

# A Design-Based Approach to Spatial Correlation

Ruonan Xu<sup>1</sup> and Jeffrey M. Wooldridge<sup>2</sup>

<sup>1</sup>Rutgers University

<sup>2</sup>Michigan State University

## Abstract

When observing spatial data, what standard errors should we report? With the finite population framework, we identify three channels of spatial correlation: sampling scheme, assignment design, and model specification. The Eicker-Huber-White standard error, the cluster-robust standard error, and the spatial heteroskedasticity and autocorrelation consistent (SHAC) standard error are compared under different combinations of the three channels. Then, we provide guidelines for whether standard errors should be adjusted for spatial correlation for both linear and nonlinear estimators. As it turns out, the answer to this question also depends on the magnitude of the sampling probability.

## 1 Introduction

A common concern among empirical researchers is spatial correlation, as policy or event treatments are often spatially correlated. A merger between two gasoline companies, for example, could affect the local gasoline retail market spatially and disproportionately in the neighborhoods close to the rebranded stations (Houde, 2012). Closure and demolition

of public housing affects the areas closest to the projects more than those farther away (Aliprantis and Hartley, 2015). Along with spatially correlated treatments, there can also be spillover effects to adjacent entities that diminish with distance. One common point of confusion is whether standard errors should be adjusted for spatial correlation. In this paper, we answer this question by disentangling sampling scheme, assignment design, and model specification under a finite population framework.

Adopting the finite population paradigm when studying spatial data has clear advantages. For one, any cases where spatial effects are of interest, the unit of observations is determined by geography or generalized distance measure, and we often sample all units in the population. For example, typically we can collect information on all counties in the United States. As pointed out by Pinkse, Shen, and Slade (2007), “with spatial data, it is common for the sample and the population to be the same (e.g., the set of all firms in a market)” (p. 216). The sampling uncertainty underlying the superpopulation approach is unnatural for thinking about uncertainty in such settings; see also Abadie, Athey, Imbens, and Wooldridge (2020).

Second, with a well-defined finite population, we can explicitly introduce different sampling schemes after imposing spatial correlations on the population units. Rather than fixing the sites and treating different realizations of the data at all sites as random draws from a superpopulation, it is more practical to treat the lattice as a finite population and draw different collections of sites from it. The latter is something we can actually do in real life, but the former is merely a thought experiment. Along with sampling, we observe only one potential outcome at each location. Therefore, our framework combines sampling- and design-based uncertainty. We can also differentiate between spatial correlation among all neighbors in the population and correlation within a subset of neighbors observed in the sample, which has a major impact on statistical inference. With that said, because we allow

the sampling probability to shrink to zero, our asymptotic theory can also accommodate sampling from superpopulations as a special case.

By means of a set of newly developed limit theorems, we derive the finite population spatial heteroskedasticity and autocorrelation consistent (SHAC) variance-covariance matrix of M-estimators. We show that, in the case where a nontrivial fraction of the population is sampled, the new asymptotic variance matrix is smaller than the superpopulation SHAC variance matrix. This finding generalizes Neyman (1923)’s result on conservativeness of the finite population variance estimation and the extension to regression settings in Abadie et al. (2020). Based on the alternate finite population variance-covariance matrix, we provide guidelines on whether standard errors should be adjusted for spatial correlation instead of relying on heuristic arguments of spatial correlation of unobserved characteristics.

To summarize, whenever there is spatial assignment, meaning that the “treatment” variables are spatially correlated, or a spillover effect is estimated, even in the absence of spillover, one should adjust standard errors for spatial correlation, for instance, by reporting the SHAC standard errors. However, there are a few exceptions. When we independently sample a negligible portion of the population, the Eicker-Huber-White (EHW) standard errors would suffice irrespective of the existence of spatial correlation. Similarly, if we sample a small fraction of clusters from all the clusters in the population, the cluster-robust standard errors would suffice.

Our paper contributes to three strands of literature. Our first contribution is to the literature on limit theorems for random fields. There is a very general asymptotic theory for spatial processes under either the mixing condition or near-epoch dependence in Jenish and Prucha (2009) and Jenish and Prucha (2012). Unfortunately, due to the introduction of sampling indicators necessary for our paper, their theory cannot incorporate sampling

from a superpopulation. Recently, Bradley and Tone (2017) develop a central limit theorem for nonstationary random fields using strong mixing conditions with certain restrictions. By borrowing the techniques in the aforementioned articles, we derive new laws of large numbers and a central limit theorem for near-epoch dependent (NED) processes. We apply these limit theorems to finite population asymptotic theory but they also accommodate superpopulations by including zero sampling probability in the limit. Meanwhile, we allow for nonstationary processes with unbounded moments on irregularly spaced lattices. We also introduce cluster correlation on top of spatial correlation by explicitly including the sampling indicators and allowing for cluster sampling.

Second, we contribute to the literature on finite population inference. Several articles under the finite population framework study the EHW standard error and the cluster-robust standard error for both linear and nonlinear estimators. See, for instance, Abadie et al. (2020), Abadie, Athey, Imbens, and Wooldridge (2017), Xu (2021b), and Xu (2021a). Bojinov, Rambachan, and Shephard (2021) extend the finite population framework to panel experiments. They allow for spillovers across time periods but maintain the assumption of no spillover across units. Sävje, Aronow, and Hudgens (2021), Sävje (2021), and Leung (2022) consider network/spatial interference in randomized experiments with access to the entire population. Their work focuses on estimating the treatment effect or exposure effect consistently. Our paper is the first to examine the necessity of spatial-correlation robust inference under different scenarios for a general class of estimators. Lastly, we contribute to the literature on spatial econometrics. A summary of recent developments can be found in Xu and Lee (2019).

The remainder of the paper is organized as follows. We derive the asymptotic distribution for M-estimators under finite populations with spatial correlation in Section 2. In Section 3, the asymptotic theory is extended to functions of M-estimators. Among the

leading examples are average partial effect (APE) estimators resulting from nonlinear models. Using simulation studies in Section 4, we compare the small sample performance of the EHW, cluster-robust, and SHAC standard errors. The research is concluded in Section 5. Proofs are collected in the appendix.

## 2 Asymptotic Properties of M-estimators

### 2.1 Setup

Let  $D \subseteq \mathbb{R}^d$ ,  $d \geq 1$ , be a lattice of (possibly) unevenly placed locations in  $\mathbb{R}^d$ . Consider a sequence of finite subsets of  $D$ ,  $\{D_M\}$ , where  $M$  indexes the sequence of finite populations.  $|D_M|$  diverges to infinity in deriving the asymptotic properties, where  $|V|$  denotes the cardinality of a finite subset  $V \subseteq D$ . We adopt the metric  $\nu(i, j) = \max_{1 \leq l \leq d} |j_l - i_l|$  in space  $\mathbb{R}^d$ , where  $i_l$  is the  $l$ -th component of  $i$ . The distance between any subsets  $K, V \subseteq D$  is defined as  $\nu(K, V) = \inf\{\nu(i, j) : i \in K \text{ and } j \in V\}$ . For any random vector  $W$ ,  $\|W\|_p = (E \|W\|^p)^{1/p}$ ,  $p \geq 1$ , denotes its  $L_p$ -norm. Let  $\mathcal{F}_{iM}(s) = \sigma(U_{jM}; j \in T_M : \nu(i, j) \leq s)$  be the  $\sigma$ -field generated by the random vectors  $U_{jM}$  located in the  $s$ -neighborhood of location  $i$ . Lastly,  $C$  denotes a generic positive constant that may be different in different circumstances.

Let  $(X, z, Y) = \{(X_{iM}, z_{iM}, Y_{iM}), i \in D_M, M \geq 1\}$  and  $U = \{U_{iM}, i \in T_M, M \geq 1\}$  be triangular arrays of random fields defined on a probability space  $(\Omega, \mathcal{F}, P)$ . For each unit  $i$ , we observe  $(X_{iM}, z_{iM}, Y_{iM})$ , where  $X_{iM}$  is the vector of assignment variables,  $z_{iM}$  is a set of attributes, and  $Y_{iM}$  is the realized outcome. The categorization of assignments and attributes is based on the posed empirical question. Typically, the key variables of interest in an empirical study could be viewed as assignment variables, and the remaining covariates as attribute variables. There is no restriction in terms of the nature of the triple

above: they can be discrete, continuous, or mixed.

We relax the stable unit treatment value assumption in the standard potential outcome framework by allowing for interference among individuals. There exists a mapping, denoted by the potential outcome function  $y_{iM}(\mathbf{x}_M)$ , from a vector of assignment variables of all population units to the potential outcomes, where  $\mathbf{x}_M = \{x_{iM}, i \in D_M, M \geq 1\}$ .<sup>1</sup> The potential outcome function,  $y_{iM}(\cdot)$ , along with the observed attributes  $z_{iM}$ , are non-stochastic. By contrast, the assignment vector  $\mathbf{X}_M$  is random, where  $\mathbf{X}_M = \{X_{iM}, i \in D_M\}$  is the realization of  $\mathbf{x}_M$ . As a result, the realized potential outcome,  $Y_{iM} = y_{iM}(\mathbf{X}_M)$ , is random. Alternatively, the finite population setting can be understood as conditioned on the potential outcomes and attributes of the  $|D_M|$  units in the population. For the most part, we denote  $W_{iM} = (\mathbf{X}_M, Y_{iM})$  for brevity.

We do not take a stance on whether an underlying model is correctly specified. There currently exist discussions on incorporating interference into causal inference and how to consistently estimate the exposure effect with misspecified exposure mappings; see, for instance, Hudgens and Halloran (2008) and Sävje (2021). With misspecified interference, we estimate some approximation of the spillover effects in the sample. Manski (2013) discusses the identification of potential outcome distributions with social interactions, which is out of the scope of the current paper. Throughout, we assume that the finite population parameters are identified whether or not any feature of the model is correctly specified.

According to the sampling scheme, the population  $M$  can be partitioned into  $G_M$  mutually exclusive clusters,  $\{D_{gM} : g = 1, 2, \dots, G_M\}$ , based on the primary sampling units.  $C_{iM} \in \{1, 2, \dots, G_M\}$  denotes the cluster that unit  $i$  belongs to. Each cluster size is denoted by  $|D_{gM}|$  with  $|D_M| = \sum_{g=1}^{G_M} |D_{gM}|$ ,  $g \in \{1, 2, \dots, G_M\}$ . When the primary sampling units are individual entities, random sampling is included as a special case.

---

<sup>1</sup>I emphasize  $\mathbf{x}_M$  as the argument of the potential outcome function because its realization,  $\mathbf{X}_M$ , is the only stochastic vector in the function.

As is the starting point in the superpopulation paradigm, we study solutions to a population minimization problem, where the estimand of interest is a  $k \times 1$  vector denoted by  $\theta_M^*$ :

$$\begin{aligned}\theta_M^* &= \arg \min_{\theta} \frac{1}{|D_M|} \sum_{g=1}^{G_M} \sum_{j \in D_{gM}} \mathbb{E}[q_{jM}(W_{jM}, \theta)] \\ &= \arg \min_{\theta} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[q_{iM}(W_{iM}, \theta)]\end{aligned}\tag{1}$$

The interpretation of the finite population parameter is up to the researcher; see for instance, Rambachan and Roth (2022), for discussion on when the finite population estimand carries a causal interpretation. Notice that the expectation  $\mathbb{E}$  in equation (1) is taken over the distribution of  $X_M$  since  $X_M$  is the source of randomness here. The function  $q_{iM}(\cdot, \cdot)$  is the objective function for a single unit. The subscripts of the objective function indicate its dependence on the non-stochastic attribute variables  $\{z_{iM}, i \in D_M\}$ .

Let  $R_{iM}$  denote the binary sampling indicator, which is equal to one if unit  $i$  is sampled. Hence, the sample size is  $|D_N| = \sum_{g=1}^{G_M} \sum_{i \in D_{gM}} R_{iM} = \sum_{i \in D_M} R_{iM}$ . Below we will be precise about the nature of the sampling scheme but, unless the sample equals the population, the sample size is random. The spatial M-estimator of  $\theta_M^*$  is denoted by  $\hat{\theta}_N$ , which solves the minimization problem in the sample.

$$\begin{aligned}\hat{\theta}_N &= \arg \min_{\theta} \frac{1}{|D_N|} \sum_{g=1}^{G_M} \sum_{j \in D_{gM}} R_{jM} q_{jM}(W_{jM}, \theta) \\ &= \arg \min_{\theta} \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} q_{iM}(W_{iM}, \theta)\end{aligned}\tag{2}$$

In order to establish desirable asymptotic properties of the spatial M-estimator, we need to impose restrictions on the spatial dependence of the stochastic components in the estimation problem. We adopt the definition of mixing coefficients in Bradley and

Tone (2017), near-epoch dependent (NED) random fields in Jenish and Prucha (2012), and  $m$ -dependent random fields in Móricz, Stadtmüller, and Thalmaier (2008).

**Definition 1** *Let  $\mathcal{A}$  and  $\mathcal{B}$  be two sub- $\sigma$ -algebras of  $\mathcal{F}$ , and let*

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup(|P(AB) - P(A)P(B)|, A \in \mathcal{A}, B \in \mathcal{B})$$

*and*

$$\rho(\mathcal{A}, \mathcal{B}) = \sup |corr(f, g)|, f \in L_{real}^2(\mathcal{A}), g \in L_{real}^2(\mathcal{B}).$$

*For  $K \subseteq D_M$  and  $V \subseteq D_M$ , let  $\sigma_M(K) = \sigma(U_{iM}, i \in K)$  and  $\alpha_M(K, V) = \alpha(\sigma_M(K), \sigma_M(V))$ .*

*Then, the  $\alpha$ -mixing coefficient for the random field  $U$  is defined as:*

$$\bar{\alpha}(r) = \sup_M \sup_{K, V} (\alpha_M(K, V), \nu(K, V) \geq r).$$

*The maximal correlation coefficient is defined as:*

$$\bar{\rho}(r) = \sup_M \sup_{K, V} (\rho_M(K, V), \nu(K, V) \geq r).$$

**Definition 2** *Let  $W = \{W_{iM}, i \in D_M, M \geq 1\}$  be a random field, let  $U = \{U_{iM}, i \in T_M, M \geq 1\}$  be another random field, where  $|T_M| \rightarrow \infty$  as  $M \rightarrow \infty$ , and let  $d = \{d_{iM}, i \in D_M, M \geq 1\}$  be an array of finite positive constants. Then the random field  $W$  is said to be  $L_p(d)$ -near-epoch dependent on the random field  $U$  if*

$$\|W_{iM} - E(W_{iM} | \mathcal{F}_{iM}(s))\|_p \leq d_{iM} \psi(s)$$

*for some sequence  $\psi(s) \geq 0$  with  $\lim_{s \rightarrow \infty} \psi(s) = 0$ . The  $\psi(s)$  are called the NED coefficients, and the  $d_{iM}$  are called the NED scaling factors.  $W$  is said to be  $L_p$ -NED on  $U$  of*



size  $-\lambda$  if  $\psi(s) = O(s^{-\mu})$  for some  $\mu > \lambda > 0$ .

**Definition 3** A random field  $U = \{U_{iM}, i \in D_M, M \geq 1\}$  is called  $m$ -dependent if for all finite subsets  $K, V \subset D$  with  $\nu(K, V) > m$  the  $\sigma$ -algebras  $\sigma(U_{iM}, i \in K)$  and  $\sigma(U_{iM}, i \in V)$  are independent.

We make the following assumptions. Detailed regularity assumptions are listed in Appendix A.

**Assumption 1** Suppose  $\{D_M\}$  is a sequence of finite subsets of  $D$  such that  $|D_M| \rightarrow \infty$  as  $M \rightarrow \infty$ , where the lattice  $D \subseteq \mathbb{R}^d$ ,  $d \geq 1$ , is infinitely countable. All elements in  $D$  are located at distances of at least  $\nu_0 > 0$  from each other, i.e., for all  $i, j \in D$ :  $\nu(i, j) \geq \nu_0$ ; w.l.o.g. we assume that  $\nu_0 > 1$ .

**Assumption 2** (i) The sampling scheme consists of two steps. In the first step, a random group of clusters is drawn according to Bernoulli sampling with probability  $\rho_{cM} > 0$ ; in the second step, units are independently sampled, according to a Bernoulli trial with probability  $\rho_{uM} > 0$ , from the subpopulation consisting of all the sampled clusters. (ii) The sequence of sampling probabilities  $\rho_{cM}$  and  $\rho_{uM}$  satisfies  $\rho_{cM} \rightarrow \rho_c \in [0, 1]$ ,  $\rho_{uM} \rightarrow \rho_u \in [0, 1]$ , and  $|D_M| \rho_{uM} \rho_{cM} \rightarrow \infty$  as  $M \rightarrow \infty$ .

**Assumption 3**  $\max_{1 \leq g \leq G_M} |D_{gM}| \leq C < \infty$  as  $M \rightarrow \infty$ .

**Assumption 4** The sampling indicators,  $R = \{R_{iM}, i \in D_M, M \geq 1\}$ , are independent of the assignment variables,  $X = \{X_{iM}, i \in D_M, M \geq 1\}$ , and the underlying mixing random fields,  $U = \{U_{iM}, i \in T_M, M \geq 1\}$ , where  $D_M \subseteq T_M \subseteq D$ .

**Assumption 5** (Mixing condition) For the input random field  $U$ : (i)  $\bar{\alpha}(r) \rightarrow 0$  as  $r \rightarrow \infty$ ; (ii)  $\lim_{r \rightarrow \infty} \bar{\rho}(r) < 1$ .

**Assumption 6** (*NED condition*) The random field  $g = \{g_{iM}(W_{iM}, \theta), i \in D_M, M \geq 1\}$  is  $L_2$ -NED on  $U = \{U_{iM}, i \in T_M, M \geq 1\}$  with the scaling factors  $d_{iM}$  and the NED coefficients  $\psi(s)$  of size  $-2d(r-1)/(r-2)$  for some  $r > 2$ . The  $g(\cdot)$  function includes  $q_{iM}(W_{iM}, \theta)$ ,  $m_{iM}(W_{iM}, \theta)$ ,  $\nabla_\theta m_{iM}(W_{iM}, \theta)$ ,  $f_{iM}(W_{iM}, \theta)$ , and  $\nabla_\theta f_{iM}(W_{iM}, \theta)$  defined in Appendix A.

**Assumption 7** The weights satisfy:  $\omega(0) = 1$ ;  $\omega\left(\frac{\nu(i,j)}{b_M}\right) = 0$  for any  $\nu(i,j) > b_M$ ;  $\left|\omega\left(\frac{\nu(i,j)}{b_M}\right)\right| < \infty, \forall \nu(i,j)$  and  $\forall M$ ;  $\lim_{M \rightarrow \infty} \frac{1}{|D_M|} \left| \sum_{i \in D_M} \sum_{j \in D_M: \nu(i,j) \leq b_M} \left[ \omega\left(\frac{\nu(i,j)}{b_M}\right) - 1 \right] \cdot Cov\left(\frac{R_{iM}}{\sqrt{\rho_{uM}\rho_{cM}}} h_{iM}, \frac{R_{jM}}{\sqrt{\rho_{uM}\rho_{cM}}} h_{jM}\right) \right| = 0$ , where  $b_M = o((|D_M|\rho_{uM}\rho_{cM})^{1/2d})$  and the  $h(\cdot)$  function includes  $m_{iM}(W_{iM}, \theta_M^*)/J_M$  and  $(f_{iM}(W_{iM}, \theta_M^*) - F_M(\theta_M^*)H_M(\theta_M^*)^{-1}m_{iM}(W_{iM}, \theta_M^*))/J_M$  as defined in Appendix A.

Assumption 1 is taken from Jenish and Prucha (2012). Consistent with the increasing domain asymptotics, the assumption of the minimum distance ensures the expansion of the sample region. In Assumption 2, we introduce the possibility of cluster correlation via the two-stage sampling scheme. It is helpful to go through various sampling schemes resulted from different values of the sampling probabilities. With  $\rho_{cM} = \rho_{uM} = 1$ , we observe the entire population;  $\rho_{cM} = 1$  and  $\rho_{uM} < 1$  means random sampling;  $\rho_{cM} < 1$  and  $\rho_{uM} \leq 1$  implies cluster sampling. In particular, sampling from a superpopulation is nested in our unified theory since the sampling probabilities in both steps are allowed to be zero in the limit. When  $\rho_c = 0$ , only a negligible fraction of clusters are sampled from a population of a large number of clusters; while when  $\rho_c = 1$  and  $\rho_u = 0$ , a negligible portion of units are randomly drawn from a large population. Regardless, the expected sample size,  $|D_M|\rho_{uM}\rho_{cM}$ , diverges to infinity.

Assumption 3 imposes boundedness of the cluster sizes whenever the population units are partitioned into clusters because of either cluster assignment or cluster sampling. As

a result, the number of clusters in the population diverges to infinity along with the population size. Assumption 4 implies that the sampling process and the assignment process are independent of each other, which rules out sample selection due to assignment status.

Assumption 5 is the key weak dependence assumption in the proof of the asymptotic properties. Assignment variables are allowed to be spatially correlated as long as the spatial correlation dies out along with distance. In addition to the mixing condition, we also require the maximal correlation coefficient to be less than one in the limit. With the extension to NED processes in Assumption 6, we can study a more generalized class of random fields.<sup>2</sup> Here, we impose NED conditions on the objective function, score function, and the Hessian matrix directly so that both continuous and discrete variables are allowed. See, for example, Chapter 4 in Gallant and White (1988) for primitive conditions that ensure preservation of the NED property under transformations. Assumption 7 is required to establish consistency of the SHAC variance estimator. The last part of Assumption 7 is a high-level condition, which requires that the kernel weights  $\omega\left(\frac{\nu(i,j)}{b_M}\right)$  converge to one sufficiently fast as  $M \rightarrow \infty$ .

