symmetric matrix of dimension K with only non-zero eigenvalue equal to $\|\alpha\|^2$, and for that reason $X_t^\top \alpha\alpha^\top X_t = \|\alpha\|^2 X_t^\top X_t = \|\alpha\|^2\|X_t\|^2$.

Condition 1
$$\frac{1}{T}\sum_{t=1}^{T} E\big(Y_t^2 \mid \mathcal{F}_{t-1}\big) = \frac{1}{T}\sum_{t=1}^{T} E\Big(\alpha^\top X_t X_t^\top \alpha \mid \mathcal{F}_{t-1}\Big)$$
$$= \alpha^\top \frac{1}{T}\sum_{t=1}^{T} E\Big(X_t X_t^\top \mid \mathcal{F}_{t-1}\Big)\alpha \xrightarrow{p} \alpha^\top V \alpha$$
(from the first assumption of Theorem A.2)

Condition 2
$$E\Big[Y_t^2 I\big(|Y_t| > \epsilon\sqrt{n}\big) \mid \mathcal{F}_{t-1}\Big] =$$
$$= E\Big[X_t^\top \alpha\alpha^\top X_t \, I\big(X_t^\top \alpha\alpha^\top X_t > \epsilon^2 n\big) \mid \mathcal{F}_{t-1}\Big]$$
$$= E\Big[\|\alpha\|^2\|X_t\|^2 \, I\big(\|\alpha\|^2\|X_t\|^2 > \epsilon^2 n\big) \mid \mathcal{F}_{t-1}\Big]$$
$$= \|\alpha\|^2 E\left[\|X_t\|^2 \, I\left(\|X_t\| > \frac{\epsilon}{\|\alpha\|}\sqrt{n}\right) \mid \mathcal{F}_{t-1}\right] \xrightarrow{p} 0$$
(From the second condition of the Theorem for $\epsilon' = \epsilon/\|\alpha\|$)

Using Theorem 4.16 from van der Vaart (2010):

$$\alpha^\top \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{d} N(0, \alpha^\top V \alpha) \overset{d}{=} \alpha^\top N(0, V).$$

□

Now that a multivariate central limit theorem (CLT) for martingale difference series is established, we prove the next result which will be crucial in obtaining the asymptotic normality of estimators for the propensity score parameters. To our knowledge, this result is also new in martingale theory, but a related result in the iid setting is given as Theorem 5.21 of Van der Vaart (1998).

Theorem A.3 (Asymptotic normality of the solution to the estimating equation)**.** *Let* $\theta \to s(x,\theta) = (s_1(x,\theta), s_2(\theta), \ldots, s_K(x,\theta))^\top \in \mathbb{R}^K$ *be twice continuously differentiable with respect to* $\theta = (\theta_1, \theta_2, \ldots, \theta_K)^\top \in \Theta$*, open subset of* $\mathbb{R}^K$*. Suppose that the following conditions hold.*

1. $s(X_t, \theta_0)$ *satisfies the conditions of Theorem A.2 under* θ_0*, in that there exists filtration* $\mathcal{F}_t$ *such that*

   (a) $E_{\theta_0}[s(X_t,\theta_0) \mid \mathcal{F}_{t-1}] = 0$ *and* $E_{\theta_0}[\|s(X_t,\theta_0)\|] < \infty$ *(and therefore it is a martingale difference series),*

   (b) $\exists\, V \in \mathbb{R}^{K\times K}$ *positive definite such that* $\frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\Big(s(X_t,\theta_0)s(X_t,\theta_0)^\top \mid \mathcal{F}_{t-1}\Big) \xrightarrow{p} V$*, and*

(c) $\frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left[\|s(X_t,\theta_0)\|^2 I\left(\|s(X_t,\theta_0)\| > \epsilon\sqrt{T}\right) \mid \mathcal{F}_{t-1}\right] \xrightarrow{p} 0$, *for all* $\epsilon > 0$,

2. $\frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left(\frac{\partial}{\partial\theta^T} s(X_t,\theta)\Big|_{\theta_0} \mid \mathcal{F}_{t-1}\right) \xrightarrow{p} V_d$, *for* $V_d \in \mathbb{R}^{K\times K}$ *invertible,*

3. *for all* k, j, *if we denote* $P_{kjt} = \frac{\partial}{\partial\theta_j} s_k(X_t,\theta)\Big|_{\theta_0}$, *we have that* $E_{\theta_0}[|P_{kjt}|] < \infty$, *and there exists* $0 < r_{kj} \leq 2$ *such that* $\sum_{t=1}^{T} \frac{1}{t^{r_{kj}}} E_{\theta_0}\left(|P_{kjt} - E_{\theta_0}\left[P_{kjt} \mid \mathcal{F}_{t-1}\right]|^{r_{kj}} \mid \mathcal{F}_{t-1}\right) \xrightarrow{p} 0$,

4. *there exists an integrable function* $\ddot{\psi}(x)$ *such that* $\ddot{\psi}(x)$ *dominates the second partial derivatives of* $s_k(x,\theta)$ *in a neighborhood of* θ_0 *for all* x, *and* $k = 1, 2, \ldots, K$.

If $\Psi_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} s(X_t,\theta)$, *and the solution to* $\Psi_T(\theta) = 0$, $\widehat{\theta}_T$, *is consistent for* θ_0, *then*

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(0, V_d^{-1} V (V_d^{-1})^\top\right).$$

Proof. We extend the proof of Theorem 5.41 of Van der Vaart (1998) from the iid to the time series setting. Since the conditions of Theorem A.2 are satisfied under θ_0, we have that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s(X_t,\theta_0) = \sqrt{T}\Psi_T(\theta_0) \to N(0, V).$$

We will use the Taylor expansion for the vector valued $\Psi_T(\widehat{\theta})$ around $\theta_0 = (\theta_{01}, \theta_{02}, \ldots, \theta_{0K})^T$. To do so, we define the matrix $\dot{\Psi}_T(\theta) \in \mathbb{R}^{K\times K}$ and array $\ddot{\Psi}_T(\theta) \in \mathbb{R}^{K\times K\times K}$ of first and second derivatives as

$$\left[\dot{\Psi}_T(\theta)\right]_{kj} = \frac{\partial}{\partial\theta_j}\Psi_{kT}(\theta)\Big|_\theta = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta_j} s_k(X_t,\theta)\Big|_\theta \qquad \text{and}$$

$$\left[\ddot{\Psi}_T(\theta)\right]_{kji} = \frac{\partial^2}{\partial\theta_j\partial\theta_i}\Psi_{kT}(\theta)\Big|_\theta = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2}{\partial\theta_j\partial\theta_i} s_k(X_t,\theta)\Big|_\theta,$$

for $i, j, k = 1, 2, \ldots, q$, where Ψ_{kT} is the k^{th} element of the Ψ_T vector. Then, we can write the Taylor expansion as

$$\Psi_T(\widehat{\theta}) = \Psi_T(\theta_0) + \dot{\Psi}_T(\theta_0)(\widehat{\theta}_T - \theta_0) + \ddot{\Psi}_T(\theta^*)(\widehat{\theta}_T - \theta_0, \widehat{\theta}_T - \theta_0), \tag{A.7}$$

where θ^* is between $\widehat{\theta}_T$ and θ_0, and $\ddot{\Psi}_T(\theta^*)(\widehat{\theta}_T - \theta_0, \widehat{\theta}_T - \theta_0)$ is a vector of length K with k^{th} entry

$$\sum_{j,i=1}^{K} \left[\ddot{\Psi}_T(\theta^*)\right]_{kji} (\widehat{\theta}_{Tj} - \theta_{0j})(\widehat{\theta}_{Ti} - \theta_{0i}),$$

and $\widehat{\theta}_{Ti}$ is the i^{th} entry of $\widehat{\theta}$. Therefore, we can write $\ddot{\Psi}_T(\theta^*)(\widehat{\theta}_T - \theta_0, \widehat{\theta}_T - \theta_0)$ as $A_T(\widehat{\theta}_T - \theta_0)$ where A_T is the $K \times K$ matrix which is the result of multiplying the tensor $\ddot{\Psi}_T(\theta^*)$ with the vector $\widehat{\theta}_T - \theta_0$ along the second mode, and it has (k, i) entry equal to

$$[A_T]_{ki} = \sum_{j=1}^{K} \left[\ddot{\Psi}_T(\theta^*)\right]_{kji} (\widehat{\theta}_{Tj} - \theta_{0j})$$

For notational simplicity, we do not include θ^* and $\widehat{\theta}_T - \theta_0$ in the notation of A_T. Since $\widehat{\theta}$ is the solution to $\Psi_T(\theta) = 0$, and based on the above, we can re-write Equation (A.7) as

$$\begin{aligned}
& 0 = \Psi_T(\theta_0) + \dot{\Psi}_T(\theta_0)(\widehat{\theta}_T - \theta_0) + A_T(\widehat{\theta}_T - \theta_0) \\
\implies & -\sqrt{T}\Psi_T(\theta_0) = \sqrt{T}\left[\dot{\Psi}_T(\theta_0) + A_T\right](\widehat{\theta}_T - \theta_0) \\
\implies & -\sqrt{T}\Psi_T(\theta_0) = \sqrt{T}\Big[\underbrace{\dot{\Psi}_T(\theta_0) - \frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left(\frac{\partial}{\partial \theta^T} s(X_t, \theta)\Big|_{\theta_0} \mid \mathcal{F}_{t-1}\right)}_{(*)} + \\
& \qquad + \frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left(\frac{\partial}{\partial \theta^T} s(X_t, \theta)\Big|_{\theta_0} \mid \mathcal{F}_{t-1}\right) + \underbrace{A_T}_{(**)}\Big](\widehat{\theta}_T - \theta_0)
\end{aligned}$$

We will show that the under-braced terms ($K \times K$ matrices) are $o_P(1)$. For the first term $(*)$, note that it involves the average over t of the P_{kjt} terms defined in Condition 3 of the theorem. Clearly, we have that $E_{\theta_0}[P_{kjt} - E_{\theta_0}[P_{kjt} \mid \mathcal{F}_{t-1}] \mid \mathcal{F}_{t-1}] = 0$, and we also have that

$$\begin{aligned}
& E_{\theta_0}[|P_{kjt} - E_{\theta_0}[P_{kjt} \mid \mathcal{F}_{t-1}]|] \\
&= E_{\theta_0}\left\{\left|\frac{\partial}{\partial \theta_j} s_k(X_t, \theta)\Big|_{\theta_0} - E_{\theta_0}\left[\frac{\partial}{\partial \theta_j} s_k(X_t, \theta)\Big|_{\theta_0} \mid \mathcal{F}_{t-1}\right]\right|\right\} \\
&\leq E_{\theta_0}\left\{\left|\frac{\partial}{\partial \theta_j} s_k(X_t, \theta)\Big|_{\theta_0}\right|\right\} + E_{\theta_0}\left\{\left|E_{\theta_0}\left[\frac{\partial}{\partial \theta_j} s_k(X_t, \theta)\Big|_{\theta_0} \mid \mathcal{F}_{t-1}\right]\right|\right\} && \text{(Triangle inequality)} \\
&\leq E_{\theta_0}\left\{\left|\frac{\partial}{\partial \theta_j} s_k(X_t, \theta)\Big|_{\theta_0}\right|\right\} + E_{\theta_0}\left\{E_{\theta_0}\left[\left|\frac{\partial}{\partial \theta_j} s_k(X_t, \theta)\Big|_{\theta_0}\right| \mid \mathcal{F}_{t-1}\right]\right\} && \text{(Jensen's inequality)} \\
&= 2E_{\theta_0}\left\{\left|\frac{\partial}{\partial \theta_j} s_k(X_t, \theta)\Big|_{\theta_0}\right|\right\} < \infty && \text{(Condition 3)}
\end{aligned}$$

So the assumptions of Chow (1965) (Theorem 5), which is also stated in Stout (1974) (Theorem 3.3.1), are satisfied and we have that

$$\dot{\Psi}_T(\theta_0) - \frac{1}{T}\sum_{t=1}^{T} E_{\theta_0}\left(\frac{\partial}{\partial\theta^\top} s(X_t,\theta)\Big|_{\theta_0} \mid \mathcal{F}_{t-1}\right)$$
$$= \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial}{\partial\theta^\top} X_{t,i}(\theta)\Big|_{\theta_0} - E_{\theta_0}\left(\frac{\partial}{\partial\theta^\top} s(X_t,\theta)\Big|_{\theta_0} \mid \mathcal{F}_{t-1}\right)\right] \xrightarrow{p} 0.$$

Then for A_T we notice that

$$\left|\left[\ddot{\Psi}_T(\theta^*)\right]_{kji}\right| \leq \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_i} s_k(X_t,\theta)\Big|_{\theta^*}\right| \leq \frac{1}{T}\sum_{t=1}^{T}\ddot{\psi}(X_t),$$

where the last inequality holds for large T because $\widehat{\theta}_T$ is consistent for θ_0 and the parameter space Θ is an open subset of $\mathbb{R}^n$ which imply that $\widehat{\theta}_T$ is within the neighborhood of θ_0 that satisfies Condition 4 of the theorem with probability that tends to 1, and therefore so will θ^*. Since $\ddot{\psi}(x)$ is integrable, the right hand side above is bounded with probability 1 from the law of large numbers. Then, using Cauchy-Schwarz on $[A_T]_{ki}$ and since $\widehat{\theta}_T$ is consistent for θ_0, we have that $[A_T]_{ki} \xrightarrow{p} 0$ for all k, i. Therefore, using Condition 2 of the theorem

$$-\sqrt{T}\Psi_T(\theta_0) = \sqrt{T}[V_d + o_P(1)](\widehat{\theta}_T - \theta_0)$$

which, since V_d is invertible, implies asymptotically that

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(0, V_d^{-1} V (V_d^{-1})^\top\right).$$

□

Theorem A.3 will be the basis for showing asymptotic normality of our estimators when the propensity score is estimated using a correctly specified parametric propensity score.

Lemma A.1 (Properties of the time series score functions.)**.** *If Assumption 1 holds, and $\boldsymbol{\psi}(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma})$ are score functions that satisfy Assumption A.2, then*

1. $E_{\boldsymbol{\gamma}_0}[\boldsymbol{\psi}(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0) \mid \mathcal{F}_{t-1}] = 0$, $E_{\boldsymbol{\gamma}_0}\left[\|\boldsymbol{\psi}(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0)\|\right] < \infty$, *and*
2. $E_{\boldsymbol{\gamma}}\left(-\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\boldsymbol{\psi}(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}) \mid \mathcal{F}_{t-1}\right) = E_{\boldsymbol{\gamma}}\left(\boldsymbol{\psi}(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma})\boldsymbol{\psi}(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma})^\top \mid \mathcal{F}_{t-1}\right)$ *which in turn implies that* $\frac{1}{T}\sum_{t=1}^{T} E_{\boldsymbol{\gamma}_0}\left(-\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\boldsymbol{\psi}(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma})\Big|_{\boldsymbol{\gamma}_0} \mid \mathcal{F}_{t-1}\right) \xrightarrow{p} V_{ps}$, *for V_{ps} positive definite, symmetric and therefore invertible.*

Proof. First, we show that $E_{\gamma_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|\Big] < \infty$. From Jensen's inequality we have that

$$E^2_{\gamma_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|\Big] \leq E_{\gamma_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\Big] < \infty,$$

so this part is shown. The remaining of the proof follows steps similar to the ones in the iid setting while conditioning on the corresponding filtration. Since $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$ are the score functions, we have that

$$\begin{aligned}
&\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big) f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma}) \\
=&\left[\frac{\partial}{\partial \boldsymbol{\gamma}} \log f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma})\right] f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma}) \\
=&\frac{\partial}{\partial \boldsymbol{\gamma}} f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma}).
\end{aligned} \tag{A.8}$$

Then,

$$\begin{aligned}
&E_{\boldsymbol{\gamma}}[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \mathcal{F}_{t-1}] = E_{\boldsymbol{\gamma}}\Big[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \overline{H}^*_{t-M}\Big] \\
&= E_{\boldsymbol{\gamma}}\Big\{E_{\boldsymbol{\gamma}}\Big[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \overline{H}^*_{t-1}\Big] \mid \overline{H}^*_{t-M}\Big\} && \text{(Since } \overline{H}^*_{t-1} \supseteq \overline{H}^*_{t-M}\text{)} \\
&= E_{\boldsymbol{\gamma}}\Big\{\Big[\int \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) f(W_t = w_t \mid \overline{H}^*_{t-1}) \mathrm{d}w_t\Big] \mid \overline{H}^*_{t-M}\Big\} \\
&= E_{\boldsymbol{\gamma}}\Big\{\Big[\int \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) f(W_t = w_t \mid \overline{H}_{t-1}) \mathrm{d}w_t\Big] \mid \overline{H}^*_{t-M}\Big\} && \text{(Assumption 1)} \\
&= E_{\boldsymbol{\gamma}}\Big\{\Big[\int \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t\Big] \mid \overline{H}^*_{t-M}\Big\} \\
&= E_{\boldsymbol{\gamma}}\Big\{\Big[\int \frac{\partial}{\partial \boldsymbol{\gamma}^\top} f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t\Big] \mid \overline{H}^*_{t-M}\Big\} && \text{(Equation A.8)} \\
&= E_{\boldsymbol{\gamma}}\Big\{\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\Big[\int f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t\Big] \mid \overline{H}^*_{t-M}\Big\} = 0,
\end{aligned}$$

where reversing the integral and derivative is valid using the Leibniz's rule which requires mild regularity conditions (continuity of the propensity score and its partial derivatives with respect to $\boldsymbol{\gamma}$). The last equation is equal to zero since the integral of the propensity score over its support is equal to 1, and the derivative to 1 is equal to 0.

To show the second part, we differentiate $E_{\boldsymbol{\gamma}}[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \mathcal{F}_{t-1}] = 0$ with respect to $\boldsymbol{\gamma}$:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \boldsymbol{\gamma}^\top} E_{\boldsymbol{\gamma}}\Big[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \mathcal{F}_{t-1}\Big] \\
&= \frac{\partial}{\partial \boldsymbol{\gamma}^\top} E_{\boldsymbol{\gamma}}\Big\{E_{\boldsymbol{\gamma}}\Big[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \overline{H}^*_{t-1}\Big] \mid \overline{H}^*_{t-M}\Big\} \\
&= \frac{\partial}{\partial \boldsymbol{\gamma}^\top} E_{\boldsymbol{\gamma}}\Big\{\int \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t \mid \overline{H}^*_{t-M}\Big\} && \text{(Assumption 1)}
\end{aligned}$$

$$
\begin{aligned}
&= E_{\boldsymbol{\gamma}}\Big\{ \int \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \Big[\boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \Big] \mathrm{d}w_t \mid \overline{H}^*_{t-M} \Big\} && \text{(Leibniz's rule)} \\
&= E_{\boldsymbol{\gamma}}\Big\{ \int \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \Big[f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \Big] \mathrm{d}w_t \mid \overline{H}^*_{t-M} \Big\} \\
&\qquad + E_{\boldsymbol{\gamma}}\Big\{ \int \Big[\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \Big] f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t \mid \overline{H}^*_{t-M} \Big\} \\
&= E_{\boldsymbol{\gamma}}\Big\{ \int \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)^\top f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t \mid \overline{H}^*_{t-M} \Big\} \\
&\qquad \text{(Equation (A.8))} \\
&\qquad + E_{\boldsymbol{\gamma}}\Big\{ \int \Big[\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \Big] f(W_t = w_t \mid \overline{H}_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t \mid \overline{H}^*_{t-M} \Big\} \\
&= E_{\boldsymbol{\gamma}}\Big\{ \int \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)^\top f(W_t = w_t \mid \overline{H}^*_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t \mid \overline{H}^*_{t-M} \Big\} \\
&\qquad + E_{\boldsymbol{\gamma}}\Big\{ \int \Big[\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \Big] f(W_t = w_t \mid \overline{H}^*_{t-1}; \boldsymbol{\gamma}) \mathrm{d}w_t \mid \overline{H}^*_{t-M} \Big\} \\
&\qquad \text{(Assumption 1)} \\
&= E_{\boldsymbol{\gamma}}\Big\{ E_{\boldsymbol{\gamma}}\Big[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)^\top \mid \overline{H}^*_{t-1} \Big] \mid \overline{H}^*_{t-M} \Big\} \\
&\qquad + E_{\boldsymbol{\gamma}}\Big\{ E_{\boldsymbol{\gamma}}\Big[\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \overline{H}^*_{t-1} \Big] \mid \overline{H}^*_{t-M} \Big\} \\
&= E_{\boldsymbol{\gamma}}\Big[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)^\top \mid \overline{H}^*_{t-M} \Big] + E_{\boldsymbol{\gamma}}\Big[\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \overline{H}^*_{t-M} \Big] \\
\implies & E_{\boldsymbol{\gamma}}\Big[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)^\top \mid \overline{H}^*_{t-M} \Big] = E_{\boldsymbol{\gamma}}\Big[-\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big) \mid \overline{H}^*_{t-M} \Big].
\end{aligned}
$$

From Condition 1b of Assumption A.2 we have the last result. □

Corollary A.1 (Asymptotic normality of spatio-temporal propensity score parameters)**.** *Consider a propensity score model that satisfies Assumption A.2 and therefore the results of Lemma A.1 hold. Theorem A.3 implies that the MLE of the propensity score parameters are asymptotically normal centered at the true value and with asymptotic variance V_{ps}^{-1}, as in the iid setting.*

Before we state our main theorem we establish a useful Lemma.

