Causal Inference under Temporal and Spatial Interference

(Appendix)

Appendix A Proofs

A1 Identification under sequential ignorability

First note that for any $\mathbf{z}^{(t-k):t}$, we have

$$
\begin{aligned}
&\mathrm{E}_{\mathbf{Z}^{1:t}}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\right]\\
=&\mathrm{E}\left[\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)}{\prod_{s=t-k}^{t}P(Z_{is}=z_{is}|\mathbf{V}_{is})}\middle|\mathbf{V}_{i,(t-k)}\right]\right]\\
=&\mathrm{E}\left[\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k+1):t}=\mathbf{z}^{(t-k+1):t}\right\}\mu_i\left(\left\{Y_{jt}\left(z_{(t-k)};\mathbf{Z}^{1:t}\setminus Z_{i,(t-k)}\right)\right\}_{j\in\mathcal{N}};d\right)}{\prod_{s=t-k+1}^{t}P(Z_{is}=z_{is}|\mathbf{V}_{is})}\middle|\mathbf{V}_{i,(t-k)},Z_{i,(t-k)}=z_{(t-k)}\right]\right]\\
=&\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k+1):t}=\mathbf{z}^{(t-k+1):t}\right\}\mu_i\left(\left\{Y_{jt}\left(z_{(t-k)};\mathbf{Z}^{1:t}\setminus Z_{i,(t-k)}\right)\right\}_{j\in\mathcal{N}};d\right)}{\prod_{s=t-k+1}^{t}P(Z_{is}=z_{is}|\mathbf{V}_{is})}\right]\\
=&\mathrm{E}\left[\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k+1):t}=\mathbf{z}^{(t-k+1):t}\right\}\mu_i\left(\left\{Y_{jt}\left(z_{(t-k)};\mathbf{Z}^{1:t}\setminus Z_{i,(t-k)}\right)\right\}_{j\in\mathcal{N}};d\right)}{\prod_{s=t-k+1}^{t}P(Z_{is}=z_{is}|\mathbf{V}_{is})}\middle|\mathbf{V}_{i,(t-k+1)}\right]\right]\\
=&\dots\\
=&\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\frac{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}\mathrm{E}\left[Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]}{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}}.
\end{aligned}
$$

The … part iterates the same step from period $t-k+1$ to t. The third equality uses Assumption 2 and the law of iterated expectation. Then,

$$\begin{aligned}
\tau_t\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d\right) =& \frac{1}{N}\sum_{i=1}^{N}\mu_i\left(\left\{\tau_{jt;i}\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}\tau_{jt;i}\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t}\right)}{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}} \\
=& \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}\mathrm{E}\left[Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]}{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}} - \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}\mathrm{E}\left[Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]}{\sum_{j=1}^{N}\mathbf{1}\{j\in\Omega_d\}} \\
=& \mathrm{E}_{\mathbf{Z}^{1:t}}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\right] - \mathrm{E}_{\mathbf{Z}^{1:t}}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right],
\end{aligned}$$

We have proved the claim in the main draft that the AME can be identified from data.

A2 Variance of the estimators

We first derive the variance of the Horvitz-Thompson estimators when the propensity scores are known. The extra uncertainty induced by estimating the nuisance parameters is discussed in Section A8. The result is stated in the following lemma:

Lemma 1. *Under Assumptions 1-3 in the main text, we have the following bound for the variance of the Horvitz-Thompson estimator $\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)$ when the propensity scores are known:*

$$\begin{aligned}
&\mathrm{Var}\left(\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right) \\
\leq& \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i^2\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)\right]}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i^2\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)\right]}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)} \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\sum_{\mathbf{a},\mathbf{b}=\mathbf{z}^{(t-k):t}}^{\tilde{\mathbf{z}}^{(t-k):t}}(-1)^{\mathbf{1}\{\mathbf{a}=\mathbf{b}\}}\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{a},\mathbf{b};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)\mu_l\left(\left\{Y_{jt}\left(\mathbf{b},\mathbf{a};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)\right] \\
&- \frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\sum_{\mathbf{a},\mathbf{b}=\mathbf{z}^{(t-k):t}}^{\tilde{\mathbf{z}}^{(t-k):t}}(-1)^{\mathbf{1}\{\mathbf{a}=\mathbf{b}\}}\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{a};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)\right]\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{b};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)\right].
\end{aligned}$$

Proof. Using the expression of the Horvitz-Thompson estimator, we have:

$$\begin{aligned}
&\mathrm{Var}\left(\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right) = \frac{1}{N^2}\mathrm{Var}\left[\sum_{i=1}^{N}\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} - \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)\right] \\
=& \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}\left[\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} - \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)\right] \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\neq i}\mathrm{Cov}\left[\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} - \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right),\right. \\
&\left.\left(\frac{\mathbf{1}\left\{\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} - \frac{\mathbf{1}\left\{\mathbf{Z}_l^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_l^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_l\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)\right] \\
=& \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}\left[\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} - \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)\right]^2 \\
&- \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}^2\left[\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} - \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)\right] \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\neq i}\sum_{\mathbf{a},\mathbf{b}=\mathbf{z}^{(t-k):t}}^{\tilde{\mathbf{z}}^{(t-k):t}}(-1)^{\mathbf{1}\{\mathbf{a}=\mathbf{b}\}}\mathrm{Cov}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{a}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{a}\right)}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right), \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{b}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{b}\right)}\mu_l\left(\left\{Y_{jt}\right\}_{j\in\mathscr{N}};d\right)\right].
\end{aligned}$$

The first two terms in the above expression can be further expanded as:

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}\left[\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}-\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]^2\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}^2\left[\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}-\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]^2+\frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}\left[\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t}=\tilde{\mathbf{z}}^{s:t}\}}{\tilde{W}_{it}^2}\mu_i^2(Y_t(\mathbf{Z}^{1:T});d)\right]^2\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}^2\left[\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}-\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right)\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i^2\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}+\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i^2\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}^2\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)-\mu_i\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]\\
\leq&\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i^2\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}+\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i^2\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}.
\end{aligned}
$$

And the first covariance term equals:

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\neq i}\mathrm{Cov}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right),\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\mu_l\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\neq i}\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\mathbf{1}\left\{\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)P\left(\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\mu_l\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\neq i}\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\mu_l\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\neq i}\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\neq i}\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]\\
&-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right]\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathcal{N}};d\right)\right].
\end{aligned}
$$

The last equality uses the fact that for $l\notin\mathscr{B}_{i;(t,d,k)}$, the expectation of the product equals the product of expectations. Other covariance terms have similar forms. We obtain the final result by combining these terms together. □

Next, we present a bound for the asymptotic variance of the Hajek estimator with known propensity scores:

Lemma 2. *Define*

$$\tilde{V}_{t,k,d} = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}\right]^2}{P\left(\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right]^2}{P\left(\mathbf{Z}_l^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}$$

$$+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\sum_{\mathbf{a},\mathbf{b}=\mathbf{z}^{(t-k):t}}^{\tilde{\mathbf{z}}^{(t-k):t}}(-1)^{\mathbf{1}\{\mathbf{a}=\mathbf{b}\}}\mathrm{E}\left[\left(\mu_i\left(\left\{Y_{jt}\left(\mathbf{a},\mathbf{b};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{a}}\right)\left(\mu_l\left(\left\{Y_{jt}\left(\mathbf{b},\mathbf{a};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{b}}\right)\right]$$

$$-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\sum_{\mathbf{a},\mathbf{b}=\mathbf{z}^{(t-k):t}}^{\tilde{\mathbf{z}}^{(t-k):t}}(-1)^{\mathbf{1}\{\mathbf{a}=\mathbf{b}\}}\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{a};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{a}}\right]\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{b};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{b}}\right].$$

where $\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}} = \frac{1}{N}\sum_{i=1}^{N}\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)$ *and* $\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}} = \frac{1}{N}\sum_{i=1}^{N}\mu_i\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)$*. Under Assumptions 1-5 in the main text, as $N\to\infty$, we have:*

$$\Pr\left(\frac{N}{\tilde{b}_{(t,d,k)}}V_{HA;(t,d,k)} \leq \frac{N}{\tilde{b}_{t,d,k}}\tilde{V}_{t,k,d}\right)\to 1.$$

where $V_{HA;(t,d,k)}$ is the asymptotic variance of the Hajek estimator and $\tilde{b}_{(t,d,k)}$ is the same constant as in Theorem 1.