## 2.2 Asymptotic Distribution

We introduce the following notation for the variance-covariance matrix of M-estimators. Define

$$\begin{aligned} S_M = & \Delta_{ehw,M}(\theta_M^*) + \rho_{uM} \Delta_{cluster,M}(\theta_M^*) + \rho_{uM} \rho_{cM} \Delta_{spatial,M}(\theta_M^*) \\ & - \rho_{uM} \rho_{cM} \Delta_{E,M} - \rho_{uM} \rho_{cM} \Delta_{EC,M} - \rho_{uM} \rho_{cM} \Delta_{ES,M} \end{aligned} \quad (3)$$

---

<sup>2</sup>Assumption 6 is more than sufficient because in some cases, we only need  $L_1$ -NED, which can be implied by  $L_2$ -NED. For certain functions, the NED coefficients are only required to be of size  $-d$ .

and

$$V_M = H_M(\theta_M^*)^{-1} S_M H_M(\theta_M^*)^{-1}, \quad (4)$$

where

$$\Delta_{ehw,M}(\theta) = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[m_{iM}(W_{iM}, \theta) m_{iM}(W_{iM}, \theta)'], \quad (5)$$

$$\Delta_{E,M} = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)] \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)']', \quad (6)$$

$$\Delta_{cluster,M}(\theta) = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} = C_{jM}) \mathbb{E}[m_{iM}(W_{iM}, \theta) m_{jM}(W_{jM}, \theta)'], \quad (7)$$

$$\Delta_{EC,M} = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} = C_{jM}) \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)] \mathbb{E}[m_{jM}(W_{jM}, \theta_M^*)']', \quad (8)$$

$$\Delta_{spatial,M}(\theta) = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} \neq C_{jM}) \mathbb{E}[m_{iM}(W_{iM}, \theta) m_{jM}(W_{jM}, \theta)'], \quad (9)$$

$$\Delta_{ES,M} = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} \neq C_{jM}) \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)] \mathbb{E}[m_{jM}(W_{jM}, \theta_M^*)']', \quad (10)$$

and

$$H_M(\theta) = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[\nabla_{\theta} m_{iM}(W_{iM}, \theta)]. \quad (11)$$

**Theorem 2.1** *Under Assumptions 1-6, and Assumption A.1 in Appendix A,  $V_M^{-1/2} |D_N|^{1/2} (\hat{\theta}_N - \theta_M^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_k)$ .*

Theorem 2.1 is derived using a set of new limit theorems given in the Appendix. It shows that the M-estimators are asymptotically normal with an alternative finite population SHAC variance-covariance matrix. The result holds for spatial assignments both at the individual level or the cluster level. For the latter assignment design, there is cluster assignment in the first place, where assignments within clusters can be strongly correlated.

On top of that, the assignments across clusters are allowed to be spatially correlated as long as the key mixing condition in Assumption 5 is satisfied. With bounded cluster sizes, units far away from each other fall into different clusters and the spatial correlation eventually can die out. For instance, while studying individual outcomes, policies are imposed at the school district level and the school districts nearby may coordinate in certain ways.

**Remark 1** *We should only adjust standard errors for spatial correlation if: (i) assignment variables are spatially correlated; or (ii) spillover effects are specified in the model.*

Let us first focus on positive sampling probabilities. Based on the variance-covariance matrix in (3) and (4), we can see that sampling scheme plays a role in cluster correlation but not in spatial correlation. The two terms involving spatial correlation,  $\Delta_{spatial,M}(\theta_M^*)$  and  $\Delta_{ES,M}$ , cancel out whenever the assignment variables are independent. Therefore, it is only necessary to adjust the standard errors for spatial correlation if the assignment variables are spatially correlated either because of the assignment design or the inclusion of the spillover effects. Also notice that, even in the absence of spillover effects in the potential outcome function, we manually introduce spatial correlation among assignment variables when we explicitly specify spillover effects in the model. Here, assignment variables include your own “treatment” and your neighbors’ “treatments.” Another implication of Remark 1 is that we can test the necessity of spatial-correlation robust inference since assignment variables are observable to us.

We report the SHAC standard errors in the simulation below as one way to adjust standard errors for spatial correlation, as this is the dominant approach in the literature so far. Nevertheless, there are alternative ways to construct standard errors and confidence intervals robust to spatial correlation; see for instance, Müller and Watson (2022a) and Müller and Watson (2022b). Regardless, Remark 1 goes through in the population variance-covariance matrix.

**Remark 2** (i) When  $\rho_u = 0$ , reporting the EHW standard error would suffice; (ii) When  $\rho_c = 0$ , reporting the cluster-robust standard error would suffice.

When we switch to zero sampling probabilities, there are exceptions. In theory, when we observe the entire population or sample a large portion from a finite population, the spatial-correlation robust standard errors should be reported to account for spatial assignments. However, if we sample a minimal amount from the population, either the cluster-robust or the EHW standard errors would suffice under spatial assignments. Intuitively, when we independently sample a small proportion of units from the population, the majority of your neighbors would not be observed. Hence, there would be little difference between the EHW, cluster-robust, and the SHAC standard errors. In another case, when we randomly draw a small fraction of clusters from the population, most of your neighbors would be contained in the clusters drawn. Therefore, the cluster-robust standard errors take the majority of the spatial correlation into account and would deviate little from the SHAC standard errors with a sufficient bandwidth. In addition, when either of the sampling probability is zero, we are essentially sampling from a superpopulation, so the usual EHW standard error or the usual cluster-robust standard error would no longer be conservative.

### 2.3 Estimation of the Variance-Covariance Matrix

Define

$$\hat{V}_{SN} = \hat{H}_N(\hat{\theta}_N)^{-1} \hat{S}_N(\hat{\theta}_N) \hat{H}_N(\hat{\theta}_N)^{-1}, \quad (12)$$

where

$$\hat{H}_N(\theta) = \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \nabla_{\theta} m_{iM}(W_{iM}, \theta) \quad (13)$$

and

$$\hat{S}_N(\theta) = \frac{1}{|D_N|} \sum_{i \in D_M} \sum_{j \in D_M} R_{iM} R_{jM} \cdot \omega\left(\frac{\nu(i,j)}{b_M}\right) m_{iM}(W_{iM}, \theta) m_{jM}(W_{jM}, \theta)'. \quad (14)$$

**Theorem 2.2** *Under Assumptions 1-7, and Assumptions A.1-A.2 in Appendix A,  $\hat{V}_{SN} - (V_M + \rho_{uM}\rho_{cM}V_E) \xrightarrow{P} \mathbf{0}$ , where  $V_E = H_M(\theta_M^*)^{-1} S_E H_M(\theta_M^*)^{-1}$  and  $S_E = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \omega\left(\frac{\nu(i,j)}{b_M}\right) \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)] \mathbb{E}[m_{jM}(W_{jM}, \theta_M^*)]'$ .*

**Remark 3** *The usual SHAC variance estimator is conservative for the finite population SHAC variance-covariance matrix unless the sampling probabilities are zero.*

In Section 2.2, we have summarized when you should and should not adjust the standard errors for spatial correlation. In terms of the estimation of the variance matrix, Theorem 2.2 shows that when spatial correlation needs to be accounted for, there is an upward bias of the usual SHAC variance estimator. Leung (2022) reaches a similar conclusion for a weighted difference-in-means estimator for network data with interference, assuming the entire population is observed. We additionally introduce sampling probabilities and study both linear and nonlinear estimators.

We do not discuss the superpopulation limit of the extra term,  $S_E$ , as this requires weak dependence assumptions on the superpopulation, which we have not touched upon yet. However, since the usual SHAC variance estimator is consistent for the superpopulation variance-covariance matrix, the finite population SHAC variance matrix should be smaller than the superpopulation version of the variance matrix, in the matrix sense.

### 3 Asymptotic Distribution of Functions of M-estimators

Given that we study M-estimators, the APE estimator from nonlinear models would be of great interest. As a result, we study the asymptotic distribution of a generic function of M-estimators to complete the discussion.

Let  $f_{iM}(W_{iM}, \theta_M^*)$  be a  $q \times 1$  function of  $W_{iM}$  and  $\theta_M^*$ . We wish to estimate  $\gamma_M^* = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[f_{iM}(W_{iM}, \theta_M^*)]$ . Let  $\hat{\gamma}_N = \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} f_{iM}(W_{iM}, \hat{\theta}_N)$  be the estimator of  $\gamma_M^*$ . Denote the finite population variance matrix by

$$\begin{aligned} V_{f,M} = & \Delta_{ehw,M}^f + \rho_{uM} \Delta_{cluster,M}^f + \rho_{uM} \rho_{cM} \Delta_{spatial,M}^f \\ & - \rho_{uM} \rho_{cM} \Delta_{E,M}^f - \rho_{uM} \rho_{cM} \Delta_{EC,M}^f - \rho_{uM} \rho_{cM} \Delta_{ES,M}^f. \end{aligned} \quad (15)$$

And the usual SHAC variance estimator is denoted by  $\hat{V}_{f,SN}$ . The detailed definition of each term can be found in Appendix A.

**Theorem 3.1** *Under Assumptions 1-7, and Assumptions A.1-A.3 in Appendix A, (1)  $V_{f,M}^{-1/2} |D_N|^{1/2} (\hat{\gamma}_N - \gamma_M^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_q)$ ; (2)  $\hat{V}_{f,SN} - (V_{f,M} + \rho_{uM} \rho_{cM} V_{f,E}) \xrightarrow{p} \mathbf{0}$ .*

We see that the same results for M-estimators carry over to functions of M-estimators. For the APE estimator, we should only report the spatial-correlation robust standard errors if assignments are spatially correlated or spillover effects are estimated. Additionally, the usual SHAC standard errors are supposed to be conservative. However, as a well-known fact, the SHAC standard errors suffer from downward bias when the spatial correlation is high. Hence, the actual finite sample performance of the usual SHAC standard errors remains unclear.



## 4 Simulation Designs

We consider an uneven lattice. The population units lie within a square of dimension  $\sqrt{M} \times \sqrt{M}$ , where  $M$  is the population size. The locations  $(s_{1,iM}, s_{2,iM})$  are drawn once and kept fixed across designs.  $s_{1,iM} \sim \mathcal{U}(0, \sqrt{M})$ ,  $s_{2,iM} \sim \mathcal{U}(0, \sqrt{M})$ , and they are independent of each other. The distance between units  $i$  and  $j$  is measured by  $\nu(i, j) = \max(|s_{1,iM} - s_{1,jM}|, |s_{2,iM} - s_{2,jM}|)$ . We have four major designs involving spatial assignments at the individual level, spatial assignments at the cluster level, spatial assignments allowing for spillover effects, and spatial assignments in nonlinear models.

### 4.1 Spatial Correlation at the Individual Level

The potential outcome function is given below:

$$y_{ig}(x_{ig}) = a \cdot \beta_{ig}x_{ig} + c_g + u_{ig}, \quad (16)$$

where half of  $\beta_{ig}$  are equal to 1 and the other half -1.  $a$  is a constant scalar. The group unobserved heterogeneity  $c_g$  is independently drawn from the standard normal distribution and remains fixed across replications. Without cluster sampling, we can ignore the  $g$  subscript. The individual unobservables are generated from the linear spatial autoregressive model below,

$$u_M = p_u W_u u_M + \epsilon_M, \quad (17)$$

where  $u_M$  and  $\epsilon_M$  are  $M \times 1$  vectors.  $\epsilon_M$  are i.i.d. draws from a standard normal distribution and kept fixed.  $W_u$  is a contiguity matrix and units  $i$  and  $j$  are neighbors if  $\nu(i, j) \leq \sqrt{2}$ . It is row-standardized with the diagonal elements being zero.

There are two sub-designs. In the first design, the assignment variables are i.i.d. draws from a Bernoulli distribution with a probability of 0.5 for each replication. The spatial

correlation of the individual unobservables (in the sense of superpopulation) depends on  $p_u$ , which varies from 0 to 0.9 with an increment of 0.1. In the second sub-design, the assignment variables are binary variables equal to one when the input value  $\xi_{iM}$  is greater than or equal to its population average  $\sum_{i=1}^M \xi_{iM}/M$ , where  $\xi_M$  is an  $M \times 1$  vector drawn from a multivariate normal distribution with mean zero and a variance-covariance matrix equal to  $p_x$  raised to the power of the distance.  $p_u$  is fixed at 0.3, while  $p_x$  takes value from 0 to 0.9. Therefore, individual assignments can be spatially correlated.

The expected size of each dimension of the lattice is 18, unless otherwise noted, which leads to an expected sample size of 324. The clusters in the sampling scheme are formed by grouping the consecutive three units by order, resulting in an expected number of 108 clusters in the sample. There are five sampling schemes: (i) we observe the entire population; (ii) we independently sample clusters from all the clusters in the population with a probability of 0.25; (iii) we independently draw units from the entire population with a probability of 0.25; (iv) we independently sample clusters from all the clusters in the population with a probability of 0.01; (v) we independently draw units from the entire population with a probability of 0.01. The last two sampling schemes mimic cluster sampling and independent sampling from the infinite population, respectively. For the last two sampling schemes, the expected size of each dimension of the lattice is decreased to 12 to reduce the computational burden. The constant  $a = 2$  for the first three sampling schemes and  $a = 1$  for the last two.

We first regress  $Y_i$  on 1 and  $X_i$  and report different standard errors of the slope coefficient estimator. For cluster sampling with probability 0.25, we also report results for fixed effect by demeaning variables within clusters. The standard errors among comparison are the EHW standard errors, the cluster-robust standard errors, and the SHAC standard errors. We use the Parzen kernel to estimate the SHAC standard errors. The SHAC

standard errors are highly sensitive to the choice of bandwidth. In the first sub-design, where assignments are independent, we report the SHAC standard errors with bandwidth  $d^* \in \{1, 2, 3\}$ . In the second sub-design, we report the SHAC standard errors with two bandwidths chosen in the following way. The SHAC standard errors are estimated using bandwidth up to 20 with a distance increment of one. Among the 20 bandwidths, We choose either the one that minimizes the mean square error or the one that minimizes the bias of the SHAC standard error with respect to the Monte Carlo standard deviation of the slope coefficient estimator. The number of iterations is 1,000.

#### 4.1.1 Independent Assignments with Spatially Correlated Unobservables

When the assignments are independent with the entire population observed, we see from Figure 1 that all standard errors are larger than the Monte Carlo standard deviation of the slope coefficient estimator and the corresponding coverage rates of the 95% confidence interval are almost all above the benchmark line of 0.95.<sup>3</sup> This is expected because the superpopulation standard errors are supposed to be conservative. Among all the standard errors reported, the EHW standard errors are the closest to the Monte Carlo standard deviation, regardless of the spatial correlation of the unobservables in the superpopulation. As well, the coverage rate of the 95% confidence interval based on the EHW standard errors is closest to the theoretical level. The cluster-robust standard errors are generally too large and the SHAC standard errors increase along with the bandwidth. With independent assignments, the latter two standard errors are unnecessarily conservative.

When we sample clusters from the population, things become slightly different. Although assignments are still independent, cluster sampling introduces cluster correlation.

---

<sup>3</sup>In the top panel of all figures below, “oracle” means the Monte Carlo standard deviation of the coefficient estimator or the APE estimator. While in the bottom panels, “oracle” stands for the benchmark coverage rate of the 95% confidence interval, 0.95.