Lemma A.2. *Aassume that Assumption 1 holds. Let $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$ be the score functions of a propensity score model that satisfies Assumption A.2 as in Lemma A.1 and $\mathcal{F}_{t-1}$ be as above. For*

$$
s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}) = \left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j; \boldsymbol{\gamma})} \right] N_B(Y_t) - N_{Bt}(F_h^M),
$$

it holds that

1. $E_{\boldsymbol{\gamma}_0}\left[s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}_0) \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big) \mid \mathcal{F}_{t-1} \right] = -E_{\boldsymbol{\gamma}_0}\left[\frac{\partial}{\partial \boldsymbol{\gamma}} s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}) \Big|_{\boldsymbol{\gamma}_0} \mid \mathcal{F}_{t-1} \right],$

2. $\frac{\partial}{\partial \boldsymbol{\gamma}_l} s(\overline{h}_{t-1}, w_t, y_t; \boldsymbol{\gamma}) = -N_B(y_t) \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \sum_{j=t-M+1}^{t} \boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big)$, *where we use* $\boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big)$ *to denote the* l^{th} *element of the* $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$ *vector, and*

3. *similarly* $\frac{\partial}{\partial \boldsymbol{\gamma}_m} \frac{\partial}{\partial \boldsymbol{\gamma}_l} s(\overline{h}_{t-1}, w_t, y_t; \boldsymbol{\gamma})$ *is equal to*

$$
\begin{aligned}
- N_B(y_t) \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \Bigg\{ \Bigg[\sum_{j=t-M+1}^{t} \frac{\partial}{\partial \boldsymbol{\gamma}_m} \boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \Bigg] - \\
- \Bigg[\sum_{j=t-M+1}^{t} \boldsymbol{\psi}_m\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \Bigg] \Bigg[\sum_{j=t-M+1}^{t} \boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \Bigg] \Bigg\}
\end{aligned}
$$

Note: $s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}_0)$ *is the term* A_{1t} *in the proof of Theorem 1.*

Proof.

1. We will show it for $M = 1$, and the proof for $M > 1$ is similar. For $M = 1$, $\mathcal{F}_{t-1} = \overline{H}^*_{t-1} = \{\overline{\boldsymbol{W}}_{t-1}, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}$, we consider

$$
\begin{aligned}
& E_{\boldsymbol{\gamma}_0} \left[s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}_0) \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big) \mid \mathcal{F}_{t-1} \right] \\
&= \int s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}_0) \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big) f(W_t = w_t \mid \mathcal{F}_{t-1}; \boldsymbol{\gamma}_0) \mathrm{d}w_t \\
&= \int s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}_0) \boldsymbol{\psi}\big(w_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big) f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma}_0) \mathrm{d}w_t \quad \text{(Assumption 1)} \\
&= \int s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}_0) \frac{\partial}{\partial \boldsymbol{\gamma}} f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma}) \Big|_{\boldsymbol{\gamma}_0} \mathrm{d}w_t \quad \text{(Equation (A.8))} \\
&= \int \frac{\partial}{\partial \boldsymbol{\gamma}} \left[s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}_0) f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma}_0) \right] \mathrm{d}w_t - \\
&\quad - \int \frac{\partial}{\partial \boldsymbol{\gamma}} s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}) \Big|_{\boldsymbol{\gamma}_0} f(W_t = w_t \mid \overline{H}_{t-1} = \overline{h}_{t-1}; \boldsymbol{\gamma}_0) \mathrm{d}w_t \\
&= \int \frac{\partial}{\partial \boldsymbol{\gamma}} \left[s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}_0) f(W_t = w_t \mid \mathcal{F}_{t-1}; \boldsymbol{\gamma}_0) \right] \mathrm{d}w_t - \\
&\quad - \int \frac{\partial}{\partial \boldsymbol{\gamma}} s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}) \Big|_{\boldsymbol{\gamma}_0} f(W_t = w_t \mid \mathcal{F}_{t-1}; \boldsymbol{\gamma}_0) \mathrm{d}w_t \quad \text{(Assumption 1)} \\
&= \frac{\partial}{\partial \boldsymbol{\gamma}} E_{\boldsymbol{\gamma}} \left[s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}) \mid \mathcal{F}_{t-1} \right] \Big|_{\boldsymbol{\gamma}_0} - E_{\boldsymbol{\gamma}_0} \left[\frac{\partial}{\partial \boldsymbol{\gamma}} s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}) \Big|_{\boldsymbol{\gamma}_0} \mid \mathcal{F}_{t-1} \right] \\
&= -E_{\boldsymbol{\gamma}_0} \left[\frac{\partial}{\partial \boldsymbol{\gamma}} s(\overline{H}_{t-1}, w_t, Y_t; \boldsymbol{\gamma}) \Big|_{\boldsymbol{\gamma}_0} \mid \mathcal{F}_{t-1} \right]
\end{aligned}
$$

where the last equation holds from Equation (A.4). This shows that the expectation is 0, so the derivative is also 0.

Note that at the second line of the proof, we would also need the distribution of Y_t given the filtration $\mathcal{F}_{t-1}$ and the treatment at time period t, $W_t = w_t$. However, given both $\mathcal{F}_{t-1}$ and W_t, the variable Y_t is no longer random, and it is equal to its potential value $Y_t(\overline{\boldsymbol{W}}_{t-1}, w_t)$, where $\overline{\boldsymbol{W}}_{t-1}$ is specified in $\mathcal{F}_{t-1}$. We refrain from explicitly including this in the proof for simplicity.

2.

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\gamma}_l} s(\overline{h}_{t-1}, w_t, y_t; \boldsymbol{\gamma}) \\
&= \frac{\partial}{\partial \boldsymbol{\gamma}_l} \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} N_B(y_t) \right] \\
&= N_B(y_t) \left[\prod_{j=t-M+1}^{t} f_h(w_j) \right] \left[\frac{\partial}{\partial \gamma_l} \frac{1}{\prod_{j=t-M+1}^{t} e_j(w_j; \boldsymbol{\gamma})} \right] \\
&= -N_B(y_t) \left[\prod_{j=t-M+1}^{t} f_h(w_j) \right] \frac{\frac{\partial}{\partial \gamma_l} \prod_{j=t-M+1}^{t} e_j(w_j; \boldsymbol{\gamma})}{\left[\prod_{j=t-M+1}^{t} e_j(w_j; \boldsymbol{\gamma}) \right]^2} \\
&= -N_B(y_t) \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \frac{\frac{\partial}{\partial \gamma_l} \prod_{j=t-M+1}^{t} e_j(w_j; \boldsymbol{\gamma})}{\prod_{j=t-M+1}^{t} e_j(w_j; \boldsymbol{\gamma})} \\
&= -N_B(y_t) \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \sum_{j=t-M+1}^{t} \frac{\frac{\partial}{\partial \gamma_l} e_j(w_j; \boldsymbol{\gamma})}{e_j(w_j; \boldsymbol{\gamma})} \\
&= -N_B(y_t) \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \sum_{j=t-M+1}^{t} \boldsymbol{\psi}_{l-1}\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big). \qquad \text{(Equation (A.8))}
\end{aligned}
$$

3. Following a similar procedure we have that

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\gamma}_m} \frac{\partial}{\partial \boldsymbol{\gamma}_l} s(\overline{h}_{t-1}, w_t, y_t; \boldsymbol{\gamma}) \\
&= -N_B(y_t) \left\{ \left[\frac{\partial}{\partial \boldsymbol{\gamma}_m} \prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \sum_{j=t-M+1}^{t} \boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \right. \\
&\qquad \left. + \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \left[\frac{\partial}{\partial \boldsymbol{\gamma}_m} \sum_{j=t-M+1}^{t} \boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \right] \right\} \\
&= -N_B(y_t) \left\{ - \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \left[\sum_{j=t-M+1}^{t} \boldsymbol{\psi}_m\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \right] \left[\sum_{j=t-M+1}^{t} \boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \right] \right. \\
&\qquad \left. + \left[\prod_{j=t-M+1}^{t} \frac{f_h(w_j)}{e_j(w_j; \boldsymbol{\gamma})} \right] \left[\sum_{j=t-M+1}^{t} \frac{\partial}{\partial \boldsymbol{\gamma}_m} \boldsymbol{\psi}_l\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big) \right] \right\}
\end{aligned}
$$

$$= -N_B(y_t)\left[\prod_{j=t-M+1}^{t}\frac{f_h(w_j)}{e_j(w_j;\boldsymbol{\gamma})}\right]\left\{\left[\sum_{j=t-M+1}^{t}\frac{\partial}{\partial\boldsymbol{\gamma}_m}\boldsymbol{\psi}_l\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\right]\right.$$
$$\left.-\left[\sum_{j=t-M+1}^{t}\boldsymbol{\psi}_m\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\right]\left[\sum_{j=t-M+1}^{t}\boldsymbol{\psi}_l\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\right]\right\}$$

□

Corollary A.2. *Part 1 of Lemma A.1 holds for any function* $s(\overline{H}_{t-1}, W_t, Y_t; \gamma)$ *for which*

$$E_{\boldsymbol{\gamma}}\left[s(\overline{H}_{t-1}, W_t, Y_t; \boldsymbol{\gamma}) \mid \mathcal{F}_{t-1}\right] = 0.$$

(The proof is identical, hence it is omitted.)

We remind one last result from real analysis which we will use in our theorem. We state it here to avoid unnecessarily complicated notation in the proof of the main theorem. The result extends to multivariate functions.

Remark A.2. *For a function* $f : \mathbb{R} \to \mathbb{R}$ *differentiable, if* $|f'(x)| \leq \alpha$ *for* $x \in (x_0 - \epsilon, x_0 + \epsilon)$ *and some* $\alpha in \mathbb{R}^+$*, then* $|f(x)|$ *is also bounded on* $(x_0 - \epsilon, x_0 + \epsilon)$.

Proof. The proof is straightforward using Taylor expansion:

$$f(x) = f(x_0) + f'(x^*)(x - x_0) \to |f(x)| \leq |f(x_0)| + \alpha\epsilon.$$

□

Now we can prove our theorem on asymptotic normality of the causal estimators using propensity scores that are estimated based on a correctly specified propensity score model.

Proof of Theorem 2. We will use Theorem A.3 to show asymptotic normality for the causal estimator based on the estimated propensity score model.

Remember that $\overline{H}_t = \{\overline{\boldsymbol{W}}_t, \overline{\boldsymbol{Y}}_t, \overline{\boldsymbol{X}}_{t+1}\}$. Then $\{\overline{H}_{t-1}, W_t, Y_t\} = \overline{H}_t \setminus \{\boldsymbol{X}_{t-1}\}$ is the set of observed variables until (and including) the t^{th} outcome. Let $\mu \in \mathbb{R}$ and $\boldsymbol{\gamma} \in \mathbb{R}^K$ be the parameters of the propensity score model with score functions $\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$, and define $\boldsymbol{\theta}^\top = (\mu, \boldsymbol{\gamma}^\top)$. Again based on Equation (A.2), we will show the asymptotic normality of the estimator that excludes spatial smoothing. We will them prove that the spatial smoothing does not affect estimation asymptotically because it

converges to zero faster than $T^{-1/2}$. Focusing on the first part of the error, define the $K+1$ vector

$$s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}) = \begin{pmatrix} \left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j;\boldsymbol{\gamma})}\right] N_B(Y_t) - N_{Bt}(F_h^M) - \mu \\ \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}\big) \end{pmatrix} = \begin{pmatrix} A_{1t} - \mu \\ \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}\big) \end{pmatrix},$$

where A_{1t} is defined in the proof of Theorem 1. We again work with the filtration $\mathcal{F}_t = \overline{H}^*_{t-M+1} = \{\overline{\boldsymbol{W}}_{t-M+1}, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T\}$. We will show that the conditions of Theorem A.3 hold.

Condition 1a We wish to show the expectation of s conditional on the filtration is 0. Since we showed in the proof of Theorem 1 that

$$\mathrm{E}\left\{\left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j;\boldsymbol{\gamma})}\right] N_B(Y_t) \mid \mathcal{F}_{t-1}\right\} = N_{Bt}(F_h^M),$$

we have $\boldsymbol{\theta}_0^\top = (\mu_0, \boldsymbol{\gamma}_0^\top) = (0, \boldsymbol{\gamma}_0^\top)$, where $\boldsymbol{\gamma}_0$ represents the true value for the parametric propensity score. Then, based on Lemma A.1, we have that $\mathrm{E}_{\boldsymbol{\theta}_0}\left[s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}_0) \mid \mathcal{F}_{t-1}\right] = 0$. Also, from Jensen's inequality we have that

$$\begin{aligned} \mathrm{E}^2_{\boldsymbol{\theta}_0}\left[\|s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}_0)\|\right] &\le \mathrm{E}_{\boldsymbol{\theta}_0}\left[\|s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}_0)\|^2\right] \\ &= \mathrm{E}_{\boldsymbol{\theta}_0}(A_{1t}^2) + \mathrm{E}_{\boldsymbol{\theta}_0}\left\{\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\|^2\right\} < \infty \end{aligned}$$

where the first term is finite because A_{1t} is bounded as shown in Equation (A.3), and the second term is finite based on Assumption A.2.

Condition 1b Since all terms are under the $\boldsymbol{\theta}_0$-law, we work with $\mu = \mu_0 = 0$. We have that

$$\begin{aligned} &E_{\boldsymbol{\theta}_0}\Big(s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}_0)s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}_0)^\top \mid \mathcal{F}_{t-1}\Big) \\ &= \begin{bmatrix} E_{\boldsymbol{\theta}_0}\left[A_{1t}^2 \mid \mathcal{F}_{t-1}\right] & E_{\boldsymbol{\theta}_0}\left[A_{1t}\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big)^\top \mid \mathcal{F}_{t-1}\right] \\ E_{\boldsymbol{\theta}_0}\left[A_{1t}\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big) \mid \mathcal{F}_{t-1}\right] & E_{\theta_0}\left[\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big)^\top \mid \mathcal{F}_{t-1}\right]. \end{bmatrix} \end{aligned}$$

Equation (A.5) implies that $(T-M+1)^{-1}\sum_{t=M}^{T} E_{\boldsymbol{\theta}_0}\left[A_{1t}^2 \mid \mathcal{F}_{t-1}\right] \xrightarrow{p} v$. In addition, due to Assumption A.2(1b), we also know that $(T-M+1)^{-1}\sum_{t=M}^{T} E_{\boldsymbol{\gamma}_0}\Big(\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big)^\top \mid \mathcal{F}_{t-1}\Big) \xrightarrow{p} V_{ps}$. Lastly, Assumption A.3 implies that $(T-M+1)^{-1}\sum_{t=M}^{T} E_{\boldsymbol{\theta}_0}\left[A_{1t}\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1};\boldsymbol{\gamma}_0\big) \mid \mathcal{F}_{t-1}\right] \xrightarrow{p} u$. Since all the entries of the matrix converge, we are left to show that the resulting matrix is positive definite.

However, since

$$M = \begin{bmatrix} A_{1t}^2 & A_{1t}\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)^\top \\ A_{1t}\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big) & \boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)^2 \end{bmatrix}$$

is positive definite (easy to check by taking vector $\boldsymbol{x} \in \mathbb{R}^k$, not all zero, and showing that $\boldsymbol{x}^\top M \boldsymbol{x} > 0$), we have that

$$\begin{bmatrix} v & u^\top \\ u & V_{ps} \end{bmatrix}$$

will also be positive definite.

Condition 1c Take $\epsilon > 0$ and write

$$\begin{aligned}
&\frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[\|s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}_0)\|^2 I\Big(\|s(\overline{H}_{t-1}, W_t, Y_t\,;\,\boldsymbol{\theta}_0)\| > \epsilon\sqrt{T}\Big) \mid \mathcal{F}_{t-1}\Big] \\
&= \frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[\big(A_{1t}^2 + \|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\big) I\Big(A_{1t}^2 + \|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T\Big) \mid \mathcal{F}_{t-1}\Big] \\
&= \frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[A_{1t}^2\, I\Big(A_{1t}^2 + \|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T\Big) \mid \mathcal{F}_{t-1}\Big] \\
&\qquad + \frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T - A_{1t}^2\Big) \mid \mathcal{F}_{t-1}\Big]
\end{aligned}$$

We start with the second term: Since A_{1t}^2 cannot exceed $(\delta_W^M \delta_Y + \delta_Y)^2$ based on Equation (A.3), we have that

$$\begin{aligned}
&\frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T - A_{1t}^2\Big) \mid \mathcal{F}_{t-1}\Big] \\
&\le \frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\| > \sqrt{\epsilon^2 T - (\delta_W^M \delta_Y + \delta_Y)^2}\Big) \mid \mathcal{F}_{t-1}\Big] \xrightarrow{p} 0,
\end{aligned}$$

based on Assumption A.2 and since $\delta_W^M \delta_Y + \delta_Y$ is fixed.

For the first term, since $I(A_{1t}^2 + \|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2) > \epsilon^2 T$ implies that at least one of A_{1t}^2 and $\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2$ is greater than $\epsilon^2 T/2$, we have that

$$I\big(A_{1t}^2 + \|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2) > \epsilon^2 T\big) \le I\big(A_{1t}^2 > \epsilon^2 T/2\big) + I\big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\big).$$

This leads to

$$E_{\theta_0}\Big[A_{1t}^2\, I\Big(A_{1t}^2 + \|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T\Big) \mid \mathcal{F}_{t-1}\Big]$$

$$\leq E_{\theta_0}\Big[A_{1t}^2\, I\Big(A_{1t}^2 > \epsilon^2 T/2\Big) \mid \mathcal{F}_{t-1}\Big] + E_{\theta_0}\Big[A_{1t}^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid \mathcal{F}_{t-1}\Big].$$

In the proof of Theorem 1 we have already shown that because A_{1t} is bounded we have that

$$\frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[A_{1t}^2\, I\Big(|A_{1t}| > \frac{\epsilon}{\sqrt{2}}\sqrt{T}\Big) \mid \mathcal{F}_{t-1}\Big] \xrightarrow{p} 0,$$

and we want to show that

$$\frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[A_{1t}^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid \mathcal{F}_{t-1}\Big] \xrightarrow{p} 0.$$

We write

$$\begin{aligned}
&E_{\theta_0}\Big[A_{1t}^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid \mathcal{F}_{t-1}\Big] \\
&= E_{\theta_0}\Big[A_{1t}^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big] \times \\
&\qquad \times P(A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2 \mid \mathcal{F}_{t-1}) + \\
&\quad + E_{\theta_0}\Big[A_{1t}^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid A_{1t}^2 \geq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big] \times \\
&\qquad \times P(A_{1t}^2 \geq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2 \mid \mathcal{F}_{t-1}) \\
&\leq E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big] \times \\
&\qquad \times P(A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2 \mid \mathcal{F}_{t-1}) + \\
&\quad + E_{\theta_0}\Big[A_{1t}^2\, I\Big(\|A_{1t}^2 > \epsilon^2 T/2\Big) \mid A_{1t}^2 \geq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big],
\end{aligned}$$

where again the average over time of the last term will be converging to zero in probability since A_{1t} is bounded and using similar arguments. Using the law of total expectation we can write

$$\begin{aligned}
&E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid \mathcal{F}_{t-1}\Big] \\
&= E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big] \times \\
&\qquad \times P(A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2 \mid \mathcal{F}_{t-1}) + \\
&\quad E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid A_{1t}^2 \geq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big] \times \\
&\qquad \times P(A_{1t}^2 \geq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2 \mid \mathcal{F}_{t-1}) \\
&\geq E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big) \mid A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big] \times \\
&\qquad \times P(A_{1t}^2 \leq \|\boldsymbol{\psi}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)\|^2 \mid \mathcal{F}_{t-1}).
\end{aligned}$$

Since all the terms in the expectations are positive and since (from Assumption A.2) we have that

$$\frac{1}{T-M+1}\sum_{t=M}^{T} E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big)\mid \mathcal{F}_{t-1}\Big] \xrightarrow{p} 0$$

we also have that

$$\frac{1}{T-M+1}\sum_{t=M}^{T}\Bigg\{ E_{\theta_0}\Big[\|\boldsymbol{\psi}\big(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\|^2\, I\Big(\|\boldsymbol{\psi}\big(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\|^2 > \epsilon^2 T/2\Big)\mid A_{1t}^2 \le \|\boldsymbol{\psi}\big(w_t,\overline{h}_{t-1};\boldsymbol{\gamma}\big)\|^2, \mathcal{F}_{t-1}\Big]\times \\ \times P(A_{1t}^2 \le \|\boldsymbol{\psi}\big(w_t,\overline{h}_{t-1};\boldsymbol{\gamma}\big)\|^2 \mid \mathcal{F}_{t-1})\Bigg\} \xrightarrow{p} 0$$

which completes the proof that Condition 1c holds.