The proof is similar to that in Wang et al. (2020) hence omitted. When $\tilde{b}_{(t,d,k)}$ increases with N, the difference between the Horvitz-Thompson estimator and the Hajek estimator diminishes asymptotically. But in any finite sample, the Hajek estimator is still more efficient. The variance expression can be further simplified under an extra assumption:

Assumption (Homophily in treatment effects)**.** *For given t,d,k, define*

$$\frac{1}{N}\sum_{i=1}^{N}\left[\mu_i\left(\left\{\tau_{jt;i}\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\tau_t\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d\right)\right]\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\left[\mu_l\left(\left\{\tau_{jt;i}\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\tau_t\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d\right)\right]\geq 0.$$

The assumption indicates that the expected treatment effect generated by unit i's history at distance d is positively correlated with those effects generated by its neighbors over the period from $t-k$ to t. There is "homophily" in treatment effects on the space $\mathscr{X}$: units that generate larger-than-average effects reside close to each other. It is often the case in reality. But researchers need to justify the assumption using their substantive knowledge.

Lemma 3. *Under Assumptions 1-5 and the extra assumption, the variance bound defined in lemma 2 can be replaced by by:*

$$\tilde{V}_{t,k,d} = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}\right]^2}{P\left(\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right]^2}{P\left(\mathbf{Z}_l^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}$$

$$+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\sum_{\mathbf{a},\mathbf{b}=\mathbf{z}^{(t-k):t}}^{\tilde{\mathbf{z}}^{(t-k):t}}(-1)^{\mathbf{1}\{\mathbf{a}=\mathbf{b}\}}\mathrm{E}\left[\left(\mu_i\left(\left\{Y_{jt}\left(\mathbf{a},\mathbf{b};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{a}}\right)\left(\mu_l\left(\left\{Y_{jt}\left(\mathbf{b},\mathbf{a};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_{i,l}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{b}}\right)\right].$$

Proof. Note that by definition,

$$\tau_t\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d\right)=\mathrm{E}\left[\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right],$$

$$\mu_i\left(\left\{\tau_{jt;i}\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)=\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\mu_l\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)\right]$$

Hence, the last term of the variance bound in lemma 2 can be simplified as

$$
\begin{aligned}
&-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}} \mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}\right]*\\
&\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\mu_l\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\left(\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right)\right]\\
&+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}} \mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right]*\\
&\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\mu_l\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\left(\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right)\right]\\
&=-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l\in\mathscr{B}_{i;(t,d,k)}} \mathrm{E}\left[\mu_i\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\mu_i\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\left(\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right)\right]*\\
&\mathrm{E}\left[\mu_l\left(\left\{Y_{jt}\left(\mathbf{z}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\mu_l\left(\left\{Y_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t};\mathbf{Z}^{1:t}\setminus\mathbf{Z}_l^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\left(\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}-\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right)\right]\\
&=-\frac{1}{N}\sum_{i=1}^{N}\left[\mu_i\left(\left\{\tau_{jt;i}\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\tau_t\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d\right)\right]\sum_{l\in\mathscr{B}_{i;(t,d,k)}}\left[\mu_l\left(\left\{\tau_{jt;i}\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t}\right)\right\}_{j\in\mathscr{N}};d\right)-\tau_t\left(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d\right)\right].
\end{aligned}
$$

Under the assumption of homophily in treatment effects, this term is non-positive. Consequently, ignoring it leads to a larger variance bound. The lemma is proved. □

For the augmented estimator, denoting its variance as $V_{aug;(t,d,k)}$, we have

$$
\Pr\left(\frac{N}{\tilde{b}_{(t,d,k)}}V_{aug;(t,d,k)}\leq\frac{N}{\tilde{b}_{(t,d,k)}}\tilde{V}^{\dagger}_{t,k,d}\right)\rightarrow 1,
$$

where

$$
\begin{aligned}
\tilde{V}^{\dagger}_{t,k,d}=\tilde{V}_{t,k,d}-&\left[\frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{P\left(\mathbf{Z}_l^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}{P\left(\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}}\left(\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}(\mathbf{z}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t})\right]\right\}_{j\in\mathscr{N}};d\right)-\mathrm{E}\left[\bar{\mu}_{t,k,d}^{\mathbf{z}^{(t-k):t}}\right]\right)\right.\\
&\left.+\sqrt{\frac{P\left(\mathbf{Z}_l^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}{P\left(\mathbf{Z}_l^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}}\left(\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}(\tilde{\mathbf{z}}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t})\right]\right\}_{j\in\mathscr{N}};d\right)-\mathrm{E}\left[\bar{\mu}_{t,k,d}^{\tilde{\mathbf{z}}^{(t-k):t}}\right]\right)\right]^2.
\end{aligned}
$$

This is a straightforward extension of the classic variance formula for doubly robust estimators (see, e.g., Lunceford and Davidian (2004)). It implies that the augmented estimator is more efficient than the Horvitz-Thompson estimator or the Hajek estimator when both the propensity score model and the diffusion model are correct. We leave the question of whether the augmented estimator achieves semiparametric efficiency under interference to future research.

A3 Asymptotic distribution of the estimators

We first show that the variances of the IPTW estimators converge to zero as $N\rightarrow\infty$. Since all the moments of the transformed outcome are bounded, the variance term in $\mathrm{Var}\left(\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right)$ has an order of $O_p\left(\frac{1}{N}\right)$ and the covariance term has an order of $O_p\left(\frac{b_{i;(t,k,d)}}{N}\right)$. From Assumption 4, we know that $\frac{b_{i;(t,k,d)}}{N}\rightarrow 0$ for any i. Therefore, $\mathrm{Var}\left(\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right)$ declines to zero as N grows to infinity. The extra variance from estimating the propensity scores is asymptotically negligible when the propensity score model is correctly specified. The same is true for the Hajek estimator's asymptotic variance. Consistency then follows from Markov's inequality. Notice that consistency only requires $\max_{i\in\{1,2,\ldots,N\}}b_{i;(t,k,d)}=o_P(N)$, as shown by Sävje, Aronow and Hudgens (2021).

The asymptotic normality of the Horvitz-Thompson estimator can be derived using the central limit theorem for dependent random variables in Ogburn et al. (2020). We first restate the key lemmas in Ogburn et al. (2020) using terms defined in this paper.

Lemma 4. *(Ogburn et al. (2020), Lemma 1 and 2) Consider a set of N units. Let $U_1,\ldots,U_N$ be bounded mean-zero random variables with finite* *fourth moments and dependency neighborhoods $\mathscr{B}_{i;(t,k,d)}$. If $\frac{\max_{i\in\{1,2,\ldots,N\}} b_{i;(t,d,k)}}{\sqrt{N}} \to 0$ for all i, then*

$$\frac{\sum_{i=1}^N U_i}{\sqrt{\mathrm{Var}\left(\sum_{i=1}^N U_i\right)}} \to N(0,1).$$

Next, we prove Theorem 1 in the main text.

Proof. Define U_i as

$$\sqrt{\frac{N}{\tilde{b}_{(t,d,k)}}}\left(\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)}{NP\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}-\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)}{NP\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}-\frac{\tau_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)}{N}\right).$$

Then, $\sum_{i=1}^N U_i = \sqrt{\frac{N}{\tilde{b}_{(t,d,k)}}}\left(\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)-\tau_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right)$ and $\mathrm{E}[U_i]=0$. We know that U_i has finite fourth moments as all the outcomes are bounded, and $\mathrm{Var}\left(\sum_{i=1}^N U_i\right)=\frac{N}{\tilde{b}_{(t,d,k)}}\mathrm{Var}\left(\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right)$ is also finite. $\frac{\max_{i\in\{1,2,\ldots,N\}} b_{i;(t,d,k)}}{\sqrt{N}} \to 0$ comes from Assumption 4. From Lemma 4, we know that $\frac{\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)-\tau_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d))}{\sqrt{\mathrm{Var}\left(\hat{\tau}_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right)}} \to N(0,1)$. The asymptotic distribution of the Hajek estimator can be obtained via the Delta method. □

For the augmented estimator, we first show that it is doubly robust when the effects from the units are additive.