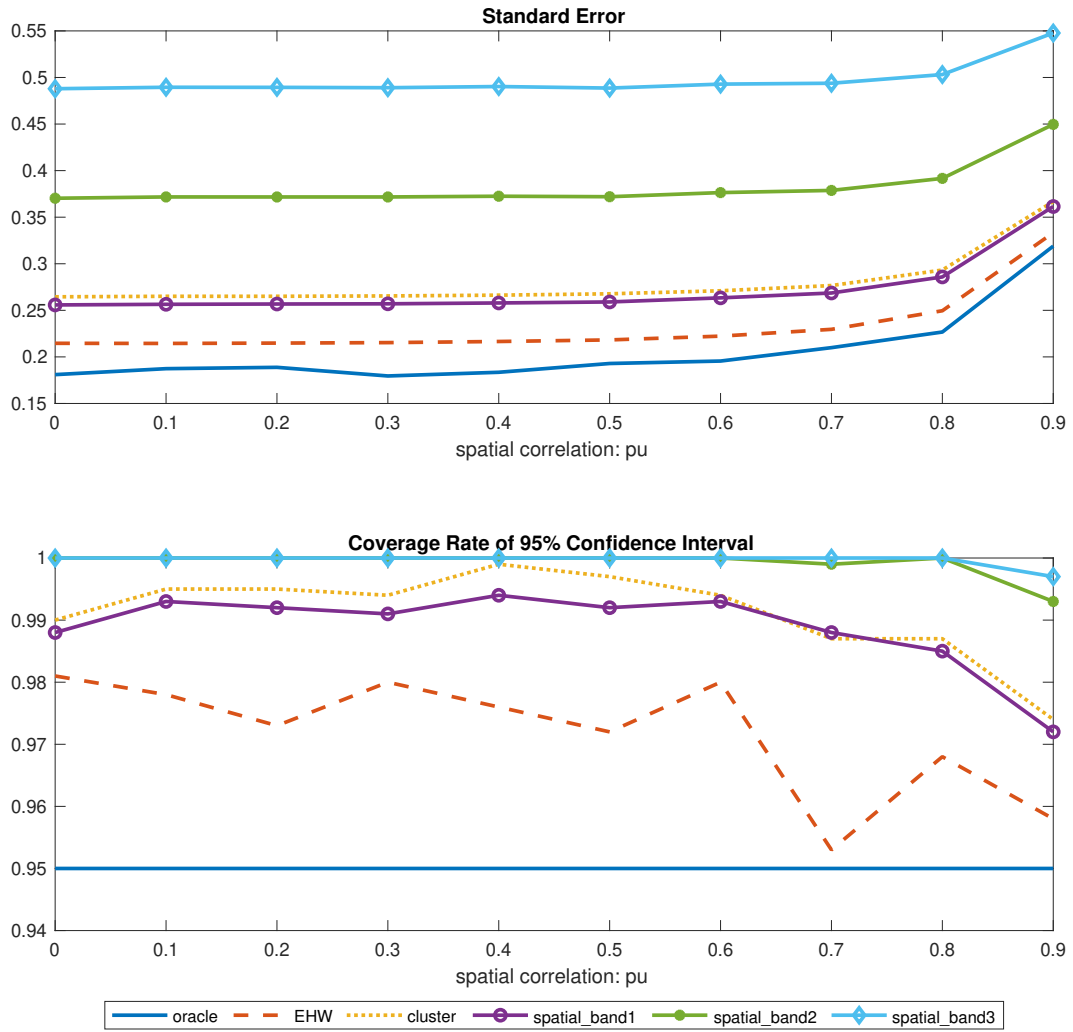


Figure 1: Independent Assignments Observing Entire Population

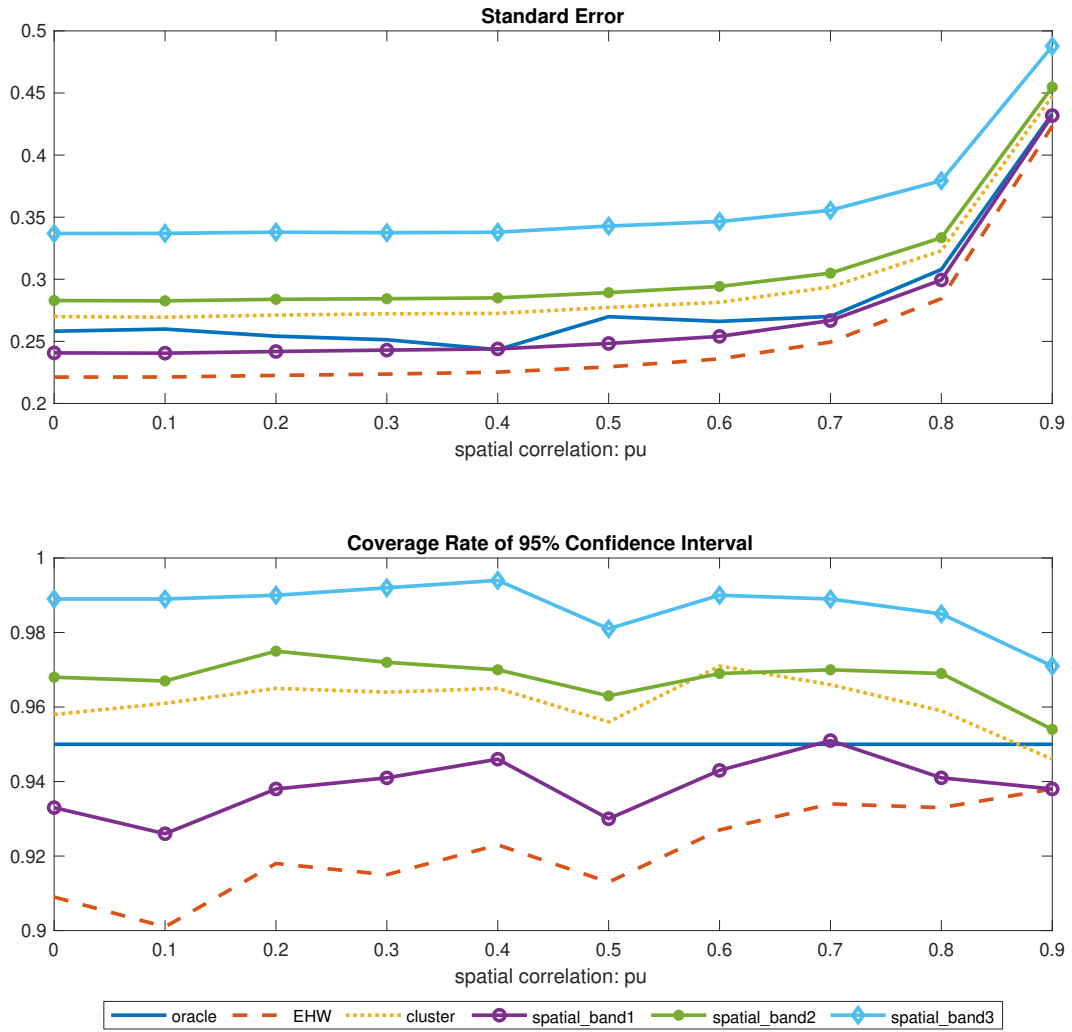


Figure 2: Independent Assignments with Cluster Sampling 0.25

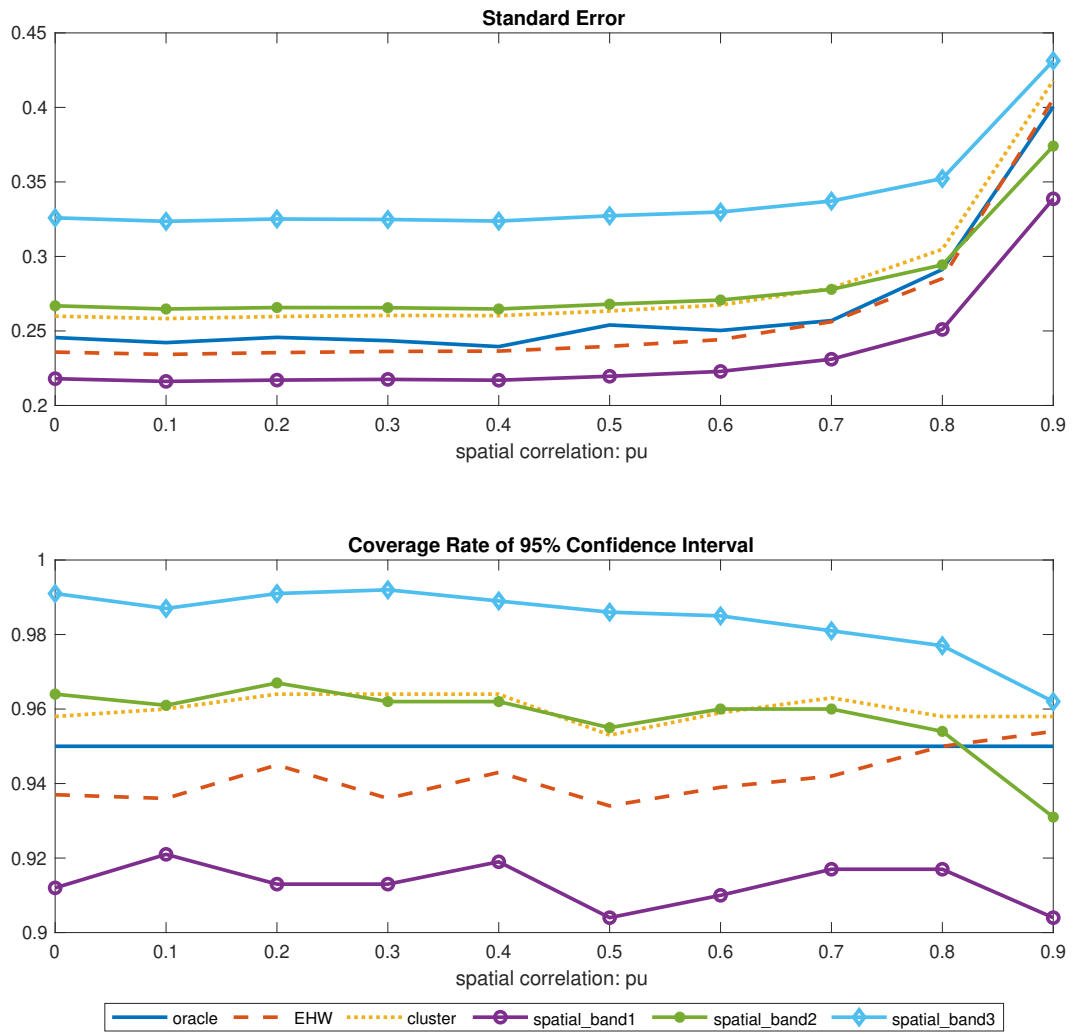


Figure 3: Independent Assignments with Cluster Sampling 0.25: fixed effects

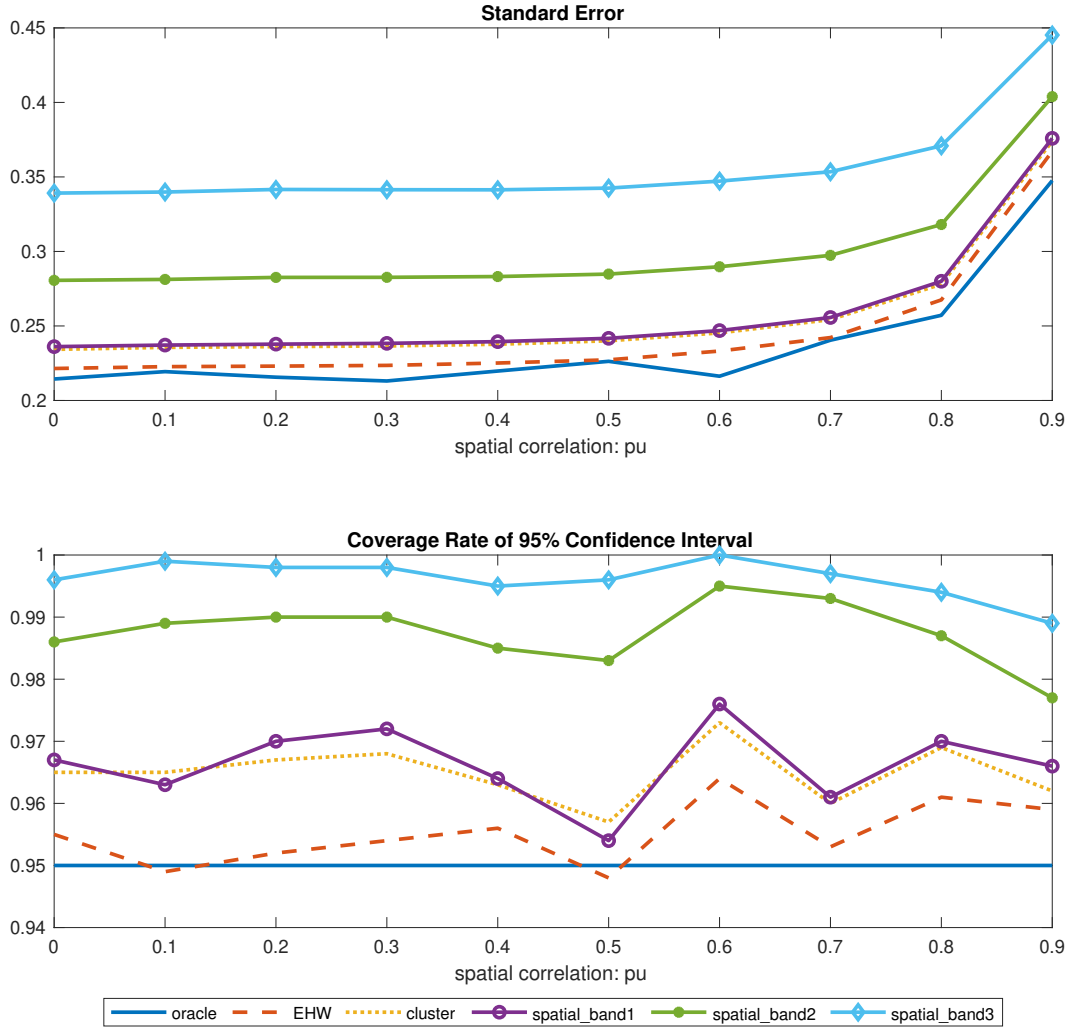


Figure 4: Independent Assignments with Independent Sampling 0.25

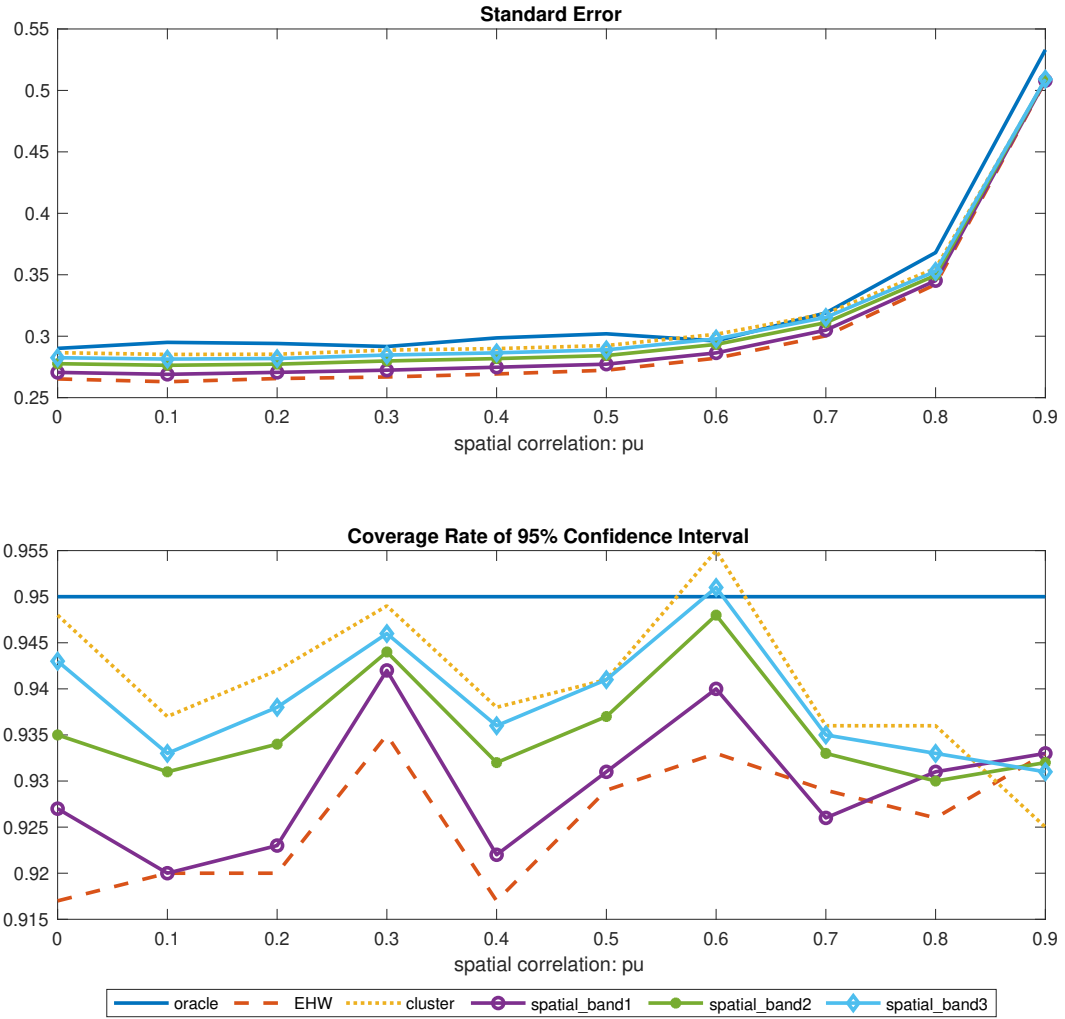


Figure 5: Independent Assignments with Cluster Sampling 0.01



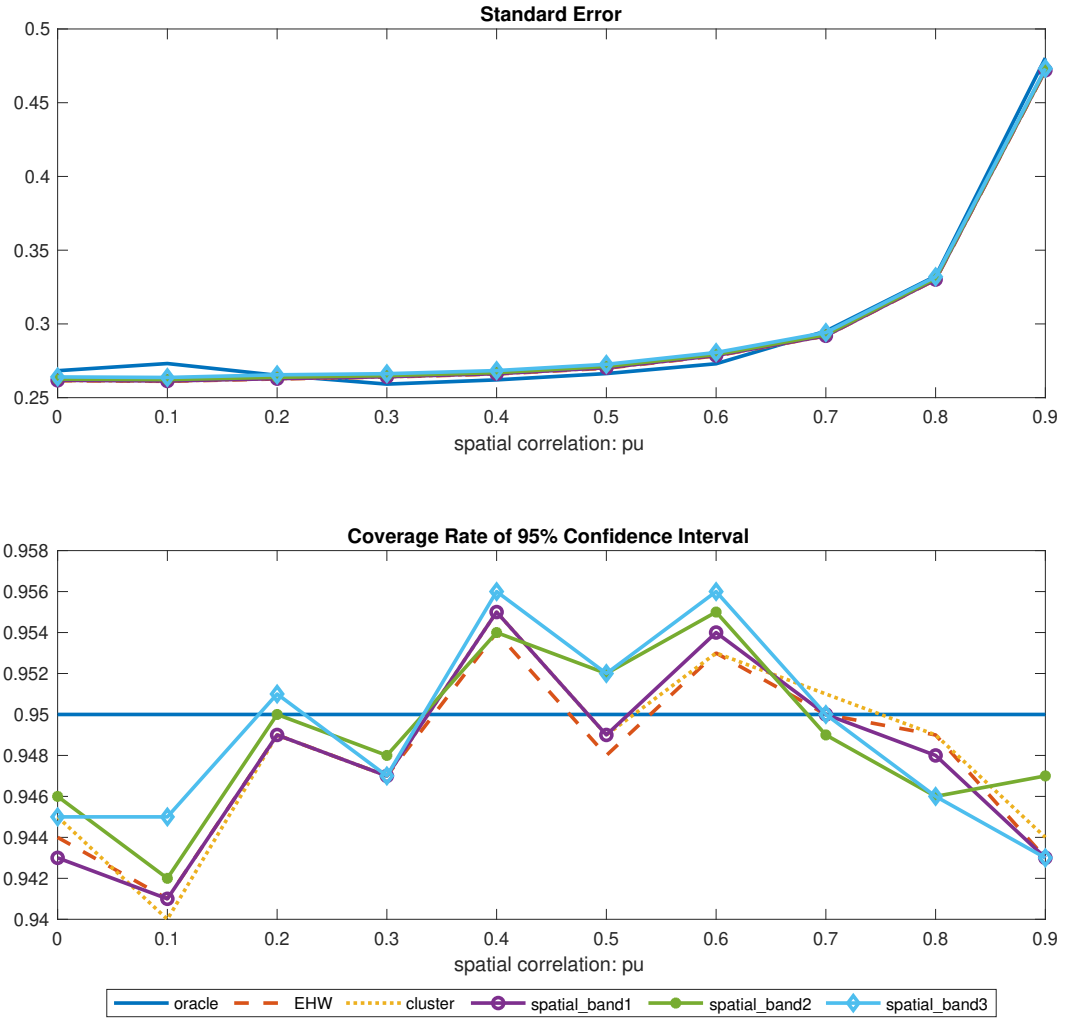


Figure 6: Independent Assignments with Independent Sampling 0.01

As a result, the cluster robust standard error is the one closest to the Monte Carlo standard deviation and the coverage rate of the corresponding confidence interval hovers around the benchmark line. The EHW standard errors are too small, as shown in Figures 2 and 5. We observe the same pattern for the standard errors of the fixed effect estimators. However, the comparison between the cluster robust standard errors and the SHAC standard errors depends on the sampling probability. When the sampling probability of each cluster is nonnegligible, the SHAC standard errors grow along with the bandwidth and eventually become too conservative. On the other hand, when the sampling probability is as small as 0.01, resembling cluster sampling from an infinite population, the difference between the cluster-robust standard errors and the SHAC standard errors narrows down. This is especially true when the chosen bandwidth contains enough nearby units.

When we independently draw units from the population, the EHW standard errors again turn out to be the appropriate one to report. Similarly, when the sampling probability is 0.25, you see a clear discrepancy among the standard errors. When the sampling probability is 0.01, which resembles the case of independent sampling from an infinite population, the EHW, cluster, and SHAC standard errors are almost identical to each other.

#### 4.1.2 Spatial Assignments

In the second sub-design, spatial assignments are introduced through the spatial correlation parameter,  $p_x$ . When  $p_x = 0$ , we are back to the case of independent assignments with  $p_u = 0.3$ . As we can see from Figures 7 and 10, the EHW standard errors are the ones with the best performance when we either observe the entire population or independently sample units from the population. With cluster sampling, the EHW standard errors underestimate the standard deviation and the SHAC standard errors are quite similar to the cluster-robust