Condition 2 We denote $\boldsymbol{\theta}^\top = (\theta_1,\theta_2,\ldots,\theta_{K+1}) = (\mu,\boldsymbol{\gamma}^\top)$ and use $s_k(\overline{H}_{t-1},W_t,Y_t\,;\,\boldsymbol{\theta})$ to denote the k^{th} entry of the $s(\overline{H}_{t-1},W_t,Y_t\,;\,\boldsymbol{\theta})$ vector. We note that

$$\frac{\partial}{\partial\boldsymbol{\theta}^T}s(\overline{H}_{t-1},W_t,Y_t\,;\,\boldsymbol{\theta}) = \begin{bmatrix} -1 & \frac{\partial}{\partial\boldsymbol{\gamma}^T}s_1(\overline{H}_{t-1},W_t,Y_t\,;\,\boldsymbol{\theta}) \\ \mathbf{0} & \frac{\partial}{\partial\boldsymbol{\gamma}^T}\boldsymbol{\psi}\big(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}\big)\end{bmatrix}$$

Lemma A.1 implies that $(T-M+1)^{-1}\sum_{t=M}^{T} E_{\boldsymbol{\gamma}_0}\left[\frac{\partial}{\partial\boldsymbol{\gamma}^T}\boldsymbol{\psi}\big(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}\big)\Big|_{\boldsymbol{\gamma}_0}\mid\mathcal{F}_{t-1}\right] \to -V_{ps}$ (invertible). Assumption A.3 and Lemma A.2 imply that

$$\begin{aligned}&(T-M+1)^{-1}\sum_{t=M}^{T} E_{\boldsymbol{\theta}_0}\left(\frac{\partial}{\partial\boldsymbol{\gamma}^T}s_1(\overline{H}_{t-1},W_t,Y_t\,;\,\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0}\mid\mathcal{F}_{t-1}\right)\\ &= -(T-M+1)^{-1}\sum_{t=M}^{T} E_{\boldsymbol{\theta}_0}\left[s_1(\overline{H}_{t-1},W_t,Y_t\,;\,\boldsymbol{\theta}_0)\boldsymbol{\psi}\big(W_t,\overline{H}_{t-1};\boldsymbol{\gamma}_0\big)\mid\mathcal{F}_{t-1}\right]\\ &\xrightarrow{p} -u^T.\end{aligned}$$

Putting these together we have that

$$(T-M+1)^{-1}\sum_{t=M}^{T} E_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\theta}^T}s(\overline{H}_{t-1},W_t,Y_t\,;\,\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0}\mid\mathcal{F}_{t-1}\right] \xrightarrow{p} \begin{bmatrix}-1 & -u\\ 0 & -V_{ps}\end{bmatrix}.$$

Since V_{ps} is invertible and the first row is the only one to have a non-zero first element we have that this limit matrix is invertible.

Condition 3 We want to show that for all $k, j = 1, 2, \ldots, K+1$, if we use P_{kjt} to denote

$$P_{kjt} = \frac{\partial}{\partial \theta_j} s_k(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0},$$

then $E_{\boldsymbol{\theta}_0}|P_{kjt}| < \infty$, and there exists $0 < r_{kj} \leq 2$ such that

$$\sum_{t=M}^{T} \frac{1}{t^{r_{kj}}} E_{\boldsymbol{\theta}_0}\left[\, |P_{kjt} - E_{\boldsymbol{\theta}_0}\left(P_{kjt} \mid \mathcal{F}_{t-1}\right)|^{r_{kj}} \mid \mathcal{F}_{t-1}\right] \xrightarrow{p} 0.$$

For $k, j \geq 2$, this is given by Condition 2 of Assumption A.2. For $j = 1$ and $k \geq 2$, we have that

$$\frac{\partial}{\partial \theta_1} s_k(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta}) = 0,$$

so the result holds for any r_{k1}. Similarly, for $k = j = 1$, we have that

$$\frac{\partial}{\partial \theta_1} s_1(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta}) = -1,$$

so the result holds for a value $r_{11} \in (1, 2]$. Therefore, it is left to show that it holds for $k = 1$ and $j \geq 2$. For $k = 1$ and $j \geq 2$, the condition that there exists $0 < r_{1j} \leq 2$ such that

$$\sum_{t=M}^{T} \frac{1}{t^{r_{1j}}} E_{\boldsymbol{\theta}_0}\left[\, |P_{1jt} - E_{\boldsymbol{\theta}_0}\left(P_{1jt} \mid \mathcal{F}_{t-1}\right)|^{r_{1j}} \mid \mathcal{F}_{t-1}\right] \xrightarrow{p} 0$$

is given by Assumption A.3. So we are left to show that $E(|P_{1jt}|) < \infty$. Lemma A.2 implies that

$$\begin{aligned}
P_{1jt} &= \frac{\partial}{\partial \boldsymbol{\gamma}_{j-1}} s_1(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0} \\
&= -N_B(Y_t) \left[\prod_{t'=t-M+1}^{t} \frac{f_h(W_{t'})}{e_{t'}(W_{t'}; \boldsymbol{\gamma}_0)}\right] \sum_{t'=t-M+1}^{t} \boldsymbol{\psi}_{j-1}\left(W_{t'}, \overline{H}_{t'-1}; \boldsymbol{\gamma}_0\right) \\
\implies |P_{1jt}| &\leq \delta_Y \delta_W^M \sum_{t'=t-M+1}^{t} \left|\boldsymbol{\psi}_{j-1}\left(W_{t'}, \overline{H}_{t'-1}; \boldsymbol{\gamma}_0\right)\right| \\
\implies E_{\boldsymbol{\theta}_0}|P_{1jt}| &\leq \delta_Y \delta_W^M \sum_{t'=t-M+1}^{t} E_{\boldsymbol{\gamma}_0}\left|\boldsymbol{\psi}_{j-1}\left(W_{t'}, \overline{H}_{t'-1}; \boldsymbol{\gamma}_0\right)\right|.
\end{aligned}$$

Since

$$E^2_{\boldsymbol{\gamma}_0}\left|\boldsymbol{\psi}_{j-1}\left(W_{t'}, \overline{H}_{t'-1}; \boldsymbol{\gamma}_0\right)\right| \leq E_{\boldsymbol{\gamma}_0}\left[\boldsymbol{\psi}_{j-1}\left(W_{t'}, \overline{H}_{t'-1}; \boldsymbol{\gamma}_0\right)^2\right] \qquad \text{(Jensen's inequality)}$$

$$\leq E_{\boldsymbol{\gamma}_0}\left[\|\boldsymbol{\psi}_{j-1}\big(W_{t'}, \overline{H}_{t'-1}; \boldsymbol{\gamma}_0\big)\|^2\right] < \infty, \qquad \text{(Assumption A.2)}$$

we have that $E|P_{1jt}| < \infty$.

Condition 4 We want to show that there exists integrable function $\ddot{\psi}(x)$ which dominates the second partial derivatives of $s(\overline{h}_{t-1}, w_t, y_t \ ; \ \boldsymbol{\theta})$ in a neighborhood of $\boldsymbol{\theta}_0$ for all $(w_t, \overline{h}_{t-1}, y_t)$. We consider derivatives of $s(\overline{h}_{t-1}, w_t, y_t \ ; \ \boldsymbol{\theta})$ with respect to θ_m, θ_l. For $k, m, l \geq 2$,

$$\frac{\partial}{\theta_m}\frac{\partial}{\theta_l} s_k(\overline{h}_{t-1}, w_t, y_t \ ; \ \boldsymbol{\theta}) = \frac{\partial}{\gamma_{m-1}}\frac{\partial}{\gamma_{l-1}}\boldsymbol{\psi}_{k-1}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big)$$

where $\boldsymbol{\psi}_{k-1}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)$ is the $k-1$ entry of the $\boldsymbol{\psi}\big(W_t, \overline{H}_{t-1}; \boldsymbol{\gamma}\big)$ vector. From Condition 3 of Assumption A.2, we know that the above is dominated by an integrable function. For $k \geq 2$ and if $l = 1$ or $m = 1$ we have that the second partial derivative is equal to 0, since

$$\frac{\partial}{\partial\theta_1} s_k(\overline{h}_{t-1}, w_t, y_t \ ; \ \boldsymbol{\theta}) = \frac{\partial}{\partial\mu}\boldsymbol{\psi}_{k-1}\big(w_t, \overline{h}_{t-1}; \boldsymbol{\gamma}\big) = 0.$$

So for $k \geq 2$, all second partial derivatives are dominated by the function in Condition 3 of Assumption A.2. Then, for $k = 1$, if at least one of $l = 1$ or $m = 1$ we have that the second partial derivative is also zero, since

$$\frac{\partial}{\partial\theta_1} s_1(\overline{h}_{t-1}, w_t, y_t \ ; \ \boldsymbol{\theta}) = \frac{\partial}{\partial\mu} s_1(\overline{h}_{t-1}, w_t, y_t \ ; \ \boldsymbol{\theta}) = -1.$$

So we need to show it only for $k = 1$, and $l, m \geq 2$. From Lemma A.2 we have that

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta_m\partial\theta_l} s_1(\overline{h}_{t-1}, w_t, y_t \ ; \ \boldsymbol{\theta}) \\
&= -N_B(y_t)\left[\prod_{j=t-M+1}^{t}\frac{f_h(w_j)}{e_j(w_j;\boldsymbol{\gamma})}\right]\left\{\left[\sum_{j=t-M+1}^{t}\frac{\partial}{\partial\boldsymbol{\gamma}_{m-1}}\boldsymbol{\psi}_{l-1}\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big)\right]\right. \\
&\qquad\qquad \left. - \left[\sum_{j=t-M+1}^{t}\boldsymbol{\psi}_{m-1}\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big)\right]\left[\sum_{j=t-M+1}^{t}\boldsymbol{\psi}_{l-1}\big(w_j, \overline{h}_{j-1}; \boldsymbol{\gamma}\big)\right]\right\}
\end{aligned}$$

Because of Assumption A.1(a) and Assumption 2 we have that

$$|N_B(y_t)| \leq \delta_Y \quad \text{and} \quad 0 \leq \prod_{j=t-M+1}^{t}\frac{f_h(w_j)}{e_j(w_j;\boldsymbol{\gamma})} \leq \delta_W^M,$$

which implies that

$$\begin{aligned}\left|\frac{\partial^2}{\partial\theta_m\partial\theta_l}s_1(\overline{h}_{t-1},w_t,y_t\,;\,\boldsymbol{\theta})\right| &\leq \delta_Y\delta_W^M\left|\sum_{j=t-M+1}^{t}\frac{\partial}{\partial\boldsymbol{\gamma}_{m-1}}\boldsymbol{\psi}_{l-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\right| + \\ &+\delta_Y\delta_W^M\left|\sum_{j,j'=t-M+1}^{t}\boldsymbol{\psi}_{m-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\boldsymbol{\psi}_{l-1}\big(w_{j'},\overline{h}_{j'-1};\boldsymbol{\gamma}\big)\right| \\ &\leq \sum_{j=t-M+1}^{t}\delta_Y\delta_W^M\left|\frac{\partial}{\partial\boldsymbol{\gamma}_{m-1}}\boldsymbol{\psi}_{l-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\right| + \\ &+\sum_{j,j'=t-M+1}^{t}\delta_Y\delta_W^M\left|\boldsymbol{\psi}_{m-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\boldsymbol{\psi}_{l-1}\big(w_{j'},\overline{h}_{j'-1};\boldsymbol{\gamma}\big)\right|\end{aligned}$$

We work first with the first term. Since the summation is over M terms with M finite, we only need to study the quantity in the absolute value. We know from Assumption A.2 that the second partial derivatives of $\boldsymbol{\psi}\big(w_t,\overline{h}_{t-1};\boldsymbol{\gamma}\big)$ are dominated by $\ddot{\psi}(w_t,\overline{h}_{t-1})$ in a neighborhood of $\boldsymbol{\gamma}_0$. Assume that this neighborhood is the $\epsilon-$ball around $\boldsymbol{\gamma}_0$ (this always exists since a neighborhood is an open set around $\boldsymbol{\gamma}_0$). Then, from Remark A.2 we know that

$$\left|\frac{\partial}{\partial\boldsymbol{\gamma}_{m-1}}\boldsymbol{\psi}_{l-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\right| \leq \left|\frac{\partial}{\partial\boldsymbol{\gamma}_{m-1}}\boldsymbol{\psi}_{l-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\Big|_{\boldsymbol{\gamma}_0}\right| + \epsilon\, K\,\ddot{\psi}(w_t,\overline{h}_{t-1}),$$

where the K appears because we consider all K second partial derivatives which are all bounded by $\ddot{\psi}$. From Assumption A.2(2), we have that the quantity on the right has finite expectation and is fixed in $\boldsymbol{\gamma}$. Therefore, it is an integrable function that dominates the first partial derivatives of $\boldsymbol{\psi}\big(w_t,\overline{h}_{t-1};\boldsymbol{\gamma}\big)$ in a neighborhood of $\boldsymbol{\gamma}_0$ for all l,m. Denote the maximum of these functions over l,m by $\ddot{\psi}_1(w_t,\overline{h}_{t-1})$.

We now turn our attention to the second term. Since (using again Remark A.2)

$$\big|\boldsymbol{\psi}_{m-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}\big)\big| \leq \big|\boldsymbol{\psi}_{m-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}_0\big)\big| + \epsilon\, K\,\ddot{\psi}_1(w_t,\overline{h}_{t-1}),$$

and $E_{\boldsymbol{\gamma}_0}\left[\big|\boldsymbol{\psi}_{m-1}\big(w_j,\overline{h}_{j-1};\boldsymbol{\gamma}_0\big)\big|\right] < \infty$ from Assumption A.2, we have that this quantity is also dominated by an integrable function that is constant in $\boldsymbol{\gamma}$. Denote the maximum of these functions over m as $\ddot{\psi}_2(w_t,\overline{h}_{t-1})$.

Putting these together we have that

$$\left|\frac{\partial^2}{\partial\theta_m\partial\theta_l}s_1(\overline{h}_{t-1},w_t,y_t\,;\,\boldsymbol{\theta})\right| \leq M\delta_Y\delta_W^M\ddot{\psi}_1(w_t,\overline{h}_{t-1}) + M^2\delta_Y\delta_W^M\left[\ddot{\psi}_2(w_t,\overline{h}_{t-1})\right]^2,$$

where the right hand side is integrable. By defining taking the maximum of the quantity on the right

hand side and $\ddot{\psi}(w_t, \overline{h}_{t-1})$ for each $(w_t, \overline{h}_{t-1})$ we have that the condition holds using this new integrable function.

Consistency of the solution The last condition of Theorem A.3 that we need to show is that the solution to $\sum_{t=M}^{T} s(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta}) = 0$ is consistent for $\boldsymbol{\theta}_0$. Since the estimator of the propensity score parameters based on the score functions are consistent, we only need to show that the solution to $\sum_{t=M}^{T} s_1(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta}) = 0$ is consistent for $\mu_0 = 0$.

Since the estimator based on the true propensity score was shown to be consistent in Theorem 1, the propensity score estimators $\widehat{\boldsymbol{\gamma}}$ are consistent for $\boldsymbol{\gamma}$, $s_1(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta})$ is a continuous function of the propensity score which is itself continuous in $\boldsymbol{\gamma}$, using Slutsky's theorem we have that the solution to $\sum_{t=M}^{T} s_1(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta}) = 0$ using the estimated propensity score parameters is also consistent.

Asymptotic normality of the estimator without spatial smoothing Since the conditions of Theorem A.3 are satisfied, we have that the solution $\widehat{\boldsymbol{\theta}}_T$ to $\sum_{t=M}^{T} s(\overline{H}_{t-1}, W_t, Y_t \,;\, \boldsymbol{\theta}) = 0$ are asymptotically normal with

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(0, V_{\boldsymbol{\theta}}\right),$$

where $V_{\boldsymbol{\theta}} = A^{-1} B \left(A^{-1}\right)^T$ for

$$A = \begin{bmatrix} -1 & -u^T \\ \mathbf{0}_K & -V_{ps} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} v & u^T \\ u & V_{ps} \end{bmatrix}. \tag{A.9}$$

As a result, focusing on the first entry of $\widehat{\theta}$ and since $\mu_0 = 0$, we have that

$$\sqrt{T}\Big\{ \underbrace{(T-M+1)^{-1} \sum_{t=M}^{T} \left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j;\boldsymbol{\gamma})} \right] N_B(Y_t)}_{\text{estimator without spatial smoothing}} - N_B(F_h^M) \Big\} \to N\left(0, v^e\right),$$

where $v^e = [V_{\boldsymbol{\theta}}]_{11}$ is the $(1,1)$ entry of $V_{\boldsymbol{\theta}}$.

Asymptotic normality of the estimator *with* spatial smoothing To prove the asymptotic normality of the estimator with spatial smoothing (our proposed estimator in Equation (8)), we again decompose the estimation error in two components like in Equation (A.2) for the proof of Theorem 1. We write

$$err_t = \underbrace{\left[\prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j;\boldsymbol{\gamma})} \right] N_B(Y_t) - N_{Bt}(F_h^M)}_{A_{1t}}$$

$$+\underbrace{\left[\prod_{j=t-M+1}^{t}\frac{f_h(W_j)}{e_j(W_j;\boldsymbol{\gamma})}\right]\left[\int_B\sum_{s\in S_{Y_t}}K_{b_T}(\omega,s)\mathrm{d}\omega-N_B(Y_t)\right]}_{A_{2t}},$$

where we use the parametric propensity score. We showed the asymptotic normality based on A_{1t}, so we are left to show that $\sqrt{T}\left((T-M+1)^{-1}\sum_{t=M}^{T}A_{2t}\right)\xrightarrow{p}0$. In the proof of Theorem 1 we already showed that the above result holds. The proof there can be directly used here also if the known propensity score is used (instead of the estimated one). By re-defining the terms c_t defined there to use the estimated propensity score as

$$c_t=\prod_{j=t-M+1}^{t}\frac{f_h(W_j)}{e_j(W_j;\boldsymbol{\gamma})},$$

it suffices to show that c_t is bounded, and the steps of the proof with the known propensity score will follow identically. But since the propensity score $e_t(w;\boldsymbol{\gamma})$ is continuous in $\boldsymbol{\gamma}$ (since it is differentiable), the function $1/x$ is continuous for $x>0$, and $f_h(w_j)/e_j(w_j;\boldsymbol{\gamma}_0)\leq\delta_W$ then c_t will be bounded in a neighborhood of $\boldsymbol{\gamma}_0$. And since $\widehat{\boldsymbol{\gamma}}\xrightarrow{p}\boldsymbol{\gamma}_0$, $\widehat{\boldsymbol{\gamma}}$ will be in the neighborhood of $\boldsymbol{\gamma}_0$ with probability 1 as T increases, so c_t will be bounded.

