*Econometric Theory*, **32**, 2016, 154–186.
doi:10.1017/S0266466614000814

# FIXED-b ASYMPTOTICS FOR SPATIALLY DEPENDENT ROBUST NONPARAMETRIC COVARIANCE MATRIX ESTIMATORS

C. ALAN BESTER and TIMOTHY G. CONLEY
*University of Western Ontario*

CHRISTIAN B. HANSEN
*University of Chicago Booth School of Business*

TIMOTHY J. VOGELSANG
*Michigan State University*

This paper develops a method for performing inference using spatially dependent data. We consider test statistics formed using nonparametric covariance matrix estimators that account for heteroskedasticity and spatial correlation (spatial HAC). We provide distributions of commonly used test statistics under "fixed-$b$" asymptotics, in which HAC smoothing parameters are proportional to the sample size. Under this sequence, spatial HAC estimators are not consistent but converge to nondegenerate limiting random variables that depend on the HAC smoothing parameters, the HAC kernel, and the shape of the spatial region in which the data are located. We illustrate the performance of the "fixed-$b$" approximation in the spatial context through a simulation example.

## 1. INTRODUCTION

Many economic analyses rely on dependent data. To conduct inference with time series data, researchers often employ heteroskedasticity and autocorrelation consistent (HAC) inference procedures. The essence of HAC procedures is the use of a nonparametric estimator for the variance of estimated model parameters, which provides valid standard errors in the presence of intertemporal correlation.[1] The use of standard error estimators that account for temporal dependence has also become common in microeconomic applications involving panel data, where researchers employ clustered covariance estimators with clusters given by the cross-sectional units of observation.[2]

We thank Silvia Gonçalves for helpful discussions. We thank seminar participants at University of Wisconsin-Madison, Harvard, MIT, Boston University, Princeton, and Hebrew University. This research was also supported by the Social Science and Humanities Research Council. We also thank anonymous referees and an editor for helpful comments. We are grateful for support from the Neubauer Family Faculty Fund and from the IBM Corporation Faculty Research Fund at the University of Chicago Graduate School of Business.
Address correspondence to Christian Hansen, University of Chicago Booth School of Business, 5807 S Woodlawn Ave, Chicago, IL 60637, USA. e-mail: chansen1@chicagobooth.edu.


© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

An often neglected feature of cross-sectional data is cross-sectional or spatial correlation: Data or residuals from one observation may be statistically related to data or residuals from other neighboring observations. As is the case with temporal correlation, inference procedures that do not account for such spatial correlation will usually be incorrect when it exists in the data. To account for this problem, spatial HAC procedures have been developed.[3] These spatial HAC estimators allow for processes with dependence indexed in more than one dimension. In this paper, we present a novel way to use spatial HAC estimators. Our approach is based on an alternate asymptotic approximation to the behavior of associated test statistics, referred to in time series contexts as "fixed-*b*" asymptotics following Kiefer and Vogelsang (2002, 2005).

Fixed-*b* approximations differ from the usual approximation in the treatment of the HAC covariance matrix estimator and the resulting reference distributions of test statistics. Researchers typically employ a limiting Gaussian approximation that depends on an unknown variance-covariance matrix. The covariance matrix is then estimated using a HAC estimator which requires the choice of a kernel and smoothing parameters. This HAC estimator is then "plugged-in" for the unknown covariance matrix in the asymptotic Gaussian approximation, which is then used to construct confidence intervals and perform statistical tests. These procedures are justified by a promise that kernel and smoothing parameters are chosen so that the HAC covariance matrix estimator is consistent. They implicitly treat the estimator as if it perfectly estimates the underlying covariance matrix, ignoring sampling uncertainty and other issues such as small sample bias that are known to arise in nonparametric covariance estimation.

In contrast, the time series fixed-*b* approach of Kiefer and Vogelsang (2002, 2005) (KV) considers the behavior of statistics under an asymptotic sequence in which the HAC smoothing parameter is proportional to the sample size (this fixed proportion is the source of the label "fixed-*b*"). This contrasts with the conventional sequence where the smoothing parameter grows more slowly than the sample size. Under the KV sequence, the HAC estimator converges in distribution to a nondegenerate random variable. The resulting limit distributions for $t$ and Wald statistics are pivotal but nonstandard, so critical values are obtained by simulation. Because they explicitly treat the HAC estimator as random, these approximations account for both sampling uncertainty and bias in the HAC estimator. Through simulations, KV provide evidence that their approximation outperforms the plug-in approach.[4]

In this paper, we extend the analysis of KV to allow for spatially dependent data. We consider the behavior of spatial HAC estimators as in Conley (1996, 1999) under asymptotic sequences in which smoothing parameters are proportional to the sample size. We follow KV and refer to this asymptotic sequence as fixed-*b* asymptotics. As in KV, this analysis produces a random variable limit for the HAC estimator and nonstandard limit distributions for $t$ and Wald statistics.

Extension of the KV results from the time series to the spatial context is both important and nontrivial. Nonparametric estimation of covariances can be

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

challenging in the time series case and is even more so when the data exhibit dependence in more than one dimension. The resulting variance estimators exhibit pronounced sampling variability and small sample biases, both of which can be captured to a large degree through the use of fixed-*b* asymptotics. The fixed-*b* reference distributions for test statistics in the spatial setting are considerably more complicated than those obtained in the time series case.

In the time series case, KV show that the fixed-*b* reference distributions depend explicitly on the choice of kernel and bandwidth. They suggest this feature is desirable, as we know that the finite sample distribution of test statistics changes based on these factors. In this paper, we find that in the spatial setting, these distributions also depend explicitly on the shape of the spatial sampling region. Like KV, we argue that this is an advantage, as in practice spatial data are observed in a wide variety of different configurations. One contribution of our paper is to make this dependence on the shape of the sampling region explicit. In particular, under the fixed-*b* sequence, the limit of the HAC estimator can be represented in terms of integrals against a set-indexed Gaussian process.

There are, of course, other ways to perform inference with spatially dependent data. Kelejian and Prucha (1999, 2001) and Lee (2004, 2007a, 2007b) consider inference based on specific models for the spatial dependence structure. Methods that do not rely on parametric covariance structures are also available. Ibragimov and Müller (2010) consider doing inference based on partitioning data into groups that are "approximately" uncorrelated and estimating parameters within each partition. Ibragimov and Müller (2010) provide properties of t-tests formed using these sets of point estimates. Bester, Conley, and Hansen (2010) consider a different, but related procedure based on using a clustered covariance estimator with researcher-defined clusters. In both Ibragimov and Müller (2010) and Bester, Conley, and Hansen (2010), the key condition required for such tests' validity is that averages of scores within clusters be asymptotically uncorrelated, which may be achieved through appropriate choice of partitions. Both papers are also in the spirit of fixed-*b* asymptotics, as both approaches advocate the use of large groups and provide asymptotic analysis that accounts for uncertainty in estimating the covariance matrix.

Our paper complements this literature. We obtain inference that is robust to general forms of spatial correlation without functional form assumptions on the dependence structure. Relative to Ibragimov and Müller (2010) and Bester, Conley, and Hansen (2010), the procedure in this paper allows finer control over assumed dependence structure through the use of downweighting via the kernel function and smoothing parameters. By allowing for smooth decay of covariance across observational units, it is less susceptible to boundary problems than the grouping schemes of Ibragimov and Müller (2010) and Bester, Conley, and Hansen (2010) as these estimators by construction ignore covariances between observations on the boundary of one group with observations in other groups. The drawback is that we derive the fixed-*b* approximation under strong stationarity and homogeneity conditions not used in either Ibragimov and Müller (2010)

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

or Bester, Conley, and Hansen (2010). Thus, we view the two approaches as complementary.

We present a set of simulation results that illustrate the performance of the fixed-$b$ approximation. We make comparisons to the approach of Ibragimov and Müller (2010) and so our results complement the simulations reported by Ibragimov and Müller (2010) for spatially correlated settings. Our simulations show that inference procedures that ignore spatial correlation are highly size-distorted when spatial correlation is present. As in Ibragimov and Müller (2010), we also see that large size distortions persist when spatial HAC estimators are employed with the usual asymptotic normal approximation. Across the simulations, correctly sized tests are obtained using the inference procedure developed in our paper along with HAC estimators using large values of the smoothing parameter or with the Ibragimov and Müller (2010) approach with a small number of groups. Further discussion on the relative merits of the fixed-$b$ HAC approach and the Ibragimov and Müller (2010) approach is presented in the simulation section. Our simulations also highlight the potential importance of accounting for the shape of the sampling region.

The remainder of this paper is organized as follows. In Section 2, we review spatial HAC estimation. We present the main asymptotic results in Section 3. Section 4 contains results from our simulation experiments, and Section 5 concludes.

## 2. SPATIAL HAC ESTIMATION

We consider estimation of a $p \times 1$ dimensional parameter $\theta_0$ identified by the following moment condition:

$$Es_t(\theta_0) = 0,$$

where the index $t$ indexes observations of a $p \times 1$ score function $s_t(\theta)$. Given a sample of $N$ observations, we are interested in the sample analog estimator of $\theta_0$, $\hat{\theta}$, implicitly defined as the solution to:

$$\frac{1}{N}\sum_{t=1}^{N} s_t(\hat{\theta}) = 0. \tag{2.1}$$

To facilitate modeling spatial dependence across observations, we take our observations to be located on a two-dimensional integer lattice with each observation $t$ having a vector of integer coordinates $d(t) = \{i, j\}$. When convenient, we index observations with coordinates $(i, j)$ directly via $i, j$ subscripts, e.g., $s_{i,j}(\theta)$.

The coordinates for observations in a sample lie within a set of integer ordered pairs denoted $B_n$, which is one of a sequence of 'sample regions' indexed by $n$.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

We present the case where all integer coordinate pairs within $B_n$ are observed so that

$$\frac{1}{N}\sum_{t=1}^{N} s_t(\theta) = \frac{1}{N}\sum_{(i,j)\in B_n} s_{i,j}(\theta).$$

Our results extend to many scenarios where only a subset of locations are observed/sampled in an exogenous manner, independent of the process $s_{i,j}(\theta_0)$. Appendix B provides further discussion.

Under suitable regularity conditions, $\widehat{\theta}_N$ will have an asymptotic normal distribution. To simplify notation define $s_{i,j} = s_{i,j}(\theta_0)$ and its sample analog $\widehat{s}_{i,j} = s_{i,j}\left(\widehat{\theta}_N\right)$. The variance covariance matrix for $\widehat{\theta}_N$ has a sandwich form with middle matrix given by

$$\Omega = \lim_{N\to\infty} \frac{1}{N} E\left[\left(\sum_{(i,j)\in B_n} s_{i,j}\right)\left(\sum_{(i,j)\in B_n} s_{i,j}\right)'\right].$$

The nonparametric HAC estimator of $\Omega$ is given by

$$\widehat{\Omega} = \frac{1}{N}\sum_{(i_1,j_1)\in B_n}\sum_{(i_2,j_2)\in B_n} K\left(\frac{i_1}{h_1}, \frac{i_2}{h_1}, \frac{j_1}{h_2}, \frac{j_2}{h_2}\right)\widehat{s}_{i_1,j_1}\widehat{s}'_{i_2,j_2}, \tag{2.2}$$

where $K(x_1, x_2, x_3, x_4) \leq 1$ is a kernel weighting function that satisfies $K(x, x, y, y) = 1$ and $h_1$ and $h_2$ are tuning parameters where larger (smaller) values of $h_1$ and $h_2$ indicate less (more) smoothing for a given kernel.

## 2.1. Sampling Regions

Part of the contribution of our paper is to explicitly characterize the dependence of the limiting distribution of $\widehat{\Omega}$ on the shape of the sampling region. The shape of our sampling region is summarized by a region $B \subseteq [-1/2, 1/2]^2$. For a scalar $c$, define $cB = \{c\mathbf{u} : \mathbf{u} \in B\}$. For each $n \in \mathbb{N}$, we formally define our set of sampled locations on the integer lattice $B_n = nB \cap \mathbb{Z}^2$. Intuitively, the sample starts at the origin for $n = 1$ and propagates outward at the same rate along each coordinate direction. The sample size is then $N(n) \equiv \|B_n\| = O\left(n^2\right)$, and we will have $N(n)/n^2 \to \lambda(B)$ where $\lambda$ is Lebesgue measure. For the remainder of the paper, we write $N$ for the sample size, suppressing the dependence on $n$, and write limits as $N \to \infty$ and $n \to \infty$ interchangeably.

We provide results for sampling regions with $B$ having boundaries with a property that we call "quadrant-wise monotone boundaries" (QMB). We define a QMB region $B \subseteq [-1/2, 1/2]^2$ as a connected region consisting of four disjoint subregions, $B = \bigcup_{i=1}^{4} B_j$, one in each quadrant of the plane. The boundary of each subregion consists of a horizontal and a vertical line segment extending outward

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

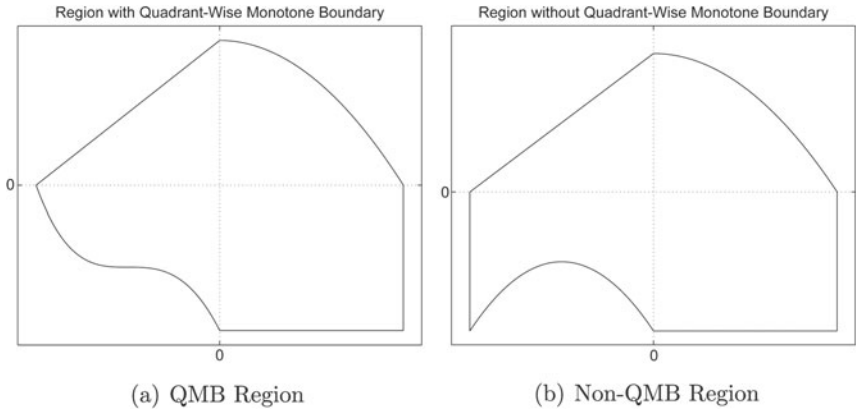



**FIGURE 1.** Sample regions satisfying (a) and violating (b) definition of quadrant-wise monotone boundary.

from the origin to an endpoint, as well as a monotone curve connecting the two endpoints. For simplicity, we assume that the boundary of the entire region $B$ is continuous. Figure 1(a) illustrates a QMB region, and Figure 1(b) illustrates a region that violates the QMB assumption. Note that the boundary of the region $B$ in the third quadrant of Figure 1(b) is nonmonotone.