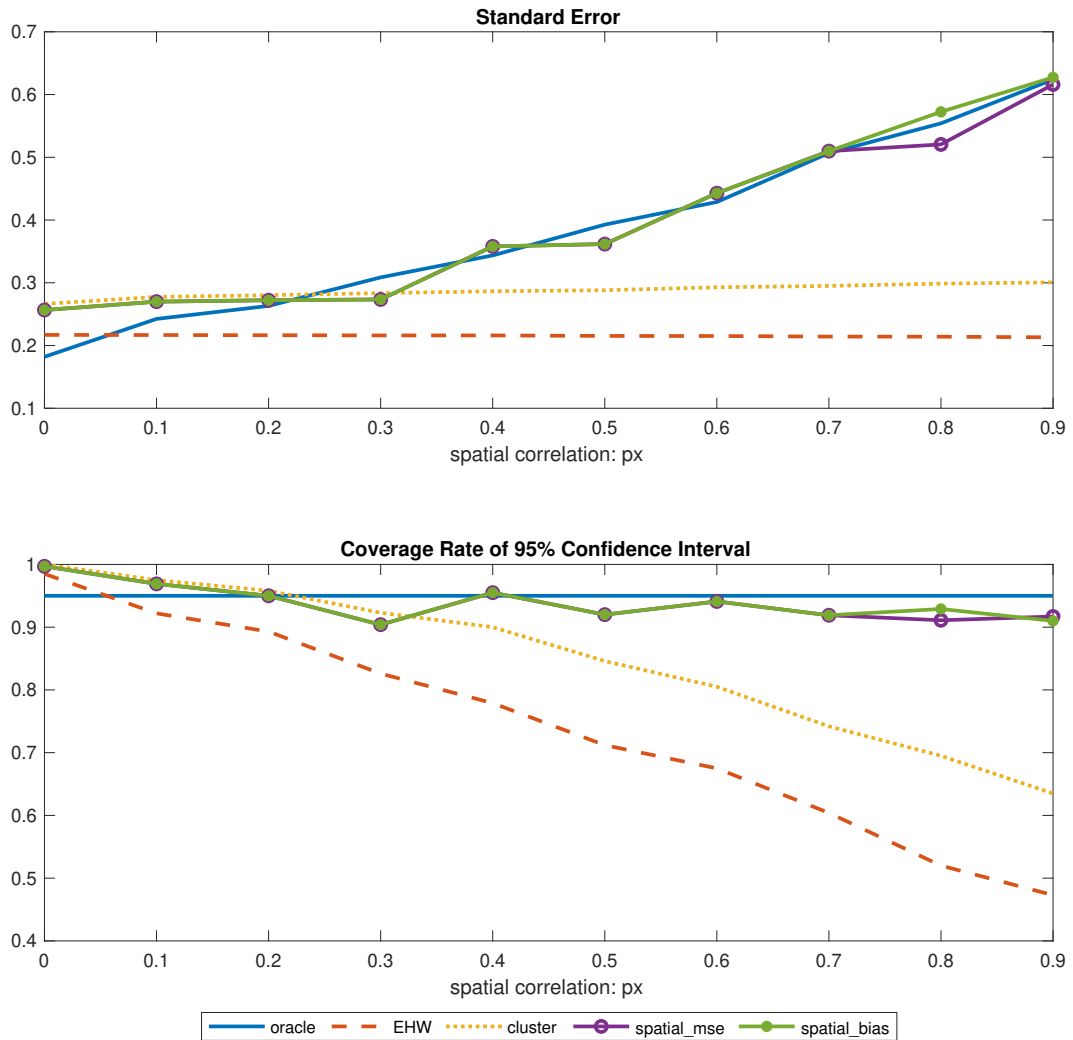


Figure 7: Spatial Assignments Observing Entire Population

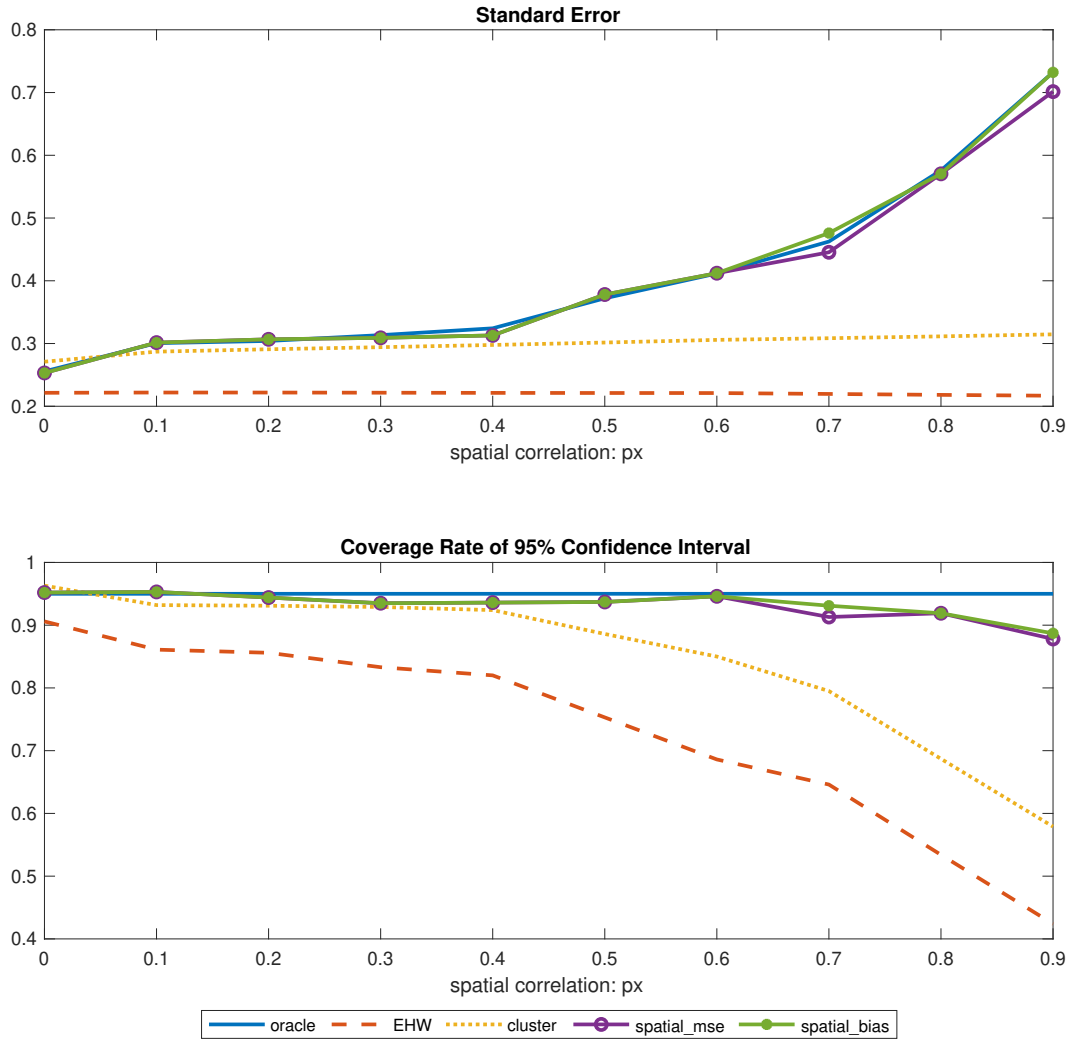


Figure 8: Spatial Assignments with Cluster Sampling 0.25

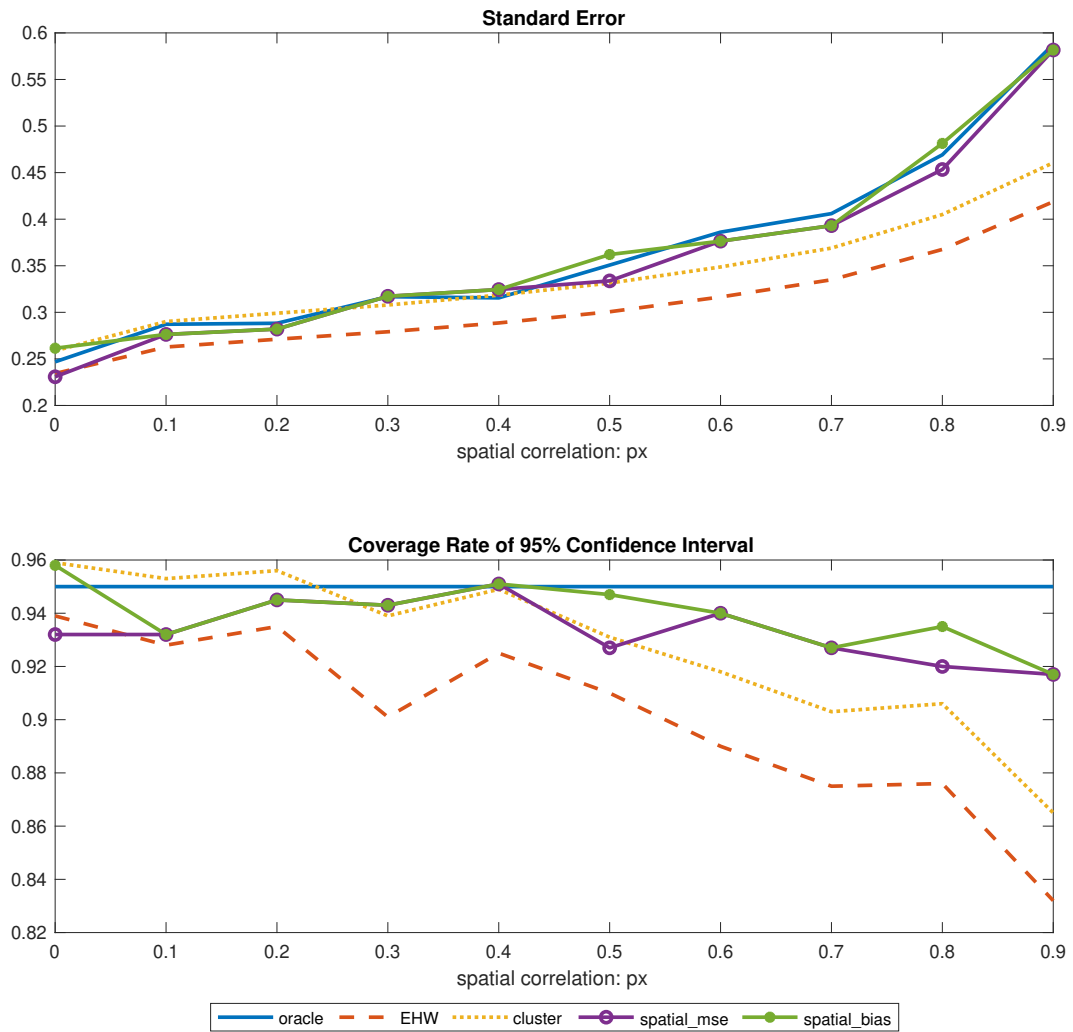


Figure 9: Spatial Assignments with Cluster Sampling 0.25: fixed effects

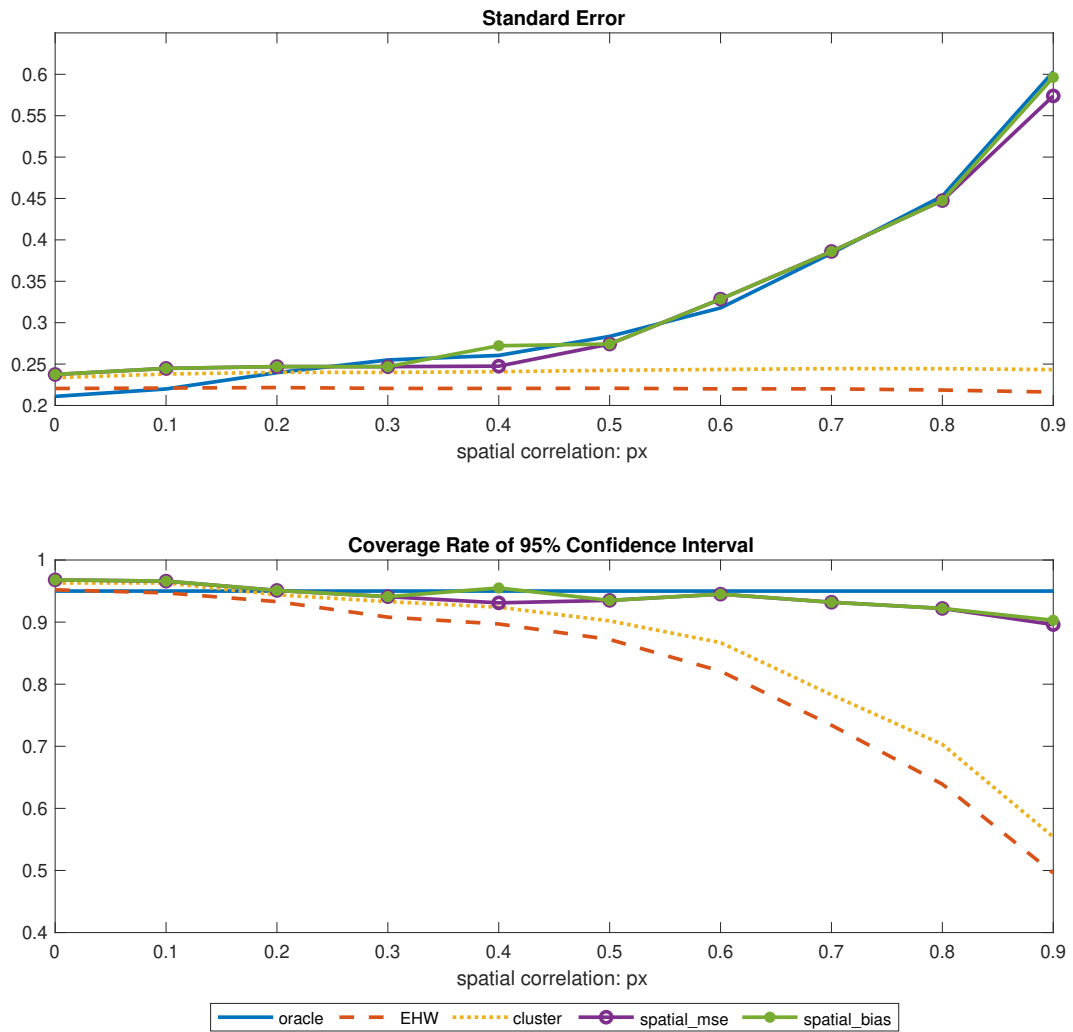


Figure 10: Spatial Assignments with Independent Sampling 0.25

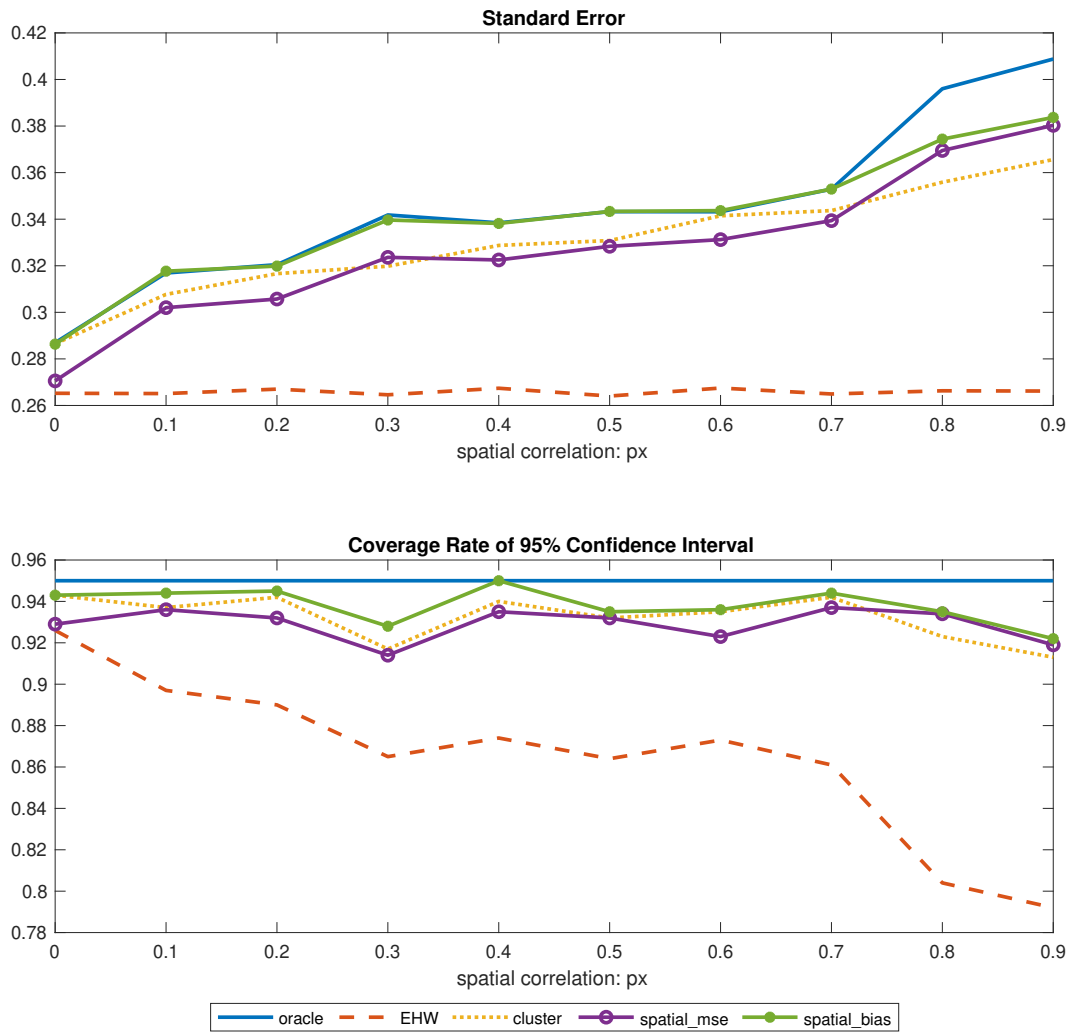


Figure 11: Spatial Assignments with Cluster Sampling 0.01

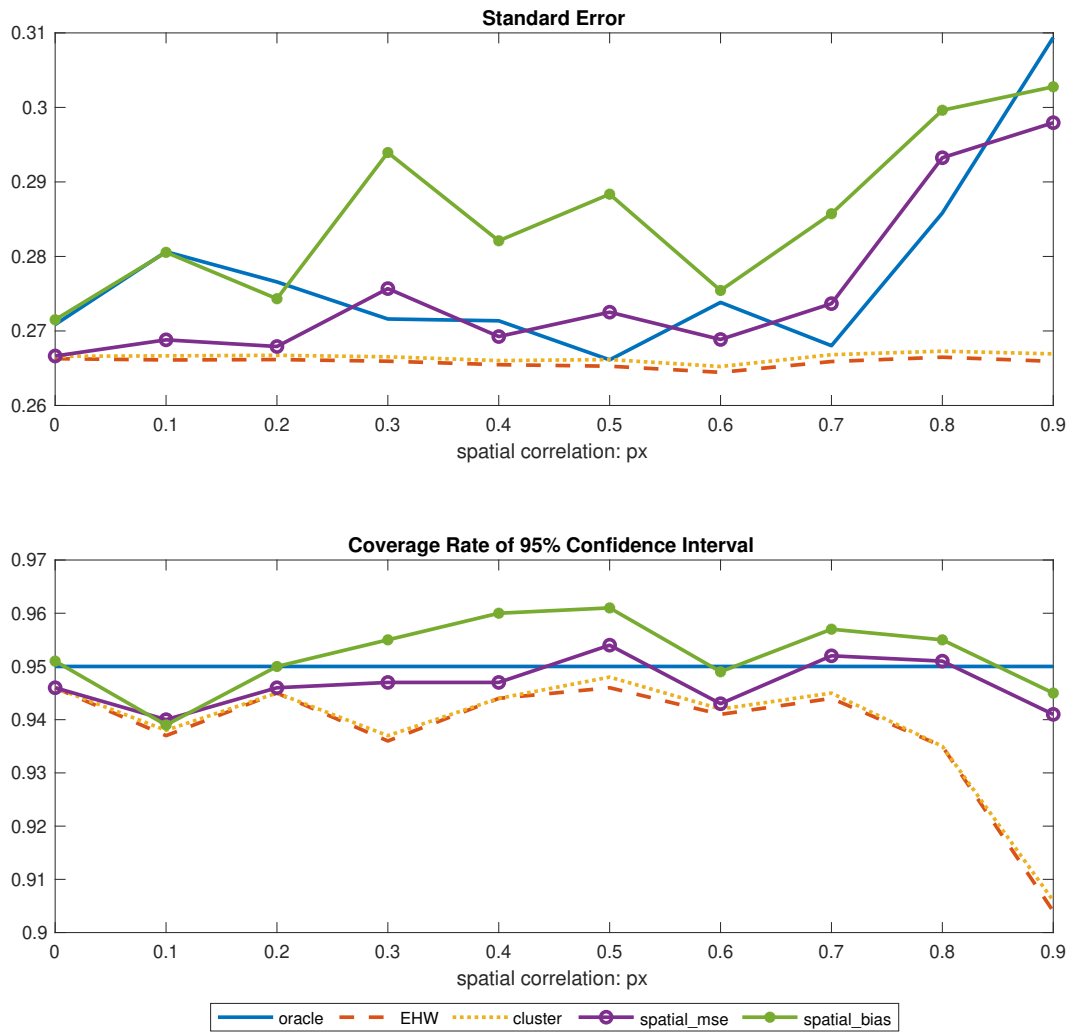


Figure 12: Spatial Assignments with Independent Sampling 0.01



standard errors. Nevertheless, with spatial assignments we see that both the EHW and the cluster-robust standard errors gradually become too small when the spatial correlation among the assignment variables increases unless the sampling probability is as small as 0.01. The same observation holds for both pooled OLS and fixed effect estimators when there is cluster partition.

Though not perfect, we see patterns along the lines of the theoretic prediction in Remark 2 when the sampling probability is 0.01. With cluster sampling, the gap between the SHAC standard errors and the cluster-robust standard errors is much smaller even with high spatial correlation compared with the case of a larger sampling probability. Similarly, with independent sampling, the difference among the EHW, cluster, and the SHAC standard errors is relatively small compared with Figures 7-10. All standard errors suffer from some downward bias. However, keep in mind that the sample size is only over 100 and we allow the correlation parameter,  $p_x$ , to be as large as 0.9, which results in the correlation between the assignment variables within a distance of one averaging at 0.85. I expect that the discrepancy among the different standard errors would gradually disappear when the sampling probability becomes even lower, especially when the spatial correlation is not too high. However, this would drastically increase the computation burden by enlarging the size of the lattice to a great extent. Hence, results are not reported here due to computation restrictions.

When the ratio of the sample to the population size is small, the intuition of reporting either the EHW or the cluster-robust standard errors remains the same no matter whether we introduce spatial correlation at the cluster level, spillover effects, or nonlinearity later on. Hence, for the simulation designs below, we omit the results for sampling probabilities of 0.01.

## 4.2 Spatial Assignments at the Cluster Level

In this design, assignments are imposed at the cluster level, and we introduce spatial correlation across clusters. The potential outcome function and the individual unobservables are the same as in equations (16) and (17) with  $a = 1$ .  $p_u$  is also fixed at 0.3. On the contrary, the assignments are generated differently. We construct a contiguity matrix among cluster pairs, where the distance between clusters is measured as the minimum distance of units in the cluster pair. Hence, the cluster contiguity matrix  $W_G$  is a  $G \times G$  matrix, where  $G$  is the number of clusters in the population and clusters  $l$  and  $m$  are neighbors if  $\nu(l, m) \leq 2$ . We adopt the first three sampling schemes as in the baseline design.

The assignment variables are generated in the following way:

$$\tilde{X}_G = p_x W_G \tilde{X}_G + \xi_G, \quad (18)$$

where  $\tilde{X}_G$  and  $\xi_G$  are  $G \times 1$  vectors, and  $\xi_g \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ ,  $g = 1, \dots, G$ . The cluster assignment  $X_g = \mathbb{1}\{\tilde{X}_g > \frac{1}{G} \sum_{l=1}^G \tilde{X}_l\}$ ,  $\forall g = 1, 2, \dots, G$ . Units within the same cluster receive the same assignment.

When the spatial correlation is imposed at the cluster assignment variables rather than the individual assignment variables, the comparison of the standard errors in the baseline design carries over. In summary, we should report the SHAC standard errors under spatial assignments. In the absence of spatial assignments, we should report the cluster-robust standard errors because cluster assignments always occur regardless of sampling scheme. However, to account for the cluster correlation in addition to the spatial correlation across clusters, I introduce a slightly different distance measure based on the distance between cluster pairs. Using this cluster distance measure, units within the same cluster receive the same weights when estimating the SHAC standard errors. The SHAC standard errors

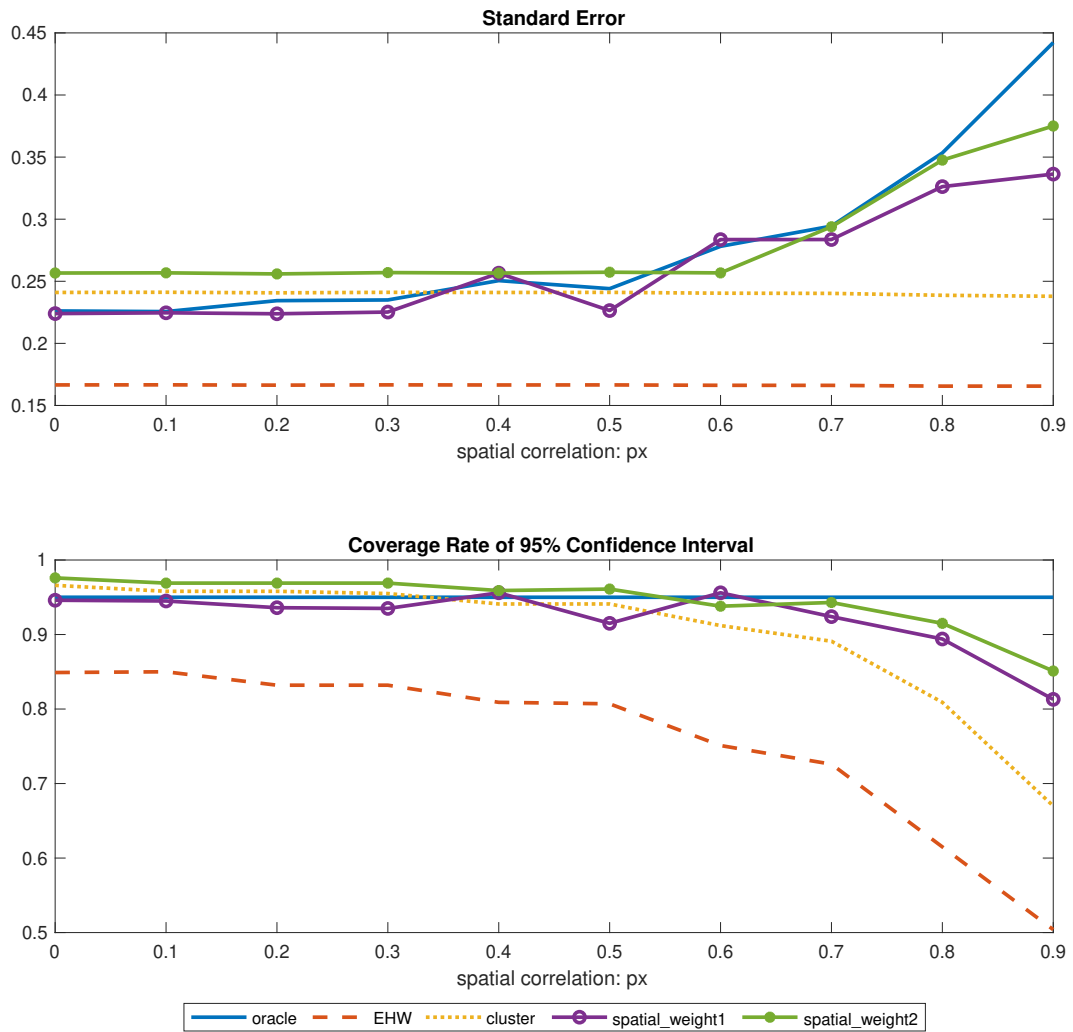


Figure 13: Spatial Assignments at the Cluster Level Observing Entire Population

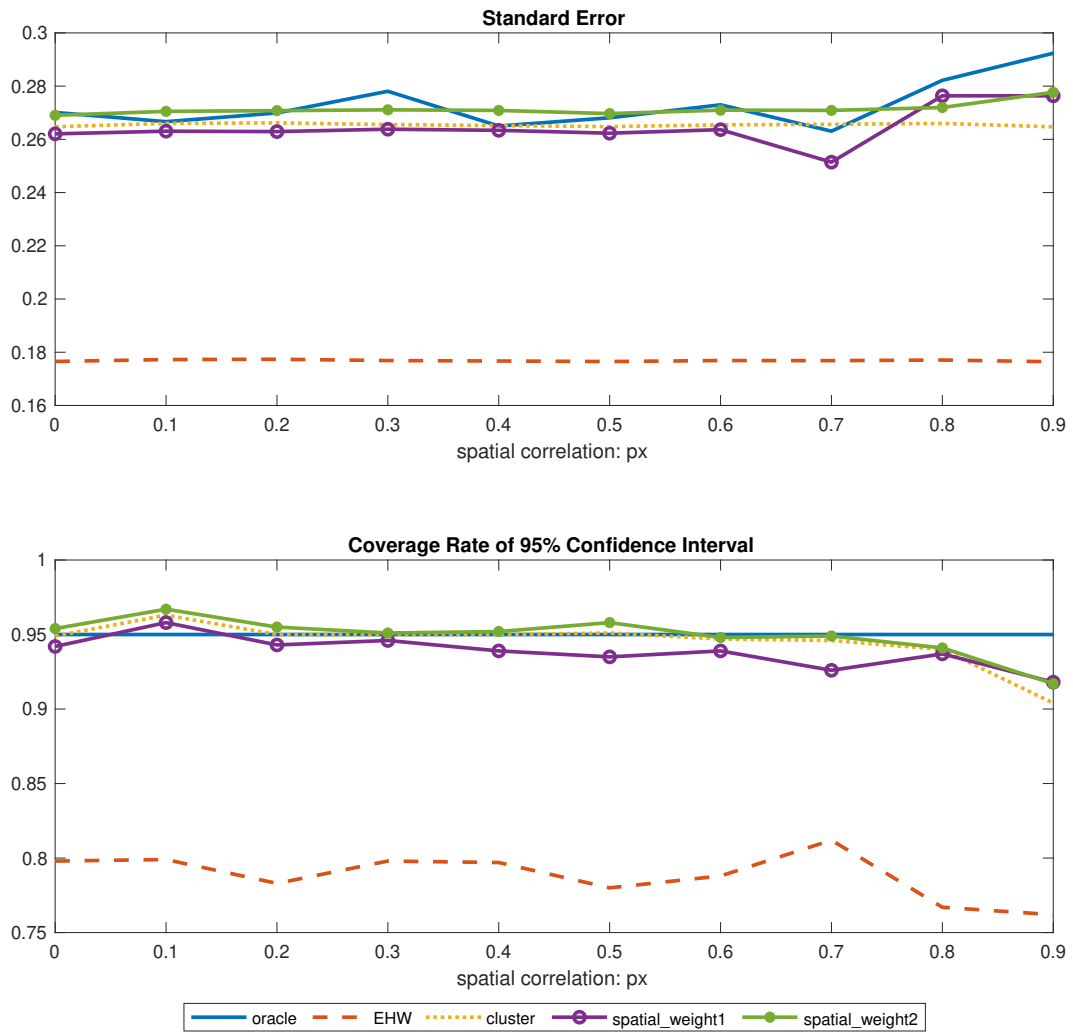


Figure 14: Spatial Assignments at the Cluster Level with Cluster Sampling 0.25

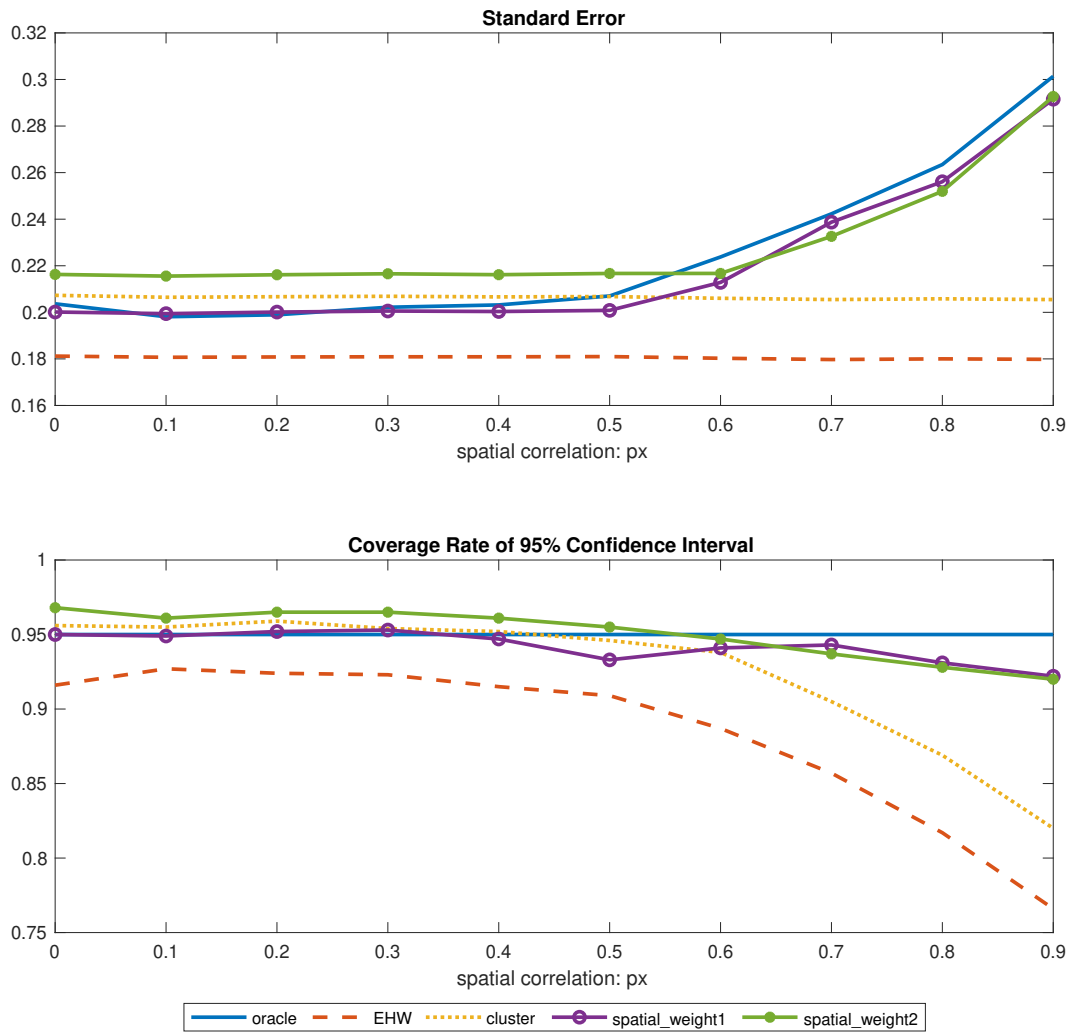


Figure 15: Spatial Assignments at the Cluster Level with Independent Sampling 0.25

based on the cluster distance slightly outperform the ones based on the unit distance. This is because the latter typically suffer from downward bias even when the spatial correlation is low or moderate. However, the former can also become too conservative when there is a lot of heterogeneity across units.