Putting these results together we have asymptotic normality of the spatially smoothed estimator and

$$\sqrt{T}\left(\widehat{N}_B(F_h^M)-N_B(F_h^M)\right)\xrightarrow{d}N\left(0,v^e\right).$$

□

Proof of Theorem 3. The asymptotic variance v^e corresponds to the (1, 1) entry of the matrix $A^{-1}B(A^{-1})^\top$, where A, B are defined in Equation (A.9).

$$\begin{aligned}
A^{-1}B(A^{-1})^\top &= \begin{bmatrix}1 & u^T\\ \mathbf{0}_K & V_{ps}\end{bmatrix}^{-1}\begin{bmatrix}v & u^\top\\ u & V_{ps}\end{bmatrix}\left\{\begin{bmatrix}1 & u^\top\\ \mathbf{0}_K & V_{ps}\end{bmatrix}^{-1}\right\}^\top\\
&= \begin{bmatrix}1 & -u^\top V_{ps}^{-1}\\ \mathbf{0}_K & V_{ps}^{-1}\end{bmatrix}\begin{bmatrix}v & u^\top\\ u & V_{ps}\end{bmatrix}\begin{bmatrix}1 & \mathbf{0}_K\\ -V_{ps}^{-1}u & V_{ps}^{-1}\end{bmatrix}\\
&= \begin{bmatrix}v-u^\top V_{ps}^{-1}u & \mathbf{0}_K^\top\\ \dots & \dots\end{bmatrix}\begin{bmatrix}1 & \dots\\ -V_{ps}^{-1}u & \dots\end{bmatrix}\\
&= \begin{bmatrix}v-u^\top V_{ps}^{-1}u & \dots\\ \dots & \dots\end{bmatrix}
\end{aligned}$$

so $v^e=v-u^\top V_{ps}^{-1}u$, and since V_{ps} is positive definite and therefore V_{ps}^{-1} is positive definite we have that $u^\top V_{ps}^{-1}u\geq 0$ and $v^e\leq v$. □

B.5 Asymptotics for an increasing number of independent regions

All the asymptotic results that have been discussed up to now correspond to the scenario where 1 region is observed repeatedly over time, and the asymptotic properties are derived when the number of time periods T increases to infinity. However, there might also be interest in situations where the number of time periods is fixed, but there exist an increasing number of independent-acting regions.

Here we consider this related but separate scenario. We start by defining relevant estimands in this setting, ensuring that these new estimands are as closely comparable to the estimands in the manuscript. We propose similar estimators, and derive the asymptotic properties of the new estimators when the number of regions R goes to infinity.

B.5.1 Estimands for independently-acting regions For this scenario, we decompose the treatments, potential outcomes, outcomes, and history over all the regions to region-specific components and write $w_t = (w_{1t}, w_{2t}, \ldots, w_{Rt})$, $\overline{\boldsymbol{w}}_t = (\overline{\boldsymbol{w}}_{1t}, \overline{\boldsymbol{w}}_{2t}, \ldots, \overline{\boldsymbol{w}}_{Rt})$, $Y_t(\overline{\boldsymbol{w}}_t) = (Y_{1t}(\overline{\boldsymbol{w}}_t), Y_{2t}(\overline{\boldsymbol{w}}_t), \ldots, Y_{Rt}(\overline{\boldsymbol{w}}_t))$, $Y_t = (Y_{1t}, Y_{2t}, \ldots, Y_{Rt})$, and $\overline{H}_t = (\overline{H}_{1t}, \overline{H}_{2t}, \ldots, \overline{H}_{Rt})$, where $\overline{H}_{rt} = \{\overline{\boldsymbol{W}}_{rt}, \overline{\boldsymbol{Y}}_{rt}, \overline{\boldsymbol{X}}_{r(t+1)}\}$. We make the following assumption that describes that the regions do not interfere spatially, and that treatment assignment is local within regions:

Assumption A.4 (Independently acting spatial regions)**.** *We assume the following:*

1. *For $\overline{\boldsymbol{w}}_t, \overline{\boldsymbol{w}}'_t$ such that $\overline{\boldsymbol{w}}_{rt} = \overline{\boldsymbol{w}}'_{rt}$, we have that $Y_{rt}(\overline{\boldsymbol{w}}_t) = Y_{rt}(\overline{\boldsymbol{w}}'_t)$ (and a similar assumption for the time-varying covariates), and*

2. *the treatment assignment of region r at time t does not depend on unobserved potential outcomes or potential time varying covariates, nor on any information from other regions, denoted as $W_{rt} \perp\!\!\!\perp \overline{H}_{t-1}, \overline{\mathcal{Y}}_T, \overline{\mathcal{X}}_T \mid \overline{H}_{r(t-1)}$.*

This assumption allows us to denote potential outcomes using their region-specific treatments *only*, and write $Y_t(\overline{\boldsymbol{w}}_t) = (Y_{1t}(\overline{\boldsymbol{w}}_{1t}), Y_{2t}(\overline{\boldsymbol{w}}_{2t}), \ldots, Y_{Rt}(\overline{\boldsymbol{w}}_{Rt}))$. It also allows us to think of the R regions as completely separately acting regions, as outcomes, covariates and treatments of one region do not depend on any information of any other region. Based on this assumption, we can use $\overline{\mathcal{Y}}_{rT}$ to denote the collection of potential outcomes for region r over all time periods and for any regional treatment path (and similarly for covariates).

For the purpose of this section *only*, we also assume the temporal carryover effect is limited to up to some lag M_Y. Specifically, we assume that the outcome at time t can only depend on treatments during the preceding M_Y time periods, formalized as

Assumption A.5 (Limited temporal carryover effect)**.** *There exists positive integer M_Y such that for $\overline{\boldsymbol{w}}_{rt}, \overline{\boldsymbol{w}}'_{rt}$ for which $w_{r\tau} = w'_{r\tau}$ for all $\tau = t - M_Y + 1, \ldots, t-1, t$, it holds that $Y_{rt}(\overline{\boldsymbol{w}}_{rt}) = Y_{rt}(\overline{\boldsymbol{w}}'_{rt})$.*

We start by defining region and time specific estimands that are as closely related to the estimands defined in Section 3. We again focus on point pattern treatments and outcomes and on estimands that represent the number of outcome active locations in each region. For simplicity we focus on the scenario where the temporal carryover lag M_Y and the intervention length M are both equal to 1, but we note that the results would also follow in all scenarios where $M \geq M_Y$.

Let F_h be a stochastic treatment assignment that is constant across regions. The stochastic intervention can depend on baseline covariates of the regions, but we refrain for explicitly denoting that for simplicity. We define the expected number of outcome active locations at region r at time t as

$$N_{rt}(F_h) = \int_{w_{rt}} N_r\Big(Y_{rt}(\overline{\boldsymbol{W}}_{r(t-1)}, w_{rt})\Big) dF_h(w_{rt}) = \int_{w_{rt}} N_r\big(Y_{rt}(w_{rt})\big) dF_h(w_{rt}),$$

where we define the estimand as in the first equation to be more closely related to the estimands in §3, and the second equation holds because of Assumption A.5 for $M_Y = 1$. We specify region-specific estimands, averaged over time, as

$$N_r(F_h) = \frac{1}{T} \sum_{t=1}^{T} N_{rt}(F_h),$$

and estimands averaged over region and time as

$$N(F_h) = \frac{1}{R} \sum_{r=1}^{R} N_r(F_h) = \frac{1}{R} \sum_{r=1}^{R} \frac{1}{T} \sum_{t=1}^{T} N_{rt}(F_h).$$

B.5.2 Estimators for independently-acting regions Like in Section 4, assume that F_h admits density f_h. Based on Assumption A.4, we can separate the treatment assignment over all regions to the treatment assignment of each region separately, as

$$e_t(w_t) = \prod_{r=1}^{R} e_{rt}(w_{rt})$$

where $e_{rt}(w_{rt}) = f(W_{rt} = w_{rt} \mid \overline{H}_{r(t-1)})$ is the region-specific propensity score. We propose corresponding region and time-specific estimator

$$\widehat{N}_{rt}(F_h) = \frac{f_h(W_{rt})}{e_{rt}(W_{rt})} N_r(Y_{rt}),$$

where we use $N_r(Y_{rt})$ to denote the number of outcome active locations in the observed outcome for region r at time t. We also propose the corresponding estimators averaged over time and over regions as

$$\widehat{N}_r(F_h) = \frac{1}{T}\sum_{t=1}^{T}\widehat{N}_{rt}(F_h) \quad \text{and} \quad \widehat{N}(F_h) = \frac{1}{R}\sum_{r=1}^{T}\widehat{N}_r(F_h).$$

B.5.3 Consistency and asymptotic normality for independent-acting regions We will show the consistency and asymptotic normality of these estimators when the propensity score is known for an increasing number of independently acting regions. The proof here follows closely the proof in Papadogeorgou *et al.* (2019) for weighting estimators under a known propensity score and for stochastic interventions. We do not show the asymptotic properties of an estimator based on a correctly specified parametric propensity score, since, once the baseline conditions for the known propensity score are established, the proof for the estimated propensity score would resemble the corresponding proof in Papadogeorgou *et al.* (2019).

To establish the asymptotic properties for an increasing number of regions, we first assume that our observed regions are a random sample from some super-population of regions. Let $(\overline{\mathcal{Y}}_{rT}, \overline{\mathcal{X}}_{rT}, \overline{\boldsymbol{W}}_{rT})$ be a draw from a super-population distribution F^{sp}. We assume that Assumption A.4 holds over F^{sp} and we make the following super-population positivity assumption for the independent regions (which resembles the one in Assumption 2):

Assumption A.6 (Positivity of treatment assignment in the super-population)**.** *There exists δ_W such that $e_{rt}(w_{rt}) > \delta_W \cdot f_h(w_{rt})$ for all treatment point patterns w_{rt}.*

We also assume that there is a bounded number of outcome active locations within each region, similarly to Assumption A.1(a), but region-specific:

Assumption A.7. *There exists $\delta_Y > 0$ such that $N_r(Y_{rt}(\overline{\boldsymbol{w}}_{rt})) < \delta_Y$ with probability 1 over F^{sp}, where $\overline{\boldsymbol{w}}_{rt}$ is any possible treatment path.*

Theorem A.4. *If Assumptions A.4, A.5 and A.6 hold, then, for $R \to \infty$, $\widehat{N}(F_h)$ is consistent for $N(F_h)$ and $\sqrt{R}\left(\widehat{N}(F_h) - N(F_h)\right) \to N(0, \sigma^2)$, for some $\sigma^2 > 0$, where $N(F_h)$ is the super-population estimand defined as $N(F_h) = E_{F^{sp}}\left[N_r(F_h)\right]$.*

Proof. Let $\mathcal{D}_r = (\overline{\boldsymbol{W}}_T, \overline{\boldsymbol{Y}}_T, \overline{\boldsymbol{X}}_T)$ denote all observed data for region r, and $\mathcal{D} = (\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_R)$. We define

$$\psi_r(\mathcal{D}_r; \mu) = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt})\right) - \mu$$

and $\Psi_R(\mathcal{D};\mu) = \sum_{r=1}^{T} \psi_r(\mathcal{D}_r;\mu)$. Then obviously the estimator $\widehat{\mu} = \widehat{N}(F_h)$ is the solution to $\Psi_R(\mathcal{D};\mu) = 0$. Then we calculate the solution to $\Psi^{sp}(\mu) = E_{F^{sp}}(\psi_r(\mathcal{D}_r;\mu)) = 0$ which is equal to

$$
\begin{aligned}
\mu_0 &= E_{F^{sp}}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt})\right] \\
&= \frac{1}{T}\sum_{t=1}^{T} E_{F^{sp}}\left[\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt})\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT}}\int_{w_{r1}}\int_{w_{r2}}\cdots\int_{w_{rt}}\frac{f_h(w_{rt})}{e_{rt}(w_{rt})}N_r(Y_{rt}(\overline{\boldsymbol{w}}_{rt}))\,\mathrm{d}F^{sp}(\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT},\overline{\boldsymbol{w}}_{rt}) \\
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT}}\int_{w_{r1}}\int_{w_{r2}}\cdots\int_{w_{r(t-1)}} \\
&\qquad\qquad \left[\int_{w_{rt}}\frac{f_h(w_{rt})}{e_{rt}(w_{rt})}N_r(Y_{rt}(\overline{\boldsymbol{w}}_{rt}))f_{W_{rt}}(w_{rt}\mid \overline{\boldsymbol{W}}_{r(t-1)} = \overline{\boldsymbol{w}}_{r(t-1)},\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT})\,\mathrm{d}w_{rt}\right] \\
&\qquad\qquad \mathrm{d}F^{sp}(\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT},\overline{\boldsymbol{w}}_{r(t-1)}) \\
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT}}\int_{w_{r1}}\int_{w_{r2}}\cdots\int_{w_{r(t-1)}}\left[\int_{w_{rt}}\frac{f_h(w_{rt})}{e_{rt}(w_{rt})}N_r(Y_{rt}(\overline{\boldsymbol{w}}_{rt}))f_{W_{rt}}(w_{rt}\mid \overline{H}_{r(t-1)})\,\mathrm{d}w_{rt}\right] \\
&\qquad\qquad \mathrm{d}F^{sp}(\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT},\overline{\boldsymbol{w}}_{r(t-1)}) \qquad \text{(From Assumption A.4)} \\
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT}}\int_{w_{r1}}\int_{w_{r2}}\cdots\int_{w_{r(t-1)}}\left[\int_{w_{rt}} f_h(w_{rt})N_r(Y_{rt}(\overline{\boldsymbol{w}}_{rt}))\,\mathrm{d}w_{rt}\right]\,\mathrm{d}F^{sp}(\overline{\mathcal{Y}}_{rT},\overline{\mathcal{X}}_{rT},\overline{\boldsymbol{w}}_{r(t-1)}) \\
&\qquad\qquad \text{(From the definition of the region-specific propensity score)} \\
&= \frac{1}{T}\sum_{t=1}^{T} E_{F^{sp}}\left[N_{rt}(F_h)\right] \\
&= E_{F^{sp}}\left[N_r(F_h)\right]
\end{aligned}
$$

Consistency We use an alteration of Lemma A in Section 7.2.1 of Serfling (1980). Since $\psi_r(\mathcal{D}_r;\mu)$ is monotone in μ with $\partial\psi_r(\mathcal{D}_r;\mu)/\partial\mu = -1 < 0$, we have that $\Psi_R(\mathcal{D};\mu)$ and $\Psi^{sp}(\mu)$ are also monotone which implies uniqueness of their roots, $\widehat{\mu}$ and μ_0. From the strong law of large numbers we have that $\Psi_R(\mathcal{D};\mu) \xrightarrow{a.s.} \Psi^{sp}(\mu)$, and

$$
|\Psi^{sp}(\widehat{\mu}) - \Psi^{sp}(\mu_0)| = |\Psi^{sp}(\widehat{\mu}) - \Psi_R(\widehat{\mu})| \leq \sup_{\mu}|\Psi^{sp}(\mu) - \Psi_R(\mu)| \to 0,
$$

which, by the uniqueness of the roots for Ψ^{sp} and Ψ_R implies that $\widehat{N}(F_h) \xrightarrow{as} N(F_h)$ and $\widehat{N}(F_h)$ is consistent for $N(F_h)$.

Asymptotic normality For asymptotic normality we will use Theorem A in Section 7.2.2 of Serfling (1980). We have already shown that μ_0 is an isolated root of $\Psi^{sp}(\mu) = 0$ (since it is unique) and that $\psi_r(\mathcal{D}_r; \mu)$ is monotone in μ. We also have that $\Psi^{sp}(\mu)$ is differentiable in μ with $\frac{\partial}{\partial \mu}\Psi^{sp}(\mu) = -1 \neq 0$. Lastly we will show that $E_{F^{sp}}\left[\psi_r^2(\mathcal{D}_r; \mu)\right]$ is finite in a neighborhood of μ_0. To do so, consider μ in an ϵ-neighborhood of μ_0, $\mu \in (\mu_0 - \epsilon, \mu_0 + \epsilon)$. Then

$$
\begin{aligned}
E_{F^{sp}}\left[\psi_r^2(\mathcal{D}_r; \mu)\right] &= E_{F^{sp}}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt}) - \mu\right]^2\right\} \\
&= E_{F^{sp}}\left\{\left|\frac{1}{T}\sum_{t=1}^{T}\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt}) - \mu\right|^2\right\} \\
&\leq E_{F^{sp}}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt}) + |\mu|\right]^2\right\} \qquad \text{(Triangle inequality)} \\
&= E_{F^{sp}}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt})\right]^2\right\} + 2|\mu|E_{F^{sp}}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{f_h(W_{rt})}{e_{rt}(W_{rt})}N_r(Y_{rt})\right] + \mu^2 \\
&\leq E_{F^{sp}}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\delta_W\delta_Y\right]^2\right\} + 2|\mu|E_{F^{sp}}\left[\frac{1}{T}\sum_{t=1}^{T}\delta_W\delta_Y\right] + |\mu|^2 \\
&= (\delta_W\delta_Y)^2 + 2|\mu|\delta_W\delta_Y + \mu^2
\end{aligned}
$$

where we used that all terms in the summation are positive along with Assumptions A.6 and A.7. Since $\mu \in (\mu_0 + \epsilon, \mu_0 + \epsilon)$ it is bounded, so the expectation above exists.

Then, since all the conditions of the theorem are satisfied we have that

$$\sqrt{R}\left(\widehat{N}(F_h) - N(F_h)\right) \to N(0, \sigma^2),$$

where $\sigma^2 = E_{F^{sp}}\left[\psi_r^2(\mathcal{D}_r; \mu_0)\right]$. □

C The Hájek Estimator

The standardization of weights used in the Hájek estimator is known to be effective in the settings where the weights are extreme. Its sample boundedness property guarantees that the resulting estimate is always within the range of the observed outcome. In our case, the Hájek estimator replaces the division by $T - M + 1$ with that by $\sum_{t=M}^{T} w_t$ where w_t is the product of fractions in Equation (6). For example,

$$\widehat{N}_B(F_h^M)_{\text{Hájek}} = \frac{1}{\sum_{t=M}^{T} w_t}\sum_{t=M}^{T}\widehat{N}_{Bt}(F_h^M)$$

The new martingale theorem stated in Theorem A.3 can be used in future research to show that the Hájek estimator is consistent and asymptotically normal, and derive the functional form of its asymptotic variance. However, for now, we use a heuristic approach to estimating the variance bound of the Hájek estimator. Since the Hájek estimator simply rescales the corresponding IPW estimator by $(T - M + 1)/\sum_{t=M}^{T} w_t$, we scale the variance bound derived for the estimator by $[(T - M + 1)/(\sum_{t=M}^{T} w_t)]^2$.

D Sensitivity analysis

In this section we discuss sensitivity analysis for the IPW estimators. In the main text of the manuscript we discuss sensitivity analysis for the Hájek estimator, which is admittedly a much harder problem due to the standardization of weights performed in the Hájek correction.

In this section we discuss sensitivity analysis based on the IPW estimator. We quickly see that bounding the estimator for different amounts of propensity score misspecification Γ can be directly achieved by solving a linear program. We can similarly bound the causal effect estimator exactly. In contrast, in the main text, bounding the value of the Hájek estimator requires additional tools to *transform* the problem to a linear program. This transformation forbids us from bounding the effect estimator exactly and forces us to acquire possibly conservative bounds for the effect estimator (see Theorem 4).

D.1 For the IPW estimator

We focus again on bounding the estimators for intervention over a single time period, though extensions to multiple time periods are direct, and discussed in more detail for the Hájek estimator in Appendix D.2. The IPW estimators that use the correct propensity score can be written as:

$$\begin{aligned}
\widehat{N}_{\boldsymbol{\rho}}(F_h) &= \frac{1}{T}\sum_{t=1}^{T} \rho_t \, w_t(F_h) \, \widetilde{N}_B(Y_t), \quad \text{and} \\
\widehat{\tau}_{\boldsymbol{\rho}}(F_{h_1}, F_{h_2}) &= \frac{1}{T}\sum_{t=1}^{T} \rho_t \, w_t(F_{h_2}) \, \widetilde{N}_B(Y_t) - \frac{1}{T}\sum_{t=1}^{T} \rho_t \, w_t(F_{h_1}) \, \widetilde{N}_B(Y_t) \\
&= \frac{1}{T}\sum_{t=1}^{T} \rho_t \left[w_t(F_{h_2}) - w_t(F_{h_1})\right] \widetilde{N}_B(Y_t)
\end{aligned}$$

where

$$w_t(F_h) = \frac{f_h(W_t)}{e_t(W_t)} \quad \text{and} \quad \widetilde{N}_B(Y_t) = \int_B \sum_{s \in S_{Y_t}} K_b(\|\omega - s\|)\mathrm{d}\omega.$$

Both of the IPW estimators $\widehat{N}_{\boldsymbol{\rho}}(F_h)$ and $\widehat{\tau}_{\boldsymbol{\rho}}(F_{h_1}, F_{h_2})$ are linear in $\boldsymbol{\rho}$, so finding their maximum/minimum over $\rho_t \in [\Gamma^{-1}, \Gamma]^T$ for each t is a linear problem and can be easily solved.