The QMB assumption simplifies our analysis because it allows interchanging the order of summation in each subregion. For $k = 1, \ldots, 4$, define the integer-valued function

$$\bar{L}_{n,k}(j) = \max_{i \in \mathbb{Z}} \{|i| : (i, j) \in nB_k\},$$

and similarly for $\bar{M}_{n,k}(i)$. The QMB assumption ensures these functions are monotone within each subregion. For example, $\bar{L}_{n,1}(j)$ and $\bar{M}_{n,1}(i)$ are both non-increasing functions, i.e., for $j' > j$, we have $\bar{L}_{n,1}(j') \leq \bar{L}_{n,1}(j)$. This means that, for any function $w : \mathbb{Z}^2 \to \mathbb{R}$, we have

$$\sum_{i,j \in B_{n,1}} w_{i,j} = \sum_{i=1}^{\bar{\mathbf{L}}_{n,1}} \sum_{j=1}^{\bar{M}_{n,1}(i)} w_{i,j} = \sum_{j=1}^{\bar{\mathbf{M}}_{n,1}} \sum_{i=1}^{\bar{L}_{n,1}(j)} w_{i,j},$$

where $\bar{\mathbf{L}}_{n,1} = \sup_j \bar{L}_{n,1}(j)$ and similarly for $\bar{\mathbf{M}}_{n,1}$.

Note that the generality of the QMB assumption can depend on the kernel. For example, consider the region depicted in Figure 1(a) rotated 45 degrees in either direction. If the coordinate axes remain as depicted in the figure, the region no longer has monotone boundaries. However, if the kernel $K$ is isotropic, the axes may be redefined so that the assumption is satisfied; in fact the HAC estimator (2.2) is invariant to this rotation.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

## 3. FIXED-b ASYMPTOTICS

This section presents asymptotic theory for the spatial HAC covariance estimator described in the previous section under a sequence where the smoothing parameters $h_1$ and $h_2$ grow quickly relative to the sample size. Under this sequence, the HAC estimator converges to a random variable whose distribution involves set-indexed Brownian motions (cf. Dedecker, 2001). Section 3.1 provides a brief overview and definitions of these processes. In Section 3.2, we provide results on the limit distributions of the HAC estimator and related test statistics.

As mentioned above, the distributions of t and Wald tests are nonstandard and explicitly reflect both the kernel and chosen values of smoothing parameters as well as the shape of the set to which indices $d(t)$ belong. Because of the many bandwidth and kernel choices available in practice and the multitude of potential shapes of sampling regions, it is impractical to tabulate all relevant critical values. Critical values must therefore typically be calculated by simulation on a case-by-case basis.

### 3.1. Set-Indexed Brownian Motion and Related Processes

In the time series case, Kiefer and Vogelsang (2002) show that, when bandwidth equals sample size,[5]

$$\widehat{\Omega}_T \Rightarrow \Lambda \left[ \int_0^1 \int_0^1 \frac{\partial^2 K(r,s)}{\partial r \partial s} Y_p(r) Y_p(s) dr ds \right] \Lambda', \tag{3.1}$$

where $Y_p$ is a $p$-vector of standard Brownian bridges. To extend their results to data observed at locations in the plane, we first need to define the analog of Brownian motion indexed by sets in the unit rectangle.

Let $\mathcal{A}$ be a collection of Borel subsets of $[0,1]^d$, and $C(\mathcal{A})$ be the space of continuous real functions on $\mathcal{A}$ equipped with the sup norm. A standard Brownian motion, $W$, indexed by $\mathcal{A}$ is a mean zero Gaussian process with sample paths in $C(\mathcal{A})$ and covariance function

$$\mathrm{Cov}(W(A), W(B)) = \lambda(A \cap B),$$

where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}^d$. With $d = 1$ and $\mathcal{A}$ including all Borel subsets of $[0,1]$, the set-indexed Brownian motion is simply the usual Brownian motion; and with $d = 2$ and $\mathcal{A}$ including all Borel subsets of $[0,1]^2$, it is the standard Brownian Sheet (cf. Deo, 1975 or Basu and Dorea, 1979). Denote by $W_p$ a $p$-vector of independent set-indexed Brownian motions. Since we want our region to be centered at the origin to simplify exposition, we consider a shifted set-indexed Brownian motion defined with $\mathcal{A}$ a collection of Borel subsets of $[-1/2, 1/2]^2$.

Fix a set $B \subseteq [-1/2, 1/2]^2$ and consider a sequence of sampling regions $B_n = nB \cap \mathbb{Z}^2$ as described above. Denote by $\partial B$ the boundary of the region $B$.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

The limiting distribution of $\widehat{\Omega}$ and related test-statistics will depend on "pinned" Brownian processes. For sets $A \subseteq B$ with $A \in \mathcal{A}$, we define the pinned process

$$W_p(A,B) = W_p(A) - \frac{\lambda(A)}{\lambda(B)} W_p(B). \tag{3.2}$$

The limit distributions in the next section will involve integrals of the pinned Brownian process $W_p(A,B)$ against derivatives of the kernel. For vectors $\mathbf{u} = (u_1,u_2) \in B$, define the set-valued functions $R : B \mapsto \mathcal{A}$ and $P : \partial B \mapsto \mathcal{A}$ as follows:[6]

$$R_B(\mathbf{u}) = (]0 \wedge u_1, 0 \vee u_1] \times ]0 \wedge u_2, 0 \vee u_2]) \cap B$$

$$P_B(\mathbf{u}) = \begin{cases} (]0,u_1] \times ]0,\overline{u}_2]) \cap B & u \in B_1 \\ \left(]u_1,\underline{u}_1] \times ]u_2,0]\right) \cap B & u \in B_2 \\ \left(]u_1,0] \times ]\underline{u}_2,u_2]\right) \cap B & u \in B_4 \\ \left(]\underline{u}_1,0] \times ]0,u_2]\right) \cap B & u \in B_3 \end{cases},$$

where $]\underline{u}_1,\overline{u}_1] \times ]\underline{u}_2,\overline{u}_2]$ is the smallest rectangle containing $B$. Note that $R_B(\mathbf{u})$ is the intersection of B with a rectangle with vertices at $\mathbf{u}$ and the origin while $P_B(\mathbf{u})$ is defined piecewise for each of the four subregions. These functions are illustrated in Figure 2.

Like KV in the time series case, the limit distributions here also involve integrals of derivatives of the kernel against the pinned process $W_p(\cdot, B)$, for example,

$$L_{p,I,I}(b_1,b_2,B) = \int_{\mathbf{u}\in B} \int_{\mathbf{v}\in B} \frac{\partial^4 K_{b_1,b_2}(\mathbf{u},\mathbf{v})}{\partial u_1 \partial u_2 \partial v_1 \partial v_2} W_p\big(R_B(\mathbf{u}),B\big) W_p\big(R_B(\mathbf{v}),B\big)' d\mathbf{v} d\mathbf{u}, \tag{3.3}$$

where $K_{b_1,b_2} = K(\frac{u_1}{b_1}, \frac{v_1}{b_1}, \frac{u_2}{b_2}, \frac{v_2}{b_2})$ and $(b_1,b_2) \in (0,1]^2$ will be related to bandwidths by $(h_1,h_2) = (b_1 n, b_2 n)$. These integrals arise as limits of partial sums taken on the interior of the rectangle.[7] Unlike the time series case, however, the limit distributions also involve integrals such as

$$L_{p,I,II}(b_1,b_2,B) = \int_{\mathbf{u}\in B} \int_{\mathbf{v}\in \partial B} -\frac{\partial^3 K_{b_1,b_2}(\mathbf{u},\mathbf{v})}{\partial u_1 \partial u_2 \partial z(\mathbf{v})} W_p\big(R_B(\mathbf{u}),B\big) W_p\big(R_B(\mathbf{v}),B\big)' d\mathbf{v} d\mathbf{u}$$

$$L_{p,I,III}(b_1,b_2,B) = \int_{\mathbf{u}\in B} \int_{\mathbf{v}\in \partial B} \frac{\partial^3 K_{b_1,b_2}(\mathbf{u},\mathbf{v})}{\partial u_1 \partial u_2 \partial t(\mathbf{v})} W_p\big(R_B(\mathbf{u}),B\big) W_p\big(P_B(\mathbf{v}),B\big)' d\mathbf{v} d\mathbf{u}, \tag{3.4}$$

where $z(\mathbf{v}) = \mathbf{1}(\mathbf{v} \in B_2 \cap B_3) v_1 + \mathbf{1}(\mathbf{v} \in B_1 \cap B_4) v_2$, and $t(\mathbf{v})$ is a unit length vector tangent to the boundary of the region, $\partial B$, in a counterclockwise direction.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

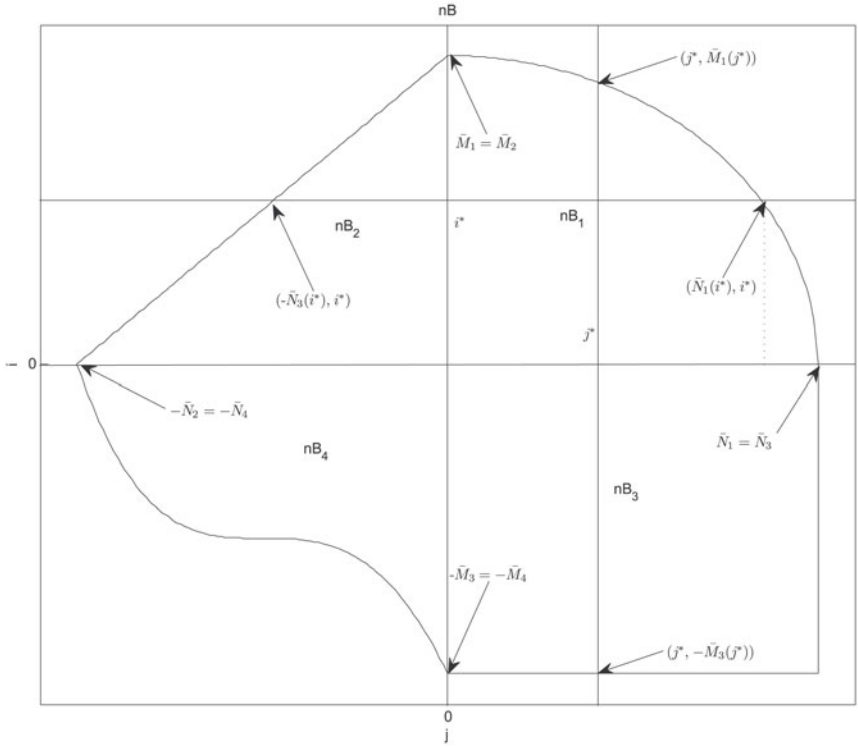



**FIGURE 2.** This figure provides an example of a region B broken into the four subregions $B_1$–$B_4$ and construction of the related index sets used in the definitions in Section 3.1.

Finally, we will also need integrals weighted against $W_p(A, B)$ evaluated over the four subregions $B_i$, e.g.,

$$L_{p,I,IV} = \sum_{i=1}^{4} \int_{\mathbf{u}\in B} \frac{\partial^2 K_{b_1,b_2}(\mathbf{u}, e_i)}{\partial u_1 \partial u_2} W_p\big(R_B(\mathbf{u}), B\big) W_p\big(B_i, B\big)' d\mathbf{u} \tag{3.5}$$

$$L_{p,IV,IV} = \sum_{i=1}^{4}\sum_{j=1}^{4} K(e_i, e_j) W_p(B_i, B) W_p(B_j, B)',$$

where $e_i$ is the 'endpoint' of each region when the boundary $\partial B$ is traversed in a counterclockwise direction starting at $(0, \overline{u}_2)$.[8] There are sixteen types of terms corresponding to possible pairs of the indices $I$ through $IV$. The remaining types of terms are defined in the appendix following the conventions established in (3.3)–(3.5); for example, $L_{II,I}$ is defined as in (3.4) by interchanging **u** and **v**.

As in the time series case (3.1), each term in the limit distribution of $\widehat{\Omega}$ is a random variable whose distribution depends explicitly on both the kernel, $K$, and

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

the 'bandwidth', $b$. In addition to the kernel, each of these terms depends on the shape of the sampling region. Even for isotropic kernels, replacing the region $B$ with a rectangle or monotone boundary polygon of equal area will generally change the limit distribution of $\widehat{\Omega}$. Restrictions on the kernel may simplify these terms. For example, if the kernel is symmetric, $K(\mathbf{u},\mathbf{v}) = K(\mathbf{v},\mathbf{u})$, then any two terms with the same indices are equivalent, e.g., $L_{p,I,II} = L_{p,II,I}$. If, in addition, the kernel evaluated at the four 'endpoints' is equal, then all terms involving the index $IV$ are zero.

### 3.2. Main Theorems

Define $\Lambda$ such that $\Lambda\Lambda' = \Omega$. The following set of assumptions are sufficient for obtaining the main results of the paper. Throughout the following, we use $\lambda(A)$ to denote the Lebesgue measure of region $A$.

**Assumption 1.** Let $\mathcal{A}$ be a collection of Borel subsets of $[-1/2, 1/2]^2$ that satisfies (i) $\int_0^1 \sqrt{H(\mathcal{A}, x)}dx < \infty$ where $H(\mathcal{A}, \epsilon)$ is the logarithm of the smallest number of open disks of radius $\epsilon$ which form a covering of $\mathcal{A}$ and (ii) $\lambda(\partial A) = 0$ for all $A \in \mathcal{A}$ where $\partial A$ is the boundary of $A$.

From Dudley (1973), Assumption 1 is sufficient to guarantee existence of standard Brownian motion over $\mathcal{A}$ as defined in the previous section.

**Assumption 2.** Suppose that $B \in \mathcal{A}$ is a QMB-region such that $A \subseteq B$ for all $A \in \mathcal{A}$. Assume that $B$ is inscribed in the rectangle $[\underline{u}_1, \overline{u}_1] \times [\underline{u}_2, \overline{u}_2]$ where $\max_{j=1,2}\left\{|\underline{u}_j|, |\overline{u}_j|\right\} = 1/2$ and $\min_{j=1,2}\left\{|\underline{u}_j|, |\overline{u}_j|\right\} = c > 0$. Let $B_n = nB \cap \mathbb{Z}^2$ for $n \in \mathbb{N}$. The sample size $N = \|B_n\|$ is such that, as $n \to \infty$, $\frac{N}{n^2} \to \lambda(B)$.

Assumption 2 defines a sequence of sampling regions $B_n$ as a function of a region $B \subseteq [-1/2, 1/2]^2$, which is assumed to have quadrant-wise monotone boundaries (QMB) as discussed in Section 2.1.[9] We note that this condition implicitly assumes roughly uniform sampling over a simple geometric region. As noted above, we suppress the dependence of the sample size, $N$, on $n$, and any limits as $N \to \infty$ are understood as limits with $n \to \infty$ where $n \in \mathbb{N}$.