Proof. Consider the scenario where the propensity scores are correctly specified while the diffusion model is not. Then,

$$\begin{aligned}
&\mathrm{E}\left[\hat{\tau}_{t,aug}(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right]=\mathrm{E}\left[\hat{\tau}_{t,HA}(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right]\\
&-\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\mu_i\left(\left\{\hat{Y}_{jt}\right\}_{j\in\mathcal{N}};d\right)}{\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}-\frac{1}{N}\sum_{i=1}^N\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}\mu_i\left(\left\{\hat{Y}_{jt}\right\}_{j\in\mathcal{N}};d\right)}{\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right]\\
&+\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)-\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&-\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)-\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)\right]\\
&+\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)-\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\tau_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d).
\end{aligned}$$

The second equality uses the linearity of $\mu_i\left(\left\{\hat{Y}_{jt}\right\}_{j\in\mathcal{N}};d\right)$. Note that the result holds even when the effects are not additive. Next, suppose the diffusion model is accurate but the propensity scores are not. Then, $Y_{jt}-\hat{Y}_{jt}=\hat{e}_{jt}$ and $\mathrm{E}[\hat{e}_{jt}]=0$. We have

$$\begin{aligned}
&\mathrm{E}\left[\hat{\tau}_{t,aug}(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d)\right]\\
=&\frac{1}{N}\sum_{i=1}^N\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\mu_i\left(\left\{\hat{e}_{jt}\right\}_{j\in\mathcal{N}};d\right)}{\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}\right]-\frac{1}{N}\sum_{i=1}^N\mathrm{E}\left[\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}\left[\mu_i\left(\left\{\hat{e}_{jt}\right\}_{j\in\mathcal{N}};d\right)\right]}{\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right]\\
&+\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)-\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\mathbf{z}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)-\frac{1}{N}\sum_{i=1}^N\mu_i\left(\left\{\mathrm{E}\left[\hat{Y}_{jt}\left(\tilde{\mathbf{z}}^{(t-k):t},\mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right]\right\}_{j\in\mathcal{N}};d\right)\right]\\
=&\tau_t(\mathbf{z}^{(t-k):t},\tilde{\mathbf{z}}^{(t-k):t};d).
\end{aligned}$$

The last equality requires the additivity of the effects. Otherwise, the propensity scores have to be accurate for us to calculate $\mathrm{E}\left[\hat{Y}_{jt}\left(\mathbf{z}^{(t-k):t}, \mathbf{Z}^{1:t} \setminus \mathbf{Z}_i^{(t-k):t}\right)\right]$. $\square$

Consistency of the augmented estimator comes from its variance expression. From the proof, we can see that the difference between the Horvitz-Thompson estimator and the augmented estimator is a sample average. Hence, it is also asymptotically normal.

A4 Variance estimation

We first present the expression of the spatial heteroscedasticity and auto-correlation consistent (HAC) variance estimator (Conley, 1999). Let's denote the diagonal weighting matrix $\left\{\frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)+\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}{\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)\hat{P}\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}\right\}_{i=1,2,\dots,N}$ as $\tilde{\mathbf{M}}_t = \{\tilde{M}_{it}\}_{i=1,2,\dots,N}$. Define $\tilde{\mathbf{X}}_t$ as $\begin{pmatrix} \mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right\}, \left\{\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right\} \\ \mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right\}, \left\{\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right\} \\ \dots \\ \mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right\}, \left\{\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right\} \end{pmatrix}$ and $\tilde{\mathbf{Y}}_t$ as $\begin{pmatrix} \mu_1\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}}; d\right) \\ \mu_2\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}}; d\right) \\ \dots \\ \mu_N\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}}; d\right) \end{pmatrix}$. Then the regression representation of the Hajek estimator has the solution $\begin{pmatrix} \hat{a}(d) \\ \hat{\tau}_{t,OLS}(\mathbf{z}^{(t-k):t}, \tilde{\mathbf{z}}^{(t-k):t}; d) \end{pmatrix} = \left(\tilde{\mathbf{X}}_t'\tilde{\mathbf{M}}_t\tilde{\mathbf{X}}_t\right)^{-1}\left(\tilde{\mathbf{X}}_t'\tilde{\mathbf{M}}_t\tilde{\mathbf{Y}}_t\right)$.

Using simple algebra, we can that $\hat{a}(d) = \frac{\sum_{i=1}^N \mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}\mu_i\left(\left\{Y_{jt}\right\}_{j\in\mathcal{N}};d\right)/P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}{\sum_{i=1}^N \mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}/P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}$ and $\hat{\tau}_{t,OLS}(\mathbf{z}^{(t-k):t}, \tilde{\mathbf{z}}^{(t-k):t}; d) = \hat{\tau}_{t,HA}(\mathbf{z}^{(t-k):t}, \tilde{\mathbf{z}}^{(t-k):t}; d)$. Let's denote the residual for each observation as $\hat{v}_{it}$ and the variance-covariance matrix of $\{\hat{v}_{it}\}_{i=1,2,\dots,N}$ as Σ_t. It is worth noting that the covariance of $\hat{v}_{it}$ and $\hat{v}_{jt}$ is non-zero if and only if $j \in \mathscr{B}_{i;(t,d,k)}$. Furthermore, we denote $\sum_{i=1}^N \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)}$ as K_1 and $\sum_{i=1}^N \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)}$ as K_0. We know the spatial HAC variance of $\begin{pmatrix} \hat{a}(d) \\ \hat{\tau}_{t,OLS}(\mathbf{z}^{(t-k):t}, \tilde{\mathbf{z}}^{(t-k):t}; d) \end{pmatrix}$ can be expressed as:

$$
\begin{aligned}
&\widehat{\mathrm{Var}}\begin{pmatrix} \hat{a}(d) \\ \hat{\tau}_{t,OLS}(\mathbf{z}^{(t-k):t}, \tilde{\mathbf{z}}^{(t-k):t}; d) \end{pmatrix} \\
=&(\tilde{\mathbf{X}}_t'\tilde{\mathbf{M}}_t\tilde{\mathbf{X}}_t)^{-1}(\tilde{\mathbf{X}}_t'\tilde{\mathbf{M}}_t\Sigma_t\tilde{\mathbf{M}}_t'\mathbf{X}_t)(\tilde{\mathbf{X}}_t'\tilde{\mathbf{M}}_t\tilde{\mathbf{X}}_t)^{-1} \\
=&\begin{pmatrix} K_0 & 0 \\ 0 & K_1 \end{pmatrix}^{-1} \left(\sum_{i=1}^N \sum_{j=1}^N \begin{pmatrix} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)} & \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} \\ \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\tilde{\mathbf{z}}^{(t-k):t}\right)} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} & \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t}=\mathbf{z}^{(t-k):t}\right)} \end{pmatrix} \hat{v}_{it}\hat{v}_{jt}\mathbf{1}\{j \in \mathscr{B}_{i;(t,d,k)}\} \right) \begin{pmatrix} K_0 & 0 \\ 0 & K_1 \end{pmatrix}^{-1}.
\end{aligned}
$$

We are interested in entry $(2,2)$ of the above expression. Rearranging the observations such that those with $\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}$ rank before those with $\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}$, the quantity of interest can be simplified as:

$$
\begin{aligned}
&\widehat{\mathrm{Var}}(\hat{\tau}_{t,OLS}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d)) \\
=&\frac{1}{K_1^2}\sum_{i=1}^N \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right)^2}\hat{v}_i^2 + \frac{1}{K_0^2}\sum_{i=1}^N \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right)}\hat{v}_i^2 + \frac{1}{K_1^2}\sum_{i=1}^N \sum_{j\in\mathscr{B}_{i;(t,d,k)}} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right)} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right)}\hat{v}_i\hat{v}_j \\
&-\frac{2}{K_1K_0}\sum_{i=1}^N \sum_{j\in\mathscr{B}_{i;(t,d,k)}} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right)} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \mathbf{z}^{(t-k):t}\right)}\hat{v}_i\hat{v}_j + \frac{1}{K_0^2}\sum_{i=1}^N \sum_{j\in\mathscr{B}_{i;(t,d,k)}} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right)} \frac{\mathbf{1}\left\{\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right\}}{P\left(\mathbf{Z}_i^{(t-k):t} = \tilde{\mathbf{z}}^{(t-k):t}\right)}\hat{v}_i\hat{v}_j
\end{aligned}
$$

The last step is to show that the variance estimate $N\widehat{\mathrm{Var}}(\hat{\tau}_{t,OLS}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d))$ is consistent for the Hajek estimator's asymptotic variance under the assumption of homophily in treatment effects. The proof is similar to that of Lemma A.6 in Wang et al. (2020) hence we omit the details to save space.