D.2 For the Hájek estimator

The standardization of the weights in the Hájek estimator implies that maximizing/minimizing the value of the estimator is no longer linear in ρ_t. This is evident in the form of the Hájek estimator for the number of points and the effect in a region, defined respectively as

$$\frac{\sum_{t=1}^{T} \rho_t \, w_t(F_h) \, \widetilde{N}_B(Y_t)}{\sum_{t=1}^{T} \rho_t \, w_t(F_h)} \quad \text{and} \quad \frac{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_2}) \, \widetilde{N}_B(Y_t)}{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_2})} - \frac{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_1}) \, \widetilde{N}_B(Y_t)}{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_1})}$$

where

$$w_t(F_h) = \frac{f_h(W_t)}{e_t(W_t)} \quad \text{and} \quad \widetilde{N}_B(Y_t) = \int_B \sum_{s \in S_{Y_t}} K_b(\|\omega - s\|) \mathrm{d}\omega.$$

Theorem 4 states that bounding the estimator for the expected number of points $\widehat{N}_{\boldsymbol{\rho}}(F_h)$ can be transformed to a linear problem. However, the standardization of weights in the Hájek estimator and the fact that our estimator is the difference of two linear fractionals forbids us to see the problem of bounding the effect estimator $\widehat{\tau}_{\boldsymbol{\rho}}(F_{h_1}, F_{h_2})$ the same way.

Proof of Theorem 4. We view the problem of bounding $\widehat{N}_{\boldsymbol{\rho}}(F_{h_2})$ as a maximization/minimization problem of a linear fractional with positive denominator. These problems have been previously studied, and it has been shown that they can be transformed to a linear programming problem using the Charnes-Cooper transformation (Charnes and Cooper, 1962). The theorem states this transformation in the context of our estimator.

For the problem of bounding the effect estimator, the objective can be written as,

$$\begin{aligned}
\widehat{\tau}_{\boldsymbol{\rho}}(F_{h_1}, F_{h_2}) &= \frac{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_2}) \, \widetilde{N}_B(Y_t)}{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_2})} - \frac{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_1}) \, \widetilde{N}_B(Y_t)}{\sum_{t=1}^{T} \rho_t \, w_t(F_{h_1})} \\
&= \widehat{N}_{\boldsymbol{\rho}}(F_{h_2}) - \widehat{N}_{\boldsymbol{\rho}}(F_{h_1}).
\end{aligned}$$

Thus, maximizing $\widehat{\tau}_{\boldsymbol{\rho}}(F_{h_1}, F_{h_2})$ over $\boldsymbol{\rho} \in [\Gamma^{-1}, \Gamma]^T$ is equivalent to maximizing $\widehat{N}_{\boldsymbol{\rho}}(F_{h_2}) - \widehat{N}_{\boldsymbol{\rho}}(F_{h_1})$ over the same region for $\boldsymbol{\rho}$. Since the space $(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) \in [\Gamma^{-1}, \Gamma]^{2T}$ includes $(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ where $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_2$ as a subspace, we have that

$$\begin{aligned}
\max_{\boldsymbol{\rho} \in [\Gamma^{-1}, \Gamma]^T} \left\{ \widehat{N}_{\boldsymbol{\rho}}(F_{h_2}) - \widehat{N}_{\boldsymbol{\rho}}(F_{h_1}) \right\} &\leq \max_{(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) \in [\Gamma^{-1}, \Gamma]^{2T}} \left\{ \widehat{N}_{\boldsymbol{\rho}_2}(F_{h_2}) - \widehat{N}_{\boldsymbol{\rho}_1}(F_{h_1}) \right\} \\
&= \max_{\boldsymbol{\rho}_2 \in [\Gamma^{-1}, \Gamma]^T} \left\{ \widehat{N}_{\boldsymbol{\rho}_2}(F_{h_2}) \right\} - \min_{\boldsymbol{\rho}_1 \in [\Gamma^{-1}, \Gamma]^T} \left\{ \widehat{N}_{\boldsymbol{\rho}_1}(F_{h_1}) \right\},
\end{aligned}$$

where the last equality holds since $\widehat{N}_{\boldsymbol{\rho}_j}(F_{h_j}) \geq 0$. Similarly, we can derive the bound for the minimum of $\widehat{\tau}_{\boldsymbol{\rho}}(F_{h_1}, F_{h_2})$. □

Next, we derive similar conservative bounds for the estimators corresponding to the interventions over multiple time periods. Recall that the Hájek estimator for the number of events in region B under a stochastic intervention is given by,

$$\frac{\sum_{t=M}^{T} w_t(F_h^M)\ \widetilde{N}_B(Y_t)}{\sum_{t=M}^{T} w_t(F_h^M)}, \quad \text{where} \quad w_t(F_h^M) = \prod_{j=t-M+1}^{t} \frac{f_h(W_j)}{e_j(W_j)}.$$

So, our sensitivity analysis would search to find the bounds of

$$\widehat{N}_{\boldsymbol{\rho}}(F_h^M) = \frac{\sum_{t=M}^{T} \left(\prod_{j=t-M+1}^{t} \rho_j\right) w_t(F_h^M)\ \widetilde{N}_B(Y_t)}{\sum_{t=M}^{T} \left(\prod_{j=t-M+1}^{t} \rho_j\right) w_t(F_h^M)}$$

over $\boldsymbol{\rho} \in [\Gamma^{-1}, \Gamma]$. Since each ρ_t can take a value in $[\Gamma^{-1}, \Gamma]$, the sensitivity analysis weights in the Hájek estimator for multiple time periods, $\prod_{j=t-M+1} \rho_j$, take a value in $[\Gamma^{-M}, \Gamma^M]$. Therefore the set $\{\boldsymbol{\alpha} \in [\Gamma^{-M}, \Gamma^M]\}$ includes all vectors of length T whose t^{th} entry can be written as $\prod_{j=t-M+1}^{t} \rho_t$ for some vector $\boldsymbol{\rho}$. Using the argument similar to that of Theorem 4, we have

$$\min_{\boldsymbol{\rho}\in[\Gamma^{-1},\Gamma]} \left\{ \widehat{N}_{\boldsymbol{\rho}}(F_h^M) \right\} \geq \min_{\boldsymbol{\alpha}\in[\Gamma^{-M},\Gamma^M]} \left\{ \frac{\sum_{t=M}^{T} \alpha_t\ w_t(F_h^M)\ \widetilde{N}_B(Y_t)}{\sum_{t=M}^{T} \alpha_t\ w_t(F_h^M)} \right\},$$
$$\max_{\boldsymbol{\rho}\in[\Gamma^{-1},\Gamma]} \left\{ \widehat{N}_{\boldsymbol{\rho}}(F_h^M) \right\} \leq \max_{\boldsymbol{\alpha}\in[\Gamma^{-M},\Gamma^M]} \left\{ \frac{\sum_{t=M}^{T} \alpha_t\ w_t(F_h^M)\ \widetilde{N}_B(Y_t)}{\sum_{t=M}^{T} \alpha_t\ w_t(F_h^M)} \right\}.$$

The quantities on the right can be computed by turning the linear fractional problem to a linear problem via the Charnes-Cooper transformation. Then, we can use these quantities as the conservative bounds for the minimum and maximum of our target quantities. Based on these bounds, we can again use Theorem 4 to acquire conservative bounds of the effect of changing the intervention for interventions over multiple time periods.

E Additional Simulation Results on the Iraq-based scenario

E.1 Asymptotic Variance and Bound, and Estimated Variance Bound

Figure 5 shows the average (over 200 simulated data sets) of the true asymptotic standard deviation and true bound as well as the estimated standard deviation bound of the IPW estimator for the average potential outcome using the true propensity score, for interventions taking place over $M \in \{1, 3\}$ time periods. Figure A.2 is a similar plot for the interventions taking place over $M = 1, 3$, and 7 (rows) time periods, and observed time series of length $T = 200, 400, 500$ (columns). These plots show the median and interquartile range of the asymptotic standard deviation, true bound, and estimated bound over 200

simulated data sets.

We begin by focusing on low uncertainty scenarios, corresponding to the interventions taking place over $M = 1$ or 3 time periods with the distribution resembling the actual data generating mechanism. We think that the intervention distribution resembles the data generating mechanism in scenarios where the intervention intensity is close to 5, which is the average number of treatment-active locations for the data generating process. In these scenarios, the asymptotic variance bound is distinctly higher than the true asymptotic variance, indicating that the inference based on the true asymptotic bound would be conservative. We find that in these low uncertainty scenarios, the estimated bound is close to the true bound. For that reason, we would expect the confidence intervals for the IPW estimator based on the estimated bound to have a higher coverage probability than its nominal coverage (see Appendix E.2 for the coverage results).

In contrast, under high uncertainty scenarios such as the interventions over longer time periods, e.g., $M = 7$, the asymptotic standard deviation and theoretical bound are essentially indistinguishable. However, under these scenarios, the estimate of the theoretical bound tends to be biased downwards, suggesting that the confidence intervals for the IPW estimator based on the estimated bound would be anti-conservative. Furthermore, we expect it to take a longer time series in order for the estimated bound to converge to its theoretical value when the intervention takes place over a longer time period.

E.2 Coverage of the Confidence Intervals for the IPW and Hájek Estimators

IPW estimator. The results in Figure 5 indicate that the coverage of confidence intervals based on the asymptotic variance bound should be similar to those based on the true variance under high uncertainty scenarios, while they should be slightly higher under low uncertainty scenarios. Furthermore, confidence intervals based on the estimated variance bound should yield coverage probability close to (lower than) the coverage achieved using the theoretical bound under low (high) uncertainty scenarios.

These expectations are indeed reflected in the coverage results shown in Figure A.3. Except when $M = 30$, the confidence interval for the IPW estimator based on either the *true* asymptotic variance or the *true* variance bound has a coverage of about 80% or higher. However, when $M = 30$, the confidence intervals based on the true asymptotic variance have a coverage below 60% or less, indicating that for interventions taking place over longer time periods, more data are needed to make use of the asymptotic approximation. However, these results are based on the true variance and variance bound, and instead inference would be based on the estimated variance bound. The under-estimation of the variance bound in high uncertainty scenarios found in Figure A.2 leads to the under-coverage of the confidence intervals based on the IPW estimator when using the estimated variance bound, especially when the interventions take place over long time periods.

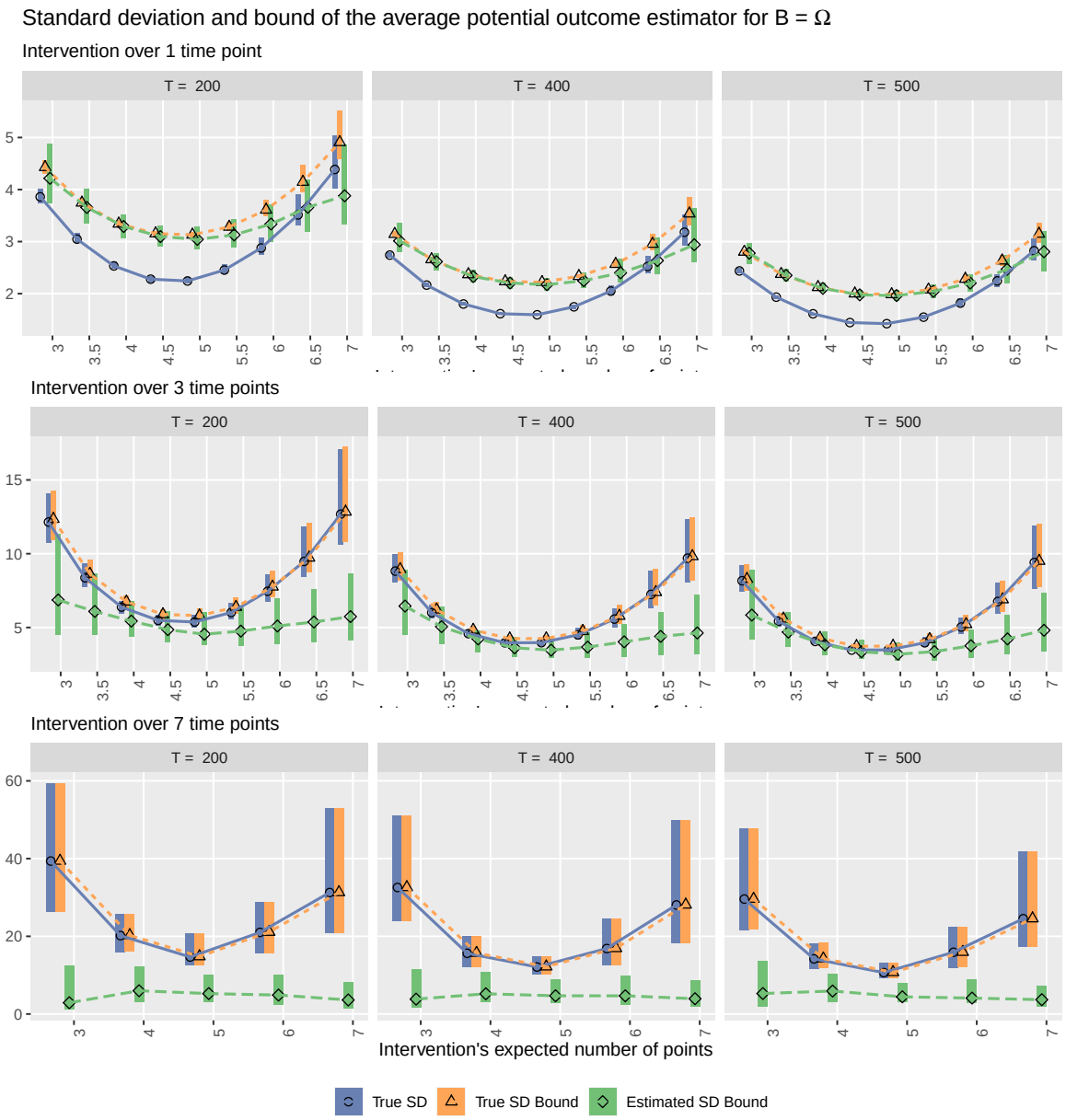


Figure A.2: Asymptotic Standard Deviation and Bound, and Estimated Bound. This figure shows the true asymptotic standard deviation (blue circles), the true asymptotic bound (orange triangles), and the estimated bound (green rhombuses) of the IPW estimator for the average potential outcome using the true propensity score, under interventions that take place over $M = 1, 3$ and 7 time periods (rows), and for increasing length of the time series (columns). The horizontal axis shows the intensity of the intervention at each time period. The points show the median value, and the rectangles show the interquartile range over 200 simulated data sets.

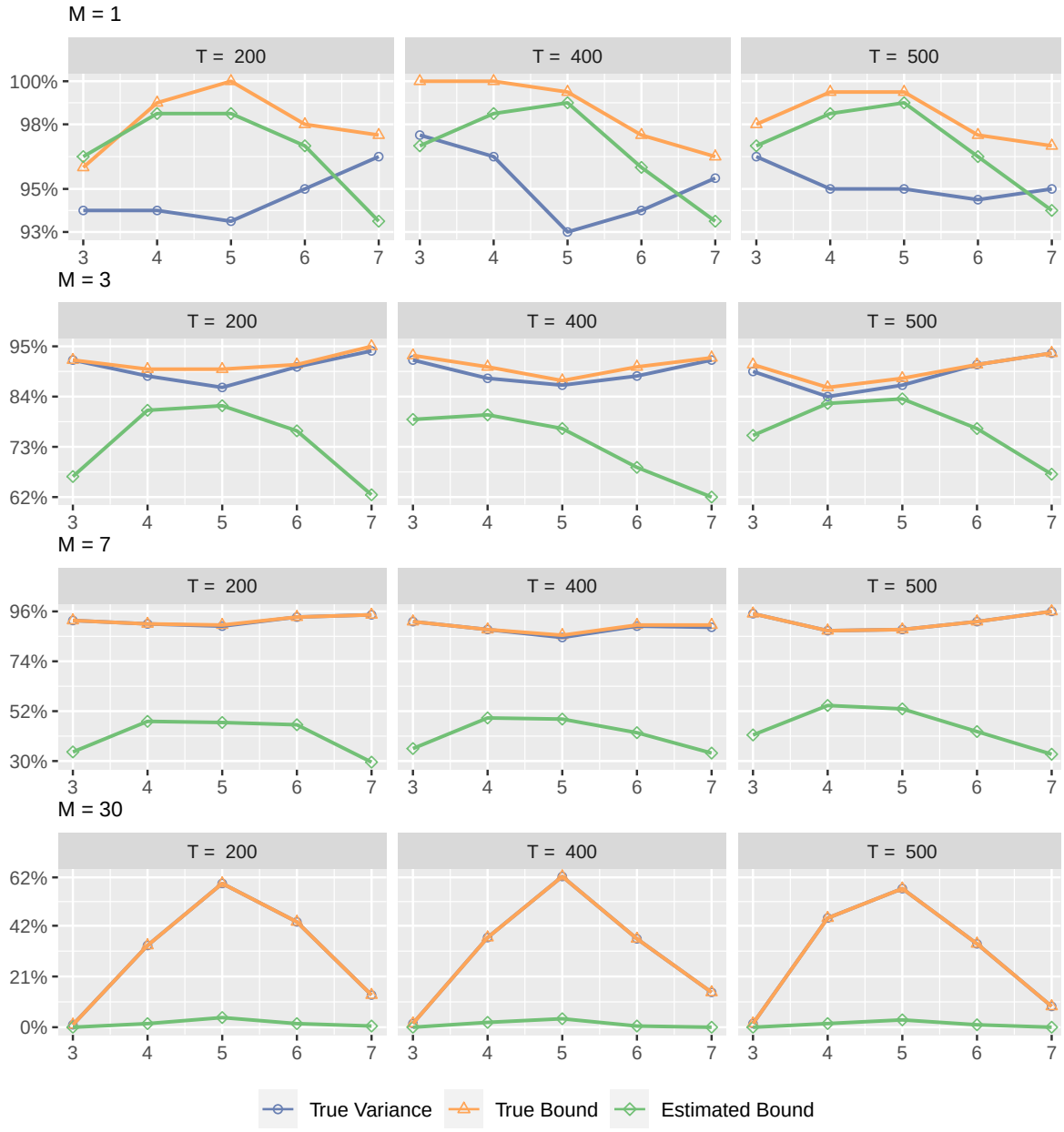


Figure A.3: Coverage of the IPW Estimator 95% Confidence Intervals. This figure shows the coverage of 95% confidence intervals for the average potential outcome over $B = \Omega$ based on the IPW estimator using the true variance (blue lines open circles), the true bound (orange lines with triangles), and the estimated bound (green lines with rhombuses), for interventions taking place over $M \in \{1, 3, 7, 30\}$ time periods (rows) and increasing length of the observed time series (columns).