The next assumption involves a set of mixing and moment conditions for the scores $s_t$. For simplicity, we work directly with the scores rather than data $y_t$ and a score function $S$.

**Assumption 3.** $\{s_{i,j}(\theta_0)\}_{(i,j)\in\mathbb{Z}^2}$ is a strictly stationary field of $\phi$-mixing random variables with mean zero, finite variance, mixing coefficients $\phi_2(k)$,[10] where (i) $\mathrm{E}\left(|s_{i,j}(\theta_0)|^4\right) < \infty$ and $\sum_{k>0} k^3\phi_2(k) < \infty$ or (ii) for some $b$ in $]2, 4[$, $\mathrm{E}\left(|s_{i,j}|^{2b/(b-2)}\right) < \infty$ and $\phi_2(k) = O\left(k^{-b}\right)$. Let $\Lambda\Lambda' = \Omega = \lim_{N\to\infty} \frac{1}{N} E\left[\left(\sum_{(i,j)\in B_n} s_{i,j}\right)\left(\sum_{(i,j)\in B_n} s_{i,j}\right)'\right]$.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

The scores $s_{i,j}$ are assumed to satisfy a set of mixing and moment conditions identical to those required for Theorem 3 in Dedecker (2001). Note we do not assume uniform mixing,[10] which is often assumed in the time series case but is restrictive in the spatial setting. In particular, it is in general too strong for application to Gibbs random fields.[11] The assumption that scores follow a strictly stationary random field rules out heteroskedasticity and various other types of heterogeneity in the score process. Because we are working directly with the scores, we state consistency of the estimator $\widehat{\theta}_N$ defined by (2.1) as a high level assumption.

**Assumption 4.** $\widehat{\theta}_N \xrightarrow{p} \theta_0$ for $\theta_0 \in \text{int}(\Theta)$ where $\Theta$ is a compact set.

Our next assumption is a functional central limit theorem for partial sums of the scores over subsets of the region $B$.

**Assumption 5.** The sequence $\left\{\frac{1}{n}\sum_{(i,j)\in\mathbb{Z}^2} \lambda\left(nA \cap R_{(i,j)}\right) s_{(i,j)} : A \in \mathcal{A}\right\}$ converges in distribution in $C(\mathcal{A})$ to $\Lambda W_p$, where $W_p$ is a set-indexed Brownian motion indexed by $\mathcal{A}$ and $R_{(i,j)} = ]i-1, i] \times ]j-1, j]$ is the unit square with upper corner at $(i, j)$.

Though we state it here as an assumption, this FCLT is proven in Theorem 3 of Dedecker (2001) under our Assumptions 1 and 3. Note that Assumption 2 is not necessary for Dedecker's result, but is assumed here to simplify the proofs, and to allow us to more explicitly characterize the limit distribution of $\widehat{\Omega}$ in terms of the integrals described in Section 3.1. This allows us to better understand the role played by both the kernel and the shape of the sampling region in determining the fixed-$b$ limit distribution. Given that the QMB assumption is in no way required for Assumptions 3–6, we conjecture that the limit distributions in our theorems would exist under much weaker restrictions on the region $B$.

**Assumption 6.**

$$\sup_{A\in\mathcal{A},\theta\in\mathcal{N}(\theta_0)} \left\| \frac{1}{N}\sum_{(i,j)\in nA} \frac{\partial s_{i,j}(\theta)}{\partial \theta'} - \frac{\lambda(A)}{\lambda(B)}\mathcal{J}(\theta) \right\| \xrightarrow{p} 0,$$

where $\mathcal{J}(\theta_0)$ is nonsingular and $\mathcal{N}(\theta_0)$ is an open neighborhood of $\theta_0$.

Assumption 6 may be interpreted as two requirements, the first of which is a uniform (over both $\mathcal{N}(\theta_0)$ and $\mathcal{A}$) law of large numbers. The second requirement is that the expectation of $\partial s_{i,j}/\partial\theta$ is homogeneous across the sampling region. For example, in the regression context, Assumption 6 requires that second moments do not depend on locations. Note that KV require an analogous assumption in the time series context.

We begin our main results with the following lemma, which gives the asymptotic distribution of the estimator $\widehat{\theta}_N$.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

LEMMA 1. *Suppose Assumptions 1–6 hold. Then, as* $N \to \infty$,

$$\sqrt{N}\left(\widehat{\theta}_N - \theta_0\right) \Rightarrow -(\lambda(B)^{1/2}\mathcal{J}(\theta_0))^{-1}\Lambda W_p(B).$$

While asymptotic normality of $\widehat{\theta}_N$ seems trivial from Assumptions 5 and 6, note that the FCLT above applies to partial sums of the form $\lambda(B_n \cap R_{i,j})s_{i,j}$, where $0 \leq \lambda(B_n \cap R_{i,j}) \leq 1$, where the estimator is defined by sums of $1_N(i,j)s_{i,j}$, where $1_N(t) = 1$ if $(i,j) \in B_n$ and zero otherwise. Assumptions 2 and 3 are used to show the difference between these two types of sums is asymptotically negligible. Proofs of Lemma 1 as well as the theorems stated below are collected in the Appendix.

The following theorem is analogous to Kiefer and Vogelsang (2005) and establishes convergence of $\widehat{\Omega}$ to a functional of the set indexed Brownian motion $W_p$.

THEOREM 1. *Suppose the four argument kernel* $K$ *with domain* $U \times U$, *with* $U \subseteq \mathbb{R}^2$, *has bandwidths* $(h_1, h_2) = (b_1 n, b_2 n)$ *where* $(b_1, b_2) \in (0,1]^2$, *and that the derivative* $\frac{\partial^4 K}{\partial x_1 \partial x_2 \partial x_3 \partial x_4}$ *is everywhere continuous. Then, under Assumptions 1–6,*

$$\widehat{\Omega} \Rightarrow \frac{1}{\lambda(B)}\Lambda Q_p(b_1, b_2, B)\Lambda' \tag{3.6}$$

$$Q_p(\cdot) = L_{p,I,I}(\cdot) + L_{p,I,II}(\cdot) + L_{p,II,II}(\cdot) + \cdots + L_{p,IV,IV}(\cdot),$$

*where* $L_{p,\cdot,\cdot}$ *are defined in (3.3)–(3.5).*

As for KV in the time series case, the limit of $\widehat{\Omega}$ is proportional to $\Lambda\Lambda' \equiv \Omega_0$, but depends on the random matrix $Q_p(b_1, b_2, B)$ which reflects the choice of kernel and bandwidths as well as the shape of the region $B$. In contrast to a consistency result, where $Q_p(b_1, b_2, B) = \lambda(B)I$, (3.6) approximates the finite sample randomness in the HAC estimator. Note that this theorem does not cover the Bartlett kernel which is widely used in practice in time series analysis and spatial contexts (see Conley, 1999). In a previous version of this paper, we showed that a result similar to Theorem 1 holds for the Bartlett kernel on a rectangular region with a slight, notationally cumbersome change in the algebraic device used to prove the theorem. Extending the result for the Bartlett kernel to a QMB region is straightforward and details are omitted.

Consider a test of the null hypothesis $H_0 : r(\theta_0) = 0$ against the alternative $H_a : r(\theta_0) \neq 0$ where $r$ is a $q$-vector of $C^1$ functions, $q \leq p$, and $R(\theta) = \frac{\partial r(\theta)}{\partial \theta'}$ is a $q \times p$ matrix of first derivatives. By Lemma 1 and the continuous mapping theorem, we have that

$$\sqrt{N}r(\widehat{\theta}_N) \Rightarrow -R(\theta_0)(\lambda(B)^{1/2}\mathcal{J})^{-1}\Lambda W_p(B). \tag{3.7}$$

Note that the right hand side follows a $q$-variate Gaussian distribution.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

Let $\widehat{J}_{\widehat{\theta}_N} = \frac{1}{N}\sum_{i,j\in B_n} \frac{\partial s_{i,j}(\widehat{\theta}_N)}{\partial \theta'}$ and define the Wald statistic

$$F = Nr(\widehat{\theta}_N)' \left( R(\widehat{\theta}_N) \left(\widehat{J}_{\widehat{\theta}_N}\right)^{-1} \widehat{\Omega} \left(\widehat{J}_{\widehat{\theta}_N}\right)^{-1} R(\widehat{\theta}_N)' \right)^{-1} r(\widehat{\theta}_N)$$

and the $t$-statistic

$$t = \frac{\sqrt{N} r(\widehat{\theta}_N)}{\sqrt{R(\widehat{\theta}_N) \left(\widehat{J}_{\widehat{\theta}_N}\right)^{-1} \widehat{\Omega} \left(\widehat{J}_{\widehat{\theta}_N}\right)^{-1} R(\widehat{\theta}_N)'}}$$

when $q = 1$. The null distributions of $F$ and $t$ are given in the following theorem.

THEOREM 2. *Suppose the conditions of Theorem 1 are satisfied. Then*

$$F \Rightarrow W_q(B)' Q_q(b_1, b_2, B)^{-1} W_q(B) \qquad \text{and} \qquad t \Rightarrow \frac{W_1(B)}{\sqrt{Q_1(b_1, b_2, B)}}.$$

As in the time series fixed-$b$ approximations of KV, the nuisance parameter $\Lambda$ appears in both numerator and denominator of $t$- and Wald tests and cancels out so that the tests do not depend on it asymptotically. The resulting reference distributions are nonstandard but can be simulated given a choice of the kernel and bandwidth in applications. We conjecture that the simple i.i.d. (independent and identically distributed) bootstrap provides a valid method for simulating these critical values as in Gonçalves and Vogelsang (2011), though verifying this conjecture is complicated by the need for a spatial FCLT of the form of Dedecker (2001) over triangular arrays. We have been unable to find such a FCLT in the literature and establishing such a result is beyond the scope of the present paper.

## 4. SIMULATION EXAMPLES

In the previous sections, we have provided approximations to the behavior of test-statistics based on spatial HAC covariance estimators under a sequence which keeps the HAC bandwidths proportional to the sample size. In this section, we provide evidence on the finite-sample inference properties of tests using the proposed procedure through a simulation experiment in a model with an endogenous explanatory variable and also provide a comparison of t-tests based on spatial HAC using 'fixed-$b$' critical values (hereafter HAC) to t-tests based on the grouping procedure of Ibragimov and Müller (2010) (hereafter IM).[12]

We report results from simulating a simple linear model with an endogenous explanatory variable. We generate instruments and errors as geometrically decaying spatial moving averages of order 15:

$$z_{s,h} = \sum_{\|j\|\le 15} \gamma^{\|j\|} w_{s+j,h} \quad \text{for } h = 1, 2, 3,$$

$$\varepsilon_s = \sum_{\|j\|\le 15} \gamma^{\|j\|} u_{s+j,1}, \quad \text{and}$$

$$v_s = \sum_{\|j\|\le 15} \gamma^{\|j\|} u_{s+j,2},$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

where $\|j\| = max(j_1, j_2)$; $u_{s,1} = .8u_{s,2} + .6e_s$; and $e_s$, $u_{s,2}$, $w_{s,1}$, $w_{s,2}$, and $w_{s,3}$ are independent and follow N(0,1) distributions. We then generate $x_s = \Pi \sum_{h=1}^{3} z_{s,h} + \sigma(z_1, z_2, z_3)v_s$ and $y_s = \beta_0 + \beta_1 x_s + \sigma(z_1, z_2, z_3)\varepsilon_s$. We set $\sigma(z_1, z_2, z_3)^2 = 1$ in our homoskedastic designs and $\sigma(z_1, z_2, z_3)^2 = \exp\{.2\sum_{h=1}^{3} z_{s,h}\}/1.3$. We use three instruments and $\Pi = .5$. In all simulations, $\beta_0 = 0$ and $\beta_1 = 1$. We consider two different values of $\gamma$, $\gamma \in \{.2, .4\}$.[13]

We consider two different designs, one in which the data are observed on a $15 \times 15$ square and the other in which the observations are observed on a cross. The cross has a length and height of 25 lattice points with the horizontal and vertical lines being five lattice points wide (see Figure 3). This design of the two regions gives an equal number of observations ($N = 225$) within each region. When $\gamma \neq 0$, we draw $u_s$ and $v_s$ on a $55 \times 55$ lattice to generate a $25 \times 25$ lattice of $x_s$ and $\varepsilon_s$. We then take the appropriate 225 observations from this $25 \times 25$ lattice for the square and cross regions.

We report results for tests based on HAC estimators and the IM procedure. For the HAC estimators, we use a Gaussian product kernel and smoothing parameter equal to $h$, i.e.,

$$K(s, \tilde{s}) = \exp\{-.5[(s_1 - \tilde{s}_1)/(h/2)]^2\}\exp\{-.5[(s_2 - \tilde{s}_2)/(h/2)]^2\}.$$

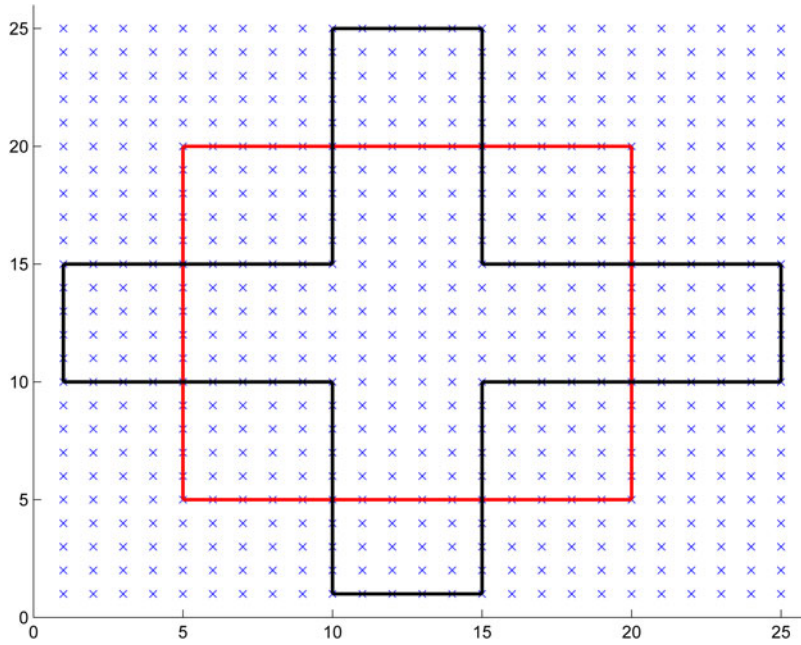



**FIGURE 3.** This figure illustrates the two sampling regions, the square and the cross, used in the simulation study.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

We consider three different values of the smoothing parameter, $h \in \{0, 8, 16\}$. Recall that the IM $t$-statistic is formed by breaking the data into $G$ groups and then estimating the model parameters within each group. In our IV model, a different point estimate of the parameter of interest, $\beta_1$ is constructed within each group, call it $\hat{\beta}_{1,g}$. Letting $\bar{\beta}_{1,G}$ denote the cross group average of the within-group point estimates, $\bar{\beta}_{1,G} = \frac{1}{G}\sum_g \hat{\beta}_{1,g}$, the IM test statistic for the null hypothesis that $\beta_1 = \beta_1^*$ is

$$t_{IM} = \frac{\bar{\beta}_{1,G} - \beta_1^*}{\frac{1}{\sqrt{G}}\sqrt{\frac{1}{G-1}\sum_g \left(\hat{\beta}_{1,g} - \bar{\beta}_{1,G}\right)^2}}.$$

Under mild regularity conditions, $t_{IM}$ converges to a limiting t-distribution with $G - 1$ degrees of freedom. In our simulations, we consider $G \in \{2, 4\}$ where groups are formed by splitting the data into the two groups formed by splitting at the vertical axis of symmetry or into the four groups formed by splitting at the vertical and horizontal axes of symmetry.