A5 Estimate the nuisance parameters

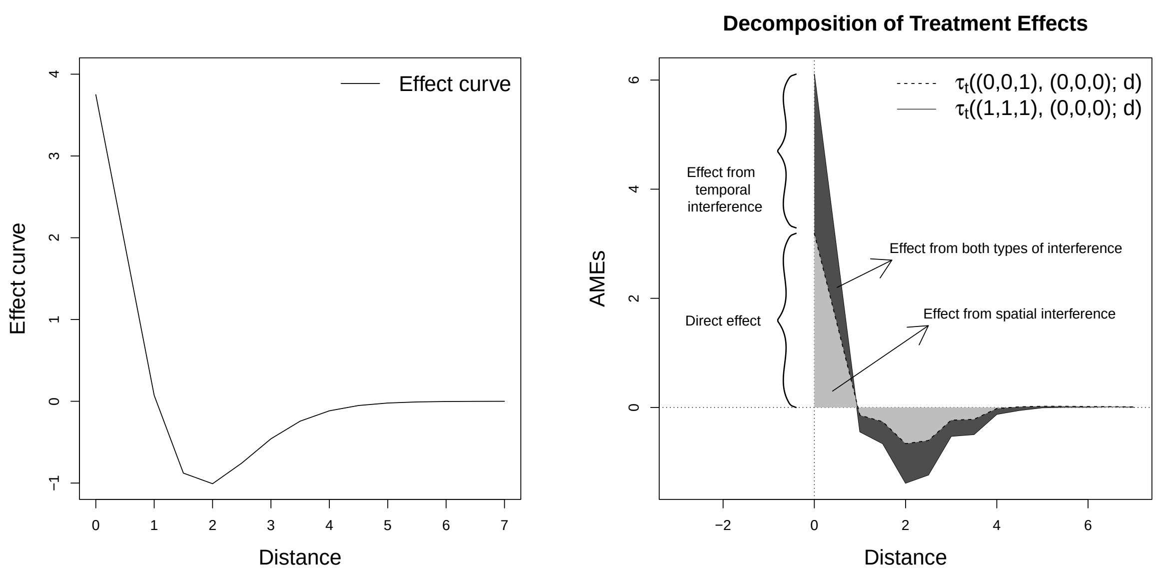
So far, we have assumed that the nuisance parameters, either the propensity scores or the diffusion model, are known to the researchers. In practice, we must estimate them from data, which impacts the variances of the proposed estimators. When the nuisance parameters are estimated using parametric models, the variances can be obtained from the standard theory of M-estimation (Lunceford and Davidian, 2004). It is known that ignoring uncertainties from estimating the nuisance parameters leads to more conservative variance estimates. Furthermore, when the convergence rate of our estimators is lower than $\sqrt{N}$, the uncertainties stemming from estimating the nuisance parameters become negligible in large samples. Researchers may also consider non-parametric estimators for the nuisance parameters, such sieve estimators (Hirano, Imbens and Ridder, 2003), covariate balancing propensity score (Imai and Ratkovic, 2015), and highly adaptive lasso (Ertefaie, Hejazi and van der Laan, 2020). The main results won't be affected as long as the estimates converge to their true values at a sufficiently fast rate, as required by Theorem 1 in the main text. When sieve estimators are adopted, the variances can be estimated following the proposal in Ackerberg, Chen and Hahn (2012). In other contexts, the variance estimates need to be discussed on a case-by-case basis.

Appendix B Extra results from simulation and applications

B1 Bias of the IPTW estimators

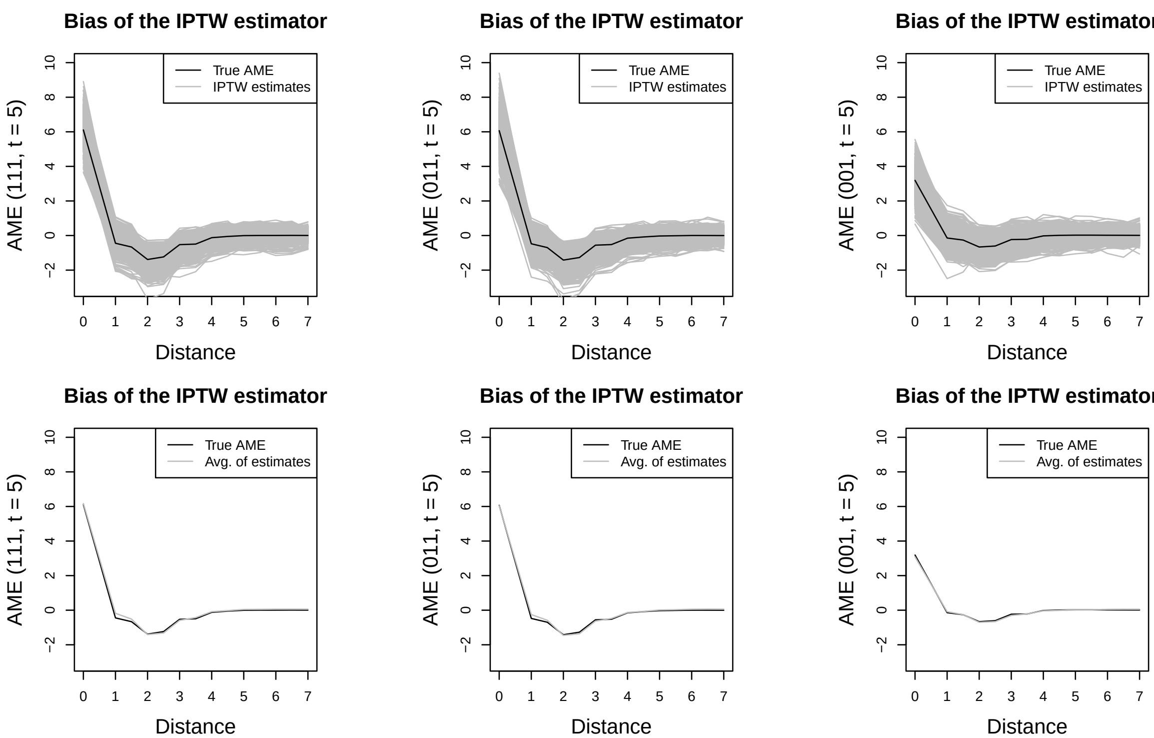
We first examine the bias of both the Horvitz-Thompson estimator and the Hajek estimator when sequential ignorability holds and the effect function is non-monotonic. The effect function and the AMEs are displayed in Figure 1. Figure 2 and Figure 3 present the estimation results. We can see that both estimators are unbiased.