### 4.3 Spillover Effects

In this simulation design, we allow for spillover effects in addition to spatial assignments by setting up a linear-in-means type of model. The expected sample size of each dimension of the lattice is 36. The potential outcome function is given below:

$$y_{ig}(\mathbf{x}_M) = 2\beta_{ig}x_{ig} + \gamma W_x \mathbf{x}_M + \epsilon_{ig}, \quad (19)$$

where  $\mathbf{x}_M$  is the  $M \times 1$  vector containing assignments for all units in the population. The realized assignments  $\mathbf{X}_M$  follows a multivariate normal distribution with mean zero and a variance-covariance matrix equal to  $p_x$  raised to the power of the distance. The contiguity matrix,  $W_x$ , is constructed in the same way as  $W_u$ . Elements in  $W_x$  equal one if the corresponding distance is less than or equal to 0.5.  $\beta_{ig}$  are defined in the same way as in Section 4.1. The unobservables  $\epsilon_{ig}$  are independently drawn from a standard normal distribution and kept fixed.

We adopt the first three sampling schemes in the baseline design. Within each sampling scheme, we have three assignment and spillover combinations: (i)  $p_x = 0$  and  $\gamma = 0$ ; (ii)  $p_x = 0$  and  $\gamma = 1$ ; (iii)  $p_x = 0.1$  and  $\gamma = 1$ . Each represents the case of no spatial assignments and no spillover effects, no spatial assignments with spillover effects, and spatial assignments combined with spillover effects, respectively. Regardless, we estimate spillover effects by regressing  $Y_i$  on 1,  $X_i$ , and  $W_s \mathbf{X}_s$ , where  $\mathbf{X}_s$  is the collection of assignments of all units in the sample and  $W_s$  is a  $N \times N$  contiguity matrix with a distance cutoff of 0.5.

Table 1: Specifying Spillover Effects:  $\gamma$ 

	entire population			cluster sampling			independent sampling		
	$p_x = 0,$ $\gamma = 0$	$p_x = 0,$ $\gamma = 1$	$p_x = 0.1,$ $\gamma = 1$	$p_x = 0,$ $\gamma = 0$	$p_x = 0,$ $\gamma = 1$	$p_x = 0.1,$ $\gamma = 1$	$p_x = 0,$ $\gamma = 0$	$p_x = 0,$ $\gamma = 1$	$p_x = 0.1,$ $\gamma = 1$
coeff	0.002	1.002	1.061	0.004	0.636	0.546	-0.001	0.500	0.472
std	0.133	0.110	0.136	0.113	0.128	0.169	0.137	0.149	0.160
EHW	0.101	0.083	0.100	0.087	0.094	0.123	0.103	0.112	0.121
EHW_CI	(0.863)	(0.860)	(0.855)	(0.860)	(0.843)	(0.836)	(0.854)	(0.853)	(0.848)
cluster	0.107	0.096	0.109	0.109	0.115	0.150	0.111	0.121	0.129
cluster_CI	(0.887)	(0.910)	(0.887)	(0.938)	(0.924)	(0.911)	(0.883)	(0.883)	(0.875)
SHAC1	0.129	0.107	0.135	0.112	0.122	0.167	0.134	0.145	0.159
SHAC1_CI	(0.934)	(0.938)	(0.950)	(0.943)	(0.941)	(0.947)	(0.936)	(0.935)	(0.935)
SHAC2	0.129	0.108	0.136	0.113	0.123	0.169	0.135	0.146	0.160
SHAC2_CI	(0.936)	(0.941)	(0.936)	(0.941)	(0.943)	(0.941)	(0.939)	(0.933)	(0.937)

Table 1 reports the coefficient estimates and the standard errors of the spillover effects coefficient. Even though all standard errors suffer from downward bias in all designs, the SHAC standard errors perform the best among the three classes of standard errors. When we include spillover effects in our models, regardless of the sampling scheme we use or whether there are spillover effects in the potential outcome function or not, we should always report the spatial-correlation robust standard errors.

It is worth mentioning that without observing the entire population, what we are identifying is the spillover effects in the sample because we can only observe our neighbors who happen to be selected into the sample, which is an imperfect measure of the spillover in the population. As a result, the coefficient estimator on the spillover effects is biased in the latter two sampling schemes. With cluster sampling, the spillover effects estimator seems to be less biased given that more neighbors are included in the cluster sample.

The coefficient estimates on the individual assignment variable are also reported in Table 2, along with the standard errors. Depending on the sampling scheme, reporting the EHW or cluster-robust standard errors is sufficient. Here, the individual assignment

Table 2: Coefficient on the Individual Assignment Variable with Independent Assignments

	entire population		cluster sampling		independent sampling	
	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
coeff	-0.002	0.000	-0.001	0.001	-0.001	0.009
std	0.087	0.084	0.116	0.118	0.094	0.098
EHW	0.100	0.100	0.100	0.101	0.100	0.101
EHW_CI	(0.977)	(0.979)	(0.908)	(0.900)	(0.958)	(0.957)
cluster	0.127	0.127	0.127	0.127	0.107	0.108
cluster_CI	(0.994)	(0.997)	(0.961)	(0.975)	(0.971)	(0.972)
SHAC1	0.130	0.122	0.112	0.113	0.106	0.108
SHAC1_CI	(0.995)	(0.995)	(0.934)	(0.938)	(0.969)	(0.974)
SHAC2	0.130	0.122	0.112	0.113	0.106	0.108
SHAC2_CI	(0.995)	(0.995)	(0.934)	(0.938)	(0.969)	(0.974)

variables and the spillover effects are independent of each other because of independent assignments. In light of this observation, it appears unnecessary to make inference robust to spatial correlation for both coefficient estimators on two assignment variables when one is assigned independently, the other is spatially correlated, and the two assignment variables are independent of each other. We can easily verify the simulation results using the Frisch-Waugh theorem in linear models.

Table 3: Coefficient on the Individual Assignment Variable Omitting Spillover Effects

	entire population			cluster sampling			independent sampling		
	$p_x = 0,$ $\gamma = 0$	$p_x = 0,$ $\gamma = 1$	$p_x = 0.3,$ $\gamma = 1$	$p_x = 0,$ $\gamma = 0$	$p_x = 0,$ $\gamma = 1$	$p_x = 0.3,$ $\gamma = 1$	$p_x = 0,$ $\gamma = 0$	$p_x = 0,$ $\gamma = 1$	$p_x = 0.3,$ $\gamma = 1$
coeff	-0.001	0.001	0.603	0.003	0.000	0.662	-0.003	-0.002	0.536
std	0.084	0.088	0.157	0.119	0.119	0.148	0.093	0.100	0.117
EHW	0.100	0.101	0.098	0.100	0.101	0.101	0.100	0.102	0.104
EHW_CI	(0.983)	(0.973)	(0.776)	(0.905)	(0.902)	(0.812)	(0.968)	(0.952)	(0.923)
cluster	0.127	0.128	0.134	0.127	0.128	0.138	0.107	0.109	0.113
cluster_CI	(0.998)	(0.997)	(0.905)	(0.960)	(0.960)	(0.934)	(0.977)	(0.965)	(0.942)
SHAC1	0.130	0.129	0.145	0.113	0.114	0.146	0.107	0.110	0.116
SHAC1_CI	(0.998)	(0.996)	(0.933)	(0.948)	(0.935)	(0.943)	(0.977)	(0.967)	(0.945)
SHAC2	0.130	0.129	0.145	0.113	0.114	0.146	0.107	0.110	0.116
SHAC2_CI	(0.998)	(0.996)	(0.933)	(0.948)	(0.935)	(0.943)	(0.977)	(0.967)	(0.945)



In Table 3, we change  $p_x$  to 0.3 for the case of spatial assignments, and omit spillover effects in the estimation by regressing  $Y_i$  on 1 and  $X_i$  only. Due to the omitted variable bias, the coefficient on the individual assignment variable is biased upward when the assignments are spatially correlated. Nevertheless, we must make inference robust to spatial correlation because of the assignment design. With independent assignments, however, the individual assignment variable and the constructed spillover effect are independent. Therefore, not only is the slope coefficient estimator unbiased, but the standard errors need not be adjusted for spatial correlation either. In other words, even with the spillover effect omitted from the estimation, correct inference depends solely on the nature of the individual assignment variable.

#### 4.4 Nonlinear Models

In the last design, we consider spatial assignments in a nonlinear model.

$$y_{ig}(x_{ig}) = \mathbb{1}\{\beta_{ig}x_{ig} + c_g + u_{ig} \geq 0\}, \quad (20)$$

where  $X_M$  is an  $M \times 1$  continuous random vector following a multivariate normal distribution with mean zero and a variance-covariance matrix equal to  $p_x$  raised to the power of the distance. Other aspects of the population generating process are the same as the baseline design. Without spatial assignments,  $p_x = 0$  and  $p_u$  ranges from 0 to 0.9; with spatial assignments,  $p_u$  is fixed at 0.3 while  $p_x$  varies from 0 to 0.9.

We run probit regressions of  $Y_i$  on 1 and  $X_i$  and report standard errors of the APE estimator of  $X_i$ . We report the results when the entire population is observed, which are identical to the linear case with or without spatial assignments. Simulation results from other sampling schemes are also similar to the linear case and hence are omitted.

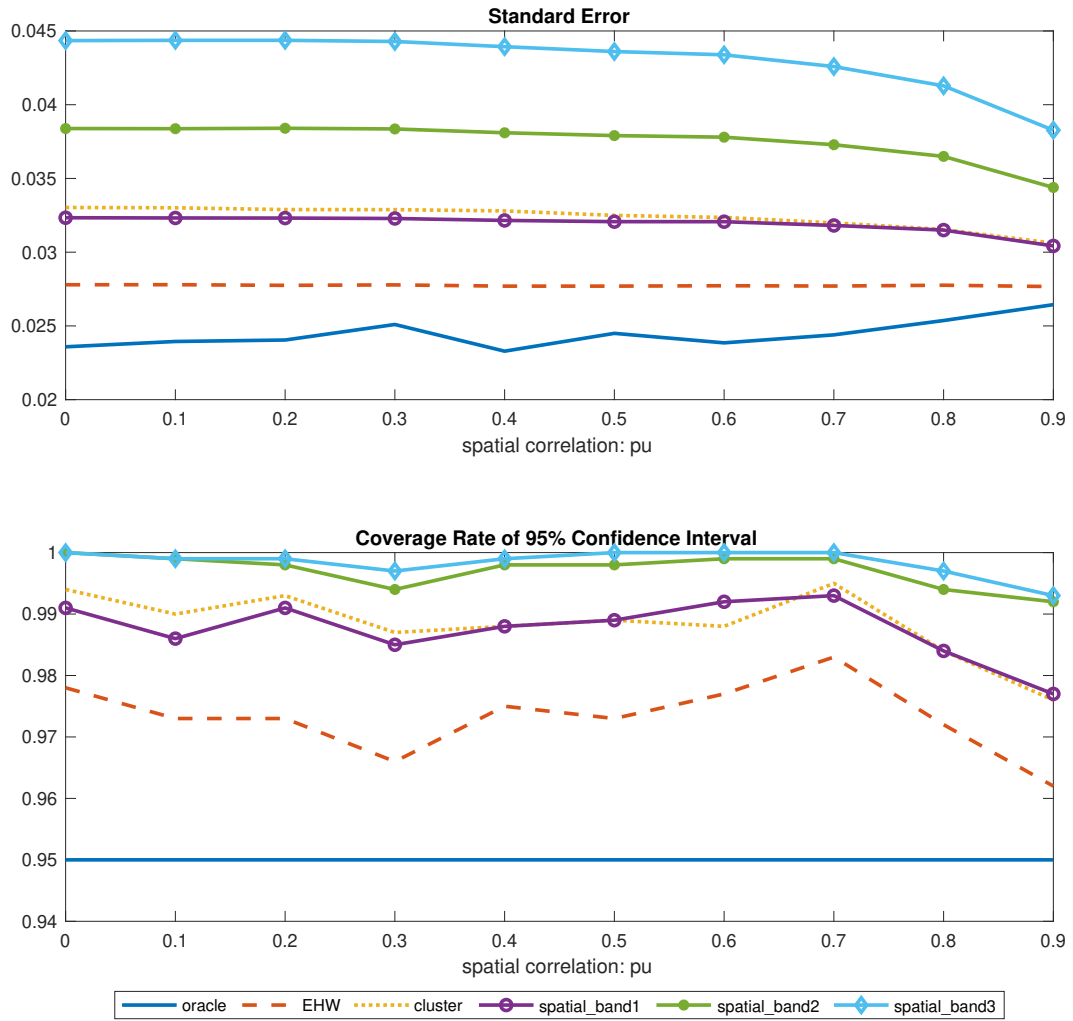


Figure 16: Independent Assignments Observing Entire Population: APE

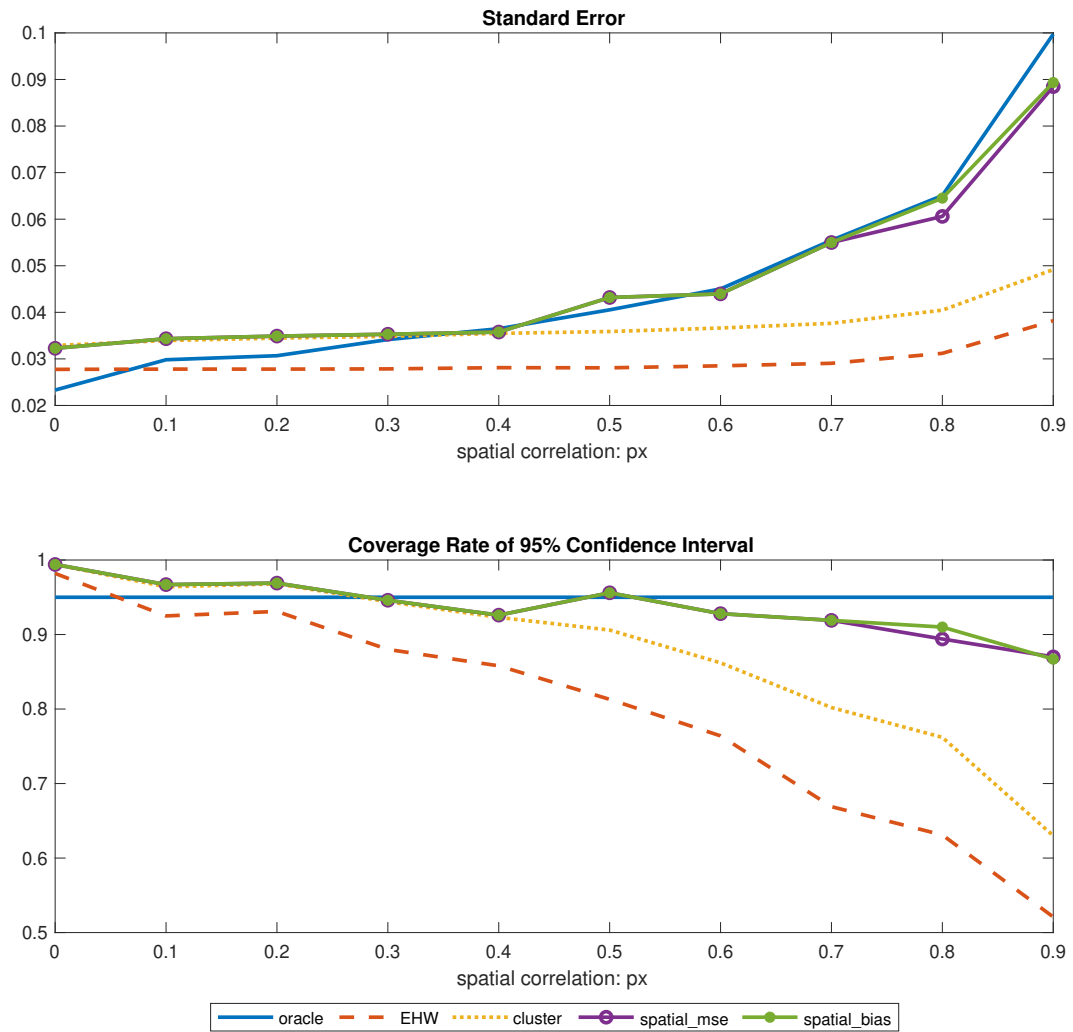


Figure 17: Spatial Assignments Observing Entire Population: APE

## 5 Practical Considerations

What are the considerations in applying our proposed methods in empirical studies with a spatial component? One must take a stand on several issues. Perhaps the most important is the modeling issue, where one needs to not only choose a functional form for a conditional mean or a conditional distribution, but the nature of spillover effects must be incorporated. Generally, a tension exists between simplicity and flexibility. It is probably much to hope that all spillover effects can be captured, but it is also unlikely that capturing minor effects is important for policy analysis. Our framework for inference allows for the spillover effects to be incorrectly modeled, and so one can obtain confidence intervals for direct effects and spillover effects even if the estimated effects are somewhat biased.

Once the model is specified – including specification of spillover effects – one needs to decide on an estimation method. In many cases, this will be pooled OLS or a pooled (quasi-) maximum likelihood procedure. In some ways, this is probably the least difficult decision.

In order to properly compute standard errors, a researcher needs to take a stand on the sampling mechanism. Ideally, this is easy to determine by the nature of the data collection. In applications with geographic units, the population and sample often coincide. But not always, as one might choose a sample from a large number of small areas. For example, there are more than 84,000 census tracts in the United States. If data are collected on all tracts to study, for example, the impacts of opportunity zones on economic development, then the population equals the sample. However, if detailed information on the tracts is desired, one might use a smaller random sample of tracts.

When applying our results to disaggregated units, such as individuals or firms, it may be reasonable to assume the the data were obtained from random sampling. Nevertheless, it could come from cluster sampling – something that is hopefully known based on the

survey design. In principle, the researcher should not have to guess about the sampling scheme used to obtain the data.

Perhaps the most difficult issue, aside from modeling spillover effects, is to determine the nature of the assignment of the key (policy) variables. In some cases, pure cluster assignment is evident if all units within a city or school are subjected to an intervention. More difficult to discern is spatially correlated assignment that is not confined to clear boundaries. For example, if a school district changes its admission policies in a state with schools of choice, do nearby districts react? If so, we can compute spatially-robust standard errors to account for this. Do districts a bit further away react? If so, we would change the calculation of our standard errors to reflect more spatial persistence. If no such spatial assignment exists, the spatially-robust standard errors generally will be conservative – perhaps very conservative.

## 6 Conclusion

Using a design-based approach, we identify the sources of uncertainty underlying spatial data. Whenever there are spatial assignments or when spillover effects are estimated, we must make inference robust to spatial correlation, unless the sampling probability is negligible. Since we allow for cluster partition, our results can be applied to short panel data as well with both spatial and temporal dependence.