Hájek estimator. Motivated by the good performance of the Hájek estimator shown in Figure 4, we also investigate the coverage probability of the 95% confidence interval as described in Appendix C. The rows of Figure A.4 show the coverage results for increasingly small regions, whereas the columns show the results for increasingly long observed time series ($T = 200, 400, 500$). Different colors correspond to the coverage results under interventions taking place over $M = 1$ (black), 3 (green), 7 (red), and 30

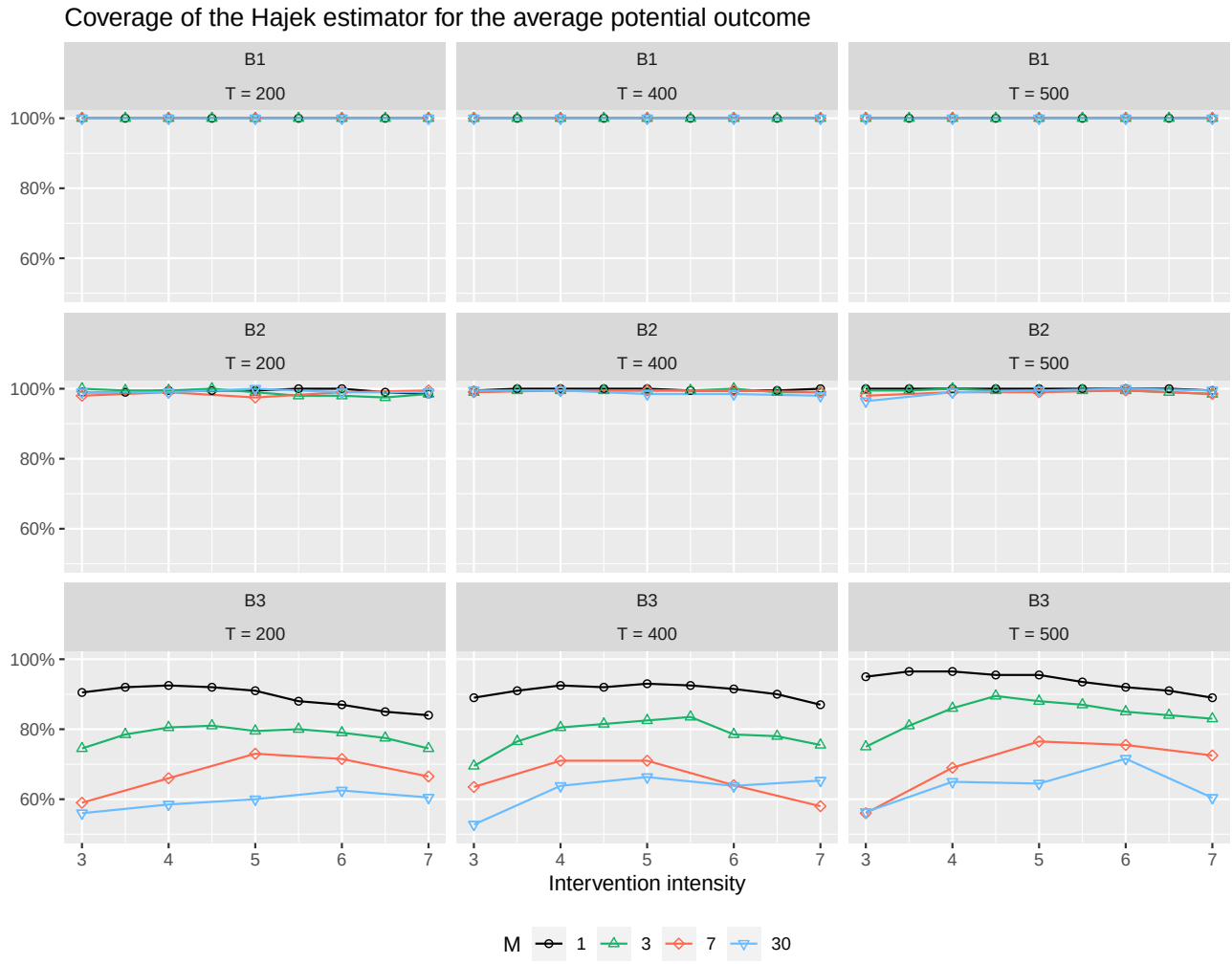


Figure A.4: Coverage of the Hájek Estimator's 95% Confidence Intervals for the Average Potential Outcomes under Various Interventions. We vary the intervention intensity h (horizontal axis), and the length of intervention $M = 1, 3, 7, 30$ (different lines). Each row represents the coverage for different regions of interest, i.e., $B_1 = [0, 1]^2$, $B_2 = [0, 0.5]^2$ and $B_3 = [0.75, 1]^2$, whereas each column represents the length of time series, i.e., $T = 200, 400$ and 500.

(blue) time periods. We find that the coverage is above 90% for all combinations of T and M for the two largest regions, even when an intervention takes place over 30 time periods. We find that the coverage is lower for the smallest region.

E.3 Uncertainty Estimates

Here, we show that the estimated standard deviation for the Hájek estimator outperforms that for the IPW estimator under many simulation scenarios.

We compute the standard deviation of the estimated average potential outcome across simulated data sets and compare it with the mean of the standard deviations, each of which is used to create the confidence intervals. The similarity of these two quantities implies the accuracy of our uncertainty

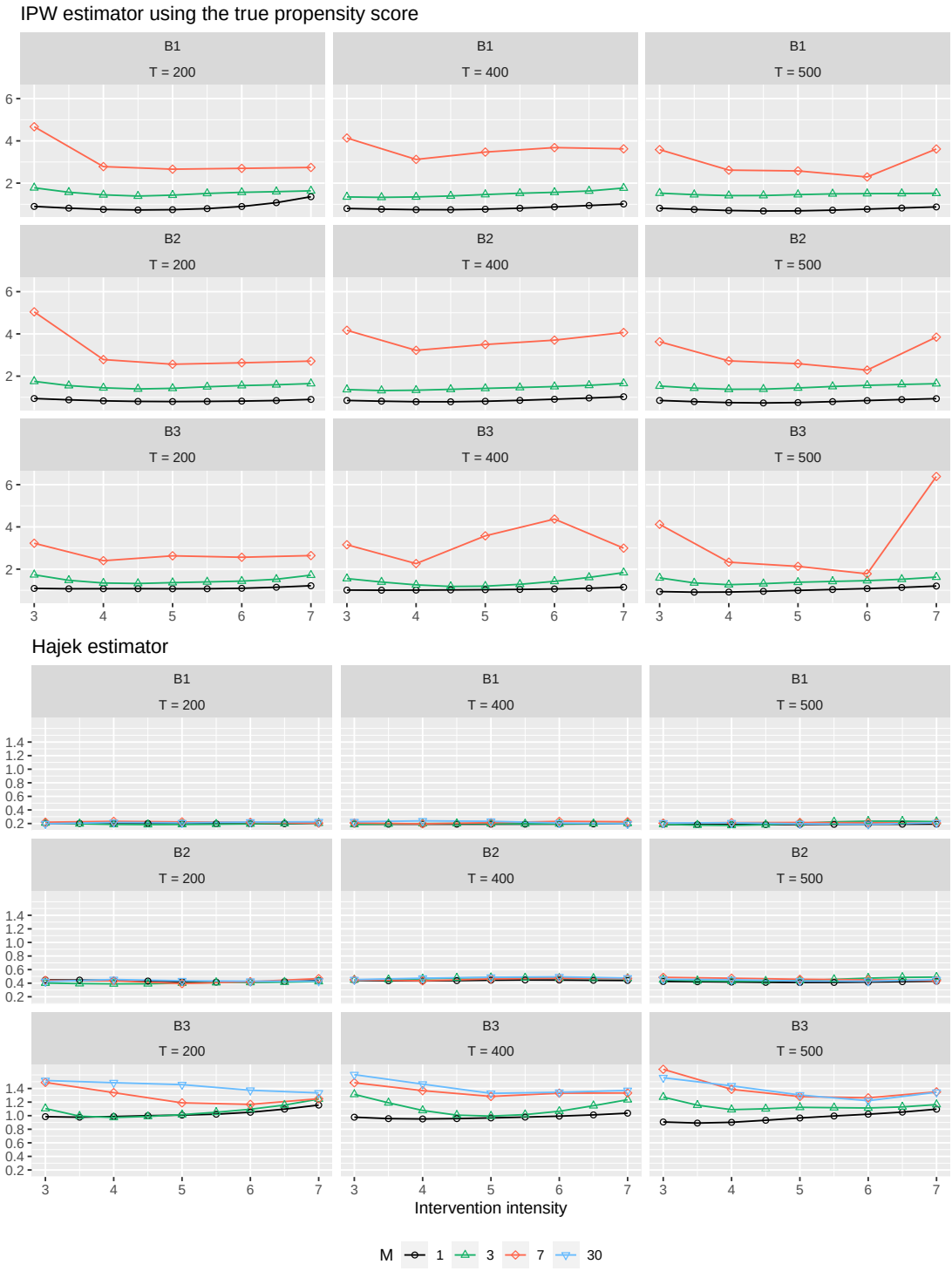


Figure A.5: Comparison of the Estimated and True Uncertainty for the Inverse Probability of Treatment and Hájek Estimators. Each plot presents the ratios between the standard deviation of each estimator and the mean estimated standard deviation across simulated data sets. A value smaller (greater) than 1 implies overestimation (underestimation) of uncertainty. The top (bottom) panel presents the results for the IPW (Hájek) estimator with the varying intensity under the intervention (horizontal axis) and for the whole country B_1 (first and forth row) and two sub-regions, B_2 (second and fifth row) and B_3 (third and sixth row). We also vary the length of intervention, $M = 1, 3, 7$ and 30 time periods (black, green, red, and blue lines, respectively). The columns correspond to different lengths of the time series $T = 200, 400$ and 500.

estimates. Figure A.5 presents the results as the ratio of these two quantities. A value below (above) 1 indicates that the true variability in our point estimates is smaller (greater) than our uncertainty estimate.

While the ratios are always below 1 for the Hájek estimator for the two largest regions B_1 and B_2, they are almost always above 1 for the IPW estimator (top panel). This is consistent with the above results, showing that we tend to overestimate (underestimate) the uncertainty for the Hájek (IPW) estimator. We find that the confidence interval for the Hájek estimator tends to be most conservative when M is small and the region of interest is large. For the IPW estimator, the degree of uncertainty underestimation decreases as the length of time series T increases but increases as the length of intervention M increases. In fact, when $M = 30$, some of the ratios are as large as 20 (hence they are not included in the figure). The results suggest that in practice the Hájek estimator should be preferred over the IPW estimator especially for stochastic interventions over a long time period.

E.4 Covariate Balance

We evaluate the balance of covariates based on the estimated propensity score by comparing their p-values in the propensity score model, and in a model with functional form as in the propensity score model but weighted by the inverse of the estimated propensity score. The left plot of Figure A.6 shows the p-value for the previous outcome-active locations, which are one of the time-varying confounders, across 200 simulated data sets. Evidently, the p-values in the unweighted model are close to 0, indicating that previous outcome-active locations form an important predictor of the treatment assignment. However, in the weighted model, the p-values of the same confounder are more evenly distributed across the $(0, 1)$ range, indicating that this confounder is better balanced in the weighted time series.

F Additional simulations on a square geometry

F.1 The Simulation Design

We also consider a time series of point patterns of length $T \in \{200, 400, 500\}$ on the unit square, $\Omega = [0, 1] \times [0, 1]$. For each time series length T, 200 data sets are generated with the following design.

Time-varying and time-invariant confounders. Our simulation study includes two time-invariant and two time-varying confounders. For the first time-invariant confounder, we construct a hypothetical road network on Ω using lines and arcs, which is highlighted by bright white lines in Figure A.7a. Then, we define $X^1(\omega) = 1.2 \exp\{-2D_1(\omega)\}$ where $D_1(\omega)$ is the distance from ω to the closest line. The second time-invariant covariate is constructed similarly, as $X^2(\omega) = \exp\{-3D_2(\omega)\}$ where $D_2(\omega)$ is the distance to the closest arc. In addition, the time-varying confounders, $X_t^3(\omega)$ and $X_t^4(\omega)$, are defined based on the exponential decay of distance to the closest point; these points are generated according to a

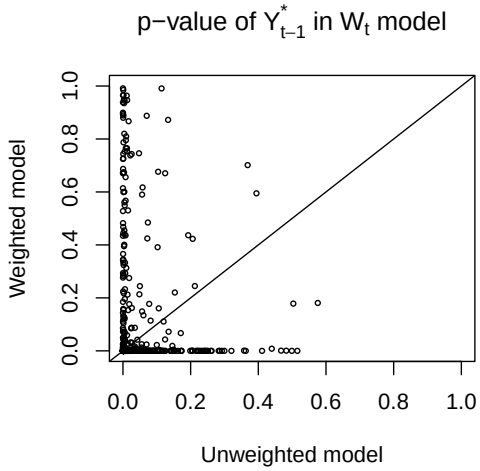


Figure A.6: Balance of the Previous Outcome-Active Locations in Treatment Model. Each point shows the relative magnitude of the p-value for the previous outcome-active locations in the unweighted propensity score model (horizontal axis) over that of the model weighted by the inverse of the estimated propensity score (vertical axis).

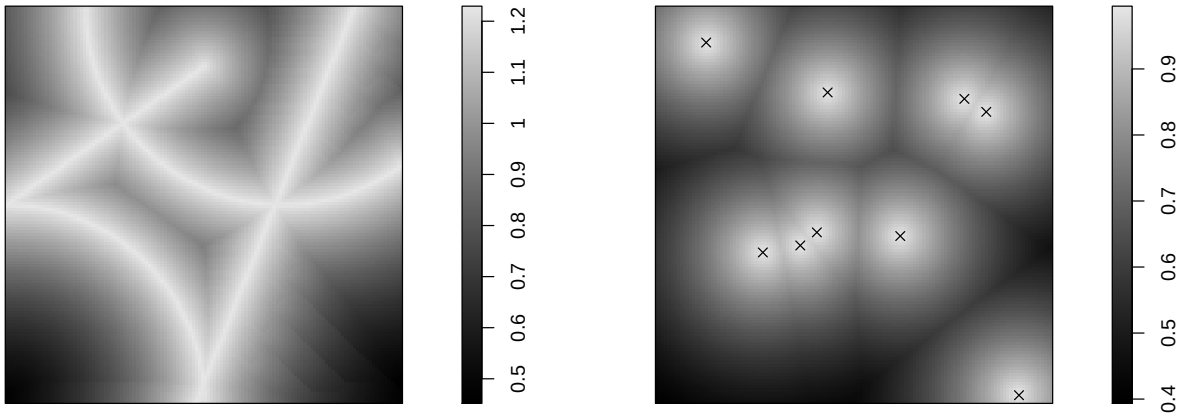
non-homogeneous Poisson point processes with the following intensity function

$$\lambda_t^{X^j}(\omega) = \exp\left\{\rho_0^j + \rho_1^j X^1(\omega)\right\},\ j = 3, 4,$$

where $\rho_1^3 = 1$, and $\rho_1^4 = 1.5$. Figure A.7b shows one realization of $X_t^3(\omega)$.

Spatio-temporal point processes for treatment and outcome variables. We again generate treatment and outcome point patterns from non-homogeneous Poisson processes that depends on all confounders, and the previous treatment and outcome realizations. The functional specification of the Poisson process intensities is the same as in Section 6. The model gives rise to an average of 5 observed treatment-active locations and 21 observed outcome-active locations within each time period.

Stochastic interventions. We consider interventions of the form F_h^M based on a homogeneous Poisson process with intensity h that is constant over Ω and ranges from 3 to 7. We consider various lengths of each intervention by setting $M \in \{1, 3, 7, 30\}$. The second intervention we consider is defined over the three time periods, i.e., $F_{\mathbf{h}} = F_{h_3} \times F_{h_2} \times F_{h_1}$ with $M = 3$. The intervention for the first time period F_{h_3} is a homogeneous Poisson process with intensity h_3 ranging from 3 to 7, whereas $F_{h_2} = F_{h_1}$ is a homogeneous Poisson process with intensity equal to 5 everywhere over Ω. For each stochastic intervention, we consider the region of interest, denoted by set B, of three different sizes: $B = \Omega = [0, 1] \times [0, 1]$, $B = [0, 0.5] \times [0, 0.5]$, and $B = [0.75, 1] \times [0.75, 1]$.



(a) Time-invariant confounder $X^1(\omega)$

(b) Realization of time-varying confounder $X_t^3(\omega)$

Figure A.7: Simulated Confounders. Panel (a) shows one of the two time-invariant confounders representing the exponential decay of distance to the road network. Panel (b) shows one realization for one of the time-varying confounders. Points $\times$ are generated from a non-homogeneous Poisson process depending on the road network in (a). Then, the time-varying confounder is defined as the exponential decay of distance to the points $\times$.

Estimand and estimation. Approximating the true values of the estimands and estimation is performed as described in Section 6. In these simulations, for $T = 500$ (the longest time series in our simulation scenario) the spatial smoothing bandwidth is approximately equal to 0.16, smaller than the size of the smallest B (which is equal to $[0.75, 1]^2$).

Variance and its upper bound. We base calculation of the theoretical variance and the variance bound on Theorems 1 and A.1, and use Monte Carlo approximations to compute these, as in Section 6. We also use Lemma 1 We use Lemma 1 to estimate the variance bound.

Covariate balance. As in Section 6, we use weighted regression by the estimated propensity score to investigate covariate balance.

F.2 Simulation Results

Estimation. Figures A.8 and A.9 present the results. In Figure A.8, the top panel shows how the (true and estimated) average potential outcomes in the whole region ($B = \Omega$) change as the intensity varies under the single time period interventions. The bottom panel shows how the true and estimated average potential outcomes in the sub-region $[0.75, 1]^2$ change under the three time period interventions when the intensity at three time periods ago ranges from 3 to 7. For both simulation scenarios, we vary the length of the time series from 200 (left plots) to 500 (right plots).

As expected, the unadjusted estimates (green crosses) are far from the true average potential outcome (black solid circles) across all simulation scenarios. In contrast, and consistent with the results of

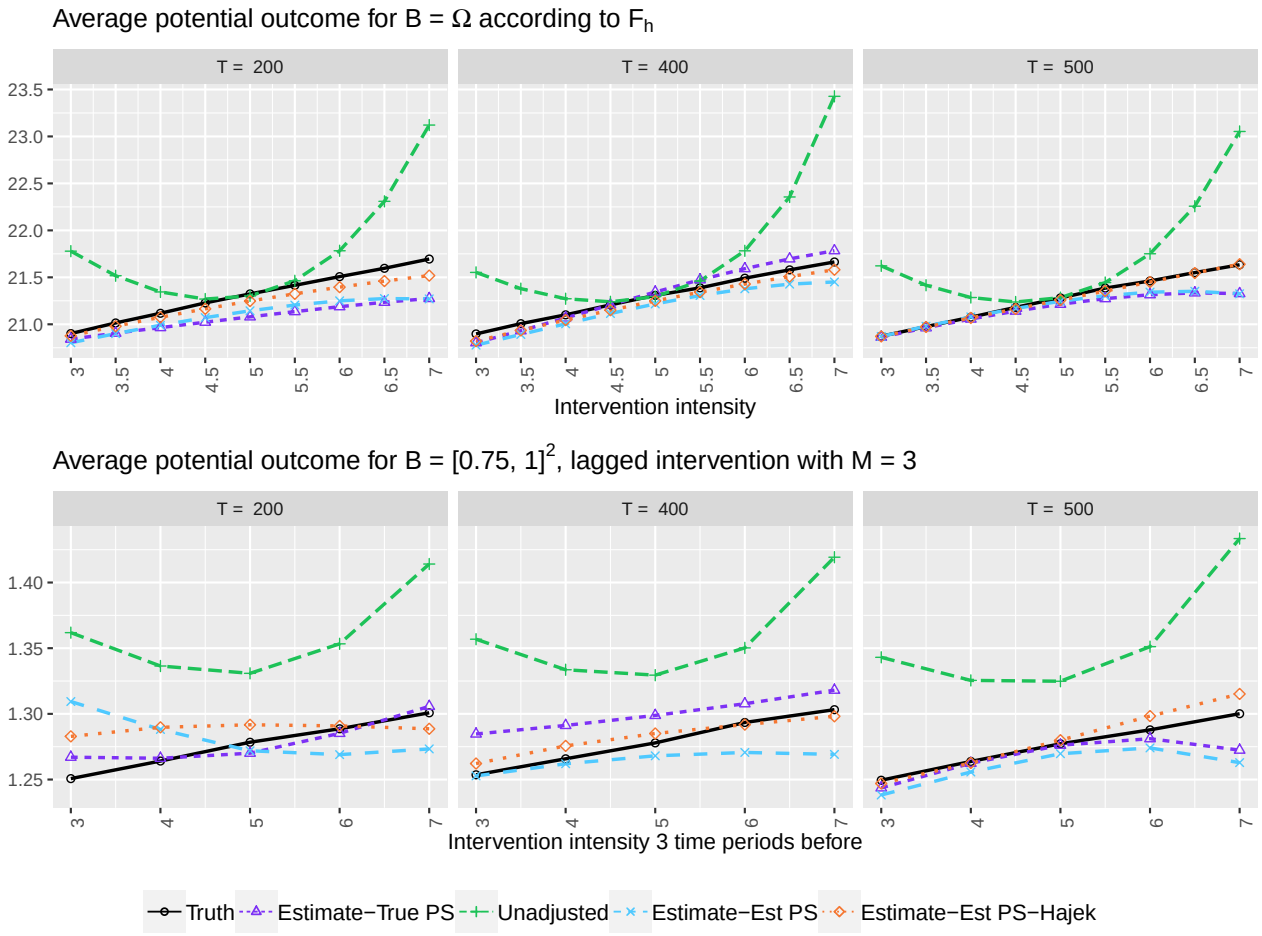


Figure A.8: Simulation Results for the True and Estimated Average Potential Outcomes. In the top panel, we present the true and estimated average potential outcomes in the entire region $B = \Omega$ under single-time interventions with the varying intensity (horizontal axis). In the bottom panel, we consider the average potential outcome in the sub-region $B = [0.75, 1]^2$ for the intervention $F_{\mathbf{h}}$, with $M = 3$, the varying intensity of F_{h_3} (horizontal axis), and F_{h_1}, F_{h_2} intensity set to 5. The black lines with solid circles represent the truths, while the other dotted or dashed lines represent the estimates; the estimator based on the true propensity score (purple triangles), the unadjusted estimator (green crosses), the estimator based on the estimated propensity score (blue x's), the Hájek estimator based on the estimated propensity score (orange rhombuses).