There are a few key differences between HAC and IM. The first is that any grouping scheme, such as IM or Bester, Conley, and Hansen (2010), ignores correlations between nearby observations that happen to belong to different groups. This ignoring of boundary terms is easily seen by looking at the form of the denominator in $t_{IM}$ which is just the sample variance obtained by treating the within-group estimators as independent. Under general types of weak dependence, treating the within-group estimators as independent is asymptotically justified, but one suspects that ignoring these boundary effects may lead to some size distortions in finite samples. The second difference between IM and HAC is that the numerator of the HAC test statistic is the usual full sample estimator of the parameter of interest while the numerator of the IM statistic is the across group average of within group estimators. Since the IM approach is using an estimator that is the average over several estimators obtained in smaller samples, it may be more susceptible to finite-sample bias problems than HAC since these biases will generally not average out. Thus, the bias in the IM numerator will generally correspond to the bias of the estimators in the smaller samples.

A key advantage of the IM procedure is that it can formally be shown to produce t-statistics that control the size of tests under weaker conditions than those used in obtaining the fixed-$b$ approximation. Importantly, the conditions of IM allow for general forms of nonexplosive nonstationarity. This type of nonstationarity is not allowed for in the conditions used to establish the validity of the type of functional central limit theorem that we use in deriving the fixed-$b$ limit.[14]

We report the simulation results in Table 1. The main results are for size based on the fixed-b approximation for HAC-based tests or the $t_{G-1}$ approximation for IM in rows labeled "Size—Fixed-b", for size based on the conventional Gaussian approximation for HAC-based tests in rows labeled "Size—Gaussian", and size-adjusted power of 5% level tests against the alternatives that $\beta$ is 0, 0.5, 1.5,

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

**TABLE 1.** Simulation results

| | $\gamma = 0.2$ | | | | | $\gamma = 0.4$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | HAC | | | IM | | HAC | | | IM | |
| | 0 | 8 | 16 | 4 | 2 | 0 | 8 | 16 | 4 | 2 |
| % Weight | 0.000 | 0.278 | 0.617 | 0.252 | 0.502 | 0.000 | 0.278 | 0.617 | 0.252 | 0.502 |
| A. Homoskedastic, square | | | | | | | | | | |
| Size—Fixed-b | 0.255 | 0.103 | 0.081 | 0.122 | 0.067 | 0.566 | 0.200 | 0.166 | 0.167 | 0.075 |
| Size—Gaussian | 0.255 | 0.201 | 0.361 | | | 0.566 | 0.309 | 0.456 | | |
| Power(0.0) | 0.998 | 0.988 | 0.976 | 0.935 | 0.433 | 0.810 | 0.792 | 0.754 | 0.735 | 0.319 |
| Power(0.5) | 0.887 | 0.813 | 0.761 | 0.647 | 0.246 | 0.482 | 0.463 | 0.432 | 0.387 | 0.182 |
| Power(1.5) | 0.997 | 0.909 | 0.783 | 0.473 | 0.220 | 0.070 | 0.164 | 0.166 | 0.072 | 0.114 |
| Power(2.0) | 1.000 | 1.000 | 0.998 | 0.961 | 0.399 | 0.913 | 0.828 | 0.745 | 0.549 | 0.237 |
| Bias(b) | 0.016 | 0.016 | 0.016 | 0.047 | 0.024 | 0.047 | 0.047 | 0.047 | 0.132 | 0.079 |
| Std(s.e.) | 0.035 | 0.049 | 0.060 | 0.102 | 0.108 | 0.154 | 0.158 | 0.175 | 0.197 | 0.256 |
| B. Homoskedastic, cross | | | | | | | | | | |
| Size—Fixed-b | 0.238 | 0.078 | 0.070 | 0.125 | 0.065 | 0.539 | 0.167 | 0.150 | 0.211 | 0.077 |
| Size—Gaussian | 0.238 | 0.156 | 0.269 | | | 0.539 | 0.257 | 0.363 | | |
| Power(0.0) | 0.999 | 0.991 | 0.981 | 0.929 | 0.471 | 0.859 | 0.820 | 0.764 | 0.701 | 0.336 |
| Power(0.5) | 0.904 | 0.854 | 0.796 | 0.686 | 0.269 | 0.524 | 0.510 | 0.440 | 0.379 | 0.207 |
| Power(1.5) | 1.000 | 0.960 | 0.871 | 0.519 | 0.228 | 0.186 | 0.225 | 0.184 | 0.077 | 0.114 |
| Power(2.0) | 1.000 | 1.000 | 1.000 | 0.968 | 0.435 | 0.959 | 0.888 | 0.789 | 0.520 | 0.279 |
| Bias(b) | 0.011 | 0.011 | 0.011 | 0.044 | 0.023 | 0.038 | 0.038 | 0.038 | 0.124 | 0.071 |
| Std(s.e.) | 0.032 | 0.047 | 0.057 | 0.069 | 0.108 | 0.111 | 0.133 | 0.148 | 0.181 | 0.235 |
| C. Heteroskedastic, square | | | | | | | | | | |
| Size—Fixed-b | 0.268 | 0.100 | 0.086 | 0.085 | 0.067 | 0.573 | 0.175 | 0.149 | 0.130 | 0.070 |
| Size—Gaussian | 0.268 | 0.208 | 0.372 | | | 0.573 | 0.307 | 0.452 | | |
| Power(0.0) | 1.000 | 0.997 | 0.991 | 0.975 | 0.463 | 0.788 | 0.780 | 0.748 | 0.743 | 0.284 |
| Power(0.5) | 0.932 | 0.894 | 0.846 | 0.785 | 0.244 | 0.462 | 0.491 | 0.462 | 0.392 | 0.166 |
| Power(1.5) | 0.999 | 0.982 | 0.902 | 0.764 | 0.224 | 0.123 | 0.237 | 0.250 | 0.125 | 0.104 |
| Power(2.0) | 1.000 | 1.000 | 1.000 | 0.994 | 0.444 | 0.895 | 0.845 | 0.769 | 0.647 | 0.239 |
| Bias(b) | 0.007 | 0.007 | 0.007 | 0.024 | 0.010 | 0.013 | 0.013 | 0.013 | 0.085 | 0.029 |
| Std(s.e.) | 0.030 | 0.045 | 0.054 | 0.068 | 0.092 | 0.203 | 0.211 | 0.226 | 0.204 | 0.264 |
| D. Heteroskedastic, cross | | | | | | | | | | |
| Size—Fixed-b | 0.239 | 0.078 | 0.066 | 0.098 | 0.060 | 0.526 | 0.149 | 0.141 | 0.170 | 0.074 |
| Size—Gaussian | 0.239 | 0.161 | 0.281 | | | 0.526 | 0.248 | 0.361 | | |
| Power(0.0) | 0.999 | 0.995 | 0.992 | 0.964 | 0.505 | 0.827 | 0.792 | 0.746 | 0.693 | 0.288 |
| Power(0.5) | 0.950 | 0.904 | 0.877 | 0.784 | 0.290 | 0.538 | 0.487 | 0.448 | 0.381 | 0.171 |
| Power(1.5) | 1.000 | 0.982 | 0.940 | 0.768 | 0.258 | 0.292 | 0.279 | 0.250 | 0.130 | 0.116 |
| Power(2.0) | 1.000 | 1.000 | 1.000 | 0.994 | 0.495 | 0.950 | 0.866 | 0.793 | 0.602 | 0.245 |
| Bias(b) | 0.004 | 0.004 | 0.004 | 0.022 | 0.009 | 0.008 | 0.008 | 0.008 | 0.083 | 0.035 |
| Std(s.e.) | 0.029 | 0.045 | 0.054 | 0.060 | 0.094 | 0.183 | 0.218 | 0.231 | 0.177 | 0.277 |

*Note*: This table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment with data generated on a two-dimensional lattice with endogenous explanatory variable. "Size—Fixed b" is the frequency of rejection of the hypothesis that the parameter is equal to the true value ($\beta = 1$) using the fixed-b approximation or the IM approximation. "Size - Gaussian" is the frequency of rejection of the hypothesis that the parameter is equal to the true value using the conventional Gaussian approximation. We also report size-adjusted power against the alternatives that $\beta = 0$, $\beta = 0.5$, $\beta = 1.5$, and $\beta = 2$ along with the bias and standard error of the point estimates. All results are based on 1,000 simulation replications. Column labels indicate the values of smoothing parameters. Columns under HAC use a Gaussian product kernel and with equal smoothing parameter in both directions equal to the number given in the column. Columns under IM use the procedure of Ibragimov and Muller (2007) with number of groups given in the column. All data were generated by a spatial moving average model as described in the text with $\gamma = .2$ or $\gamma = .4$ and three instruments.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

or 2.0 based on HAC or IM. Columns labeled "HAC" give results for HAC with the smoothing parameter given by the number in the column label. Similarly, IM results are reported in the columns labeled "IM" where the number of groups is given by the number in the column label. We also provide the bias of the 2SLS and of the IM point estimators of the slope coefficient in rows labeled "Bias(b)." Finally, we report two additional statistics for each covariance estimator considered: the empirical standard deviation of the estimated variance of the slope coefficient (Std(s.e.))[15] and the proportion of the total number of potential cross-products that are used in forming the covariance matrix estimator (% Weight).[16]

First, comparing the fixed-*b* approximation to the usual Gaussian approximation, we see that the fixed-*b* approximation produces tests that are less size distorted than the Gaussian approximation across all of the designs considered. This reduction in size-distortion is most pronounced when a large bandwidth is used. We also see that the size for HAC is closest to the nominal level in all cases for the largest value of the bandwidth parameter. It is worth noting that the fixed-*b* approximation improves upon the usual approximation under the model of heteroskedasticity considered.

We should also note that in previous versions of the paper, we used the i.i.d. bootstrap to obtain critical values for the spatial HAC tests. Those results exhibited very similar patterns to the results we report here using the simulated asymptotic fixed-*b* critical values. This further supports our conjecture that it should be possible to show that the i.i.d. bootstrap can be used to obtain valid fixed-*b* critical values once appropriate spatial FCLTs are developed in the literature.

Next, we compare the performance of HAC and IM. As mentioned above, a key difference between the procedures is that they use different numerators in forming test statistics. The simulation results demonstrate that the IM numerator has substantially more bias than the HAC numerator and that the bias in the IM numerator increases rapidly as the number of groups (observations per group) increases (decreases). In terms of testing properties, IM with two groups does the best job of controlling size across all settings. It is interesting that the size distortions of the HAC procedure are much more constant across the different smoothing parameter values than those of the IM procedure. This rapid increase in the size distortion of IM as one increases the number of groups, which was also found by Ibragimov and Müller (2010) in a simple location model with spatial data, is likely generated by the coarseness of the grouping structure which leads to a rapid accumulation of ignored boundary terms as the number of groups gets large coupled with the rapid increase in finite sample bias as one decreases the number of observations per group. Looking at size-adjusted power, we see that HAC with the bandwidths considered clearly dominates IM with the groupings considered. The least powerful HAC procedure with a smoothing parameter of 16 has size-adjusted power greater than IM using four groups while doing a generally better job at controlling size.

Overall, these simulation results are favorable to HAC using 'fixed-*b*' critical values relative to HAC using critical values from the Gaussian approximation.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

We also see the trade-offs between HAC and IM. IM tends to be more biased and less-powerful than the HAC-based procedure. However, the most conservative IM procedure possible, with only two groups, controls size better than the most conservative HAC estimator we considered. It is clear that both approaches have favorable features and a more systematic comparison may be interesting. As with most procedures, we do not believe that one will uniformly dominate the other, and thus the two approaches are complementary.

## 5. CONCLUSION

In this paper, we have provided asymptotic approximation results for inference procedures based on spatial HAC covariance matrix estimators under "fixed-$b$" asymptotics, where the HAC smoothing parameters are proportional to the sample size. Similar to the results in Kiefer and Vogelsang (2002, 2005) for the time series case, the spatial HAC estimator is not consistent but instead converges in distribution to a nondegenerate random variable whose distribution depends on the kernel and smoothing parameters. We also find that, in a spatial setting, these distributions depend explicitly on the shape of the sampling region. We provide a characterization of 'boundary effects' in terms of integrals against a set-indexed Gaussian process. Under the fixed-$b$ sequence, asymptotic distributions of test statistics based on the spatial HAC estimator account for finite sample variability in the HAC estimator and explicitly reflect the choice of kernel and smoothing parameters. The resulting approximate distributions for test statistics are nonstandard, so critical values must be obtained via simulation.

We illustrate the use of spatial HAC estimators and our approximation through simulations. In the simulations, we find that inference based on the usual asymptotic approximation tends to be badly size distorted. We also find that inference based on spatial HAC estimators and the approximate distributions we develop in this paper improves size control relative to this baseline across all simulations we consider.

There are important questions that deserve more attention. One issue that we do not address is smoothing parameter selection. Our simulation results do suggest that one needs to use quite large values for smoothing parameters to control the size of tests based on HAC estimators. This is consistent with Sun, Phillips, and Jin (2008), who consider smoothing parameter selection in the context of Gaussian location model in a time series and show that the rate of increase for the optimal smoothing parameter chosen by trading off size and power using higher-order asymptotics is much faster than the rate for minimizing MSE of the variance estimator. An interesting direction for future research would be to adapt the arguments of Sun, Phillips, and Jin (2008) to the present context. The theoretical results we obtain are also provided under very strong homogeneity assumptions. It may be useful to extend these results to allow for heteroskedasticity as well as heterogeneity in the sampling process and region.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

## NOTES

1. The most popular implementation is in Newey and West (1987). See also Andrews (1991) for a thorough treatment of time series HAC estimation.