## A Notation and Regularity Conditions

The following notation provides details of the variance-covariance matrix and the variance estimator in Section 3:

$$\Delta_{ehw,M}^f = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E} \left\{ [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \cdot [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)]' \right\}, \quad (\text{A.1})$$

$$\Delta_{E,M}^f = \frac{1}{|D_M|} \sum_{i \in D_M} \left\{ \mathbb{E} [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \cdot \mathbb{E} [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)]' \right\}, \quad (\text{A.2})$$

$$\Delta_{cluster,M}^f = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} = C_{jM}) \mathbb{E} \left\{ [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \cdot [f_{jM}(W_{jM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{jM}(W_{jM}, \theta_M^*)]' \right\}, \quad (\text{A.3})$$

$$\Delta_{EC,M}^f = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} = C_{jM}) \left\{ \mathbb{E} [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \cdot \mathbb{E} [f_{jM}(W_{jM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{jM}(W_{jM}, \theta_M^*)]' \right\}, \quad (\text{A.4})$$

$$\Delta_{spatial,M}^f = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} \neq C_{jM}) \mathbb{E} \left\{ [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \cdot [f_{jM}(W_{jM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{jM}(W_{jM}, \theta_M^*)]' \right\}, \quad (\text{A.5})$$

$$\Delta_{ES,M}^f = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M, j \neq i} \mathbb{1}(C_{iM} \neq C_{jM}) \left\{ \mathbb{E}[f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \right. \\ \left. \mathbb{E}[f_{jM}(W_{jM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{jM}(W_{jM}, \theta_M^*)] \right\}, \quad (\text{A.6})$$

$$\hat{V}_{f,SN} = \frac{1}{|D_N|} \sum_{i \in D_M} \sum_{j \in D_M} R_{iM} R_{jM} \cdot \omega\left(\frac{\nu(i,j)}{b_M}\right) \left\{ [f_{iM}(W_{iM}, \hat{\theta}_N) - \hat{\gamma}_N - \hat{F}_N(\hat{\theta}_N) \hat{H}_N(\hat{\theta}_N)^{-1} m_{iM}(W_{iM}, \hat{\theta}_N)] \right. \\ \left. [f_{jM}(W_{jM}, \hat{\theta}_N) - \hat{\gamma}_N - \hat{F}_N(\hat{\theta}_N) \hat{H}_N(\hat{\theta}_N)^{-1} m_{jM}(W_{jM}, \hat{\theta}_N)] \right\}', \quad (\text{A.7})$$

and

$$V_{f,E} = \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \omega\left(\frac{\nu(i,j)}{b_M}\right) \left\{ \mathbb{E}[f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \right. \\ \left. \mathbb{E}[f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] \right\}', \quad (\text{A.8})$$

where

$$F_M(\theta) = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[\nabla_{\theta} f_{iM}(W_{iM}, \theta)] \quad (\text{A.9})$$

and

$$\hat{F}_N(\theta) = \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \nabla_{\theta} f_{iM}(W_{iM}, \theta). \quad (\text{A.10})$$

**Definition 4** The random function  $g_{iM}(W_{iM}, \theta)$  is said to be Lipschitz in the parameter  $\theta$  on  $\Theta$  if there is  $h(u) \downarrow 0$  as  $u \downarrow 0$  and  $b(\cdot) : \mathcal{W} \rightarrow \mathbb{R}$  such that  $\sup_{M, i \in D_M} \mathbb{E}[|b_{iM}(W_{iM})|] < \infty$ , and for all  $\tilde{\theta}, \theta \in \Theta$ ,  $|g_{iM}(W_{iM}, \tilde{\theta}) - g_{iM}(W_{iM}, \theta)| \leq b_{iM}(W_{iM}) h(\|\tilde{\theta} - \theta\|)$ ,  $i \in D_M, M \geq 1$ .

We impose the following regularity conditions for the theorems in the paper.

**Assumption A.1** Suppose that  $\frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \cdot m_{iM}(W_{iM}, \hat{\theta}_N) = o_p(|D_N|^{-1/2})$  and (i) let  $Q_M(\theta) = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[q_{iM}(W_{iM}, \theta)]$ .  $\{Q_M(\theta)\}$  has identifiably unique minimizers  $\{\theta_M^*\}$  on  $\Theta$  as in Definition 3.2 in Gallant and White (1988); (ii)  $\Theta$  is compact; (iii)  $\theta_M^* \in \text{int}(\Theta)$  uniformly in  $M$ ; (iv)  $\sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} |q_{iM}(W_{iM}, \theta)/c_{iM}|^2 \right] < \infty$ , where  $\{c_{iM}, i \in D_M, M \geq 1\}$  is a sequence of nonrandom positive constants; (v)  $q_{iM}(W_{iM}, \theta)/c_{iM}$  is Lipschitz in  $\theta$  on  $\Theta$ ; (vi)  $q_{iM}(w, \theta)$  is twice continuously differentiable on  $\text{int}(\Theta)$  for all  $w$  in the support of  $W_{iM}$ ,  $\forall i, M$ ; (vii)  $\sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} \|\nabla_{\theta} m_{iM}(W_{iM}, \theta)/c_{iM}\|^2 \right] < \infty$ ; (viii)  $\sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} \|m_{iM}(W_{iM}, \theta)/c_{iM}\|^{2+\delta} \right] < \infty$  for some  $\delta > 0$ ; (ix)  $\inf_M |D_M|^{-1} J_M^{-2} \lambda_{\min}(S_M) > 0$ , where  $J_M = \max_{i \in D_M} \{c_{iM}, d_{iM}\}$  and  $\lambda_{\min}(\cdot)$  stands for the smallest eigenvalue; (x)  $\nabla_{\theta} m_{iM}(W_{iM}, \theta)/c_{iM}$  is Lipschitz in  $\theta$  on  $\Theta$ ; (xi)  $H_M(\theta_M^*)/J_M$  is nonsingular; (xii) for any fixed  $s > 0$ , there exists a positive constant  $C$  such that for any  $M$  and every nonempty set  $K \subseteq D_M$ ,  $\mathbb{E} \left[ \left( \sum_{i \in K} m_{iM}^s \right)^2 \right] \geq C \sum_{i \in K} \mathbb{E}[(m_{iM}^s)^2]$ , where  $m_{iM}^s = \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \mathbb{E}(m_{iM}(W_{iM}, \theta_M^*)/J_M | \mathcal{F}_{iM}(s))$ .

**Assumption A.2** (i) The NED scaling factors of the random field  $\{m_{iM}(W_{iM}, \theta) m_{jM}(W_{jM}, \theta), i, j \in D_M, M \geq 1\}$  satisfy  $\sup_{M, i \in D_M} c_{iM}^{-2} d'_{iM} \leq C < \infty$ ; (ii)  $\sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} \|m_{iM}(W_{iM}, \theta)/c_{iM}\|^{2r} \right] < \infty$  for  $r$  in Assumption 6.

**Assumption A.3** (i)  $f_{iM}(w, \theta)$  is continuously differentiable on  $\text{int}(\Theta)$  for all  $w$  in the support of  $W_{iM}$ ,  $\forall i, M$ ; (ii)  $\sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} \|f_{iM}(W_{iM}, \theta)/c_{iM}\|^{2r} \right] < \infty$  for  $r$  in Assumption 6; (iii) the NED scaling factors of the random field  $\{f_{iM} f_{jM}, i, j \in D_M, M \geq 1\}$  satisfy  $\sup_{M, i \in D_M} c_{iM}^{-2} d'_{iM} \leq C < \infty$ , where  $f_{iM} = f_{iM}(W_{iM}, \theta) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta)$ ; (iv)  $\inf_M |D_M|^{-1} J_M^{-2} \lambda_{\min}(V_{f,M}) > 0$ , where  $J_M = \max_{i \in D_M} \{c_{iM}, d_{iM}\}$ ; (v)  $\sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} \|\nabla_{\theta} f_{iM}(W_{iM}, \theta)\|^2 \right] < \infty$ ; (vi)  $f_{iM}(W_{iM}, \theta)/c_{iM}$  is Lipschitz in  $\theta$  on  $\Theta$ ; (vii)  $\nabla_{\theta} f_{iM}(W_{iM}, \theta)/c_{iM}$  is Lipschitz in  $\theta$  on  $\Theta$ ; (viii) for any fixed  $s > 0$ , there exists a positive constant  $C$  such that for any  $M$  and every nonempty set  $K \subseteq D_M$ ,  $\mathbb{E} \left[ \left( \sum_{i \in K} f_{iM}^s \right)^2 \right] \geq C \sum_{i \in K} \mathbb{E}[(f_{iM}^s)^2]$ , where  $f_{iM}^s = \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \mathbb{E}[(f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* -$



$$F_M(\theta_M^*)H_M(\theta_M^*)^{-1}m_{iM}(W_{iM}, \theta_M^*)/J_M|\mathcal{F}_{iM}(s)].$$

## B Proof

**Lemma B.1** Under Assumptions 1-3,  $\frac{|D_N|}{|D_M|\rho_{uM}\rho_{cM}} \xrightarrow{p} 1$ .

**Proof:**

$$\frac{|D_N|}{|D_M|\rho_{uM}\rho_{cM}} = \frac{\sum_{i \in D_M} R_{iM}}{|D_M|\rho_{uM}\rho_{cM}} \quad (\text{B.1})$$

Since

$$\mathbb{E}\left(\frac{\sum_{i \in D_M} R_{iM}}{|D_M|\rho_{uM}\rho_{cM}}\right) = 1 \quad (\text{B.2})$$

and

$$\begin{aligned} \mathbb{V}\left(\frac{\sum_{i \in D_M} R_{iM}}{|D_M|\rho_{uM}\rho_{cM}}\right) &\leq \frac{\sum_{g=1}^{G_M} \mathbb{E}[(\sum_{i \in D_{gM}} R_{iM})^2]}{|D_M|^2 \rho_{uM}^2 \rho_{cM}^2} \\ &= \frac{\sum_{g=1}^{G_M} \mathbb{E}[\sum_{i \in D_{gM}} R_{iM} + \sum_{i \in D_{gM}, j \in D_{gM}, j \neq i} R_{iM} R_{jM}]}{|D_M|^2 \rho_{uM}^2 \rho_{cM}^2} \\ &\leq \frac{C \sum_{g=1}^{G_M} \sum_{i \in D_{gM}} \rho_{uM} \rho_{cM}}{|D_M|^2 \rho_{uM}^2 \rho_{cM}^2} \\ &= \frac{C}{|D_M|\rho_{uM}\rho_{cM}} \rightarrow 0, \end{aligned} \quad (\text{B.3})$$

Lemma B.1 follows from the mean square convergence.

**Lemma B.2** (Weak Law of Large Numbers) Suppose that  $W = \{W_{iM}, i \in D_M, M \geq 1\}$  is  $L_1$ -NED on  $U = \{U_{iM}, i \in T_M, M \geq 1\}$  with the scaling factors  $d_{iM}$  and  $\sup_{M, i \in D_M} \mathbb{E} \|W_{iM}/c_{iM}\|^2 < \infty$ , where  $\{c_{iM}, i \in D_M, M \geq 1\}$  is a sequence of nonrandom positive constants. Under

Assumptions 1-4 and Assumption 5(ii),

$$\frac{1}{J_M|D_N|} \sum_{i \in D_M} R_{iM} W_{iM} - \frac{1}{J_M|D_M|} \sum_{i \in D_M} E W_{iM} \xrightarrow{L_1} 0,$$

where  $J_M = \max_{i \in D_M} \max\{c_{iM}, d_{iM}\}$ .

**Proof:** Define  $Y_{iM} = W_{iM}/J_M$ .

$$\begin{aligned} & \frac{1}{J_M|D_N|} \sum_{i \in D_M} R_{iM} W_{iM} - \frac{1}{J_M|D_M|} \sum_{i \in D_M} E W_{iM} \\ &= \frac{|D_M| \rho_{uM} \rho_{cM}}{|D_N|} \frac{1}{|D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} Y_{iM} - \frac{1}{|D_M|} \sum_{i \in D_M} E Y_{iM} \\ &= (1 + o_p(1)) \frac{1}{|D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} Y_{iM} - \frac{1}{|D_M|} \sum_{i \in D_M} E Y_{iM} \end{aligned} \quad (\text{B.4})$$

Hence, it suffices to show that

$$\frac{1}{|D_M|} \sum_{i \in D_M} \left( \frac{R_{iM}}{\rho_{uM} \rho_{cM}} Y_{iM} - E Y_{iM} \right) \xrightarrow{L_1} 0. \quad (\text{B.5})$$

We can write

$$\lim_{M \rightarrow \infty} \left\| |D_M|^{-1} \sum_{i \in D_M} \left( \frac{R_{iM}}{\rho_{uM} \rho_{cM}} Y_{iM} - E Y_{iM} \right) \right\|_1 \quad (\text{B.6})$$

$$= \lim_{s \rightarrow \infty} \lim_{M \rightarrow \infty} \left\| |D_M|^{-1} \sum_{i \in D_M} \left( \frac{R_{iM}}{\rho_{uM} \rho_{cM}} Y_{iM} - E Y_{iM} \right) \right\|_1 \quad (\text{B.7})$$

$$\leq \lim_{s \rightarrow \infty} \limsup_{M \rightarrow \infty} \left\| |D_M|^{-1} \sum_{i \in D_M} \left( \frac{R_{iM}}{\rho_{uM} \rho_{cM}} Y_{iM} - \frac{R_{iM}}{\rho_{uM} \rho_{cM}} E(Y_{iM} | \mathcal{F}_{iM}(s)) \right) \right\|_1 \quad (\text{B.8})$$

$$+ \lim_{s \rightarrow \infty} \lim_{M \rightarrow \infty} \left\| |D_M|^{-1} \sum_{i \in D_M} \left( \frac{R_{iM}}{\rho_{uM} \rho_{cM}} E(Y_{iM} | \mathcal{F}_{iM}(s)) - E Y_{iM} \right) \right\|_1, \quad (\text{B.9})$$

where  $\mathcal{F}_{iM}(s) = \sigma(U_{jM}; j \in T_M : \nu(i, j) \leq s)$ .

Note that

$$\begin{aligned}
& \mathbb{E} \left\| |D_M|^{-1} \sum_{i \in D_M} \left( \frac{R_{iM}}{\rho_{uM} \rho_{cM}} Y_{iM} - \frac{R_{iM}}{\rho_{uM} \rho_{cM}} E(Y_{iM} | \mathcal{F}_{iM}(s)) \right) \right\| \\
& \leq |D_M|^{-1} \sum_{i \in D_M} E \left\| \frac{R_{iM}}{\rho_{uM} \rho_{cM}} [Y_{iM} - E(Y_{iM} | \mathcal{F}_{iM}(s))] \right\| \\
& = |D_M|^{-1} \sum_{i \in D_M} E \|Y_{iM} - E(Y_{iM} | \mathcal{F}_{iM}(s))\| \\
& \leq \sup_{M, i \in D_M} E \|Y_{iM} - E(Y_{iM} | \mathcal{F}_{iM}(s))\| \\
& \leq \sup_{M, i \in D_M} J_M^{-1} d_{iM} \psi(s) \leq \psi(s).
\end{aligned} \tag{B.10}$$

Therefore, (B.8) is zero as  $\lim_{s \rightarrow \infty} \psi(s) = 0$ .

Next, we show that for fixed  $s > 0$ , (B.9) is zero. Denote  $V_{iM}^s = \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))$ . Obverse that  $\mathbb{E}(V_{iM}^s) = \mathbb{E}(Y_{iM})$ . For each fixed  $s$ ,  $\mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))$  is a measurable function of  $\{U_{jM}; j \in T_M : \nu(i, j) \leq s\}$ . Since the cluster sizes are bounded  $\{R_{iM}\}$  is  $m$ -dependent. As a result, the maximal correlation coefficient on  $\{V_{iM}^s\}$  also satisfies  $\lim_{r \rightarrow \infty} \bar{\rho}(r) < 1$  for each fixed  $s$ .

By Theorem 2.1 in Bradley and Tone (2017),

$$\begin{aligned}
\mathbb{V} \left( \sum_{i \in D_M} V_{iM}^s \right) &= \mathbb{E} \left[ \left\| \sum_{i \in D_M} (V_{iM}^s - \mathbb{E}(Y_{iM})) \right\|^2 \right] \\
&\leq C \sum_{i \in D_M} \mathbb{E} \left[ \|V_{iM}^s - \mathbb{E}(Y_{iM})\|^2 \right] \leq C \sum_{i \in D_M} \mathbb{E}(\|V_{iM}^s\|^2) \\
&= C \sum_{i \in D_M} \mathbb{E} \left\{ \frac{R_{iM}}{\rho_{uM}^2 \rho_{cM}^2} \|\mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))\|^2 \right\} \\
&\leq C \sum_{i \in D_M} \frac{1}{\rho_{uM} \rho_{cM}} \mathbb{E}(\|Y_{iM}\|^2).
\end{aligned} \tag{B.11}$$

Therefore,

$$\mathbb{V}\left(\frac{1}{|D_M|} \sum_{i \in D_M} V_{iM}^s\right) \leq \frac{C}{|D_M| \rho_{uM} \rho_{cM}} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}(\|Y_{iM}\|^2) \rightarrow 0, \quad (\text{B.12})$$

as  $|D_M| \rho_{uM} \rho_{cM} \rightarrow \infty$ . The result follows by the Liapunov's inequality.

**Lemma B.3** *Under Assumptions 1, 2, and 4, suppose (i)  $a_{iM}(W_{iM}, \theta)$  is Lipschitz in the  $\theta$  on  $\Theta$ ; (ii)  $\sup_{M, i \in D_M} \mathbb{E}[\sup_{\theta \in \Theta} \|a_{iM}(W_{iM}, \theta)\|^p] < \infty$  for some  $p > 1$ . Then (1) Let  $A_N(\theta) = \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} a_{iM}(W_{iM}, \theta)$ .  $\|A_N(\tilde{\theta}) - A_N(\theta)\| \leq B_N h(\|\tilde{\theta} - \theta\|)$ , where  $B_N = \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \cdot b_{iM}(W_{iM}) = O_p(1)$ ; and (2)  $A_M(\theta) = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}[a_{iM}(W_{iM}, \theta)]$  is uniformly equicontinuous.*

**Proof:** We first show result (1).

$$\begin{aligned} & \|A_N(\tilde{\theta}) - A_N(\theta)\| \\ &= \left\| \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} [a_{iM}(W_{iM}, \tilde{\theta}) - a_{iM}(W_{iM}, \theta)] \right\| \\ &\leq \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \|a_{iM}(W_{iM}, \tilde{\theta}) - a_{iM}(W_{iM}, \theta)\| \\ &\leq \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \cdot b_{iM}(W_{iM}) h(\|\tilde{\theta} - \theta\|) \\ &= B_N h(\|\tilde{\theta} - \theta\|), \end{aligned} \quad (\text{B.13})$$

where

$$B_N = \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \cdot b_{iM}(W_{iM}) = \frac{|D_M| \rho_{uM} \rho_{cM}}{|D_N|} \frac{1}{|D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} b_{iM}(W_{iM}). \quad (\text{B.14})$$

For all  $\epsilon > 0$ , let  $b_\epsilon = C/\epsilon$  for some  $C < \infty$ ,

$$\begin{aligned}
& P\left(\left|\frac{1}{|D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM}\rho_{cM}} b_{iM}(W_{iM})\right| \geq b_\epsilon\right) \\
& \leq \mathbb{E}\left(\left|\frac{1}{|D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM}\rho_{cM}} b_{iM}(W_{iM})\right|\right)/b_\epsilon \\
& \leq \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}\left(\frac{R_{iM}}{\rho_{uM}\rho_{cM}}\right) \mathbb{E}\left[|b_{iM}(W_{iM})|\right]/b_\epsilon \\
& \leq \sup_{M, i \in D_M} \mathbb{E}\left[|b_{iM}(W_{iM})|\right]/b_\epsilon < C/b_\epsilon = \epsilon.
\end{aligned} \tag{B.15}$$

Because of Lemma B.1 and the continuous mapping theorem,  $\frac{|D_M|\rho_{uM}\rho_{cM}}{|D_N|} \xrightarrow{p} 1$ . Hence,  $B_N = O_p(1)$ .

Next, we show  $\{A_M(\theta)\}$  is uniformly equicontinuous. The proof is based on slight modification of the proof of Theorem 2 in Jenish and Prucha (2009).

$$\begin{aligned}
& \sup_{\theta \in \Theta} \sup_{\tilde{\theta} \in B(\theta, \delta)} \left\| A_M(\tilde{\theta}) - A_M(\theta) \right\| \\
& \leq \frac{1}{|D_M|} \sum_{i \in D_M} \sup_{\theta \in \Theta} \sup_{\tilde{\theta} \in B(\theta, \delta)} \left\| \mathbb{E}[a_{iM}(W_{iM}, \tilde{\theta}) - a_{iM}(W_{iM}, \theta)] \right\| \\
& \leq \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E} \sup_{\theta \in \Theta} \sup_{\tilde{\theta} \in B(\theta, \delta)} \left\| a_{iM}(W_{iM}, \tilde{\theta}) - a_{iM}(W_{iM}, \theta) \right\| \\
& = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E} Y_{iM}(\delta),
\end{aligned} \tag{B.16}$$

where  $Y_{iM}(\delta) = \sup_{\theta \in \Theta} \sup_{\tilde{\theta} \in B(\theta, \delta)} \|a_{iM}(W_{iM}, \tilde{\theta}) - a_{iM}(W_{iM}, \theta)\|$ . Define  $f_{iM} = \sup_{\theta \in \Theta} \|a_{iM}(W_{iM}, \theta)\|$ . Given condition (ii), there exists  $k = k(\epsilon) < \infty$  for some  $\epsilon > 0$  such that

$$\limsup_{M \rightarrow \infty} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}(f_{iM} \mathbb{1}(f_{iM} > k)) < \frac{\epsilon}{6}. \tag{B.17}$$

Under condition (i),  $a_{iM}(W_{iM}, \theta)$  is  $L_0$  stochastically equicontinuous on  $\Theta$  by Proposition 1 in Jenish and Prucha (2009). Hence, we can find some  $\delta = \delta(\epsilon)$  such that

$$\limsup_{M \rightarrow \infty} \frac{1}{|D_M|} \sum_{i \in D_M} P(Y_{iM}(\delta) > \epsilon/3) \leq \frac{\epsilon}{6k}. \quad (\text{B.18})$$

$$\begin{aligned} & \limsup_{M \rightarrow \infty} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E} Y_{iM}(\delta) \\ & \leq \epsilon/3 + \limsup_{M \rightarrow \infty} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E} Y_{iM}(\delta) \mathbb{1}(Y_{iM}(\delta) > \epsilon/3, f_{iM} > k) \\ & \quad + \limsup_{M \rightarrow \infty} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E} Y_{iM}(\delta) \mathbb{1}(Y_{iM}(\delta) > \epsilon/3, f_{iM} \leq k) \\ & \leq \epsilon/3 + 2 \limsup_{M \rightarrow \infty} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E} f_{iM} \mathbb{1}(f_{iM} > k) + 2k \frac{1}{|D_M|} \sum_{i \in D_M} \limsup_{M \rightarrow \infty} P(Y_{iM}(\delta) > \epsilon/3) = \epsilon \end{aligned} \quad (\text{B.19})$$

As a result,  $\limsup_{M \rightarrow \infty} \sup_{\theta \in \Theta} \sup_{\tilde{\theta} \in B(\theta, \delta)} \|A_M(\tilde{\theta}) - A_M(\theta)\| \rightarrow 0$  as  $\delta \rightarrow 0$ .