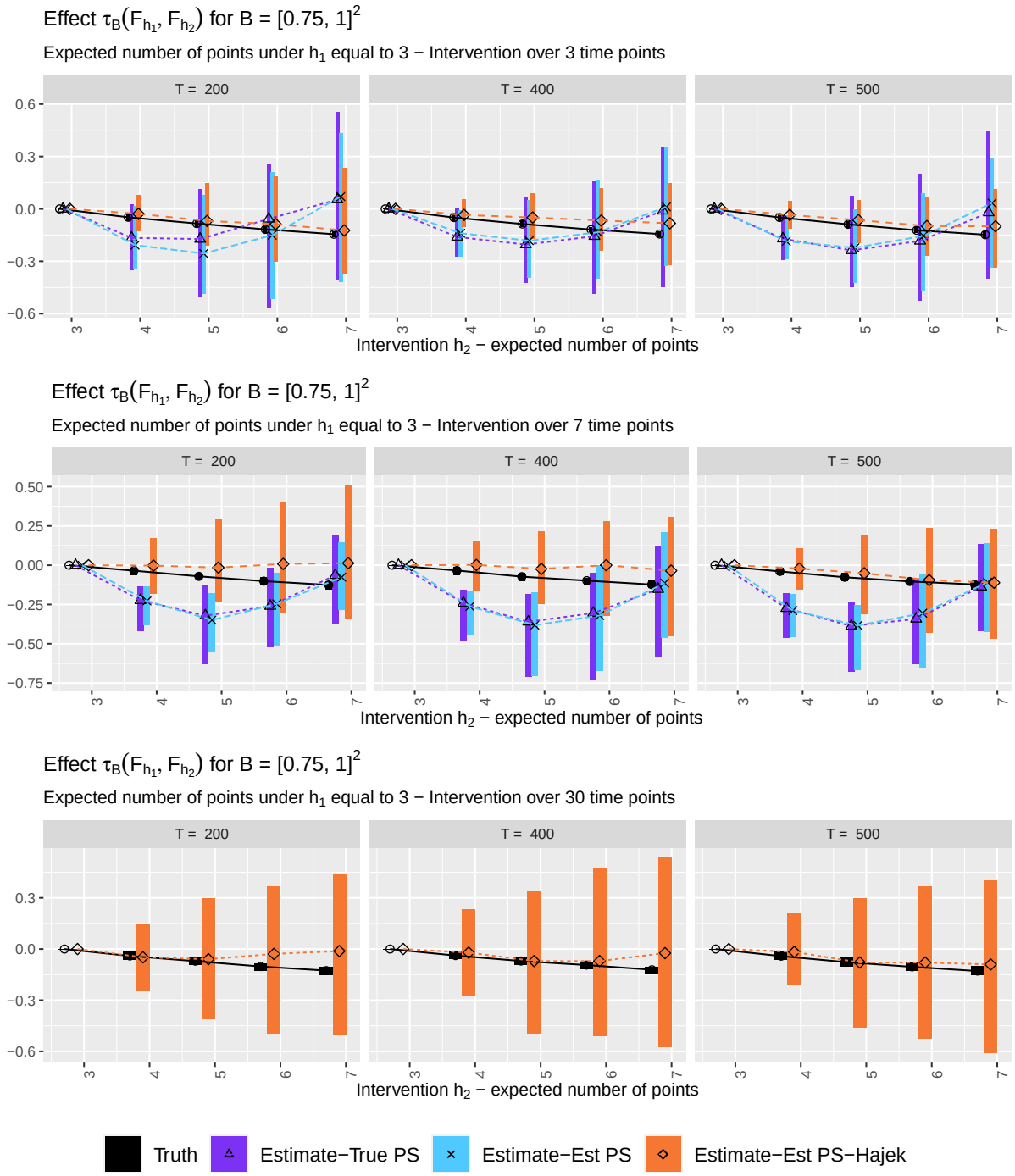


Figure A.9: Simulation Results for the Interventions of Increasing Time Lengths. Rows correspond to the interventions taking place over $M = 3, 7$, and 30 time periods. Columns correspond to the increasing length of the time series from 200 (left plots) to 500 (right plots). The vertical axis shows the change in the expected number of the outcome active locations over $[0.75, 1]^2$ for a change in the intervention intensity from 3 under h_1 to the value shown in the horizontal axis under h_2, for M time periods. The points in the plot show the median estimate over 200 data sets, and the rectangles show the interquartile range of estimates. Only the Hájek estimates are shown for $M = 30$ as the extremely small weights arising from a large number of time periods make the estimates from the other estimators close to zero.

Theorems 1 and A.1, the accuracy of the proposed estimator (purple triangles based on the true propensity score, blue x's based on the estimated propensity score) improves as the number of time periods increases. We note that the convergence is slower when $M = 3$ than $M = 1$.

Figure A.9 shows the performance of the estimators for the interventions over many time periods. The plots show the estimated change in the number of outcome-active locations over the sub-region $B = [0.75, 1]$ for a change in the stochastic intervention from 3 per time period to the value on the horizontal axis. The rows correspond to the interventions over $M = 3, 7$, and 30 time periods, respectively, whereas the columns represent the different lengths of time series, i.e., $T = 200, 400$ and 500. The results are shown for the IPW estimators based on the true propensity score (purple lines with open triangles) and the estimated propensity score (blue lines with x's) as well as the Hájek estimator based on the estimated propensity score (orange lines with open rhombuses). Only the Hájek estimates are shown for $M = 30$ as the extremely small weights arising from a large number of time periods make the estimates from the other estimators essentially equal to zero. The lines and points in the plot show the median estimate and the rectangles show the interquartile range of estimates across 200 simulated data sets.

Again, as in the simulations of Section 6, we find that the Hájek estimator performs well across all simulation scenarios, whereas the IPW estimator tends to suffer from extreme weights.

The variance and its bound. Next, we compare the true theoretical variance, v/T, with the variance bound v^*/T and its consistent estimator (see Lemma 1). We again focus on the proposed estimators with the true propensity score. Figure A.10 shows the results of an intervention F_h^M for $M = 1, 3$ and 7, for region $B = [0, 0.5]^2$, and observed time series of length $T = 200, 400, 500$. These plots show the median and interquartile range of the asymptotic standard deviation, true bound, and estimated bound over 200 simulated data sets.

The conclusions are similar to the main manuscript. As expected, the true variance decreases as the total number of time periods increases. We start by focusing on low uncertainty scenarios, corresponding to the interventions taking place over $M = 1$ or 3 time periods with the distribution resembling the actual data generating mechanism. We think that the intervention distribution resembles the data generating mechanism in scenarios where the intervention intensity is close to 5, which is the average number of treatment-active locations for the data generating process. In these scenarios, the asymptotic variance bound is distinctly higher than the true asymptotic variance, indicating that the inference based on the true asymptotic bound would be conservative. We find that in these low uncertainty scenarios, the estimated bound is close to the true bound. For that reason, we would expect the confidence intervals for the IPW estimator based on the estimated bound to have a higher coverage probability than its nominal coverage.

In contrast, under high uncertainty scenarios such as the interventions over longer time periods, e.g., $M = 7$, the asymptotic standard deviation and theoretical bound are essentially indistinguishable. However, under these scenarios, the estimate of the theoretical bound tends to be biased downwards,

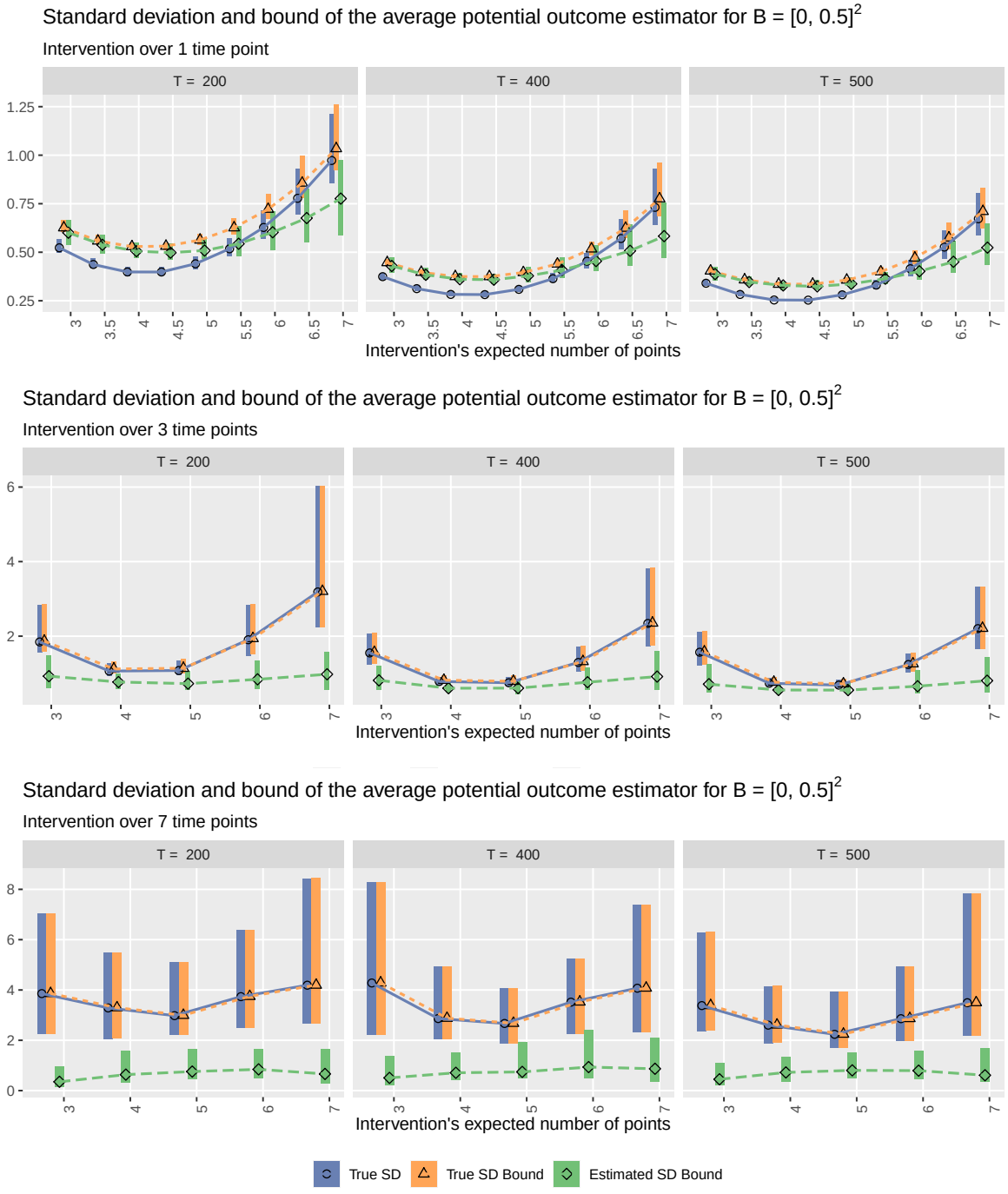


Figure A.10: Asymptotic Standard Deviation and Bound, and Estimated Bound. This figure shows the true asymptotic standard deviation (blue circles), the true asymptotic bound (orange triangles), and the estimated bound (green rhombuses) of the IPW estimator for the average potential outcome using the true propensity score, under interventions that take place over $M = 1, 3$ and 7 time periods (rows), and for increasing length of the time series (columns). The horizontal axis shows the intensity of the intervention at each time period. The points show the median value, and the rectangles show the interquartile range over 200 simulated data sets.

suggesting that the confidence intervals for the IPW estimator based on the estimated bound would be anti-conservative. As the length of time series increases, the estimated variance bound more closely approximates its theoretical value (consistent with Lemma 1), but we expect it to take a longer time series in order for the estimated bound to converge to its theoretical value when the intervention takes place over a longer time period.

Coverage. These results on the asymptotic variance and variance bound lead to similar conclusions with respect to the coverage of 95% confidence intervals of the IPW estimator.

The coverage results are shown in Figure A.11. We find that, except when $M = 30$, the confidence interval for the IPW estimator based on either the true asymptotic variance or the true variance bound has a coverage of about 80% or higher. This implies that the asymptotic normality established in Theorem 1 provides an adequate approximation to the estimator's sampling distribution for small or moderate values of M. However, for $M = 30$, the confidence interval for the IPW estimator is anti-conservative due to the fact that the weights, which equal the product of ratios across many time periods, become extremely small. In addition, the underestimation of the variance bound in high uncertainty scenarios found in Figure A.10 leads to the under-coverage of the confidence intervals based on the IPW estimator and using the estimated variance bound, especially when the interventions take place over long time periods.

We also investigate the coverage probability of the 95% confidence interval for the Hájek estimator. The rows of Figure A.12 show the coverage results for increasingly small regions, i.e., $B_1 = [0, 1]^2$, $B_2 = [0, 0.5]^2$, and $B_3 = [0.75, 1]^2$, whereas the columns show the results for increasingly long observed time series ($T = 200, 400, 500$). Different colors correspond to the coverage results under interventions taking place over $M = 1$ (black), 3 (green), 7 (red), and 30 (blue) time periods. We find that the coverage is above 85% for all cases, even when an intervention takes place over 30 time periods. As expected, the coverage is higher for smaller values of M, since these correspond to lower-uncertainty situations. We also find that the coverage is lower for smaller regions.

Comparison of Monte Carlo and estimated variance We find that the confidence interval for the Hájek estimator has a better coverage probability even for the interventions over long time periods. Here, we show that the estimated standard deviation for the Hájek estimator outperforms that for the IPW estimator under many simulation scenarios.

Figure A.13 shows the ratio of the standard deviation of the estimated average potential outcome across simulated data sets over the mean of the standard deviations. A value below (above) 1 indicates that the true variability in our point estimates is smaller (greater) than our uncertainty estimate. While the ratios are always below 1 for the Hájek estimator (bottom panel), they are almost always above 1 for the IPW estimator (top panel). This shows that we tend to overestimate (underestimate) the uncertainty for the Hájek (IPW) estimator. Further, we find that the confidence interval for the Hájek estimator tends to be most conservative when M is small and the region of interest is large. For the IPW estimator, the

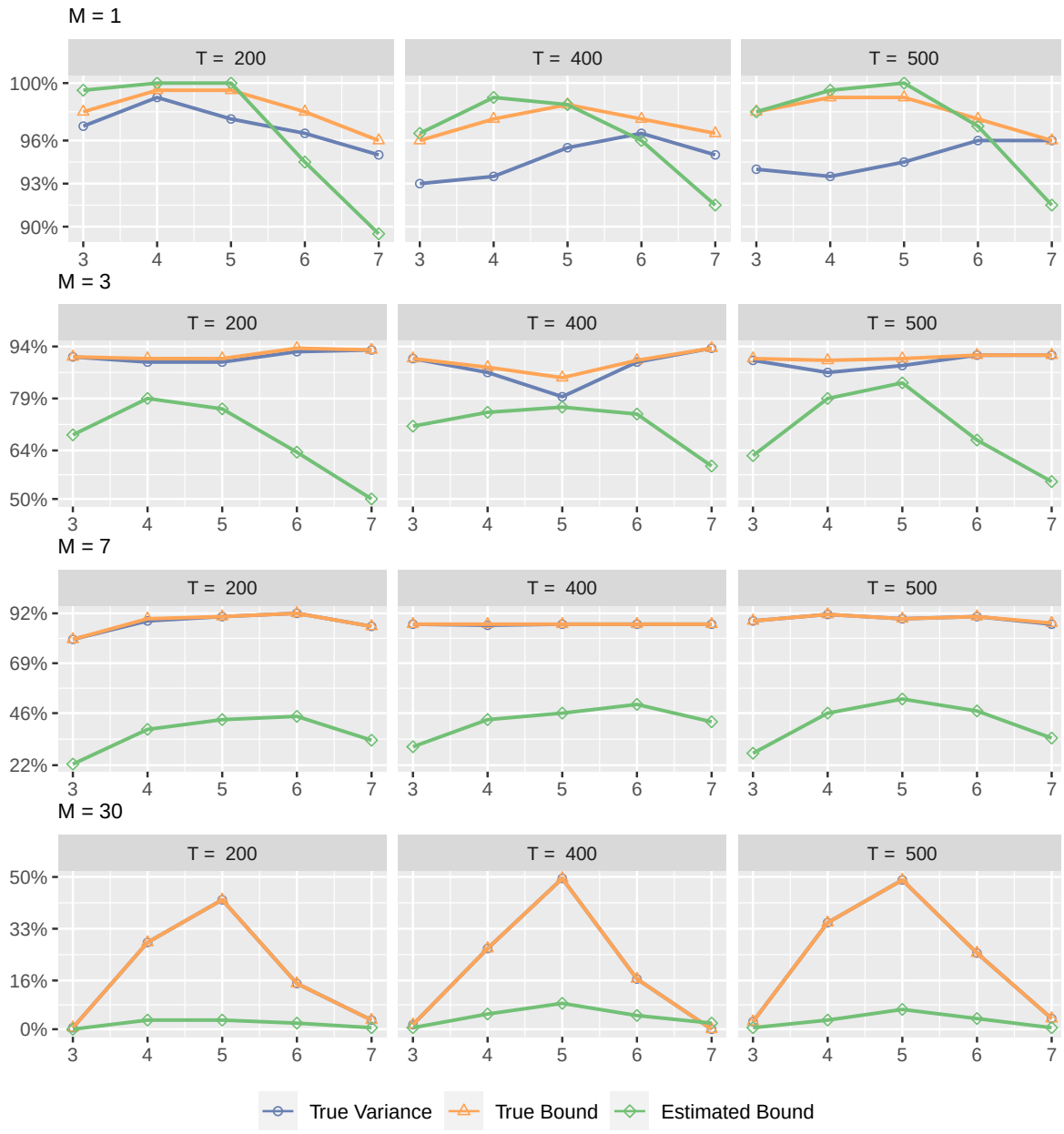


Figure A.11: Coverage of the IPW Estimator 95% Confidence Intervals. This figure shows the coverage of 95% confidence intervals for the average potential outcome over $B = \Omega$ based on the IPW estimator using the true variance (blue lines open circles), the true bound (orange lines with triangles), and the estimated bound (green lines with rhombuses), for interventions taking place over $M \in \{1, 3, 7, 30\}$ time periods (rows) and increasing length of the observed time series (columns).

degree of underestimation decreases as the length of time series T increases but increases as the length of intervention M increases. In fact, when $M = 30$, some of the ratios are as large as 20 (hence they are not included in the figure). The results suggest that in practice the Hájek estimator should be preferred over the IPW estimator especially for stochastic interventions over a long time period.