2. See Wooldridge (2003) for a concise review of this literature. See also Liang and Zeger (1986), Arellano (1987), Bertrand, Duflo, and Mullainathan (2004), Hansen (2007), and Bester, Conley, and Hansen (2010).

3. See Conley (1996, 1999) for early examples and Kelegian and Prucha (2007) and Kim and Sun (2011) for more recent examples.

4. Jansson (2004) and Sun, Phillips, and Jin (2008) also provide theoretical evidence that the fixed-$b$ approximation outperforms the standard approximation in Gaussian location models.

5. Note that KV assume that $K(r,s) = k(r-s)$ so that $-k''(r-s)$ appears in place of the partial derivative of $K$ in their expression.

6. For the remainder of this section and in the Appendix, $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ are points in $\mathbb{R}^2$ while $u_1, u_2$ are scalars denoting the *vertical* and *horizontal* coordinates of $\mathbf{u}$, respectively. We also use $a \wedge b = \min\{a,b\}$ and $a \vee b = \max\{a,b\}$.

7. See Lemma 2 in the Appendix.

8. That is, $e_1 = (\overline{u}_1, 0)$, $e_2 = (0, \underline{u}_2)$, $e_3 = (\underline{u}_1, 0)$, and $e_4 = (0, \overline{u}_2)$.

9. We assume that one of the 'endpoints' $e_j$ defined in footnote 8 is at distance exactly $1/2$ from the origin for convenience. Any $\widetilde{B} = cB$ for $0 < c < 1$ would give the same limit distribution.

10. For a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and two $\sigma$-algebras $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{A}$, define $\phi(\mathcal{U},\mathcal{V}) = \sup\{\|\mathbb{P}(V|\mathcal{U}) - \mathbb{P}(V)\|_\infty, V \in \mathcal{V}\}$. For a real random field $(\mathbb{R}^{\mathbb{Z}^d}, \mathcal{B}^{\mathbb{Z}^d}, \mathbb{P})$ with $|\Gamma|$ denoting the cardinality of any subset $\Gamma$ of $\mathbb{Z}^d$, we can define the $\phi$-mixing coefficient for any $(k,m) \in \mathbb{N}^2$ as $\phi_k(m) = \sup\{\phi(\mathcal{F}_{\Gamma_1}, \mathcal{F}_{\Gamma_2}), |\Gamma_2| \leq k, d(\Gamma_1,\Gamma_2) \geq m\}$ for $d(\Gamma_1,\Gamma_2) = \min\{|j-i|, i \in \Gamma_1, j \in \Gamma_2\}$. Note that these mixing coefficients differ from the uniform mixing coefficients, $\phi_\infty(m)$ typically considered in the time series case, where the $\sigma$-algebra $\mathcal{V}$ may be generated by an infinite number of random variables. When $S$ is a matrix, $|S| = \text{trace}(SS')^{\frac{1}{2}}$.

11. See comments and references in Dedecker (2001). A previous version of this paper referenced a FCLT in Goldie and Greenwood (1986), which requires uniform mixing. We thank an anonymous referee for bringing this issue to our attention.

12. We chose not to report results based on OLS or using clustering as discussed in Bester, Conley, and Hansen (2010) to keep the amount of reported results manageable. OLS results for the Ibragimov and Müller (2010) approach in a simple location model with spatial data can be found in their Table 3.

13. Note that in this setting, a given value of $\gamma$ induces a higher level of dependence in the data versus the time series setting where each point would have exactly two points in the sample located a given integer distance away. In this design, each point has eight neighbors located at distance 1, 16 at distance 2, and so on.

14. It is worth noting that the robustness IM demonstrate for their t-statistics has not been shown to carry over to joint tests of multiple hypotheses.

15. To make this object more meaningful, we report the empirical standard deviation scaled by the amount one would need to multiply the standard error estimator to produce an unbiased estimate of the empirical standard deviation of the slope coefficient estimators.

16. Specifically, % Weight reports the weighted sum of the cross-products used where weights are determined by the weight assigned to each cross-product by the kernel; i.e., % Weight = $\frac{1}{225^2}\sum_{s_1=1}^{15}\sum_{s_2=1}^{15}\sum_{\widetilde{s}_1=1}^{15}\sum_{\widetilde{s}_2=1}^{15} K(s,\widetilde{s})$. For HAC, $K(s,\widetilde{s})$ is defined above. For IM, $K(s,\widetilde{s}) = 1$ whenever $s$ and $\widetilde{s}$ belong to the same group; and $K(s,\widetilde{s}) = 0$ otherwise.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

## REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59(3), 817–858.

Arellano, M. (1987) Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49(4), 431–434.

Basu, A.K. & C.C.Y. Dorea (1979) On functional central limit theorem for stationary martingale random fields. *Acta Mathematica Academia Scientiarum Hungaricae* 33, 307–316.

Bertrand, M., E. Duflo, & S. Mullainathan (2004) How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119, 249–275.

Bester, C.A., T.G. Conley, & C.B. Hansen (2010) Inference for dependent data using cluster covariance estimators. Available at SSRN: http://ssrn.com/abstract=1708263.

Conley, T.G. (1996) Econometric modelling of cross-sectional dependence. Ph.D. dissertation, University of Chicago.

Conley, T.G. (1999) GMM estimation with cross sectional dependence. *Journal of Econometrics* 92, 1–45.

Dedecker, J. (2001) Exponential inequalities and functional central limit theorems for random fields. *ESAIM: Probability and Statistics* 5, 77–104.

Deo, C. (1975) A functional central limit theorem for stationary random fields. *Annals of Probability* 3(4), 708–715.

Dudley, R.M. (1973) Sample functions of the Gaussian process. *Annals of Probability* 1, 66–103.

Goldie, C.M. & P.E. Greenwood (1986) Variance of set-indexed sums of mixing random variables and weak convergence of set-indexed processes. *Annals of Probability* 14(3), 817–839.

Gonçalves, S. & T.J. Vogelsang (2011) Block bootstrap HAC robust tests: The sophistication of the naive bootstrap. *Econometric Theory* 27(4), 745–791.

Hansen, C.B. (2007) Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics* 141, 597–620.

Ibragimov, R. & U.K. Müller (2010) t-Statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics* 28, 453–468.

Jansson, M. (2004) The error in rejection probability of simple autocorrelation robust tests. *Econometrica* 72(3), 937–946.

Kelegian, H. & I. Prucha (2007) HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131–154.

Kelejian, H.H. & I. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.

Kelejian, H.H. & I. Prucha (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219–257.

Kiefer, N.M. & T.J. Vogelsang (2002) Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* 18, 1350–1366.

Kiefer, N.M. & T.J. Vogelsang (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

Kim, M.S. & Y. Sun (2011) Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics* 160, 349–371.

Lee, L.-F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* 72, 1899–1926.

Lee, L.-F. (2007a) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489–514.

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

Liang, K.-Y. & S. Zeger (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73(1), 13–22.

Newey, W.K. & K.D. West (1987) A simple, positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

Sun, Y., P.C.B. Phillips, & S. Jin (2008) Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76(1), 175–194.

Wooldridge, J.M. (2003) Cluster-sample methods in applied econometrics. *American Economic Review* 93(2), 133–188.

# APPENDIX A: Proofs

## A.1. Lemmas

**Proof of Lemma 1.** We have that $\widehat{\theta}_N$ solves the estimating equation $\sum_{i,j\in B_n} s_{i,j}\left(\widehat{\theta}_N\right)=0$. From a mean-value expansion, we then have

$$\sum_{i,j\in B_n} s_{i,j}(\theta_0)+\Delta_{B_n}\left(\widehat{\theta}_N,\theta_0,\delta_{B_n}\right)\left(\widehat{\theta}_N-\theta_0\right)=0,$$

where $\Delta_{B_n}\left(\widehat{\theta}_N,\theta_0,\delta_{B_n}\right)$ is a $p\times p$ matrix with $l^{th}$ row given by the $l^{th}$ row of the matrix $\left(\sum_{i,j\in B_n}\frac{\partial s_{i,j}(\theta)}{\partial\theta'}\right)|_{\theta=\bar{\theta}^{(l)}_{B_n}}$ with $\bar{\theta}^{(l)}_{B_n}=\delta^{(l)}_{B_n}\theta_0+\left(1-\delta^{(l)}_{B_n}\right)\widehat{\theta}_N$ where $\delta^{(l)}_{B_n}$ is the $l^{th}$ element of the vector $\delta_{B_n}$ and $0\leq\delta^{(l)}_{B_n}\leq 1$. It follows that

$$n\left(\widehat{\theta}_N-\theta_0\right)=-\left(\frac{1}{n^2}\Delta_{B_n}\left(\widehat{\theta}_N,\theta_0,\delta_{B_n}\right)\right)^{-1}\frac{1}{n}\sum_{i,j\in B_n} s_{i,j}(\theta_0).$$

Now note that

$$\begin{aligned}\frac{1}{n}\sum_{i,j\in B_n} s_{i,j}(\theta_0)&=\frac{1}{n}\sum_{i=-n/2}^{n/2}\sum_{j=-n/2}^{n/2}1_{i,j}s_{i,j}(\theta_0)\\&=\frac{1}{n}\sum_{i=-n/2}^{n/2}\sum_{j=-n/2}^{n/2}\lambda^n_{i,j}s_{i,j}(\theta_0)+\frac{1}{n}\sum_{i=-n/2}^{n/2}\sum_{j=-n/2}^{n/2}(1_{i,j}-\lambda^n_{i,j})s_{i,j}(\theta_0),\end{aligned}$$

where $1_{i,j}$ is an indicator which equals one when an observation occurs at lattice point $(i,j)$, $\lambda^n_{i,j}=\lambda(B_n\cap]i-1,i]\times]j-1,j])$, and $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}^2$ as in Dedecker (2001).

For the first term, we have $\frac{1}{n}\sum_{i=-n/2}^{n/2}\sum_{j=-n/2}^{n/2}\lambda^n_{i,j}s_{i,j}(\theta_0)\Rightarrow\Lambda W_p(B)$ by Assumption 5.

For the second term, we define $dB^+=\{(i,j):[(i,j)\in B_n$ and $((i-1,j)\notin B_n$ or $(i,j-1)\notin B_n$ or $(i-1,j-1)\notin B_n)]$ or $[(i,j)\notin B_n$ and $((i-1,j)\in B_n$ or $(i,j-1)\in B_n$ or $(i-1,j-1)\in B_n)]$. By Assumption 2 (specifically, that the boundary of each subregion is monotone), the number of points in $dB^+$ is less than or equal to $Cn$. Write

$$\frac{1}{n}\sum_{i=-n/2}^{n/2}\sum_{j=-n/2}^{n/2}(1_{i,j}-\lambda^n_{i,j})s_{i,j}(\theta_0)=\frac{1}{n}\sum_{i,j\in dB^+}(1_{i,j}-\lambda^n_{i,j})s_{i,j}(\theta_0).$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

We then have

$$\begin{aligned}
&\mathrm{E}\Big\|\frac{1}{n}\sum_{i,j\in dB^+}(1_{i,j}-\lambda^n_{i,j})s_{i,j}(\theta_0)\Big\|^2\\
&\quad\le \frac{1}{n^2}\sum_{i,j\in dB^+}\sum_{k,l\in dB^+}|1_{i,j}-\lambda^n_{i,j}||1_{k,l}-\lambda^n_{k,l}|\|\mathrm{E}s_{i,j}(\theta_0)s_{k,l}(\theta_0)'\|\\
&\quad\le \frac{1}{n^2}Cn\to 0.
\end{aligned}$$

The first inequality follows from the Cauchy–Schwarz and Triangle inequalities. The second inequality follows from the mixing and moment conditions in Assumption 3, noting that the largest number of $d^{th}$-order neighbors of a set of $k$ points in $\mathbb{R}^m$ is $C(m)kd$ with $C(m)$ a finite constant that depends on the size of the space, and the condition on the size of the boundary set $dB^+$.

Thus, we have

$$\frac{1}{n}\sum_{i,j\in B_n}s_{i,j}(\theta_0)\Rightarrow \Lambda W_p(B).$$

For the remaining term, we have that

$$\frac{1}{n^2}\Delta_{B_n}(\widehat{\theta}_N,\theta_0,\delta_{B_n})=\left(\frac{N}{n^2}\right)\frac{1}{N}\Delta_{B_n}(\widehat{\theta}_N,\theta_0,\delta_{B_n})\xrightarrow{p}\lambda(B)\mathcal{J}(\theta_0)$$

from Assumptions 2 and 6.

Combining these results gives

$$n\big(\widehat{\theta}_N-\theta_0\big)\Rightarrow-(\lambda(B)\mathcal{J}(\theta_0))^{-1}\Lambda W_p(B),$$

and it immediately follows using $\sqrt{\frac{N}{n^2}}\to\lambda(B)^{1/2}$ that

$$\sqrt{N}\big(\widehat{\theta}_N-\theta_0\big)=\sqrt{\frac{N}{n^2}}n\big(\widehat{\theta}_N-\theta_0\big)\Rightarrow\Big(-\lambda(B)^{1/2}\mathcal{J}(\theta_0)\Big)^{-1}\Lambda W_p(B).$$ ■

It is also convenient to establish the limiting behavior of partial sums of the realized scores of the form $\frac{1}{n}\sum_{i,j\in A_n}\widehat{s}_{i,j}$, where $A_n\equiv nA$.