**Lemma B.4** (*Central Limit Theorem*) Let  $W = \{W_{iM}, i \in D_M, M \geq 1\}$  be a real valued zero-mean random field that is  $L_2$ -NED on  $U = \{U_{iM}, i \in T_M, M \geq 1\}$  with the scaling factors  $d_{iM}$  and the NED coefficients  $\psi(s)$ . Define  $L_M = \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} W_{iM}$  and  $\sigma_M^2 = \mathbb{V}(L_M)$ . Suppose  $W$  satisfies: (i)  $\sup_{M, i \in D_M} \mathbb{E}|W_{iM}/c_{iM}|^{2+\delta} < \infty$  for some  $\delta > 0$ , where  $\{c_{iM}, i \in D_M, M \geq 1\}$  is a sequence of nonrandom positive constants; (ii) for any fixed  $s > 0$ , there exists a positive constant  $C$  such that for any  $M$  and every nonempty set  $K \subseteq D_M$ ,  $\mathbb{E}\left[\left(\sum_{i \in K} \xi_{iM}^s\right)^2\right] \geq C \sum_{i \in K} \mathbb{E}[(\xi_{iM}^s)^2]$ , where  $\xi_{iM}^s = \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \mathbb{E}(W_{iM}/J_M | \mathcal{F}_{iM}(s))$  and  $J_M = \max_{i \in D_M} \{c_{iM}, d_{iM}\}$ ; (iii)  $\inf_M |D_M|^{-1} J_M^{-2} \sigma_M^2 > 0$ ; (iv) the NED coefficient

$\psi(s)$  is of size  $-d$ . Under Assumptions 1-5,

$$\sigma_M^{-1} L_M \xrightarrow{d} \mathcal{N}(0, 1).$$

**Proof:** Define  $Y_{iM} = W_{iM}/J_M$ .

$$\sigma_M^{-1} L_M = J_M \sigma_M^{-1} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} W_{iM} / J_M = \sigma_{Y,M}^{-1} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} Y_{iM} \quad (\text{B.20})$$

$$= \sigma_{Y,M}^{-1} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s)) \quad (\text{B.21})$$

$$+ \sigma_{Y,M}^{-1} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \left[ Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s)) \right] \quad (\text{B.22})$$

From now on, let  $L_M$  denote  $\sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} Y_{iM}$  and  $\sigma_M^2$  denote the variance of  $\sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} Y_{iM}$ .

Define  $\xi_{iM}^s = \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))$  and  $\eta_{iM}^s = \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \left[ Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s)) \right]$ .

Let  $\|Y\|_2 = \sup_{M, i \in D_M} \|Y_{iM}\|_2$ .  $\forall i \in D_M$ ,

$$\|\xi_{iM}^s\|_2 = \left( \mathbb{E} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \right)^{1/2} \|\mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))\|_2 \leq \|Y_{iM}\|_2 \leq \|Y\|_2 < \infty \quad (\text{B.23})$$

and

$$\begin{aligned} \|\eta_{iM}^s\|_2 &= \left( \mathbb{E} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \right)^{1/2} \|Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))\|_2 \\ &\leq \|Y_{iM}\|_2 + \|\mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))\|_2 \\ &\leq 2 \|Y_{iM}\|_2 < \infty. \end{aligned} \quad (\text{B.24})$$

Furthermore,

$$\sup_{M, i \in D_M} \|\eta_{iM}^s\|_2 = \sup_{M, i \in D_M} \left( \mathbb{E} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \right)^{1/2} \|Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))\|_2$$

$$\leq \sup_{M, i \in D_M} J_M^{-1} d_{iM} \psi(s) \leq \psi(s). \quad (\text{B.25})$$

For each fixed  $s > 0$ ,  $\mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))$  is a measurable function of  $U_{iM}$ . Also,  $R_{iM}$  is  $m$ -dependent. Hence, for the random field  $\{\xi_{iM}^s\}$ ,  $\bar{\alpha}(r) \rightarrow 0$  as  $r \rightarrow \infty$  and  $\lim_{r \rightarrow \infty} \bar{\rho}(r) < 1$ .

Now, decompose  $\frac{R_{iM}}{\sqrt{\rho_{uM}\rho_{cM}}} Y_{iM}$  as

$$\frac{R_{iM}}{\sqrt{\rho_{uM}\rho_{cM}}} Y_{iM} = \xi_{iM}^s + \eta_{iM}^s. \quad (\text{B.26})$$

Applying Theorem 2.1 in Bradley and Tone (2017),

$$\mathbb{E} \left| \sum_{i \in D_M} \xi_{iM}^s \right|^2 \leq C \sum_{i \in D_M} \mathbb{E}(\xi_{iM}^s)^2 \leq C |D_M| \|Y\|_2^2. \quad (\text{B.27})$$

By the Cauchy-Schwartz inequality,

$$|Cov(\xi_{iM}^s, \eta_{jM}^s)| \leq \|\xi_{iM}^s\|_2 \|\eta_{jM}^s\|_2 \leq C\psi(s) \quad (\text{B.28})$$

and

$$|Cov(\eta_{iM}^s, \eta_{jM}^s)| \leq \|\eta_{iM}^s\|_2 \|\eta_{jM}^s\|_2 \leq C\psi(s). \quad (\text{B.29})$$

Let  $h = r/3$  with  $r$  in the definition of the maximal correlation coefficient in Assumption 5. Then, for sufficiently large  $r$ ,

$$\begin{aligned} \sigma_M^2 &= \mathbb{V}(L_M) = \mathbb{V} \left( \sum_{i \in D_M} \xi_{iM}^h + \sum_{i \in D_M} \eta_{iM}^h \right) \\ &= \mathbb{V} \left( \sum_{i \in D_M} \xi_{iM}^h \right) + \mathbb{V} \left( \sum_{i \in D_M} \eta_{iM}^h \right) + 2Cov \left( \sum_{i \in D_M} \xi_{iM}^h, \sum_{i \in D_M} \eta_{iM}^h \right) \\ &\leq \mathbb{E} \left[ \left( \sum_{i \in D_M} \xi_{iM}^h \right)^2 \right] + \sum_{i \in D_M} \mathbb{V}(\eta_{iM}^h) + \sum_{i, j \in D_M, i \neq j} |Cov(\eta_{iM}^h, \eta_{jM}^h)| \end{aligned}$$



$$\begin{aligned}
& + 2 \sum_{i,j \in D_M, i \neq j} |Cov(\xi_{iM}^h, \eta_{jM}^h)| + 2 \sum_{i \in D_M} |Cov(\xi_{iM}^h, \eta_{iM}^h)| \\
& \leq C_1 |D_M| + C_2 \sum_{i \in D_M} \left[ \sum_{j \in D_M: j \neq i, \nu(i,j) \leq h} \psi(h) + \sum_{j \in D_M: \nu(i,j) > h} \psi(h) \right] \\
& \leq C_1 |D_M| + C_2 \sum_{i \in D_M} \left[ C_3 h^d \psi(h) + \sum_{j \in D_M: \nu(i,j) > h} \psi(h) \right] \\
& \leq C |D_M|,
\end{aligned} \tag{B.30}$$

given the last term  $\sum_{j \in D_M: \nu(i,j) > h} \psi(h)$  approaches zero as  $h \rightarrow \infty$ .

Under condition (iii),

$$\inf_M |D_M|^{-1} \sigma_M^2 > 0. \tag{B.31}$$

Hence, there exists  $0 < B < \infty$  such that for all  $M$ ,

$$B |D_M| \leq \sigma_M^2. \tag{B.32}$$

As a result, we have

$$B |D_M| \leq \sigma_M^2 \leq C |D_M|. \tag{B.33}$$

Similarly, let  $\lambda_{iM}^m = \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} \mathbb{E} \left[ Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s)) | \mathcal{F}_{iM}(m) \right]$  and  $\gamma_{iM}^m = \eta_{iM}^s - \lambda_{iM}^m$ .

We have  $\|\lambda_{iM}^m\|_2 \leq \|\eta_{iM}^s\|_2 < \infty$  and  $\|\gamma_{iM}^m\|_2 \leq 2 \|\eta_{iM}^s\|_2 < \infty$ . In addition,

$$\begin{aligned}
\sup_{M, i \in D_M} \|\gamma_{iM}^m\|_2 &= \sup_{M, i \in D_M} \left( \mathbb{E} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \right)^{1/2} \left\| Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s)) \right. \\
&\quad \left. - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(m)) + \mathbb{E} \left[ \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s)) | \mathcal{F}_{iM}(m) \right] \right\|_2 \\
&= \begin{cases} \sup_{M, i \in D_M} \|Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(m))\|_2 \leq C\psi(m), & m \geq s \\ \sup_{M, i \in D_M} \|Y_{iM} - \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s))\|_2 \leq C\psi(s) \leq C\psi(m), & m < s. \end{cases}
\end{aligned} \tag{B.34}$$

$$\begin{aligned}
\tilde{\sigma}_{M,s}^2 &= \mathbb{V}\left(\sum_{i \in D_M} \eta_{iM}^s\right) = \mathbb{V}\left(\sum_{i \in D_M} \lambda_{iM}^h + \sum_{i \in D_M} \gamma_{iM}^h\right) \\
&= \mathbb{V}\left(\sum_{i \in D_M} \lambda_{iM}^h\right) + \mathbb{V}\left(\sum_{i \in D_M} \gamma_{iM}^h\right) + 2\text{Cov}\left(\sum_{i \in D_M} \lambda_{iM}^h, \sum_{i \in D_M} \gamma_{iM}^h\right) \\
&\leq \mathbb{E}\left[\left(\sum_{i \in D_M} \lambda_{iM}^h\right)^2\right] + \sum_{i \in D_M} \mathbb{V}(\gamma_{iM}^h) + \sum_{i,j \in D_M, i \neq j} |\text{Cov}(\gamma_{iM}^h, \gamma_{jM}^h)| \\
&\quad + 2 \sum_{i,j \in D_M, i \neq j} |\text{Cov}(\lambda_{iM}^h, \gamma_{jM}^h)| + 2 \sum_{i \in D_M} |\text{Cov}(\lambda_{iM}^h, \gamma_{iM}^h)| \\
&\leq C_1 \sum_{i \in D_M} \|\lambda_{iM}^h\|_2^2 + \sum_{i \in D_M} \|\gamma_{iM}^h\|_2^2 + 2 \sum_{i \in D_M} \|\lambda_{iM}^h\|_2 \|\gamma_{iM}^h\|_2 \\
&\quad + 2 \sum_{i,j \in D_M, i \neq j} \|\lambda_{iM}^h\|_2 \|\gamma_{jM}^h\|_2 + \sum_{i,j \in D_M, i \neq j} \|\gamma_{iM}^h\|_2 \|\gamma_{jM}^h\|_2 \\
&\leq C_2 |D_M| \|\eta_{iM}^s\|_2 + C_3 \sum_{i \in D_M} \left[ \sum_{j \in D_M: j \neq i, \nu(i,j) \leq h} \psi(h) \|\eta_{iM}^s\|_2 + \sum_{j \in D_M: \nu(i,j) > h} \psi(h) \|\eta_{iM}^s\|_2 \right] \\
&\leq C |D_M| \psi(s)
\end{aligned} \tag{B.35}$$

Hence,

$$\lim_{s \rightarrow \infty} \limsup_{M \rightarrow \infty} \frac{\tilde{\sigma}_{M,s}^2}{\sigma_M^2} \leq C \lim_{s \rightarrow \infty} \psi(s) = 0. \tag{B.36}$$

We now show that for any fixed  $s > 0$ ,  $\xi_{iM}^s$  satisfies the central limit theorem for an  $\alpha$ -mixing process. The proof is based on the modification of the proof of Theorem 1.1 in Bradley and Tone (2017). We first verify the Lindeberg condition. It suffices to verify the Liapunov's condition. Let  $\sigma_{M,s}^2 = \mathbb{V}\left(\sum_{i \in D_M} \xi_{iM}^s\right)$ . By repeated use of the Minkowski's inequality, we have

$$|\sigma_M - \tilde{\sigma}_{M,s}| \leq \sigma_{M,s}. \tag{B.37}$$

Because of (B.35), there exists  $s_*$  such that for all  $s > s_*$ ,

$$|D_M|^{-1} \tilde{\sigma}_{M,s}^2 \leq C\psi(s) \leq B/2. \quad (\text{B.38})$$

Therefore,

$$\inf_M |D_M|^{-1} \sigma_{M,s}^2 \geq \inf_M |D_M|^{-1} \sigma_M^2 - \sup_M |D_M|^{-1} \tilde{\sigma}_{M,s}^2 \geq B - B/2 = B/2 > 0. \quad (\text{B.39})$$

Consider  $M$  large enough so that  $\sigma_{M,s}^2/|D_M| > 0$ .

$$\begin{aligned} \sum_{i \in D_M} \mathbb{E}[|\xi_{iM}^s / \sigma_{M,s}|^{2+\delta}] &= \sum_{i \in D_M} \mathbb{E} \left[ \left| \frac{R_{iM}}{(\sigma_{M,s} / \sqrt{|D_M|}) \sqrt{|D_M| \rho_{uM} \rho_{cM}}} \mathbb{E}(Y_{iM} | \mathcal{F}_{iM}(s)) \right|^{2+\delta} \right] \\ &\leq \frac{\sum_{i \in D_M} \rho_{uM} \rho_{cM} \mathbb{E}(|Y_{iM}|^{2+\delta})}{(\sigma_{M,s} / \sqrt{|D_M|})^{2+\delta} (|D_M| \rho_{uM} \rho_{cM})^{1+\delta/2}} \\ &= \frac{1}{(\sigma_{M,s} / \sqrt{|D_M|})^{2+\delta} (|D_M| \rho_{uM} \rho_{cM})^{\delta/2}} \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}(|Y_{iM}|^{2+\delta}) \rightarrow 0 \end{aligned} \quad (\text{B.40})$$

as  $|D_M| \rho_{uM} \rho_{cM} \rightarrow \infty$  as  $M \rightarrow \infty$ .

To allow for irregularly spaced lattice, we modify the proof of Theorem 1.1 in Bradley and Tone (2017) in the following way. Change (4.10) in Step 5 to  $L_1^{(n_q)} > q^2 m_q^d$ . Correspondingly, change (4.12) to

$$\left| \sum_{j=q+1}^{q^2 m_q^d} \left( s_{p(n_q, j)}^{(n_q)} \right)^2 - \sum_{j=q+1}^{q^2 m_q^d} \lambda_j \right| \leq \frac{1}{q}. \quad (\text{B.41})$$

In (4.13), replace  $q^2 m_q$  with  $q^2 m_q^d$ . In Step 10, using Lemma A.1 (ii) in Jenish and Prucha (2009), the cardinality of  $\Gamma_3^{(q)}$  is  $Cq m_q^d$ .

The rest of the proof is the same as part 5 of the proof of Theorem 2 in Jenish and Prucha (2012).

**Proof of Theorem 2.1:**

We first show that  $\hat{\theta}_N - \theta_M^* \xrightarrow{p} \mathbf{0}$ .

Denote  $Q_N(\theta) = \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} q_{iM}(W_{iM}, \theta)$ . First note that

$$\begin{aligned} Q_N(\theta) &= \frac{|D_M| \rho_{uM} \rho_{cM}}{|D_N|} \frac{1}{|D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} q_{iM}(W_{iM}, \theta) \\ &= (1 + o_p(1)) \frac{1}{|D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} q_{iM}(W_{iM}, \theta). \end{aligned} \quad (\text{B.42})$$

Define  $J_M = \max_{i \in D_M} \{c_{iM}, d_{iM}\}$ . Hence, it is sufficient to show that for each  $\theta \in \Theta$

$$\left\| \frac{1}{J_M |D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} q_{iM}(W_{iM}, \theta) - \frac{1}{J_M} Q_M(\theta) \right\| \xrightarrow{p} 0, \quad (\text{B.43})$$

which holds by applying Lemma B.2. Next,

$$\sup_{\theta \in \Theta} \frac{1}{J_M} |Q_N(\theta) - Q_M(\theta)| = o_p(1) \quad (\text{B.44})$$

follows from Corollary 2.2 in Newey (1991), Lemma B.3, and (B.43). As a result, consistency follows, e.g., from Gallant and White (1988).

For asymptotic normality, I start by verifying that

$$\sum_{i \in D_M} \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*) / J_M] = \mathbf{0}, \quad (\text{B.45})$$

which holds by Lemma 3.6 in Newey and McFadden (1994) and Jensen's inequality.

By the element-by-element mean value expansion around  $\theta_M^*$ ,

$$\begin{aligned} o_p(|D_N|^{-1/2}) &= \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \cdot m_{iM}(W_{iM}, \hat{\theta}_N) \\ &= \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \cdot m_{iM}(W_{iM}, \theta_M^*) + J_M \frac{1}{J_M |D_N|} \sum_{i \in D_M} R_{iM} \nabla_{\theta} m_{iM}(W_{iM}, \check{\theta})(\hat{\theta}_N - \theta_M^*), \end{aligned} \quad (\text{B.46})$$

where  $\check{\theta}$  lies on the line segment connecting  $\theta_M^*$  and  $\hat{\theta}_N$ .

I first show

$$\frac{1}{J_M} \hat{H}_N(\check{\theta}) = \frac{1}{J_M |D_N|} \sum_{i \in D_M} R_{iM} \nabla_{\theta} m_{iM}(W_{iM}, \check{\theta}) = \frac{1}{J_M} H_M(\theta_M^*)(I_k + o_p(1)). \quad (\text{B.47})$$

We can write

$$\begin{aligned} \frac{1}{J_M} \hat{H}_N(\theta) &= \frac{|D_M| \rho_{uM} \rho_{cM}}{|D_N|} \frac{1}{J_M |D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \nabla_{\theta} m_{iM}(W_{iM}, \theta) \\ &= (1 + o_p(1)) \frac{1}{J_M |D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \nabla_{\theta} m_{iM}(W_{iM}, \theta). \end{aligned} \quad (\text{B.48})$$

Applying Lemma B.2,

$$\left\| \frac{1}{J_M |D_M|} \sum_{i \in D_M} \frac{R_{iM}}{\rho_{uM} \rho_{cM}} \nabla_{\theta} m_{iM}(W_{iM}, \theta) - \frac{1}{J_M} H_M(\theta) \right\| \xrightarrow{p} \mathbf{0}. \quad (\text{B.49})$$

By Lemma B.3 above and Corollary 2.2 in Newey (1991),

$$\begin{aligned} &\left\| \frac{1}{J_M} \hat{H}_N(\check{\theta}) - \frac{1}{J_M} H_M(\theta_M^*) \right\| \\ &= \left\| \frac{1}{J_M} (\hat{H}_N(\check{\theta}) - H_M(\check{\theta}) + H_M(\check{\theta}) - H_M(\theta_M^*)) \right\| \\ &\leq \sup_{\theta \in \Theta} \frac{1}{J_M} \left\| \hat{H}_N(\theta) - H_M(\theta) \right\| + \frac{1}{J_M} \left\| H_M(\check{\theta}) - H_M(\theta_M^*) \right\| \xrightarrow{p} \mathbf{0}. \end{aligned} \quad (\text{B.50})$$

(B.47) implies

$$J_M \hat{H}_N(\check{\theta})^{-1} = J_M H_M(\theta_M^*)^{-1}(I_k + o_p(1)). \quad (\text{B.51})$$

Using (B.51), (B.46) can be written as

$$\begin{aligned} S_M^{-1/2} H_M(\theta_M^*) \sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*) &= -S_M^{-1/2} \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} \cdot m_{iM}(W_{iM}, \theta_M^*) \\ &\quad - S_M^{-1/2} o_p(1) \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} \cdot m_{iM}(W_{iM}, \theta_M^*) + o_p(1). \end{aligned} \quad (\text{B.52})$$

We can write

$$\begin{aligned} &\frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} \cdot m_{iM}(W_{iM}, \theta) \\ &= \sqrt{\frac{|D_M| \rho_{uM} \rho_{cM}}{|D_N|}} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta) \\ &= (1 + o_p(1)) \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta) \end{aligned} \quad (\text{B.53})$$

Plug (B.53) into (B.52), we have

$$\begin{aligned} &S_M^{-1/2} H_M(\theta_M^*) \sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*) \\ &= -S_M^{-1/2} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta_M^*) \\ &\quad - S_M^{-1/2} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta_M^*) \cdot o_p(1) + o_p(1). \end{aligned} \quad (\text{B.54})$$

By Lemma B.4 above and Corollary 1 in Jenish and Prucha (2012),

$$\begin{aligned}
& S_M^{-1/2} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta_M^*) \\
&= S_M^{-1/2} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \left( \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta_M^*) - \sqrt{\rho_{uM} \rho_{cM}} \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)] \right) \\
&\xrightarrow{d} \mathcal{N}(\mathbf{0}, I_k)
\end{aligned} \tag{B.55}$$

Because of (B.55),

$$\begin{aligned}
S_M^{-1/2} H_M(\theta_M^*) \sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*) &= - S_M^{-1/2} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta_M^*) \\
&\quad + o_p(1) O_p(1) + o_p(1) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_k).
\end{aligned} \tag{B.56}$$

**Lemma B.5** Suppose (i)  $X = \{X_{iM}, i \in D_M, M \geq 1\}$  is  $L_2$ -NED on  $U = \{U_{iM}, i \in T_M, M \geq 1\}$  with the scaling factors  $d_{iM}$  and the NED coefficients  $\psi(s)$  of size  $-2d(r - 1)/(r - 2)$  for some  $r > 2$ ; (ii)  $\sup_{M, i \in D_M} \mathbb{E}(|X_{iM}/c_{iM}|^{2r}) < \infty$ , where  $\{c_{iM}, i \in D_M, M \geq 1\}$  is a sequence of nonrandom positive constants.

Define  $W_{iM} = \sum_{j \in D_M: j \neq i, \nu(i, j) \leq b_M} \omega\left(\frac{\nu(i, j)}{b_M}\right) \left[ \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} X_{iM} X_{jM} - \mathbb{E}\left(\frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} X_{iM} X_{jM}\right) \right]$ .

Under Assumptions 1-4, Assumption 5(ii), and Assumption 7,

$$\frac{1}{J_M^2 |D_M|} \sum_{i \in D_M} W_{iM} \xrightarrow{p} 0,$$

where  $J_M = \max_{i \in D_M} \max\{c_{iM}, \sqrt{d'_{iM}}\}$  and  $d'_{iM}$  is the NED scaling factor of the random field  $\{X_{iM} X_{jM}, i, j \in D_M, M \geq 1\}$ .

**Proof:**

Denote

$$Y_{iM} = \sum_{j \in D_M: j \neq i, \nu(i,j) \leq b_M} \omega\left(\frac{\nu(i,j)}{b_M}\right) \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} X_{iM} X_{jM} / J_M^2,$$

$$\lambda_{iM}^s = \sum_{j \in D_M: j \neq i, \nu(i,j) \leq b_M} \omega\left(\frac{\nu(i,j)}{b_M}\right) \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} \mathbb{E}[X_{iM} X_{jM} / J_M^2 | \mathcal{F}_{iM}(s)],$$

and  $\eta_{iM}^s = Y_{iM} - \lambda_{iM}^s$ .  $\forall i \in D_M$ ,

$$\begin{aligned} \|\lambda_{iM}^s\|_2 &\leq \sum_{j \in D_M: j \neq i, \nu(i,j) \leq b_M} \left| \omega\left(\frac{\nu(i,j)}{b_M}\right) \right| \left[ \mathbb{E} \frac{R_{iM} R_{jM}}{\rho_{uM}^2 \rho_{cM}^2} \right]^{1/2} \left[ \mathbb{E} \left( \mathbb{E}[X_{iM} X_{jM} / J_M^2 | \mathcal{F}_{iM}(s)] \right)^2 \right]^{1/2} \\ &\leq C_1 \sum_{r=1}^{b_M} \sum_{\nu(i,j) \in [r, r+1)} \frac{1}{\sqrt{\rho_{cM}}} \leq C_2 \sum_{r=1}^{b_M} r^{d-1} \frac{1}{\sqrt{\rho_{cM}}} \leq C b_M \cdot b_M^{d-1} \frac{1}{\sqrt{\rho_{cM}}} = C \frac{b_M^d}{\sqrt{\rho_{cM}}} \end{aligned} \quad (\text{B.57})$$

and

$$\|\eta_{iM}^s\|_2 \leq C \frac{b_M^d}{\sqrt{\rho_{cM}}}. \quad (\text{B.58})$$

Applying Corollary 4.3 in Gallant and White (1988),

$$\sup_{M, i \in D_M} \|\eta_{iM}^s\|_2 \leq C \frac{b_M^d}{\sqrt{\rho_{cM}}} \psi(s), \quad (\text{B.59})$$

where  $\psi(s)$  is of order  $-d$ .