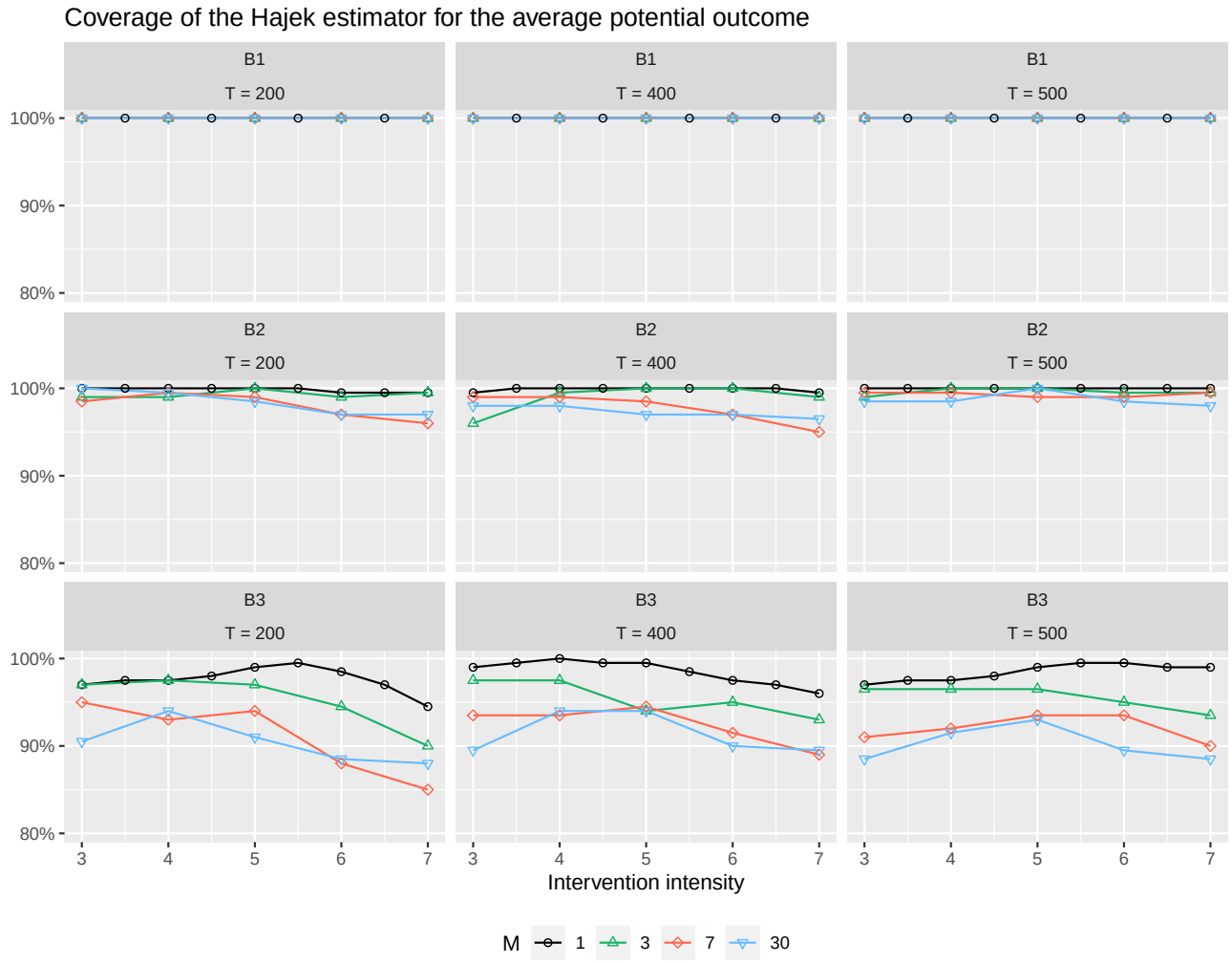


Figure A.12: Coverage of the Hájek Estimator's 95% Confidence Intervals for the Average Potential Outcomes under Various Interventions. We vary the intervention intensity h (horizontal axis), and the length of intervention $M = 1, 3, 7, 30$ (different lines). Each row represents the coverage for different regions of interest, i.e., $B_1 = [0, 1]^2$, $B_2 = [0, 0.5]^2$ and $B_3 = [0.75, 1]^2$, whereas each column represents the length of time series, i.e., $T = 200, 400$ and 500.

Balance. As in the simulations in the main manuscript, we find that the p-values of one of the confounders (Y^*_{t-1} in Equation(12)) are substantially greater in the weighted propensity score model than in the unweighted model, where the weights are given by the inverse of the estimated propensity score (shown in Figure A.14).

G Additional Empirical Results

G.1 Visualization

As discussed in Section 7.1, we consider a stochastic intervention whose focal point is the center of Baghdad. The degree of concentration is controlled by the precision parameter α whose greater value,

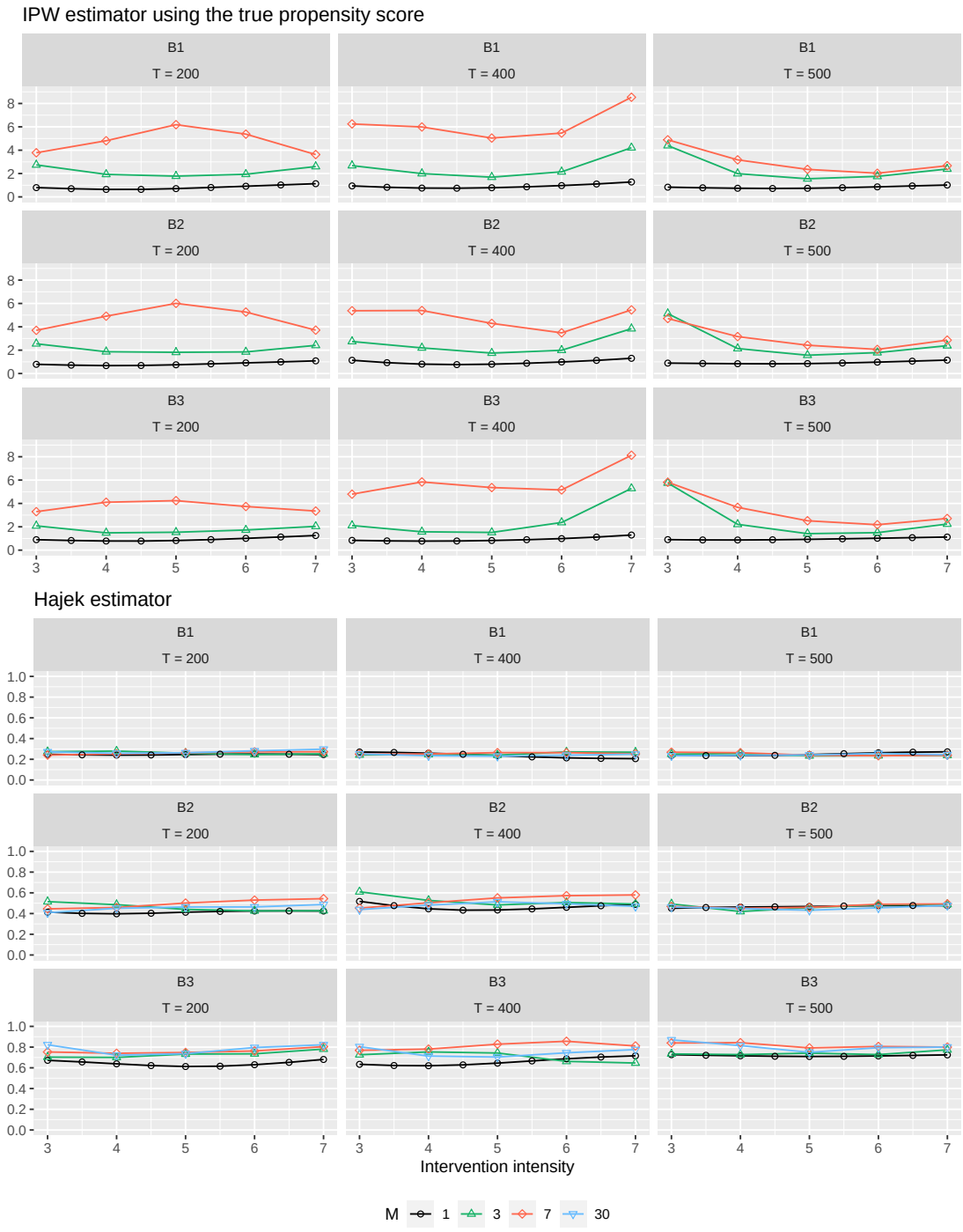


Figure A.13: Comparison of the Estimated and True Uncertainty for the Inverse Probability of Treatment and Hájek Estimators. Each plot presents the ratios between the standard deviation of each estimator and the mean estimated standard deviation across simulated data sets. A value smaller (greater) than 1 implies overestimation (underestimation) of uncertainty. The top (bottom) panel presents the results for the IPW (Hájek) estimator with the varying intensity under the intervention (horizontal axis) and for the whole region B_1 (first and forth row) and two sub-regions, $B_2 = [0, 0.5]^2$ (second and fifth row) and $B_3 = [0.75, 1]^2$ (third and sixth row). We also vary the length of intervention, $M = 1, 3, 7$ and 30 time periods (black, green, red, and blue lines, respectively). The columns correspond to different lengths of the time series $T = 200, 400$ and 500.

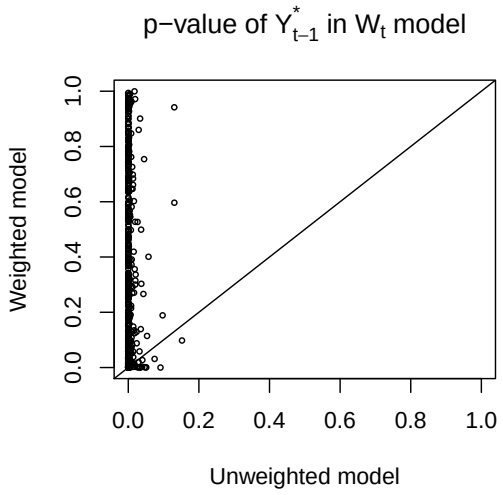


Figure A.14: Balance of the Previous Outcome-Active Locations in Treatment Model. In the left plot, each point shows the relative magnitude of the p-value for the previous outcome-active locations in the unweighted propensity score model (horizontal axis) over that of the model weighted by the inverse of the estimated propensity score (vertical axis). The right plot shows the distribution of the estimated coefficient of the previous outcome-active locations in the weighted propensity score model.

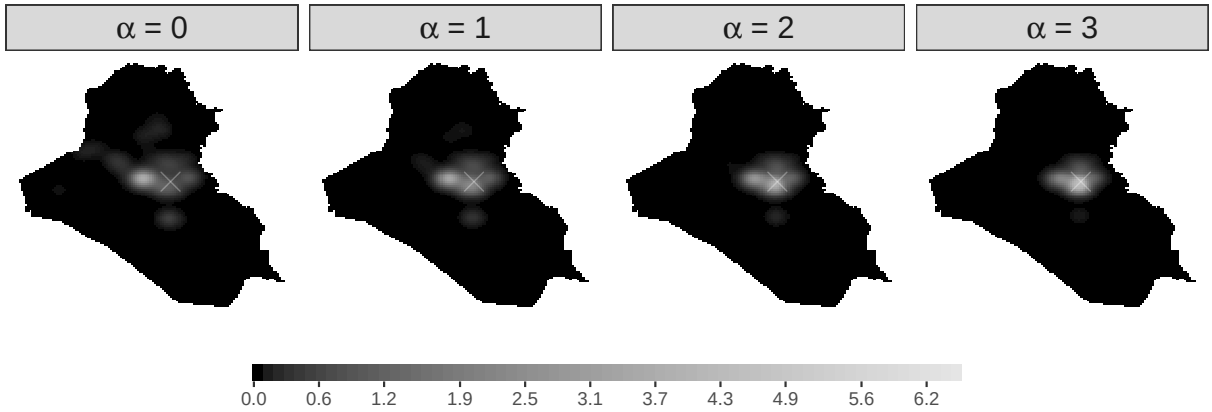


Figure A.15: Visualization of Intensity under Stochastic Interventions whose Focal Point is the Center of Baghdad. Across plots, we vary the degree to which the airstrikes are concentrated around the focal point using the precision parameter, while the expected number of airstrikes is held constant at 3 per day.

implying that more airstrikes are occurring near the focal point. We vary the value of α from 0 to 3, while keeping the expected number of airstrikes constant at 3 per day. Figure A.15 illustrates intensities for the different values of α. The first plot in the figure does not focus on Baghdad at all, representing the baseline spatial distribution ϕ_0. As the value of α increases, the spatial distribution of airstrikes becomes concentrated more towards the center of Baghdad.

Type ($F_{\mathbf{h}'}, F_{\mathbf{h}''}$)	M	Outcome	Iraq	Baghdad	Outside Baghdad
Increasing the intensity (1, 3)	3	IED	-1.3 (-8.2, 5.7)	-0.2 (-2.2, 1.8)	-1.1 (-6.2, 4)
		SAF	-1.9 (-8.9, 5.2)	-1.2 (-3.8, 1.4)	-0.7 (-5.2, 3.9)
		Other Attack	-1.8 (-19.9, 16.4)	0.1 (-6.2, 6.5)	-1.9 (-13.9, 10)
	7	IED	5 (-2.6, 12.7)	1.2 (-0.8, 3.3)	3.8 (-1.9, 9.5)
		SAF	**10 (1.7, 18.2)**	**3.1 (0.4, 5.8)**	**6.9 (1.1, 12.6)**
		Other Attack	14 (-5.2, 33.3)	5.6 (-0.4, 11.6)	8.4 (-5.1, 21.9)
	30	IED	11 (-1.1, 23)	3.2 (-0.1, 6.5)	7.8 (-1.1, 16.6)
		SAF	**14.8 (2.9, 26.7)**	**5.9 (1.2, 10.5)**	**8.9 (1.5, 16.3)**
		Other Attack	**33.1 (3.3, 62.9)**	**13.6 (2.7, 24.6)**	**19.5 (0.3, 38.6)**
Changing the focal points (0, 3)	3	IED	2.6 (-7.2, 12.5)	0.6 (-2.4, 3.5)	2.1 (-4.9, 9)
		SAF	2.6 (-6.7, 12)	0.9 (-2.2, 4.1)	1.7 (-4.8, 8.1)
		Other Attack	7.5 (-16, 31)	3.8 (-4.3, 11.8)	3.7 (-12, 19.4)
	7	IED	2 (-6.9, 10.8)	1.1 (-1.7, 3.9)	0.9 (-5.2, 7)
		SAF	0.2 (-9.7, 10.1)	0.6 (-2.8, 3.9)	-0.4 (-7.1, 6.4)
		Other Attack	3.5 (-18.8, 25.8)	1.7 (-6, 9.4)	1.8 (-13, 16.7)
	30	IED	-1.2 (-15.9, 13.4)	-0.7 (-4.5, 3.1)	-0.5 (-11.6, 10.6)
		SAF	5.7 (-10, 21.4)	-1.3 (-6.4, 3.8)	7 (-4.1, 18.1)
		Other Attack	-3.5 (-37.8, 30.7)	-6.6 (-17.5, 4.2)	3.1 (-21.1, 27.3)
Lagged effects (1, 5)	3	IED	-2.3 (-10.1, 5.5)	-0.6 (-2.7, 1.5)	-1.7 (-7.5, 4.1)
		SAF	-1 (-9.9, 8)	-0.7 (-4.5, 3)	-0.2 (-5.5, 5)
		Other Attack	-3.9 (-23.6, 15.8)	-1.2 (-8.2, 5.9)	-2.8 (-15.5, 10)
	7	IED	6.8 (-0.7, 14.3)	2.2 (-0.2, 4.6)	4.6 (-0.6, 9.8)
		SAF	**9.4 (1.6, 17.2)**	**3.6 (1, 6.2)**	**5.8 (0.4, 11.2)**
		Other Attack	**20.9 (2.3, 39.4)**	**8.2 (1.8, 14.6)**	**12.7 (0.4, 24.9)**
	30	IED	1.5 (-3.8, 6.8)	0.3 (-1, 1.5)	1.2 (-2.8, 5.3)
		SAF	2.8 (-1.8, 7.3)	1.1 (-0.6, 2.8)	1.6 (-1.2, 4.5)
		Other Attack	5.8 (-6.2, 17.8)	2.2 (-1.9, 6.4)	3.6 (-4.3, 11.4)

Table A.2: Causal Effect Estimates and 95% Confidence Intervals for Various Stochastic Interventions. We present the results for three interventions discussed in the main text: increasing the expected number of airstrikes from 1 to 3 per day for M days, changing the focal points of airstrikes from $\alpha = 0$ to $\alpha = 3$ for M days, and the lagged effects of increasing the expected number of airstrikes from 1 to 5 per day M days ago. The range of M we consider is $\{3, 7, 30\}$. The regions of interest are Iraq, Baghdad, and the area outside Baghdad. The results in bold represent statistically significant estimates.

G.2 Empirical Results

Table A.2 presents the numerical effect estimates and 95% confidence intervals for various interventions, including those shown in the main text. We also show the effect estimates for the whole Iraq, Baghdad only, and the area outside Baghdad.

G.3 Single time point adaptive interventions

Adaptive intervention strategies are often of interest in longitudinal settings, where previous outcomes might drive future treatment assignments. In our setting, these adaptive interventions would correspond to military strategies that depend on the observed history, such as the locations of previous insurgent at-

tacks. Although we leave full development of adaptive strategies to future research, we consider adaptive strategies that take place over a single time period, and then discuss the challenges of further extending it to the multiple time period interventions.

Here, we design adaptive dosage interventions over a single time period that closely resemble the observed data in terms of the expected number of airstrikes over time and their location. Using the observed number of airstrikes over time, we fitted a smooth function of time to obtain an estimate of the expected number of airstrikes over time, which is denoted by $\widehat{n}_t$. We used the estimated expected number of points to define adaptive interventions under which (1) the spatial distribution under the intervention is equal to the spatial distribution of airstrikes according to the propensity score, and (2) the expected number of points under the intervention is set to $c\widehat{n}_t$, with c varying from 0.5 to 2 (representing a change in the number of airstrikes ranging from half to double the observed values). Formally, this intervention that depend on the observed history is given by:

$$h_{t+1}(\omega; \overline{H}_t) = \frac{c\,\widehat{n}_t}{\int_\Omega h^{ps}_{t+1}(s; \overline{H}_t)\mathrm{d}s} h^{ps}_{t+1}(\omega; \overline{H}_t),$$

where $h^{ps}_{t+1}(\omega; \overline{H}_t)$ is the estimated propensity score intensity function. This definition of intensity ensures that that expected number of airstrikes at time t is equal to $c\widehat{n}_t$ using the ratio term, and the relative likelihood of each location ω being treated is as specified in the estimated propensity score. This approach is related to the incremental propensity score of Kennedy (2019) who considered non-spatial and non-temporal settings.

Figure A.16 shows the effect estimates for number of IED and SAF attacks in Iraq for these interventions. The result shows that the estimates are too imprecise to lead to a definitive conclusion.

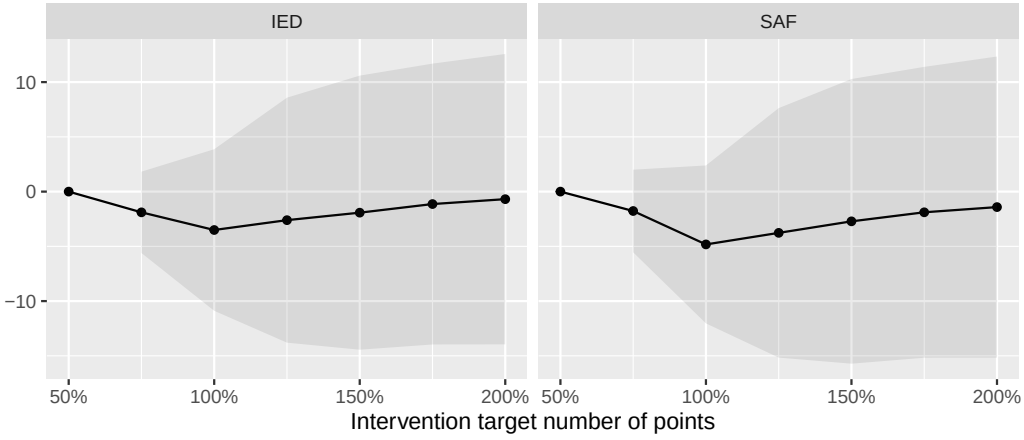


Figure A.16: Effect estimates for a change in the expected number of airstrikes from 50% $\widehat{n}_t$ to $c\widehat{n}_t$, for c shown in the x-axis. Left plot shows effect estimates for IEDs and right plot shows effect estimates for SAF attacks in Iraq.

Unfortunately, the evaluation of adaptive strategies over multiple time periods rapidly becomes complicated. Specifically, for interventions over multiple time periods that depend on the most recent history, we would need to have access to intermediate potential outcomes which are unobserved. Therefore, we would have to model the outcome process in order to predict the counterfactual outcomes that would then inform the adaptive treatment assignment in the subsequent time periods. One advantage of our proposed framework is its ability to incorporate unstructured spillover and carryover effects. This is possible because our framework does not require researchers to model the outcome process. Given this difficulty, we will leave the complete investigation of adaptive spatio-temporal treatment strategies to future work.

References

Charnes, A. and Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research logistics quarterly* **9**, 3-4, 181–186.

Chow, Y. S. (1965). Local convergence of martingales and the law of large numbers. *The Annals of Mathematical Statistics* **36**, 2, 552–558.

Crimaldi, I. and Pratelli, L. (2005). Convergence results for multivariate martingales. *Stochastic processes and their applications* **115**, 4, 571–577.

Csörgö, M. (1968). On the Strong Law of Large Numbers and the Central Limit Theorem for Martingales. *Transactions of the American Mathematical Society* **131**, 1, 259–275.

Kennedy, E. H. (2019). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association* **114**, 526, 645–656.

Küchler, U., Sørensen, M., *et al.* (1999). A note on limit theorems for multivariate martingales. *Bernoulli* **5**, 3, 483–493.

Papadogeorgou, G., Mealli, F., and Zigler, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics* **75**, 3, 778–787.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Stout, W. F. (1974). *Almost sure convergence*, vol. 24. Academic press.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge university press.

van der Vaart, A. W. (2010). Time Series. *VU University Amsterdam, lecture notes* .