LEMMA 2. *Suppose Assumptions 1–6 hold. Then, as $N\to\infty$,*

$$\frac{1}{n}\sum_{i,j\in A_n}\widehat{s}_{i,j}\Rightarrow\Lambda\left(W_p(A)-\frac{\lambda(A)}{\lambda(B)}W_p(B)\right)\equiv W_p(A,B)$$

*for $A\subseteq B$.*

**Proof of Lemma 2.** Note that

$$\frac{1}{n}\sum_{i,j\in A_n}\widehat{s}_{i,j}=\frac{1}{n}\sum_{i,j\in A_n}s_{i,j}(\theta_0)+\frac{1}{n^2}\Delta_{A_n}(\widehat{\theta}_N,\theta_0,\delta_{A_n})n\big(\widehat{\theta}_N-\theta_0\big),$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

where $\Delta_{A_n}(\widehat{\theta}_N, \theta_0, \delta_{A_n})$ is defined similarly to $\Delta_{B_n}(\widehat{\theta}_N, \theta_0, \delta_{B_n})$. By an argument similar to that used in Lemma 1, we can show that

$$\frac{1}{n}\sum_{i,j\in A_n} s_{i,j}(\theta_0) \Rightarrow \Lambda W_p(A).$$

Similar to the argument for $\Delta_{B_n}(\widehat{\theta}_N, \theta_0, \delta_{B_n})$, we have

$$\frac{1}{n^2}\Delta_{A_n}(\widehat{\theta}_N, \theta_0, \delta_{A_n}) \xrightarrow{p} \lambda(A)\mathcal{J}(\theta_0).$$

Plugging these two results in and using Lemma 1 then gives

$$\begin{aligned}\frac{1}{n}\sum_{i,j\in A_n}\widehat{s}_{i,j} &= \frac{1}{n}\sum_{i,j\in A_n} s_{i,j}(\theta_0) + \frac{1}{n^2}\Delta_{A_n}(\widehat{\theta}_N, \theta_0, \delta_{A_n})\sqrt{\frac{n^2}{N}}\sqrt{N}(\widehat{\theta}_N - \theta_0)\\ &\Rightarrow \Lambda W_p(A) - \lambda(A)\mathcal{J}(\theta_0)\lambda(B)^{-1/2}\left(\lambda(B)^{1/2}\mathcal{J}(\theta_0)\right)^{-1}\Lambda W_p(B)\\ &\equiv \Lambda\left(W_p(A) - \frac{\lambda(A)}{\lambda(B)}W_p(B)\right).\end{aligned}$$

■

The next lemma provides a representation of sums over QMB regions in terms of partial sums over subregions. To make things concrete, we will define the subregions and boundaries of the QMB region as in Figure 4.

In stating the result, we will make use of the following regions:

$$\begin{aligned}
A_1 &= \{i,j : 0 \le i \le \bar{M}_1(j+1)-1,\ 0 \le j \le \bar{N}_1 - 1\},\\
A_2 &= \{i,j : i = \bar{M}_1(j+1),\ 0 \le j \le \bar{N}_1 - 1\},\\
A_3 &= \{i,j : 0 \le i \le \bar{M}_1 - 1, j = \bar{N}_1(i)\},\\
A_4 &= \{i,j : i = \bar{M}_1, j = \bar{N}_1(\bar{M}_1)\},\\
A_5 &= \{i,j : 0 \le i \le \bar{M}_2 - 1,\ 1 \le j \le \bar{N}_2(i+1)-1\},\\
A_6 &= \{i,j : i = 0 \le i \le \bar{M}_2 - 1,\ j = \bar{N}_2(i+1)\},\\
A_7 &= \{i,j : i = \bar{M}_2(j), 1 \le j \le \bar{N}_2 - 1\},\\
A_8 &= \{i,j : i = \bar{M}_2(\bar{N}_2), j = \bar{N}_2\},\\
A_9 &= \{i,j : 1 \le i \le \bar{M}_3 - 1,\ 0 \le j \le \bar{N}_3(i+1)-1\},\\
A_{10} &= \{i,j : i = 1 \le i \le \bar{M}_3 - 1,\ j = \bar{N}_3(i+1)\},\\
A_{11} &= \{i,j : i = \bar{M}_3(j), 0 \le j \le \bar{N}_3 - 1\},\\
A_{12} &= \{i,j : i = \bar{M}_3(\bar{N}_3), j = \bar{N}_3\},\\
A_{13} &= \{i,j : 1 \le i \le \bar{M}_4(j+1)-1,\ 1 \le j \le \bar{N}_4 - 1\},\\
A_{14} &= \{i,j : i = \bar{M}_4(j+1),\ 1 \le j \le \bar{N}_4 - 1\},\\
A_{15} &= \{i,j : 1 \le i \le \bar{M}_4 - 1, j = \bar{N}_4(i)\},\\
A_{16} &= \{i,j : i = \bar{M}_4, j = \bar{N}_4(\bar{M}_4)\}.
\end{aligned}$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

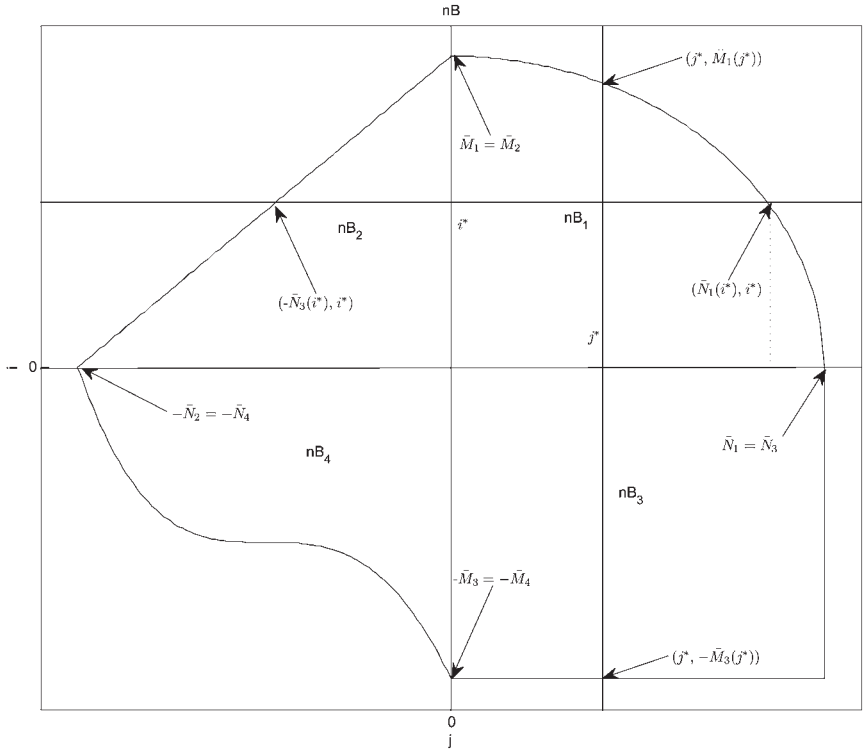



**FIGURE 4.** This figure illustrates the regions, boundaries, and indexing used in establishing the formal results and can be used to aid visualization of the partial sums $b_1$–$b_{16}$ and regions $A_1$–$A_{16}$.

For a function of four arguments, $W(i,j,k,l)$, we will also make use of the differences

$$
\begin{aligned}
D^{i,j,1}W(i,j,k,l) &= (W(i,j,k,l) - W(i,j+1,k,l)) \\
&\quad - (W(i+1,j,k,l) - W(i+1,j+1,k,l)) \\
D^{i,j,2}W(i,j,k,l) &= W(\bar{M}_1(j+1),j,k,l) - W(\bar{M}_1(j+1),j+1,k,l) \\
D^{i,j,3}W(i,j,k,l) &= W(i,\bar{N}_1(i),k,l) - W(i+1,\bar{N}_1(i+1),k,l) \\
D^{i,j,4}W(i,j,k,l) &= W(\bar{M}_1,\bar{N}_1(\bar{M}_1),k,l) \\
D^{i,j,5}W(i,j,k,l) &= (W(i,-j,k,l) - W(i+1,-j,k,l)) \\
&\quad - (W(i,-(j+1),k,l) - W(i+1,-(j+1),k,l)) \\
D^{i,j,6}W(i,j,k,l) &= W(i,-\bar{N}_2(i+1),k,l) - W(i+1,-\bar{N}_2(i+1),k,l) \\
D^{i,j,7}W(i,j,k,l) &= W(\bar{M}_2(j),-j,k,l) - W(\bar{M}_2(j+1),-(j+1),k,l) \\
D^{i,j,8}W(i,j,k,l) &= W(\bar{M}_2(\bar{N}_2),-\bar{N}_2,k,l) \\
D^{i,j,9}W(i,j,k,l) &= (W(-i,j,k,l) - W(-(i+1),j,k,l))
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

$$\begin{aligned}
&\quad -(W(-i,j+1,k,l)-W(-(i+1),j+1,k,l))\\
D^{i,j,10}W(i,j,k,l) &= W(-i,\bar{N}_3(i+1),k,l)-W(-(i+1),\bar{N}_3(i+1),k,l)\\
D^{i,j,11}W(i,j,k,l) &= W(-\bar{M}_3(j),j,k,l)-W(-\bar{M}_3(j+1),j+1,k,l)\\
D^{i,j,12}W(i,j,k,l) &= W(-\bar{M}_3(\bar{N}_3),\bar{N}_3,k,l)\\
D^{i,j,13}W(i,j,k,l) &= (W(-i,-j,k,l)-W(-i,-(j+1),k,l))\\
&\quad -(W(-(i+1),-j,k,l)-W(-(i+1),-(j+1),k,l))\\
D^{i,j,14}W(i,j,k,l) &= W(-\bar{M}_4(j+1),-j,k,l)-W(-\bar{M}_4(j+1),-(j+1),k,l)\\
D^{i,j,15}W(i,j,k,l) &= W(-i,-\bar{N}_4(i),k,l)-W(-(i+1),-\bar{N}_4(i+1),k,l)\\
D^{i,j,16}W(i,j,k,l) &= W(-\bar{M}_4,-\bar{N}_4(\bar{M}_4),k,l).
\end{aligned}$$

Note that differences $D^{k,l,x}$ will be defined similarly with the third and fourth indices being incremented appropriately. Finally, for a set of random elements indexed in $\mathbb{Z}^2$, say $\{b_{i,j}\}$, we will make use of the following set of partial sums:

$$b^1(i,j)=\sum_{p=0}^{i}\sum_{q=0}^{j}b_{p,q},\qquad b^2(i,j)=\sum_{p=0}^{\bar{M}_1(j+1)}\sum_{q=0}^{j}b_{p,q}$$

$$b^3(i,j)=\sum_{p=0}^{i}\sum_{q=0}^{\bar{N}_1(p)}b_{p,q},\qquad b^4(i,j)=\sum_{p=0}^{\bar{M}_1}\sum_{q=0}^{\bar{N}_1(p)}b_{p,q}$$

$$b^5(i,j)=\sum_{p=0}^{i}\sum_{q=1}^{j}b_{p,-q},\qquad b^6(i,j)=\sum_{q=1}^{\bar{N}_2(i+1)}\sum_{p=0}^{i}b_{p,-q}$$

$$b^7(i,j)=\sum_{q=1}^{j}\sum_{p=0}^{\bar{M}_2(q)}b_{p,-q},\qquad b^8(i,j)=\sum_{q=1}^{\bar{N}_2}\sum_{p=0}^{\bar{M}_2(q)}b_{p,-q}$$

$$b^9(i,j)=\sum_{p=1}^{i}\sum_{q=0}^{j}b_{-p,q},\qquad b^{10}(i,j)=\sum_{q=0}^{\bar{N}_3(i+1)}\sum_{p=1}^{i}b_{-p,q}$$

$$b^{11}(i,j)=\sum_{q=0}^{j}\sum_{p=1}^{\bar{M}_3(q)}b_{-p,q},\qquad b^{12}(i,j)=\sum_{q=0}^{\bar{N}_3}\sum_{p=1}^{\bar{M}_3(q)}b_{-p,q}$$

$$b^{13}(i,j)=\sum_{p=1}^{i}\sum_{q=1}^{j}b_{-p,-q},\qquad b^{14}(i,j)=\sum_{p=1}^{\bar{M}_4(j+1)}\sum_{q=1}^{j}b_{-p,-q}$$

$$b^{15}(i,j)=\sum_{p=1}^{i}\sum_{q=1}^{\bar{N}_4(p)}b_{-p,-q},\qquad b^{16}(i,j)=\sum_{p=1}^{\bar{M}_4}\sum_{q=1}^{\bar{N}_4(p)}b_{-p,-q}.$$

LEMMA 3. *Suppose $B$ is a QMB region. Then*

$$\sum_{i,j\in B_n}W(i,j,k,l)b_{i,j}=\sum_{x=1}^{16}\sum_{i,j\in A_x}D^{i,j,x}W(i,j,k,l)b^x(i,j)$$

*for $A_x$, $D^{i,j,x}W(i,j,k,l)$, and $b^x(i,j)$ defined above.*

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

**Proof of Lemma 3.** The conclusion of the lemma follows from the definition of a QMB region and decomposing $\sum_{i,j\in B_n} W(i,j,k,l)b_{i,j}$ as

$$\sum_{i,j\in B_n} W(i,j,k,l)b_{i,j} = \sum_{i,j\in B_{n,1}} W(i,j,k,l)b_{i,j} + \sum_{i,j\in B_{n,2}} W(i,j,k,l)b_{i,j} + \sum_{i,j\in B_{n,3}} W(i,j,k,l)b_{i,j} + \sum_{i,j\in B_{n,4}} W(i,j,k,l)b_{i,j}.$$

We then note that by repeated application of Abel's lemma we can write the double sum of the product of $a_{i,j}$ and $b_{i,j}$ as

$$\begin{aligned}\sum_{i,j\in A} a_{i,j}b_{i,j} &= \sum_{i=0}^{\bar{L}}\sum_{j=0}^{\bar{M}(i)} a_{i,j}b_{i,j} \\ &= \sum_{j=0}^{\bar{M}-1}\sum_{i=0}^{\bar{L}(j+1)-1} [(a_{i,j}-a_{i,j+1})-(a_{i+1,j}-a_{i+1,j+1})]\sum_{p=0}^{i}\sum_{q=0}^{j} b_{p,q} \\ &\quad + \sum_{j=0}^{\bar{M}-1}\left(a_{\bar{L}(j+1),j}-a_{\bar{L}(j+1),j+1}\right)\sum_{p=0}^{\bar{L}(j+1)}\sum_{q=0}^{j} b_{p,q} \\ &\quad + \sum_{i=0}^{\bar{L}-1}\left(a_{i,\bar{M}(i)}-a_{i+1,\bar{M}(i+1)}\right)\sum_{p=0}^{i}\sum_{q=0}^{\bar{M}(p)} b_{p,q} + a_{\bar{L},\bar{M}(\bar{L})}\sum_{p=0}^{\bar{L}}\sum_{q=0}^{\bar{M}(p)} b_{p,q}\end{aligned}$$

for any region $A$ that satisfies $\sum_{i,j\in A} f_{i,j} = \sum_{i=0}^{\bar{L}}\sum_{j=0}^{\bar{M}(i)} f_{i,j} = \sum_{j=0}^{\bar{M}}\sum_{i=0}^{\bar{L}(j)} f_{i,j}$. Since the subregions in the QMB region satisfy this condition, the conclusion follows from applying this identity to the sum over each subregion. ■

The next lemma provides convergence of partial sums of scores of the form given in Lemma 3.