Using a similar argument of (B.30),

$$\begin{aligned} \mathbb{V}\left(\sum_{i \in D_M} Y_{iM}\right) &= \mathbb{V}\left(\sum_{i \in D_M} \lambda_{iM}^h + \sum_{i \in D_M} \eta_{iM}^h\right) \\ &= \mathbb{V}\left(\sum_{i \in D_M} \lambda_{iM}^h\right) + \mathbb{V}\left(\sum_{i \in D_M} \eta_{iM}^h\right) + 2Cov\left(\sum_{i \in D_M} \lambda_{iM}^h, \sum_{i \in D_M} \eta_{iM}^h\right) \\ &\leq \sum_{i \in D_M} \mathbb{E}[(\lambda_{iM}^h)^2] + \sum_{i \in D_M} \mathbb{V}(\eta_{iM}^h) + \sum_{i, j \in D_M, i \neq j} |Cov(\eta_{iM}^h, \eta_{jM}^h)| \end{aligned}$$



$$\begin{aligned}
& + 2 \sum_{i,j \in D_M, i \neq j} |Cov(\lambda_{iM}^h, \eta_{jM}^h)| + 2 \sum_{i \in D_M} |Cov(\lambda_{iM}^h, \eta_{iM}^h)| \\
& \leq C_1 |D_M| \frac{b_m^{2d}}{\rho_{cM}} + C_2 \frac{b_m^{2d}}{\rho_{cM}} \sum_{i \in D_M} \left[ \sum_{j \in D_M: j \neq i, \nu(i,j) \leq h} \psi(h) + \sum_{j \in D_M: \nu(i,j) > h} \psi(h) \right] \\
& \leq C |D_M| \frac{b_m^{2d}}{\rho_{cM}}. \tag{B.60}
\end{aligned}$$

Using Chebyshev's inequality, for any  $\eta > 0$ ,

$$\begin{aligned}
& P\left(\left|\frac{1}{J_M^2 |D_M|} \sum_{i \in D_M} W_{iM}\right| \geq \eta\right) \leq \frac{1}{\eta^2 |D_M|^2} \mathbb{V}\left(\sum_{i \in D_M} Y_{iM}\right) \\
& \leq \frac{C}{\eta^2} \frac{b_M^{2d}}{|D_M| \rho_{cM}} = o(1). \tag{B.61}
\end{aligned}$$

Hence, the proof.

**Proof of Theorem 2.2:**

Since (B.51) holds by replacing  $\check{\theta}$  with  $\hat{\theta}_N$ ,

$$J_M \hat{H}_N(\hat{\theta}_N)^{-1} = J_M H_M(\theta_M^*)^{-1} (I_k + o_p(1)), \tag{B.62}$$

where  $J_M = \max_{i \in D_M} \{c_{iM}, d_{iM}\}$  and  $d_{iM}$  is the maximum of the NED scaling factors of the random fields used in the proof.

Notice that

$$\begin{aligned}
S_M &= \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \mathbb{E} \left\{ \left( \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{iM}(W_{iM}, \theta_M^*) - \sqrt{\rho_{uM} \rho_{cM}} \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)] \right) \right. \\
& \quad \cdot \left. \left( \frac{R_{jM}}{\sqrt{\rho_{uM} \rho_{cM}}} m_{jM}(W_{jM}, \theta_M^*) - \sqrt{\rho_{uM} \rho_{cM}} \mathbb{E}[m_{jM}(W_{jM}, \theta_M^*)] \right) \right\} \\
&= \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM}), \tag{B.63}
\end{aligned}$$

where

$$\tilde{m}_{iM} = \frac{R_{iM}}{\sqrt{\rho_{uM}\rho_{cM}}} m_{iM}(W_{iM}, \theta_M^*) - \sqrt{\rho_{uM}\rho_{cM}} \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)]. \quad (\text{B.64})$$

The proof below is a generalization of Theorem 6.8 in Gallant and White (1988) to random fields under finite populations.

$$\begin{aligned} & \left\| \hat{S}_N(\hat{\theta}_N) - S_M - \rho_{uM}\rho_{cM}S_E \right\| / J_M^2 \\ & \leq \left\| \hat{S}_N(\hat{\theta}_N) - \hat{S}_N(\theta_M^*) \right\| / J_M^2 + \left\| \hat{S}_N(\theta_M^*) - \tilde{S}_M - \rho_{uM}\rho_{cM}S_E \right\| / J_M^2 \\ & \quad + \left\| \tilde{S}_M - S_M \right\| / J_M^2, \end{aligned} \quad (\text{B.65})$$

where

$$\tilde{S}_M = \frac{1}{|D_M|} \sum_{i \in D_M} \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{iM}') + \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M: j \neq i, \nu(i,j) \leq b_M} \omega\left(\frac{\nu(i,j)}{b_M}\right) \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM}'). \quad (\text{B.66})$$

Since any sequence of symmetric matrices  $\{A_N\}$  converges to a symmetric matrix  $\{A_0\}$  if and only if  $c'A_Nc \rightarrow c'A_0c$  for any vectors  $c$ , we can reach our conclusion by taking an arbitrary linear combination of  $m_{iM}(W_{iM}, \theta)$ . From now on, we focus on the case of scalar  $m_{iM}(W_{iM}, \theta)$ .

For the last term in the right-hand side of (B.65),

$$\begin{aligned} |\tilde{S}_M - S_M| / J_M^2 & \leq \frac{1}{|D_M|} \left| \sum_{i \in D_M} \sum_{j \in D_M: j \neq i, \nu(i,j) \leq b_M} \left[ \omega\left(\frac{\nu(i,j)}{b_M}\right) - 1 \right] \cdot \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM}) \right| / J_M^2 \\ & \quad + \frac{1}{|D_M|} \left| \sum_{i \in D_M} \sum_{j \in D_M: \nu(i,j) > b_M} \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM}) \right| / J_M^2. \end{aligned} \quad (\text{B.67})$$

The first term in the right-hand side of (B.67) goes to zero because of Assumption 7. For the second term in the right-hand side of (B.67),

$$\begin{aligned}
& \frac{1}{|D_M|} \left| \sum_{i \in D_M} \sum_{j \in D_M: \nu(i,j) > b_M} \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM}) \right| / J_M^2 \\
&= \frac{1}{|D_M|} \left| \sum_{i \in D_M} \sum_{j \in D_M} \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM}) \right| / J_M^2 - \frac{1}{|D_M|} \left| \sum_{i \in D_M} \sum_{j \in D_M: \nu(i,j) \leq b_M} \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM}) \right| / J_M^2.
\end{aligned} \tag{B.68}$$

Applying a similar argument in (B.30),  $\frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \mathbb{E}(\tilde{m}_{iM} \tilde{m}_{jM} / J_M^2)$  is bounded. As  $b_M \rightarrow \infty$  along with  $M \rightarrow \infty$ , the two sums on the right-hand side of (B.68) approach to the same limit. As a result,

$$|\tilde{S}_M - S_M| / J_M^2 = o(1). \tag{B.69}$$

In terms of the first term in the right hand side of (B.65), take a mean value expansion of  $\hat{S}_N(\hat{\theta}_N)$  around  $\theta_M^*$ . Let  $\check{\theta}$  denote the mean value from this expansion.

$$\begin{aligned}
& |\hat{S}_N(\hat{\theta}_N) - \hat{S}_N(\theta_M^*)| / J_M^2 \\
&= \left| (\hat{\theta}_N - \theta_M^*) \frac{1}{|D_N|} \sum_{i \in D_M} \sum_{j \in D_M} R_{iM} R_{jM} \cdot \omega\left(\frac{\nu(i,j)}{b_M}\right) \right. \\
&\quad \cdot \left. [\nabla_{\theta} m_{iM}(W_{iM}, \check{\theta}) m_{jM}(W_{jM}, \check{\theta}) + m_{iM}(W_{iM}, \check{\theta}) \nabla_{\theta} m_{jM}(W_{jM}, \check{\theta})] / J_M^2 \right| \\
&\leq C_1 |\sqrt{|D_N|}(\hat{\theta}_N - \theta_M^*)| \frac{1}{|D_N|} \sum_{i \in D_N} \frac{1}{\sqrt{|D_N|}} \sum_{r=1}^{b_M} \sum_{j \in D_N: \nu(i,j) \in [r, r+1)} \sup_{\theta \in \Theta} |\nabla_{\theta} m_{iM}(W_{iM}, \theta) \cdot m_{jM}(W_{jM}, \theta) / J_M^2| \\
&\leq C |\sqrt{|D_N|}(\hat{\theta}_N - \theta_M^*)| \frac{1}{\sqrt{|D_N|}} \sum_{r=1}^{b_M} r^{d-1} \frac{1}{|D_N|} \sum_{i \in D_N} \sup_{\theta \in \Theta} |\nabla_{\theta} m_{iM}(W_{iM}, \theta) \cdot m_{jM}(W_{jM}, \theta) / J_M^2|
\end{aligned} \tag{B.70}$$

Since

$$\begin{aligned}
& \mathbb{E} \left[ \frac{1}{|D_N|} \sum_{i \in D_N} \sup_{\theta \in \Theta} |\nabla_{\theta} m_{iM}(W_{iM}, \theta) \cdot m_{jM}(W_{jM}, \theta) / J_M^2| \right] \\
& \leq \sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} |\nabla_{\theta} m_{iM}(W_{iM}, \theta) \cdot m_{jM}(W_{jM}, \theta) / J_M^2| \right] \\
& \leq \sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} |\nabla_{\theta} m_{iM}(W_{iM}, \theta) / c_{iM}|^2 \right]^{1/2} \cdot \sup_{M, i \in D_M} \mathbb{E} \left[ \sup_{\theta \in \Theta} |m_{iM}(W_{iM}, \theta) / c_{iM}|^2 \right]^{1/2} < \infty,
\end{aligned} \tag{B.71}$$

$$\frac{1}{|D_N|} \sum_{i \in D_N} \sup_{\theta \in \Theta} |\nabla_{\theta} m_{iM}(W_{iM}, \theta) \cdot m_{jM}(W_{jM}, \theta) / J_M^2| = O_p(1) \tag{B.72}$$

by Markov's inequality. Given  $b_M = o((|D_M| \rho_{uM} \rho_{cM})^{1/2d})$ ,  $\frac{1}{\sqrt{|D_N|}} \sum_{r=1}^{b_M} r^{d-1} = o_p(1)$ . Also,  $\sqrt{|D_N|}(\hat{\theta}_N - \theta_M^*) = O_p(1)$  by Theorem 2.1. Hence,  $|\hat{S}_N(\hat{\theta}_N) - \hat{S}_N(\theta_M^*)| / J_M^2 = o_p(1)$ .

Let us focus on the second term in the right-hand side of (B.65) now.

$$\begin{aligned}
\hat{S}_N(\theta_M^*) &= \frac{1}{|D_N|} \sum_{i \in D_M} \sum_{j \in D_M} R_{iM} R_{jM} \cdot \omega \left( \frac{\nu(i, j)}{b_M} \right) m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*) \\
&= \frac{|D_M| \rho_{uM} \rho_{cM}}{|D_N|} \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} \omega \left( \frac{\nu(i, j)}{b_M} \right) m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*) \\
&= (1 + o_p(1)) \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} \omega \left( \frac{\nu(i, j)}{b_M} \right) m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*)
\end{aligned} \tag{B.73}$$

by Lemma B.1. It suffices to prove that

$$\begin{aligned}
& \left| \frac{1}{|D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} \omega \left( \frac{\nu(i, j)}{b_M} \right) m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*) - \tilde{S}_M - \rho_{uM} \rho_{cM} S_E \right| / J_M^2 \\
&= \left| \frac{1}{J_M^2 |D_M|} \sum_{i \in D_M} \sum_{j \in D_M} \omega \left( \frac{\nu(i, j)}{b_M} \right) \left\{ \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*) \right. \right.
\end{aligned}$$

$$\begin{aligned}
& - \mathbb{E} \left[ \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*) \right] \Bigg\} \Bigg| \\
& \leq \left| \frac{1}{J_M^2 |D_M|} \sum_{i \in D_M} \left\{ \frac{R_{iM}}{\rho_{uM} \rho_{cM}} m_{iM}(W_{iM}, \theta_M^*)^2 - \mathbb{E}[m_{iM}(W_{iM}, \theta_M^*)^2] \right\} \right| \\
& + \left| \frac{1}{J_M^2 |D_M|} \sum_{i \in D_M} \sum_{j \in D_M: j \neq i, \nu(i, j) \leq b_M} \omega\left(\frac{\nu(i, j)}{b_M}\right) \left\{ \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*) \right. \right. \\
& \left. \left. - \mathbb{E} \left[ \frac{R_{iM} R_{jM}}{\rho_{uM} \rho_{cM}} m_{iM}(W_{iM}, \theta_M^*) m_{jM}(W_{jM}, \theta_M^*) \right] \right\} \right| = o_p(1). \tag{B.74}
\end{aligned}$$

Applying Lemma B.2, the first term in the right-hand side of the last inequality of (B.74) converges to zero in probability. Applying Lemma B.5, the second term in the right-hand side of the last inequality of (B.74) converges to zero in probability as well.

Hence,

$$\begin{aligned}
& |\hat{V}_{SN} - (V_M + \rho_{uM} \rho_{cM} V_E)| \\
& = |(H_M(\theta_M^*)^{-1} + o_p(1))(S_M + \rho_{uM} \rho_{cM} S_E + o_p(1))(H_M(\theta_M^*)^{-1} + o_p(1)) - (V_M + \rho_{uM} \rho_{cM} V_E)| \\
& = o_p(1). \tag{B.75}
\end{aligned}$$

### Proof of Theorem 3.1:

To simplify notation, from now on, we treat all functions as the original functions divided by  $J_M$ .

First, using similar arguments in the proof of Theorem 2.1,

$$(\hat{\gamma}_N - \gamma_M^*) \xrightarrow{p} \mathbf{0}. \tag{B.76}$$

By the mean value expansion around  $\theta_M^*$ ,

$$\begin{aligned}
& V_{f,M}^{-1/2} \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} f_{iM}(W_{iM}, \hat{\theta}_N) \\
&= V_{f,M}^{-1/2} \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} f_{iM}(W_{iM}, \theta_M^*) \\
&\quad + V_{f,M}^{-1/2} \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \nabla_{\theta} f_{iM}(W_{iM}, \check{\theta}) \sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*),
\end{aligned} \tag{B.77}$$

where  $\check{\theta}$  lies on the line segment connecting  $\theta_M^*$  and  $\hat{\theta}_N$ . Given Theorem 2.1,

$$\sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*) = O_p(1). \tag{B.78}$$

Further,

$$\hat{F}_N(\check{\theta}) = F_M(\theta_M^*) + o_p(1). \tag{B.79}$$

Therefore,

$$\begin{aligned}
& V_{f,M}^{-1/2} \frac{1}{|D_N|} \sum_{i \in D_M} R_{iM} \nabla_{\theta} f_{iM}(W_{iM}, \check{\theta}) \sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*) \\
&= V_{f,M}^{-1/2} F_M(\theta_M^*) \sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*) + o_p(1).
\end{aligned} \tag{B.80}$$

According to the mean value expansion in the proof of the asymptotic normality of  $\sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*)$ ,

$$\sqrt{|D_N|} (\hat{\theta}_N - \theta_M^*) = - \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*) + o_p(1). \tag{B.81}$$

Combining (B.77), (B.80), and (B.81),

$$\begin{aligned}
& V_{f,M}^{-1/2} \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} f_{iM}(W_{iM}, \hat{\theta}_N) \\
&= V_{f,M}^{-1/2} \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} [f_{iM}(W_{iM}, \theta_M^*) - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] + o_p(1).
\end{aligned} \tag{B.82}$$

Subtract  $V_{f,M}^{-1/2} \sqrt{|D_N|} \gamma_M^*$  from both sides of (B.82).

$$\begin{aligned}
& V_{f,M}^{-1/2} \sqrt{|D_N|} (\hat{\gamma} - \gamma_M^*) \\
&= V_{f,M}^{-1/2} \frac{1}{\sqrt{|D_N|}} \sum_{i \in D_M} R_{iM} [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] + o_p(1) \\
&= V_{f,M}^{-1/2} \sqrt{\frac{|D_M| \rho_{uM} \rho_{cM}}{|D_N|}} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* \\
&\quad - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] + o_p(1) \\
&= (1 + o_p(1)) V_{f,M}^{-1/2} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \frac{R_{iM}}{\sqrt{\rho_{uM} \rho_{cM}}} [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* \\
&\quad - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta_M^*)] + o_p(1)
\end{aligned} \tag{B.83}$$

Observe that  $\forall \theta \in \Theta$

$$\begin{aligned}
& \sup_{M, i \in D_M} \mathbb{E} \left[ \left\| [f_{iM}(W_{iM}, \theta) - \gamma_M^* - F_M(\theta_M^*) H_M(\theta_M^*)^{-1} m_{iM}(W_{iM}, \theta)] \right\|^{2+\delta} \right] \\
&\leq \sup_{M, i \in D_M} \left\{ \left[ \mathbb{E} \left( \sup_{\theta \in \Theta} \|f_{iM}(W_{iM}, \theta)\|^{2+\delta} \right) \right]^{1/(2+\delta)} + \|\gamma_M^*\| \right. \\
&\quad \left. + C \|F_M(\theta_M^*)\| \left[ \mathbb{E} \left( \sup_{\theta \in \Theta} \|m_{iM}(W_{iM}, \theta)\|^{2+\delta} \right) \right]^{1/(2+\delta)} \right\}^{2+\delta} < \infty
\end{aligned} \tag{B.84}$$

for some  $\delta > 0$  by Minkowski's inequality and Jensen's inequality. By Lemma B.4 above

and Corollary 1 in Jenish and Prucha (2012),

$$V_{f,M}^{-1/2} \frac{1}{\sqrt{|D_M|}} \sum_{i \in D_M} \left\{ \frac{R_{iM}}{\sqrt{\rho_{uM}\rho_{cM}}} [f_{iM}(W_{iM}, \theta_M^*) - \gamma_M^* - F_M(\theta_M^*)H_M(\theta_M^*)^{-1}m_{iM}(W_{iM}, \theta_M^*)] \right. \\ \left. - \sqrt{\rho_{uM}\rho_{cM}} \left[ \mathbb{E}(f_{iM}(W_{iM}, \theta_M^*)) - \gamma_M^* - F_M(\theta_M^*)H_M(\theta_M^*)^{-1}\mathbb{E}(m_{iM}(W_{iM}, \theta_M^*)) \right] \right\} \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_q). \quad (\text{B.85})$$

The proof of Theorem 3.1(2) follows closely from the proof of Theorem 2.2 and hence is omitted.

## References

- Abadie, A., Athey, S., Imbens, G.W., and Wooldridge, J. (2017), When should you adjust standard errors for clustering? Tech. rep., National Bureau of Economic Research.
- Abadie, A., Athey, S., Imbens, G.W., and Wooldridge, J.M. (2020), Sampling-based versus design-based uncertainty in regression analysis. *Econometrica* 88(1), 265–296.
- Aliprantis, D. and Hartley, D. (2015), Blowing it up and knocking it down: The local and city-wide effects of demolishing high concentration public housing on crime. *Journal of Urban Economics* 88, 67–81.
- Bojinov, I., Rambachan, A., and Shephard, N. (2021), Panel experiments and dynamic causal effects: A finite population perspective. *Quantitative Economics* 12(4), 1171–1196.
- Bradley, R.C. and Tone, C. (2017), A central limit theorem for non-stationary strongly mixing random fields. *Journal of Theoretical Probability* 30(2), 655–674.
- Gallant, A.R. and White, H. (1988), *A unified theory of estimation and inference for nonlinear dynamic models*. Blackwell.



- Houde, J.F. (2012), Spatial differentiation and vertical mergers in retail markets for gasoline. *American Economic Review* 102(5), 2147–2182.
- Hudgens, M.G. and Halloran, M.E. (2008), Toward causal inference with interference. *Journal of the American Statistical Association* 103(482), 832–842.
- Jenish, N. and Prucha, I.R. (2009), Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150(1), 86–98.
- Jenish, N. and Prucha, I.R. (2012), On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170(1), 178–190.
- Leung, M.P. (2022), Causal inference under approximate neighborhood interference. *Econometrica* 90(1), 267–293.
- Manski, C.F. (2013), Identification of treatment response with social interactions. *The Econometrics Journal* 16(1), S1–S23.
- Móricz, F., Stadtmüller, U., and Thalmaier, M. (2008), Strong laws for blockwise  $m$ -dependent random fields. *Journal of Theoretical Probability* 21(3), 660–671.
- Müller, U.K. and Watson, M.W. (2022a), Spatial correlation robust inference. *Econometrica* .
- Müller, U.K. and Watson, M.W. (2022b), Spatial correlation robust inference in linear regression and panel models. *Journal of Business & Economic Statistics* pp. 1–15.
- Newey, W.K. (1991), Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59, 1161–1167.
- Newey, W.K. and McFadden, D. (1994), Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4, 2111–2245.

- Neyman, J. (1923), On the application of probability theory to agricultural experiments. essay on principles. section 9. *Annals of Agricultural Sciences* 10, 1–51.
- Pinkse, J., Shen, L., and Slade, M. (2007), A central limit theorem for endogenous locations and complex spatial interactions. *Journal of Econometrics* 140(1), 215–225.
- Rambachan, A. and Roth, J. (2022), Design-based uncertainty for quasi-experiments. *arXiv preprint arXiv:2008.00602* .
- Sävje, F. (2021), Causal inference with misspecified exposure mappings. Tech. rep., Yale University.
- Sävje, F., Aronow, P.M., and Hudgens, M.G. (2021), Average treatment effects in the presence of unknown interference. *The Annals of Statistics* 49(2), 673–701.
- Xu, R. (2021a), Asymptotic properties of m-estimators with finite populations under cluster sampling and cluster assignment. Tech. rep., Rutgers University.
- Xu, R. (2021b), Potential outcomes and finite-population inference for m-estimators. *The Econometrics Journal* 24(1), 162–176.
- Xu, X. and Lee, L.f. (2019), Theoretical foundations for spatial econometric research. *Regional Science and Urban Economics* 76, 2–12.