Figure 1: Effect function and the AMEs



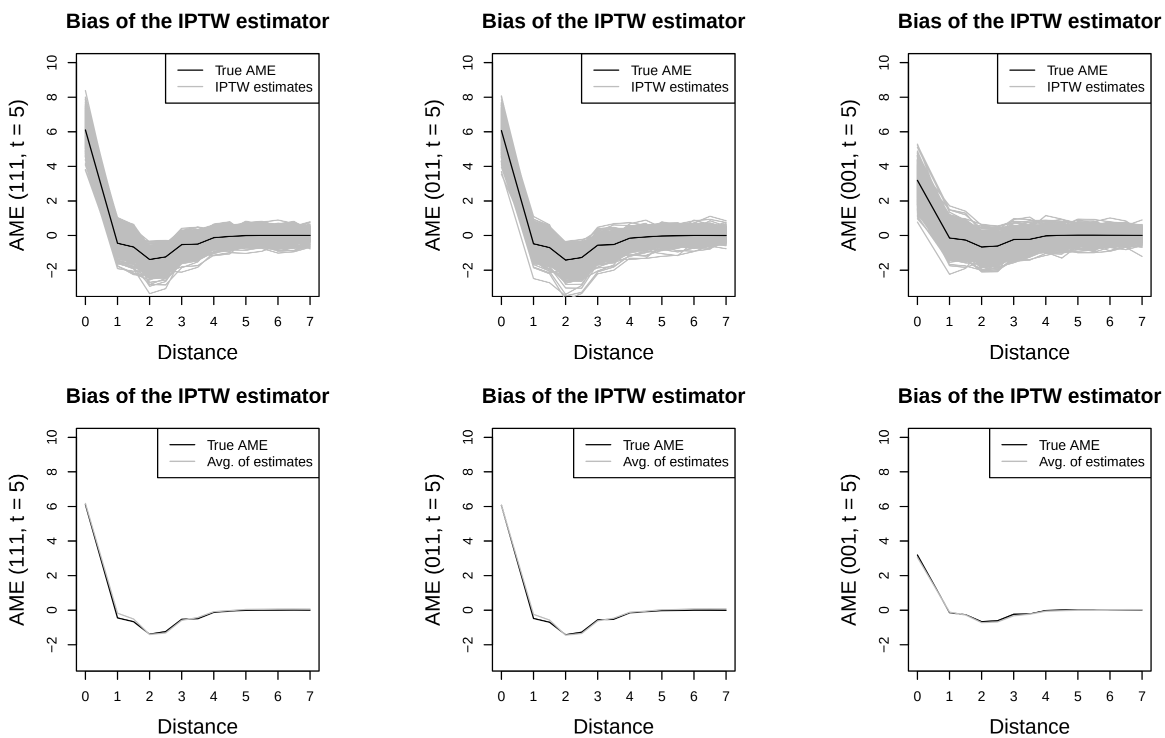
Notes: The left plot shows how the effect function that emanates from each unit varies over distance. The right plot presents the estimands $\tau_5\left((0,0,1),(0,0,0);d\right)$ and $\tau_5\left((1,1,1),(0,0,0);d\right)$ over the same range of distance values.

Figure 2: Bias of the Horvitz-Thompson estimator



Notes: The top figures display the estimates obtained from the Horvitz-Thompson estimator for all $1,000$ assignments, while the bottom figures compare the averages of these estimates with the true AMEs. The bias of the estimator is illustrated as the difference between the gray and black curves in the bottom figures. Note that the effect function is non-monotonic and that sequential ignorability holds in this scenario.

Figure 3: Bias of the Hajek estimator

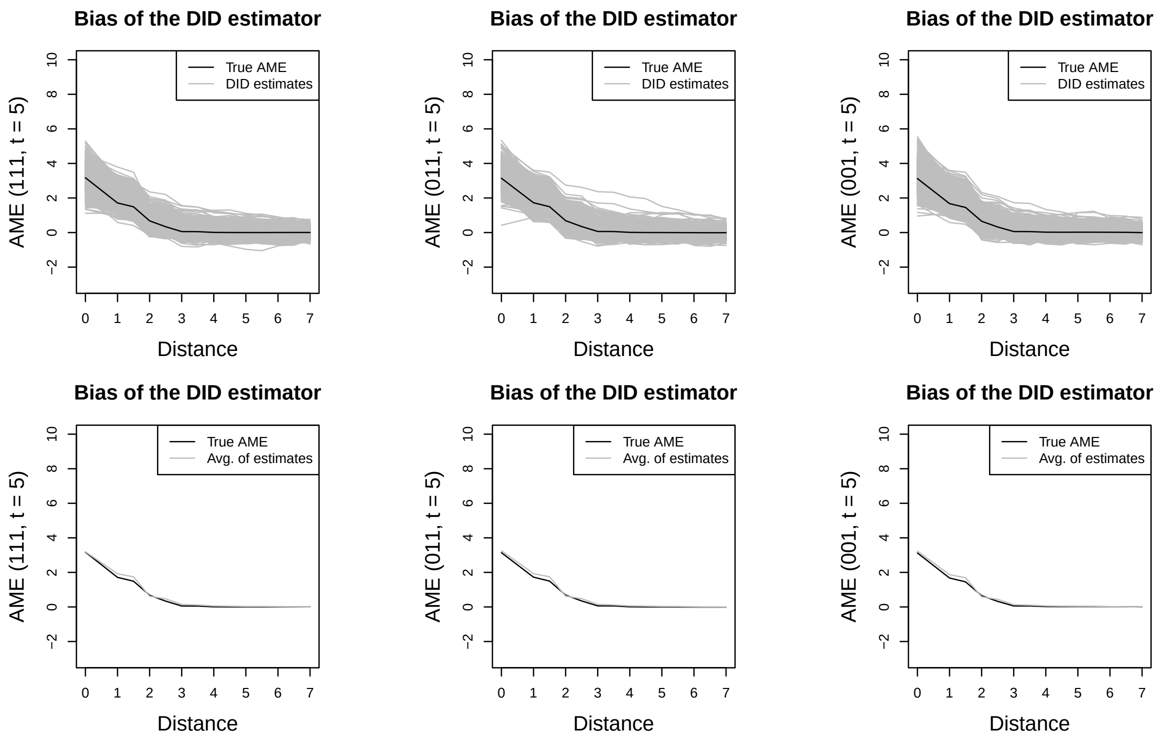


Notes: The top figures display the estimates obtained from the Hajek estimator for all $1{,}000$ assignments, while the bottom figures compare the averages of these estimates with the true AMEs. The bias of the estimator is illustrated as the difference between the gray and black curves in the bottom figures. Note that the effect function is non-monotonic and that sequential ignorability holds in this scenario.

B2 Bias of the DID estimator under homogeneous treatment effect

Figure 4 shows the bias of the DID estimator when sequential ignorability does not hold and unit fixed effects become confounders. The DGP is the same as the one in the main text. The only difference is that the effects are homogeneous across the units and do not accumulate over periods. As predicted by our discussion on the impossible trilemma, the DID estimator is now unbiased for the AMEs.

Figure 4: Bias of the DID estimator under homogeneous treatment effect



Notes: The top figures display the estimates obtained from the DID estimator for all $1,000$ assignments, while the bottom figures compare the averages of these estimates with the true AMEs. The bias of the estimator is illustrated as the difference between the gray and black curves in the bottom figures. Note that the effect function is monotonic and homogeneous and that sequential ignorability does not hold in this scenario.

B3 Bias of the augmented estimator

To test the performance of the augmented estimator, we rely on the same data generating process in the main text. We adopt the following model to predict the value of each Y_{it}:

$$Y_{it} = \sum_{d\in\mathscr{D}} \alpha_d \sum_{j=1}^{N} \mathbf{1}\{d_{ij}=d\} + \sum_{d\in\mathscr{D}} \beta_d \sum_{j=1}^{N} Z_{jt}\mathbf{1}\{d_{ij}=d\} + \mathbf{X}_i^{'}\psi + \varepsilon_{it}$$

The model can be estimated via OLS. Then, we have the predicted value for each Y_{it}:

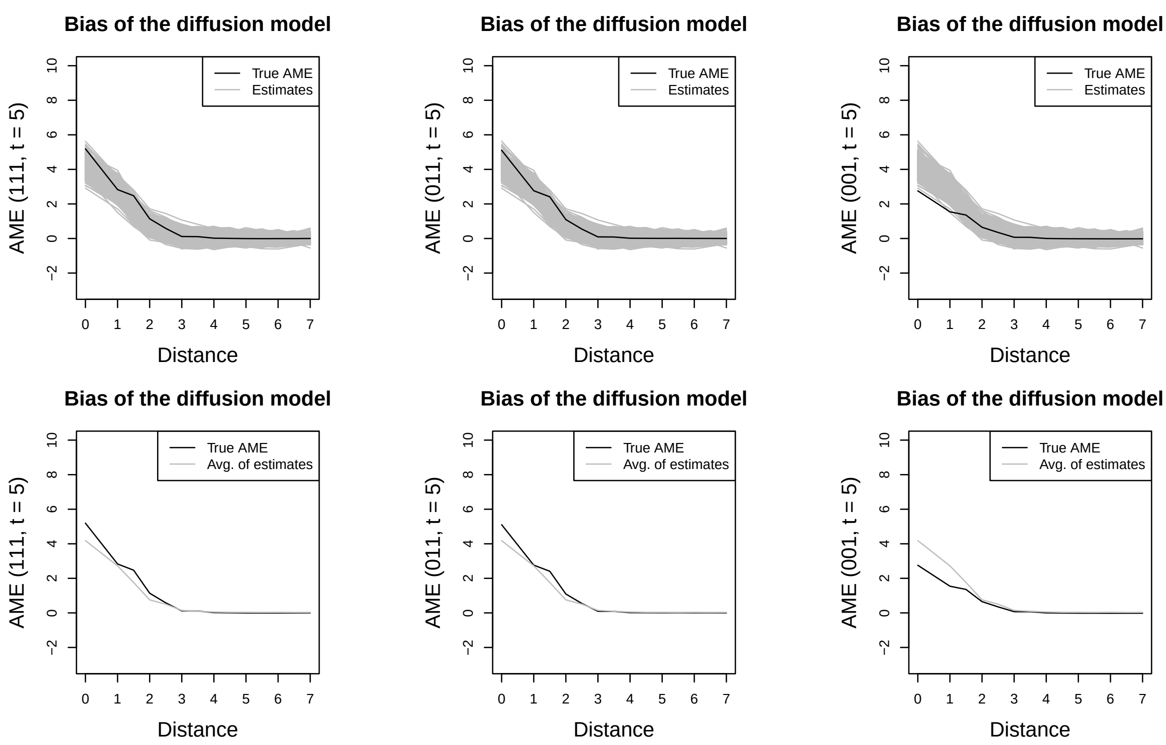
$$\hat{Y}_{it} = \sum_{d\in\mathscr{D}} \hat{a}_d \sum_{j=1}^{N} \mathbf{1}\{d_{ij}=d\} + \sum_{d\in\mathscr{D}} \hat{\beta}_d \sum_{j=1}^{N} Z_{jt}\mathbf{1}\{d_{ij}=d\} + \mathbf{X}_i^{'}\hat{\psi}.$$