LEMMA 4. *Let $\widehat{s}^x(i,j)$ for $x = 1,\ldots,16$ be defined as the partial sums in Lemma 3. Then $\frac{1}{n}\widehat{s}^x(i,j) \Rightarrow W_p(A^x(u_1,u_2),B)$ for*

$$\begin{aligned} A^1(u_1,u_2) &= (]0,u_1]\times]0,u_2])\cap B_1, & A^2(u_1,u_2) &= (]0,f_1(u_2)]\times]0,u_2])\cap B_1 \\ A^3(u_1,u_2) &= (]0,u_1]\times]0,\bar{g}_1])\cap B_1 & A^4(u_1,u_2) &= B_1 \\ A^5(u_1,u_2) &= (]0,u_1]\times]-u_2,0])\cap B_2, & A^6(u_1,u_2) &= (]0,u_1]\times]-g_2(u_1),0])\cap B_2 \\ A^7(u_1,u_2) &= (]0,\bar{f}_2]\times]-u_2,0])\cap B_2, & A^8(u_1,u_2) &= B_2 \\ A^9(u_1,u_2) &= (]-u_1,0]\times]0,u_2])\cap B_3 & A^{10}(u_1,u_2) &= (]-u_1,0]\times]0,g_3(-u_1)])\cap B_3 \\ A^{11}(u_1,u_2) &= (]-\bar{f}_3,0]\times]0,u_2])\cap B_3, & A^{12}(u_1,u_2) &= B_3 \\ A^{13}(u_1,u_2) &= (]-u_1,0]\times]-u_2,0])\cap B_4, & A^{14}(u_1,u_2) &= (]-f_4(-u_2),0]\times]-u_2,0])\cap B_4 \\ A^{15}(u_1,u_2) &= (]-u_1,0]\times]-\bar{g}_4,0])\cap B_4, & A^{16}(u_1,u_2) &= B_4 \end{aligned}$$

*where $u_1 = f_1(u_2)$ is the upper boundary of $B_1$; $u_2 = g_1(u_1)$ is the right boundary of $B_1$ with $\bar{g}_1 = \sup_{u_1\in[0,1/2]} g_1(u_1)$; $u_1 = f_2(-u_2)$ is the upper boundary of $B_2$ with*

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

*$\bar{f}_2 = \sup_{u_2 \in [0,1/2]} f_2(-u_2)$; $-u_2 = -g_2(u_1)$ is the left boundary of $B_2$; $-u_1 = -f_3(u_2)$ is the lower boundary of $B_3$ with $-\bar{f}_3 = -\sup_{u_2 \in [0,1/2]} f_3(u_2)$; $u_2 = g_3(-u_1)$ is the right boundary of $B_3$; $-u_1 = -f_4(-u_2)$ is the lower boundary of $B_4$; and $-u_2 = -g_4(-u_1)$ is the left boundary of $B_4$ with $-\bar{g}_4 = -\sup_{u_1 \in [0,1/2]} g_4(-u_1)$.*

**Proof of Lemma 4.** The result is immediate from Lemma 2 and the definition of partial sums used in Lemma 3. ■

## A.2. Proof of Theorem 1

Recall that

$$\widehat{\Omega} = \frac{1}{N} \sum_{i,j \in B_n} \sum_{k,l \in B_n} K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right) \widehat{s}_{i,j} \widehat{s}'_{k,l}$$
$$= \frac{n^2}{N} \sum_{k,l \in B_n} \left[ \sum_{i,j \in B_n} K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right) (\widehat{s}_{i,j}/n) \right] (\widehat{s}'_{k,l}/n).$$

Applying Lemma 3 to this expression twice, first to the inner sum then to the outer sum, gives

$$\widehat{\Omega} = \frac{n^2}{N} \sum_{y=1}^{16} \sum_{k,l \in A_y} \sum_{x=1}^{16} \sum_{i,j \in A_x} \left[ D^{k,l,y} \left( D^{i,j,x} K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right)\right)\right]$$
$$\times (\widehat{s}^x(i,j)/n)(\widehat{s}^y(k,l)/n)'. \tag{A.1}$$

The expression (A.1) contains 256 terms where each term is the sum over two different subregions of $B_n$ of the product of two partial sums with the kernel or a first, second, third, or fourth difference of the kernel. Convergence of each element in the sum to a Brownian functional can be shown by applying the result in Lemma 4 and the continuous mapping theorem. Rather than providing the argument for each term, we note that the analysis of each term is similar, provide the argument for several terms, and then state the result.

Consider the term in (A.1) with $x = y = 1$. Applying the definitions of the terms given in Lemma 3 and letting $T_{y^*,x^*}$ denote the element of (A.1) with $y = y*$ and $x = x^*$, we have that

$$T_{1,1} = \frac{n^2}{N} \sum_{l=0}^{\bar{N}_1 - 1} \sum_{k=0}^{\bar{M}_1(l+1)-1} \sum_{j=0}^{\bar{N}_1 - 1} \sum_{i=0}^{\bar{M}_1(j+1)-1} \left[ D^{k,l,1} \left( D^{i,j,1} K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right)\right)\right]$$
$$\times \left( \sum_{p=0}^{i} \sum_{q=0}^{j} \widehat{s}_{p,q}/n \right) \left( \sum_{r=0}^{k} \sum_{t=0}^{l} \widehat{s}'_{r,t}/n \right),$$

where

$$\left[ D^{k,l,1} \left( D^{i,j,1} K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right)\right)\right]$$
$$= \left[ \left( K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right) - K\left(\frac{i}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right)\right)\right.$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

$$
\begin{aligned}
&- \left(K\left(\frac{i+1}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right) - K\left(\frac{i+1}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right)\right) \\
&- \left(K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l+1}{b_2 n}\right) - K\left(\frac{i}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k}{b_1 n}, \frac{l+1}{b_2 n}\right)\right) \\
&- \left(K\left(\frac{i+1}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l+1}{b_2 n}\right) - K\left(\frac{i+1}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k}{b_1 n}, \frac{l+1}{b_2 n}\right)\right)\Bigg] \\
&- \Bigg[\left(K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l}{b_2 n}\right) - K\left(\frac{i}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l}{b_2 n}\right)\right) \\
&- \left(K\left(\frac{i+1}{b_1 n}, \frac{j}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l}{b_2 n}\right) - K\left(\frac{i+1}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l}{b_2 n}\right)\right) \\
&- \left(K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l+1}{b_2 n}\right) - K\left(\frac{i}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l+1}{b_2 n}\right)\right) \\
&- \left(K\left(\frac{i+1}{b_1 n}, \frac{j}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l+1}{b_2 n}\right) - K\left(\frac{i+1}{b_1 n}, \frac{j+1}{b_2 n}, \frac{k+1}{b_1 n}, \frac{l+1}{b_2 n}\right)\right)\Bigg].
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&n^4\left[D^{k,l,1}\left(D^{i,j,1} K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right)\right)\right] \\
&\quad \to \frac{1}{b_1^2 b_2^2} \frac{\partial K(x_1, x_2, x_3, x_4)}{\partial x_1 \partial x_2 \partial x_3 \partial x_4}\Big|_{x_1 = u_1/b_1,\, x_2 = u_2/b_2,\, x_3 = v_1/b_1,\, x_4 = v_2/b_2} \\
&\quad \equiv K^4_{b_1, b_2}(u_1, u_2, v_1, v_2).
\end{aligned}
$$

Using Assumption 2, Lemma 4, and the continuous mapping theorem, we have that

$$
\begin{aligned}
T_{1,1} = \frac{n^2}{N} \frac{1}{n^4} &\sum_{l=0}^{\bar{N}_1 - 1} \sum_{k=0}^{\bar{M}_1 (l+1) - 1} \sum_{j=0}^{\bar{N}_1 - 1} \sum_{i=0}^{\bar{M}_1 (j+1) - 1} n^4 \left[D^{k,l,1}\left(D^{i,j,1} K\left(\frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n}\right)\right)\right] \\
&\left(\sum_{p=0}^{i} \sum_{q=0}^{j} \widehat{s}_{p,q}/n\right)\left(\sum_{r=0}^{k} \sum_{t=0}^{l} \widehat{s}'_{r,t}/n\right) \\
&\Rightarrow \Lambda Q_p^{1,1}(b_1, b_2, B) \Lambda',
\end{aligned}
$$

where

$$
Q_p^{1,1}(b_1, b_2, B) = \frac{1}{\lambda(B)} \int_{\mathbf{u} \in B_1} \int_{\mathbf{v} \in B_1} K^4_{b_1, b_2}(\mathbf{u}, \mathbf{v}) W_p(A^1(\mathbf{u}), B) W_p(A^1(\mathbf{v}), B) d\mathbf{v} d\mathbf{u},
$$

$\mathbf{u} = (u_1, u_2)$, $\mathbf{v} = (v_1, v_2)$, and $\int_{\mathbf{u} \in B_1} h(\mathbf{u}) d\mathbf{u} \equiv \int_0^{\bar{f}_1} \int_0^{g_1(u_1)} h(\mathbf{u}) du_2 du_1$. This argument is identical to Kiefer and Vogelsang (2005) except that the convergence is to functionals involving set-indexed Brownian processes rather than Brownian motion.

By identical arguments, we can also handle the terms with $y = 1$ and $x = 5$, $x = 9$, or $x = 13$ to obtain

$$
\begin{aligned}
&T_{1,5} \Rightarrow \Lambda Q_p^{1,5}(b_1, b_2, B) \Lambda', \; T_{1,9} \Rightarrow \Lambda Q_p^{1,9}(b_1, b_2, B) \Lambda', \\
&\text{and } T_{1,13} \Rightarrow \Lambda Q_p^{1,13}(b_1, b_2, B) \Lambda',
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

where

$$Q_p^{1,5}(b_1,b_2,B) = -\frac{1}{\lambda(B)}\int_{\mathbf{u}\in B_1}\int_{\mathbf{v}\in B_2} K^4_{b_1,b_2}(\mathbf{u},v_1,-v_2)W_p(A^1(\mathbf{u}),B)W_p(A^5(\mathbf{v}),B)d\mathbf{v}d\mathbf{u}$$

$$Q_p^{1,9}(b_1,b_2,B) = -\frac{1}{\lambda(B)}\int_{\mathbf{u}\in B_1}\int_{\mathbf{v}\in B_3} K^4_{b_1,b_2}(\mathbf{u},-v_1,v_2)W_p(A^1(\mathbf{u}),B)W_p(A^9(\mathbf{v}),B)d\mathbf{v}d\mathbf{u}$$

$$Q_p^{1,13}(b_1,b_2,B) = \frac{1}{\lambda(B)}\int_{\mathbf{u}\in B_1}\int_{\mathbf{v}\in B_4} K^4_{b_1,b_2}(\mathbf{u},-\mathbf{v})W_p(A^1(\mathbf{u}),B)W_p(A^{13}(\mathbf{v}),B)d\mathbf{v}d\mathbf{u}$$

with $\int_{\mathbf{u}\in B_2} h(\mathbf{u})d\mathbf{u} \equiv \int_0^{\bar{f}_2}\int_0^{g_2(u_1)} h(\mathbf{u})du_2du_1$, $\int_{\mathbf{u}\in B_3} h(\mathbf{u})d\mathbf{u} \equiv \int_0^{\bar{f}_3}\int_0^{g_3(u_1)} h(\mathbf{u})du_2du_1$, and $\int_{\mathbf{u}\in B_4} h(\mathbf{u})d\mathbf{u} \equiv \int_0^{\bar{f}_4}\int_0^{g_4(u_1)} h(\mathbf{u})du_2du_1$. Changing variables in $Q_p^{1,5}$, $Q_p^{1,9}$, and $Q_p^{1,13}$ paying attention to the limits of integration and then summing yields

$$\begin{aligned} Q_p^{1,\text{Type}_1}(b_1,b_2,B) &\equiv Q_p^{1,1}(b_1,b_2,B) + Q_p^{1,5}(b_1,b_2,B) + Q_p^{1,9}(b_1,b_2,B) + Q_p^{1,13}(b_1,b_2,B) \\ &\equiv \frac{1}{\lambda(B)}\int_{\mathbf{u}\in B_1}\int_{\mathbf{v}\in B} K^4_{b_1,b_2}(\mathbf{u},\mathbf{v})W_p(A^1(\mathbf{u}),B)W_p(R(\mathbf{v}),B)d\mathbf{v}d\mathbf{u}, \quad \textbf{(A.2)} \end{aligned}$$

where

$$R(v_1,v_2) = (]\min(v_1,0),\max(v_1,0)[\times]\min(v_2,0),\max(v_2,0)[) \cap B. \quad \textbf{(A.3)}$$

Using similar definitions and arguments, one can define

$$Q_p^{2,\text{Type}_1}(b_1,b_2,B) \equiv Q_p^{5,1}(b_1,b_2,B) + Q_p^{5,5}(b_1,b_2,B) + Q_p^{5,9}(b_1,b_2,B) + Q_p^{5,13}(b_1,b_2,B) \quad \textbf{(A.4)}$$

$$Q_p^{3,\text{Type}_1}(b_1,b_2,B) \equiv Q_p^{9,1}(b_1,b_2,B) + Q_p^{9,5}(b_1,b_2,B) + Q_p^{9,9}(b_1,b_2,B) + Q_p^{9,13}(b_1,b_2,B) \quad \textbf{(A.5)}$$

$$Q_p^{4,\text{Type}_1}(b_1,b_2,B) \equiv Q_p^{13,1}(b_1,b_2,B) + Q_p^{13,5}(b_1,b_2,B) + Q_p^{13,9}(b_1,b_2,B) + Q_p^{13,13}(b_1,b_2,B). \quad \textbf{(A.6)}$$

Adding the expressions given in (A.2) and (A.4)– (A.6) gives

$$L_1(b_1,b_2,B) = \frac{1}{\lambda(B)}\int_{\mathbf{u}\in B}\int_{\mathbf{v}\in B} K^4_{b_1,b_2}(\mathbf{u},\mathbf{v})W_p(R(\mathbf{u}),B)W_p(R(\mathbf{v}),B)'d\mathbf{v}d\mathbf{u}. \quad \textbf{(A.7)}$$

There are 15 additional types of terms all of which involve the boundary of $B$ in some fashion. We consider one such case here by taking the term in (A.1) with $y = 4$ and $x = 3$:

$$\begin{aligned} T_{4,3} = \frac{n^2}{N}\sum_{i=0}^{\bar{M}_1-1}&\left[D^{k,l,4}\left(D^{i,j,3}K\left(\frac{i}{b_1n},\frac{j}{b_2n},\frac{k}{b_1n},\frac{l}{b_2n}\right)\right)\right] \\ &\times\left(\sum_{p=0}^{i}\sum_{q=0}^{\bar{N}_1(p)}\widehat{s}_{p,q}/n\right)\left(\sum_{r=0}^{\bar{M}_1}\sum_{t=0}^{\bar{N}_1(r)}\widehat{s}'_{r,t}/n\right), \end{aligned}$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

where

$$D^{k,l,4}\left(D^{i,j,3}K\left(\frac{i}{b_1 n},\frac{j}{b_2 n},\frac{k}{b_1 n},\frac{l}{b_2 n}\right)\right)$$
$$= K\left(\frac{i}{b_1 n},\frac{\bar{N}_1(i)}{b_2 n},\frac{\bar{M}_1}{b_1 n},\frac{\bar{N}_1(\bar{M}_1)}{b_2 n}\right) - K\left(\frac{i+1}{b_1 n},\frac{\bar{N}_1(i+1)}{b_2 n},\frac{\bar{M}_1}{b_1 n},\frac{\bar{N}_1(\bar{M}_1)}{b_2 n}\right).$$

Once again, observe that

$$nD^{k,l,4}\left(D^{i,j,3}K\left(\frac{i}{b_1 n},\frac{j}{b_2 n},\frac{k}{b_1 n},\frac{l}{b_2 n}\right)\right)$$
$$\to -\frac{1}{b_1}\frac{\partial K}{\partial x_1}\left(\frac{u_1}{b_1},\frac{g_1(u_1)}{b_2},\frac{\bar{f}_1}{b_1},0\right) - \frac{1}{b_2}\frac{\partial K}{\partial x_2}\left(\frac{u_1}{b_1},\frac{g_1(u_1)}{b_2},\frac{\bar{f}_1}{b_1},0\right)\frac{\partial g_1(z)}{\partial z}|_{z=u_1}$$
$$\equiv -\frac{\partial K}{\partial t_{b_1,b_2}(\mathbf{u})}\left(\frac{u_1}{b_1},\frac{g_1(u_1)}{b_2},\frac{\bar{f}_1}{b_1},0\right)$$

converges to the derivative of the kernel function in the direction tangent to the boundary of the space $B$.