We can further estimate the marginalized outcomes by:

$$\mathrm{E}\left[\hat{Y}_{lt}\left(\tilde{\mathbf{z}}^{(t-k):t}, \mathbf{Z}^{1:t}\setminus\mathbf{Z}_i^{(t-k):t}\right)\right] = \sum_{d\in\mathscr{D}} \hat{a}_d \sum_{j=1}^{N} \mathbf{1}\{d_{ij}=d\} + \sum_{d\in\mathscr{D}} \hat{\beta}_d \left(z_{lt}\mathbf{1}\{d_{ik}=d\} + \sum_{j\neq l}^{N} P(Z_{jt}=1)\mathbf{1}\{d_{ij}=d\}\right) + \mathbf{X}_i^{'}\hat{\psi}.$$

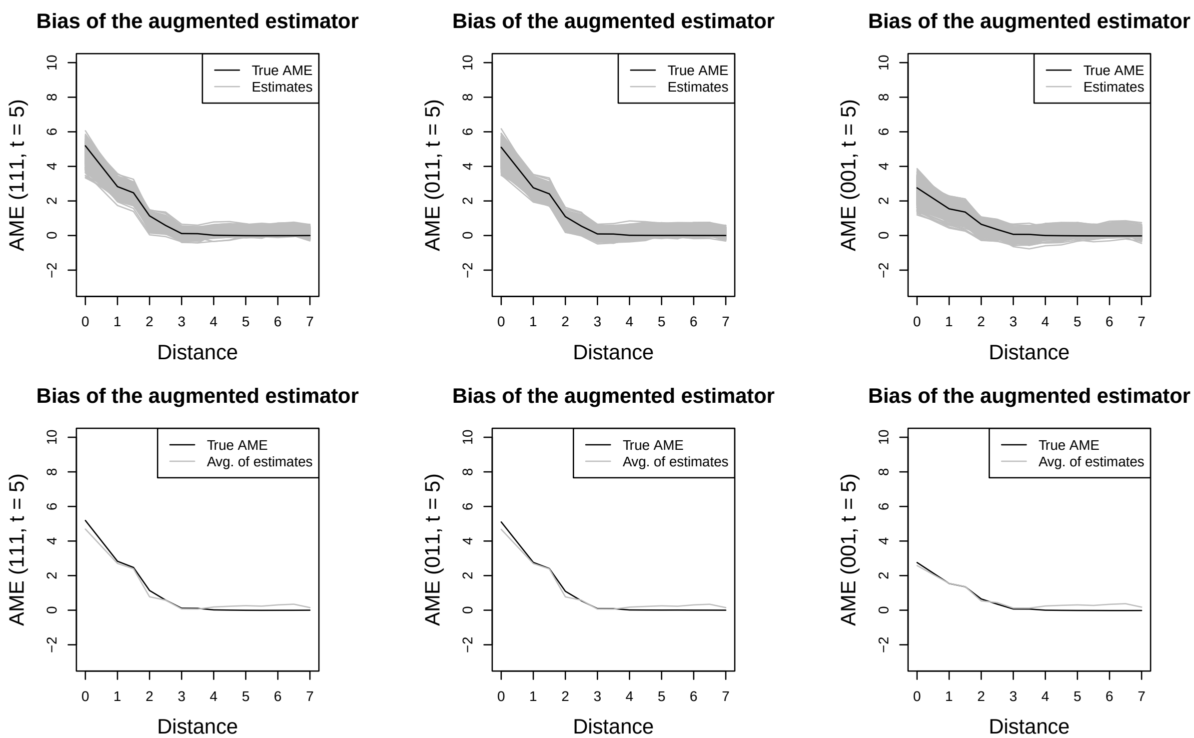
The model has ignored the effects from treatments in the previous period as well as the interaction among the effects. In Figure 5, we compare the estimated coefficients $\hat{\beta}_d$ with the AMEs. They are obviously biased. Figure 6 shows that the bias disappears once we augment the diffusion model with the propensity scores. Estimates from the augmented estimator have smaller variances than those from either the Horvitz-Thomson estimator or the Hajek estimator. Across all the distance values, the variances decline by more than 40%.

Figure 5: Bias of the diffusion model



Notes: The top figures display the estimates obtained from the diffusion model for all $1,000$ assignments, while the bottom figures compare the averages of these estimates with the true AMEs. The bias of the estimator is illustrated as the difference between the gray and black curves in the bottom figures. Note that the effect function is monotonic and that sequential ignorability holds in this scenario.

Figure 6: Bias of the augmented estimator

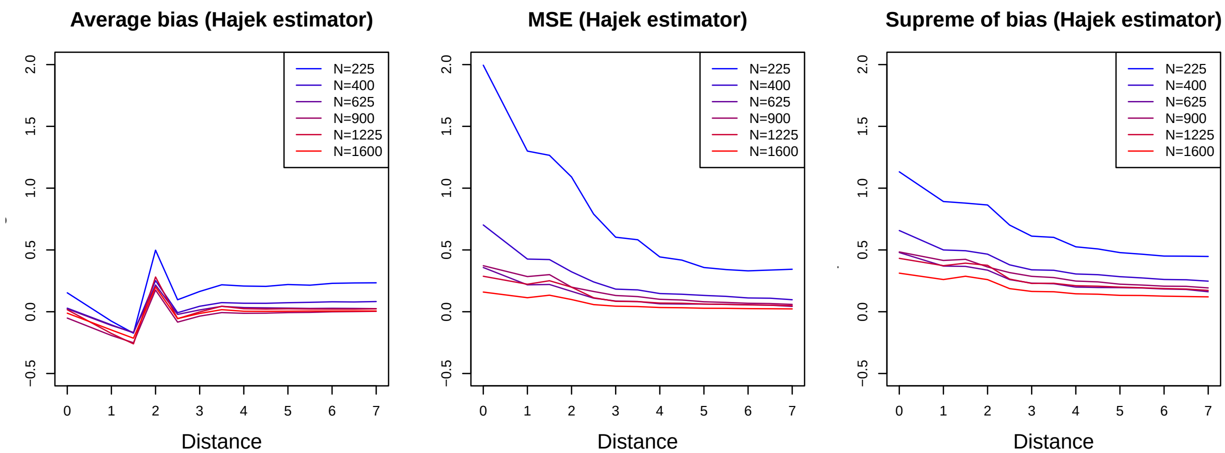


Notes: The top figures display the estimates obtained from the augmented estimator for all $1,000$ assignments, while the bottom figures compare the averages of these estimates with the true AMEs. The bias of the estimator is illustrated as the difference between the gray and black curves in the bottom figures. Note that the effect function is monotonic and that sequential ignorability holds in this scenario.

B4 Consistency of the IPTW estimators

Figure 7 illustrates how the bias of the Hajek estimator varies over sample sizes. We can see that when the number of units grows, the average bias, the mean squared error (MSE), and the supreme of bias all decline to zero across distance values, suggesting that the Hajek estimator is consistent.

Figure 7: Consistency of the Hajek estimator

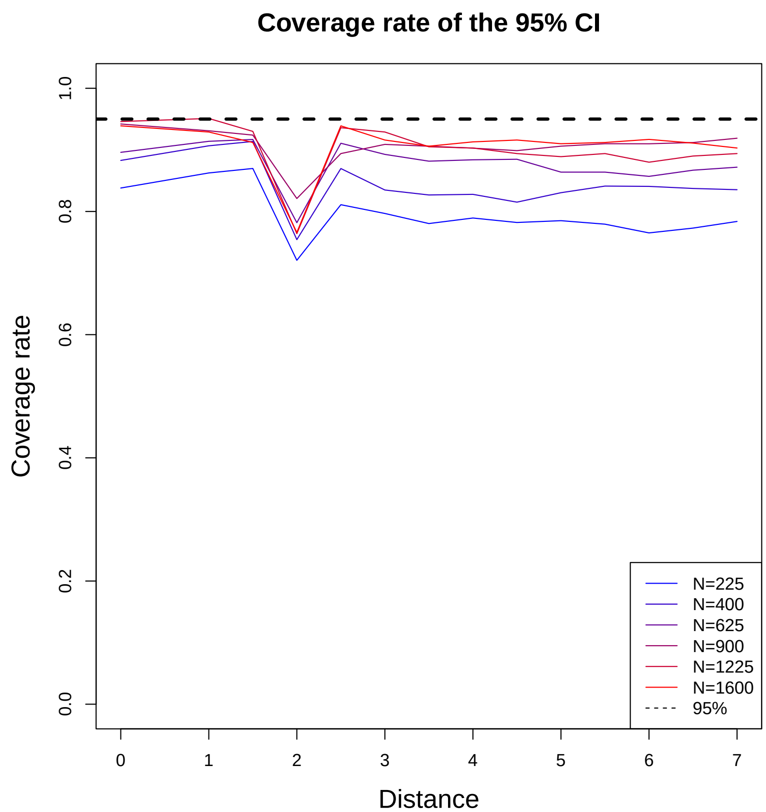


Notes: Figures from left to right show the average bias, the average MSE, and the supreme of bias obtained from the Hajek estimator for all $1,000$ assignments. Different colors indicate varying sample sizes.

B5 Coverage rate of the proposed confidence interval

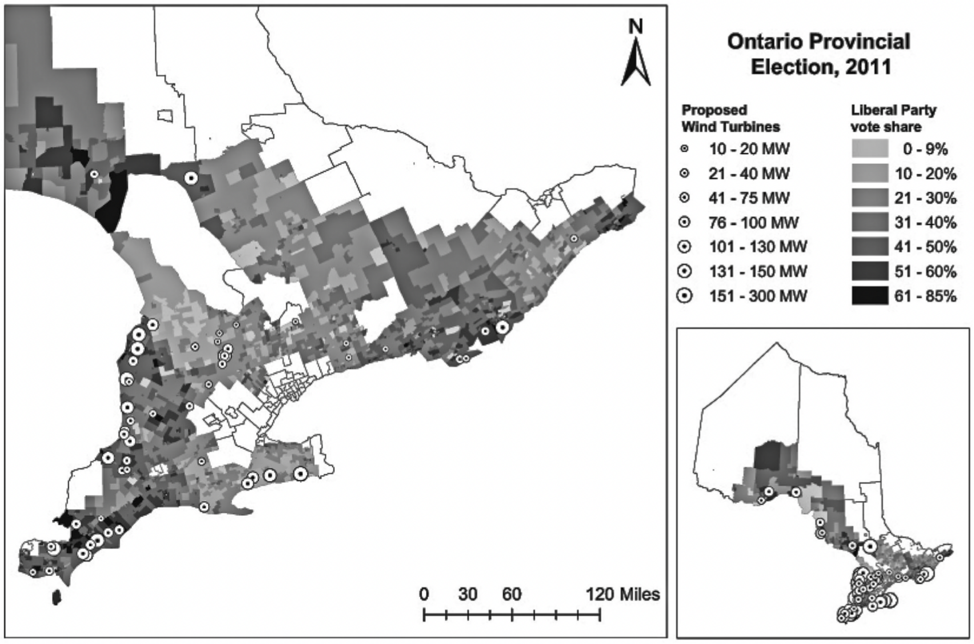
Figure 8 presents how the coverage rate of the proposed 95% confidence interval, based on the spatial HAC variance estimator varies over sample sizes. The dashed line represents the nominal level of 95%. We can see that as N grows, the coverage rate approaches to the nominal level across distance values.

Figure 8: Consistency of the Hajek estimator



Notes: The figure shows the average coverage rate of the proposed 95% confidence interval across all $1,000$ assignments for each of the distance values. Different colors indicate varying sample sizes.

B6 Map from Stokes (2016)



Notes: The figure is taken from @stokes2016electoral. It shows the locations of the precincts and the proposed wind turbines in Ontario, Canada, in 2011.

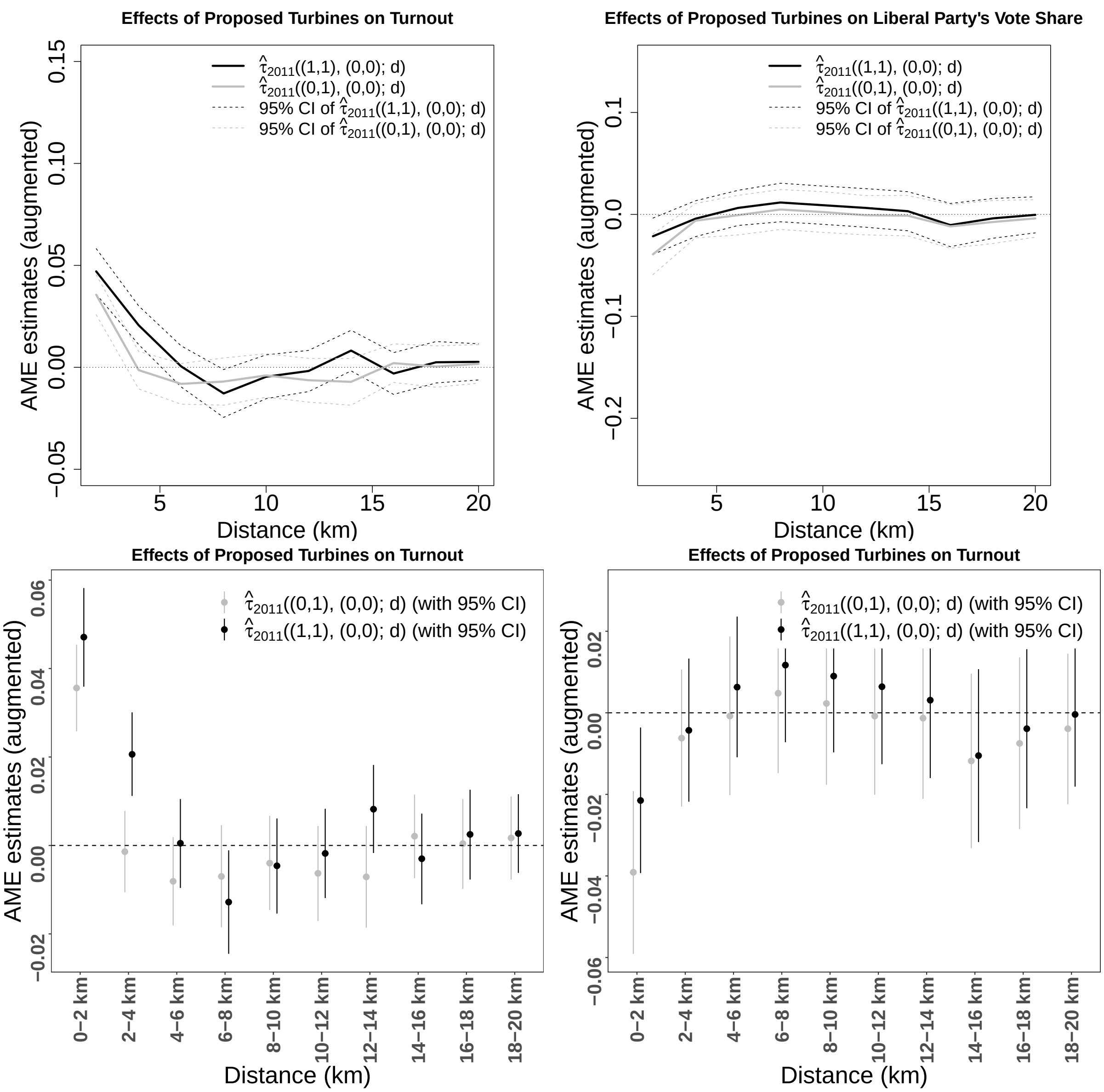
B7 Replication of Stokes (2016) using the augmented estimator