As with the previous argument, we have that

$$T_{4,3} = \frac{n^2}{N}\frac{1}{n}\sum_{i=0}^{\bar{M}_1-1} n\left[D^{k,l,4}\left(D^{i,j,3}K\left(\frac{i}{b_1 n},\frac{j}{b_2 n},\frac{k}{b_1 n},\frac{l}{b_2 n}\right)\right)\right]$$
$$\times\left(\sum_{p=0}^{i}\sum_{q=0}^{\bar{N}_1(p)}\widehat{s}_{p,q}/n\right)\left(\sum_{r=0}^{\bar{M}_1}\sum_{t=0}^{\bar{N}_1(r)}\widehat{s}'_{r,t}/n\right)$$
$$\Rightarrow \Lambda Q_p^{4,3}(b_1,b_2,B)\Lambda',$$

where

$$Q_p^{4,3}(b_1,b_2,B) = -\frac{1}{\lambda(B)}\int_0^{\bar{f}_1}\frac{\partial K}{\partial t_{b_1,b_2}(\mathbf{u})}\left(\frac{u_1}{b_1},\frac{g_1(u_1)}{b_2},\frac{\bar{f}_1}{b_1},0\right)$$
$$\times W_p\left(A^3(u_1,u_2),B\right)W_p(B_1,B)'du_1$$

using Assumption 2, Lemma 4, and the continuous mapping theorem.

Define the set

$$P(v_1,v_2) = (]1(v_1<0)(1(v_2\geq 0)(-\bar{f}_3)+1(v_2<0)(v_1)),1(v_1\geq 0)(1(v_2\geq 0)v_1$$
$$+1(v_2<0)\bar{f}_2)]\times]1(v_2<0)(1(v_1\geq 0)(v_2)+1(v_1<0)(-\bar{g}_4)),$$
$$1(v_2\geq 0)(1(v_1\geq 0)\bar{g}_1+1(v_1<0)v_2])$$
$$\cap B. \tag{A.8}$$

Also, let

$$\int_{\mathbf{v}\in\partial B^-} h(\mathbf{v})d\mathbf{v} = \int_{\mathbf{v}\in B_1:v_1\geq 0,v_2=g_1(v_1)} h(\mathbf{v})dv_1 + \int_{\mathbf{v}\in B_2:v_1=f_2(-v_2),v_2\leq 0} h(\mathbf{v})dv_2$$
$$+\int_{\mathbf{v}\in B_3:v_1=-f_3(v_2),v_2\geq 0} h(\mathbf{v})dv_2 + \int_{\mathbf{v}\in B_4:v_1\leq 0,v_2=-g_4(-v_1)} h(\mathbf{v})dv_1. \tag{A.9}$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

Then applying the same argument as for the term with $y = 4$ and $x = 3$ to the terms with $y = 4$ and $x = 7$, $x = 11$, or $x = 15$ and summing the results yield

$$\begin{aligned} Q_p^{1,\text{Type}_1 5}(b_1, b_2, B) &\equiv Q_p^{4,3}(b_1, b_2, B) + Q_p^{4,7}(b_1, b_2, B) + Q_p^{4,11}(b_1, b_2, B) \\ &\quad + Q_p^{4,13}(b_1, b_2, B) \\ &\equiv -\frac{1}{\lambda(B)} \int_{\mathbf{u}\in\partial B^-} \frac{\partial K}{\partial t_{b_1,b_2}(\mathbf{u})} \left(\frac{u_1}{b_1}, \frac{u_2}{b_2}, \frac{\bar{f}_1}{b_1}, 0\right) \\ &\quad \times W_p(P(u_1, u_2), B) W_p(B_1, B)' d\mathbf{u}. \end{aligned} \tag{A.10}$$

Obtaining the result follows from applying similar arguments to the remaining terms and combining as demonstrated above. To make stating the final result more tractable, we define the following objects in addition to (A.3), (A.8), and (A.9):

$$\begin{aligned} \int_{\mathbf{v}\in\partial B^+} h(\mathbf{v}) d\mathbf{v} &= \int_{\mathbf{v}\in B_1: v_1 = f_1(v_2), v_2 \geq 0} h(\mathbf{v}) dv_2 + \int_{\mathbf{v}\in B_2: v_1 \geq 0, v_2 = -g_2(v_1)} h(\mathbf{v}) dv_1 \\ &\quad + \int_{\mathbf{v}\in B_3: v_1 \leq 0, v_2 = g_3(-v_2)} h(\mathbf{v}) dv_1 \\ &\quad + \int_{\mathbf{v}\in B_4: v_1 = -f_4(-v_2), v_2 \leq 0} h(\mathbf{v}) dv_2, \end{aligned} \tag{A.11}$$

and

$$\overrightarrow{v}_1 = (\bar{f}_1, 0), \quad \overrightarrow{v}_2 = (0, -\bar{g}_2), \quad \overrightarrow{v}_3 = (0, \bar{g}_3), \quad \overrightarrow{v}_4 = (-\bar{f}_4, 0). \tag{A.12}$$

Given these objects and following the steps above, we have that $\widehat{\Omega} \Rightarrow \frac{1}{\lambda(B)} \Lambda Q_p(b_1, b_2, B) \Lambda'$, where $Q_p(b_1, b_2, B) = L_{p,I,I}(b_1, b_2, B) + L_{p,I,II}(b_2, b_2, B) + L_{p,II,II}(b_1, b_2, B) + \cdots + L_{p,IV,IV}(b_1, b_2, B)$ where $L_{p,\cdot,\cdot}$ are defined in (3.3)–(3.5). ■

## A.3. Proof of Theorem 2

Given (3.7) and Theorem 1, it is easy to show that $F$ and $t$ converge to functionals of $W_p(A, B)$, $\Lambda$, $\mathcal{J}$, and $R(\theta_0)$ using the continuous mapping theorem. Straightforward extensions of arguments used by Kiefer and Vogelsang (2005) in the proof of their Theorem 3 show that the limiting functionals can be rewritten in terms of $W_q(A, B)$ with $\Lambda$, $\mathcal{J}$, and $R(\theta_0)$ canceling from the representations.

# APPENDIX B: Missing Locations

This appendix describes one way to accommodate missing locations that is closely related to the framework in Conley (1999). Define $\ell_{i,j}$ as an indicator for whether a given location is sampled: $\ell_{i,j} = 1$ if there is an observation at a point $\{i, j\}$, $\ell_{i,j} = 0$ otherwise. Sample size $N = \sum_{(i,j)\in B_n}$ which will no longer equal $\|B_n\|$ the integer coordinate pairs within the set $B_n$. Our sample moment condition written just as a sum over $\ell_{i,j} = 1$ locations is

$$\frac{1}{N} \sum_{(i,j)\in B_n : \ell_{i,j}=1} s_{i,j}(\hat{\theta}) = 0. \tag{B.1}$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

An observationally equivalent representation can easily be obtained using a sum over all locations in $B_n$. Define a random variable $s^*_{i,j}(\theta) = \begin{cases} 0 \text{ if } & \ell_{i,j} = 0 \\ s_{i,j}(\theta) \text{ if } & \ell_{i,j} = 1 \end{cases}$ that is observable at all integer coordinate locations in $B_n$. This new process simply replaces 'missing' observations with zeros. The corresponding sample moment condition is:

$$\frac{1}{\|B_n\|} \sum_{(i,j)\in B_n} s^*_{i,j}(\hat{\theta}) = 0. \tag{B.2}$$

The estimators implicitly defined by equations (B.1) and (B.2) are numerically identical because these sets of equations differ only in scale, the sums are the same:

$$\sum_{(i,j)\in B_n:\ell_{i,j}=1} s_{i,j}(\theta) = \sum_{(i,j)\in B_n} s^*_{i,j}(\theta).$$

The test statistics we consider in this paper are also identical using either the representations involving $s_{i,j}(\theta)$ or $s^*_{i,j}(\theta)$. For example, take our Wald statistic using $s^*_{i,j}(\theta)$ notation:

$$F_1 = \|B_n\| r(\hat{\theta})' \times \left( \begin{array}{c} \left[\frac{1}{\|B_n\|}\sum_{(i,j)\in B_n} \frac{\partial}{\partial\theta'} s^*_{i,j}(\hat{\theta})\right]^{-1} \times \\ \left[\frac{1}{\|B_n\|}\sum_{(i_1,j_1)\in B_n}\sum_{(i_2,j_2)\in B_n} K(\frac{i_1}{h_1},\frac{i_2}{h_1},\frac{j_1}{h_2},\frac{j_2}{h_2}) s^*_{i_1,j_1}(\hat{\theta}) s^*_{i_2,j_2}(\hat{\theta})'\right] \times \\ \left[\frac{1}{\|B_n\|}\sum_{(i,j)\in B_n} \frac{\partial}{\partial\theta'} s^*_{i,j}(\hat{\theta})\right]^{-1} \end{array} \right)^{-1} r(\hat{\theta}).$$

The corresponding Wald statistic using the $s_{i,j}(\theta)$ representation is:

$$F_2 = Nr(\hat{\theta})' \left( \begin{array}{c} \left[\frac{1}{N}\sum_{(i,j)\in B_n:\ell_{i,j}=1} \frac{\partial}{\partial\theta'} s_{i,j}(\hat{\theta})\right]^{-1} \times \\ \left[\frac{1}{N}\sum_{(i_1,j_1)\in B_n:\ell_{i_1,j_1}=1}\sum_{(i_2,j_2)\in B_n:\ell_{i_2,j_2}=1} K(\frac{i_1}{h_1},\frac{i_2}{h_1},\frac{j_1}{h_2},\frac{j_2}{h_2}) s_{i_1,j_1}(\hat{\theta}) s_{i_2,j_2}(\hat{\theta})'\right] \times \\ \left[\frac{1}{N}\sum_{(i,j)\in B_n:\ell_{i,j}=1} \frac{\partial}{\partial\theta'} s_{i,j}(\hat{\theta})\right]^{-1} \end{array} \right)^{-1} r(\hat{\theta}).$$

The scale factors of $\|B_n\|$ and $N$, respectively, cancel out of the statistics $F_1$ and $F_2$. $F_1$ and $F_2$ are numerically identical because:

$$\frac{1}{N}\sum_{(i,j)\in B_n:\ell_{i,j}=1} \frac{\partial}{\partial\theta'} s_{i,j}(\hat{\theta}) = \sum_{(i,j)\in B_n} \frac{\partial}{\partial\theta'} s^*_{i,j}(\hat{\theta})$$

and

$$\sum_{(i_1,j_1)\in B_n}\sum_{(i_2,j_2)\in B_n} K(\frac{i_1}{h_1},\frac{i_2}{h_1},\frac{j_1}{h_2},\frac{j_2}{h_2}) s^*_{i_1,j_1}(\hat{\theta}) s^*_{i_2,j_2}(\hat{\theta})'$$

$$= \sum_{(i_1,j_1)\in B_n:\ell_{i_1,j_1}=1}\sum_{(i_2,j_2)\in B_n:\ell_{i_2,j_2}=1} K(\frac{i_1}{h_1},\frac{i_2}{h_1},\frac{j_1}{h_2},\frac{j_2}{h_2}) s_{i_1,j_1}(\hat{\theta}) s_{i_2,j_2}(\hat{\theta})'.$$

Thus our results apply to t and Wald tests involving $\hat{\theta}$ defined via

$$\frac{1}{N}\sum_{(i,j)\in B_n:\ell_{i,j}=1} s_{i,j}(\hat{\theta}) = 0$$

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press

provided the associated $s^*_{i,j}(\theta)$ process satisfies our regularity conditions in Assumptions 1–6. Repeating its definition for convenience

$$s^*_{i,j}(\theta) = \begin{cases} 0 \text{ if } & \ell_{i,j} = 0 \\ s_{i,j}(\theta) \text{ if } & \ell_{i,j} = 1 \end{cases}.$$

Our maintained assumptions are of course that the process $s_{i,j}(\theta_0)$ satisfies our regularity conditions in Assumptions 1–6. Thus the issue reduces to what conditions on the $\ell_{i,j}$ will imply that $s^*_{i,j}(\theta_0)$ also meets our regularity conditions. There are many possible assumptions regarding $\ell_{i,j}$ that would be sufficient. The mixing conditions in Assumption 3 will be satisfied for example by a stationary $\ell_{i,j}$ process that is independent of $s_{i,j}(\theta_0)$ and exhibits weak spatial dependence satisfying the same mixing rate conditions as $s_{i,j}(\theta_0)$. Our high-level assumptions of consistency (Assumption 2) and homogeneity and Uniform LLN for $\frac{\partial}{\partial \theta'} s_{i,j}(\theta)$ are cumbersome to directly extend to $s^*_{i,j}(\theta_0)$. It is however clear that these assumptions will follow from a set of primitive conditions that allow for stationary and weakly spatially dependent $\ell_{i,j}$. We leave an exposition of these primitive conditions for future research.

https://doi.org/10.1017/S0266466614000814 Published online by Cambridge University Press