We rely on the following diffusion model to predict the value of each Y_{it}:

$$Y_{it} = \sum_{d\in\mathscr{D}} a_d \sum_{j=1}^{N} \mathbf{1}\{d_{ij} = d\} + \sum_{d\in\mathscr{D}} \beta_d \sum_{j=1}^{N} Z_{jt}\mathbf{1}\{d_{ij} = d\} + \sum_{d\in\mathscr{D}} \gamma_d \sum_{j=1}^{N} Z_{j,t-1}\mathbf{1}\{d_{ij} = d\} + \lambda Y_{i,t-1} + \mathbf{X}_i^{'}\psi + \varepsilon_{it}$$

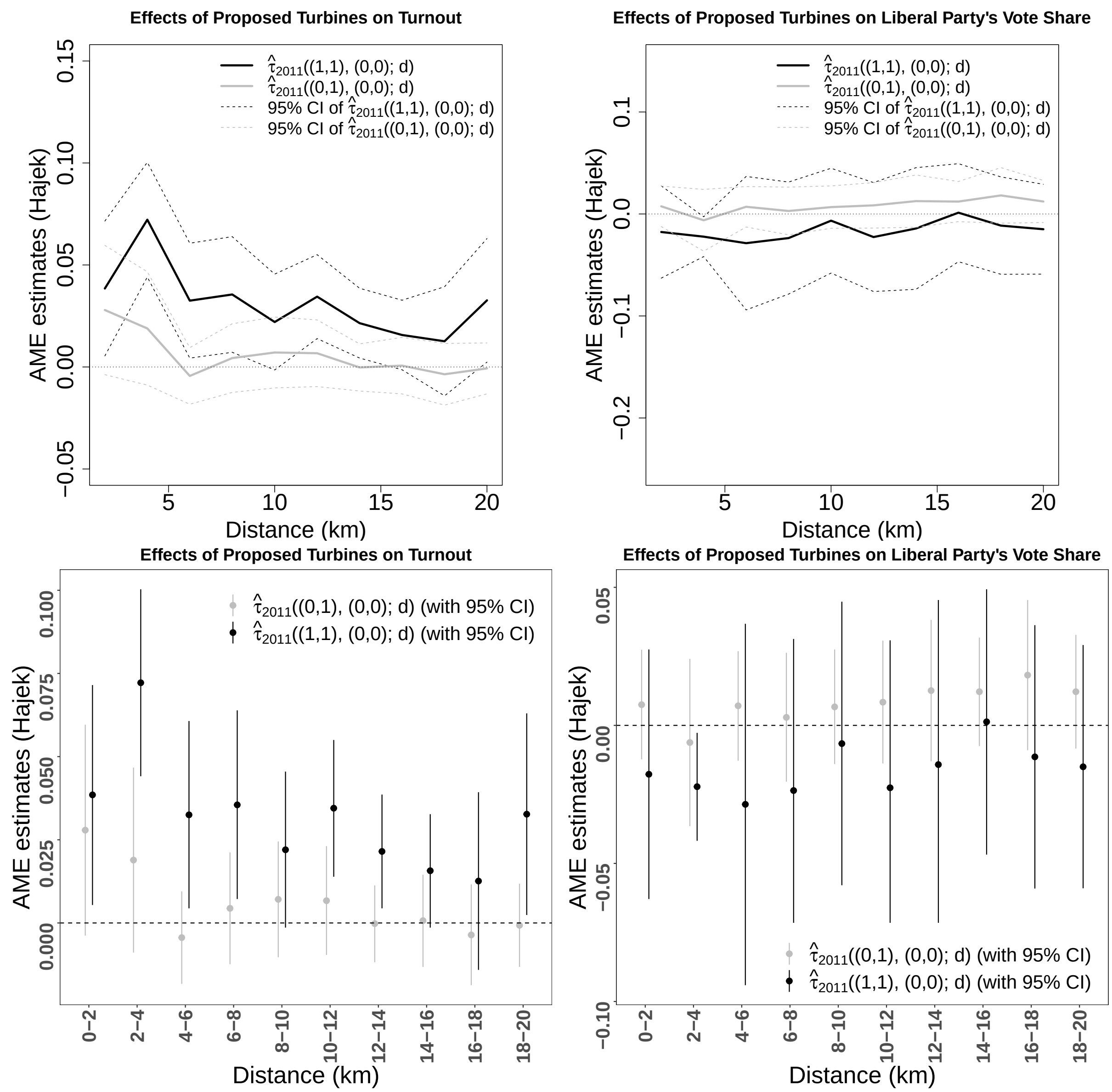
Estimates from the augmented estimator are presented in Figure 9. We can see that the main patterns are similar to what we have detected in the main text using the Hajek estimator. But the magnitude of the estimates becomes smaller for both outcomes, as do their standard errors estimates. Results in Figure 10 are based on units whose propensity score estimates are between 0.1 and 0.9. We can see that the findings are robust to the potential violation of positivity.

Figure 9: Replication of Stokes (2016)



Notes: The figures in the left column display the estimates for the turnout rate, while those in the right column show the estimates for the Liberal Party's vote share in election year t. The top row depicts replication results obtained using the augmented estimator for each donut, with radii ranging from $0-2$ km to $18-20$ km. The black solid curve represents estimates of $\tau_{2011}((1,1),(0,0);d)$, while the gray solid curve represents estimates of $\tau_{2011}((0,1),(0,0);d)$. The black and gray dotted lines represent their respective 95% confidence intervals, calculated using the spatial HAC variance estimator. The bottom row presents the same estimates as coefficient plots.

Figure 10: Replication of Stokes (2016) under positivity



Notes: The figures in the left column display the estimates for the turnout rate, while those in the right column show the estimates for the Liberal Party's vote share in election year t. The top row depicts replication results obtained using the Hajek estimator for each donut, with radii ranging from $0-2$ km to $18-20$ km. The black solid curve represents estimates of $\tau_{2011}((1,1),(0,0);d)$, while the gray solid curve represents estimates of $\tau_{2011}((0,1),(0,0);d)$. The black and gray dotted lines represent their respective 95% confidence intervals, calculated using the spatial HAC variance estimator. The bottom row presents the same estimates as coefficient plots. The sample only includes units whose propensity score estimates are between 0.1 and 0.9.

References

Ackerberg, Daniel, Xiaohong Chen and Jinyong Hahn. 2012. "A practical asymptotic variance estimator for two-step semiparametric estimators." *Review of Economics and Statistics* 94(2):481–498.

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92:1–45.

Ertefaie, Ashkan, Nima S Hejazi and Mark J van der Laan. 2020. "Nonparametric inverse probability weighted estimators based on the highly adaptive lasso." *arXiv preprint arXiv:2005.11303* .

Hirano, Keisuke, Guido W Imbens and Geert Ridder. 2003. "Efficient estimation of average treatment effects using the estimated propensity score." *Econometrica* 71(4):1161–1189.

Imai, Kosuke and Marc Ratkovic. 2015. "Robust estimation of inverse probability weights for marginal structural models." *Journal of the American Statistical Association* 110(511):1013–1023.

Lunceford, Jared K and Marie Davidian. 2004. "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study." *Statistics in medicine* 23(19):2937–2960.

Ogburn, Elizabeth L, Oleg Sofrygin, Ivan Diaz and Mark J van der Laan. 2020. "Causal inference for social network data." *arXiv preprint arXiv:1705.08527* .

Sävje, Fredrik, Peter M Aronow and Michael G Hudgens. 2021. "Average treatment effects in the presence of unknown interference." *The Annals of Statistics* 49(2):673–701.

Wang, Ye, Cyrus Samii, Haoge Chang and Peter M Aronow. 2020. "Design-based inference for spatial experiments with interference." *arXiv preprint arXiv:2010.13599* .