Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators

C. Alan Bester, Timothy G. Conley, Christian B. Hansen and Timothy J. Vogelsang

First Version: August 2008. This Version: September 2011

Abstract. 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 non-degenerate 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 show that critical values for commonly-used test statistics may be obtained through the i.i.d. bootstrap. We illustrate the potential gains of the “fixed-b” approximation in the spatial context through a simulation example.

Keywords: HAC, serial correlation, bootstrap

JEL Codes: C12, C21, C22, C23

*We thank Silvia Gonçalves for helpful discussions. We thank seminar participants at University of Wisconsin - Madison, Harvard, MIT, Boston University, Princeton, and Hebrew University. 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.

†University of Chicago Booth School of Business, cbester@chicagobooth.edu

‡University of Western Ontario, tconley3@uwo.ca

§University of Chicago Booth School of Business, chansen1@chicagobooth.edu

¶Department of Economics, Michigan State University, tjv@msu.edu
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 consistent 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\(^2\).

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).

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


\(^3\)See Conley (1996, 1999) for early examples.
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 normal 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 non-degenerate 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.

\footnote{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.}
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 challenging in the time series case, and is even more so when the data exhibits dependence in more than one dimension. The resulting variance estimates 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 this 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 change 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 is 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.
This dependence on the sampling region complicates implementation relative to the time
series case as critical values must be recalculated for each different sampling region encoun-
tered in practice. Direct calculation requires approximating functions of Brownian sheets
with non-trivial boundary conditions and are cumbersome at best. A contribution of this
paper is showing that a simple procedure based on an i.i.d. bootstrap is valid for the fixed-$b$
asymptotic distribution and can be used to obtain critical values quite generally.\footnote{Note that the bootstrap is used in this paper purely as a simulation device and that other Monte Carlo procedures could be used instead.} Our paper thus delivers a simple, practical way to conduct inference which accounts for the wide array
of configurations in which spatial data is observed.

We present a set of simulation results that illustrate the potential advantages of our
approach. The simulations show that inference procedures that ignore spatial correlation
are highly size-distorted when spatial correlation is present. 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. Our simulations also highlight the potential importance of
accounting for the shape of the sampling region.

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 paramet-
ric covariance structures are also available. Ibragimov and Müller (2010) considers doing
inference based on partitioning data into groups that are “approximately” uncorrelated and
estimating parameters within each partition. Ibragimov and Müller (2010) provides 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.\(^6\) 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 conditions not used in Ibragimov and Müller (2010). Thus, we view the two approaches as complementary.

\(^6\)This is trivially achieved with time series data and may also be achieved with more general spatial dependence under conditions given in Bester, Conley, and Hansen (2010).
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, and show how the simple i.i.d. bootstrap can be used as a simulation procedure to obtain critical values. Section 4 contains results from our simulation experiments, and Section 5 concludes.

2. Spatial HAC Estimation

Consider an estimator, \( \hat{\theta}_N \), of a \( p \times 1 \) parameter vector \( \theta_0 \) that satisfies
\[
\frac{1}{N} \sum_{t=1}^{N} s_t(\hat{\theta}_N) = 0
\]
where \( N \) is the sample size and \( s_t(\theta) \) is a \( p \times 1 \) score vector such that
\[
E(s_t(\theta_0)) = 0.
\]

We allow the data and score vector to exhibit spatial dependence where the strength of the correlation between any two observations depends on some observable distance measure. In particular, we assume that it is possible to map the data onto a region on a two-dimensional integer lattice such that the distance measure can be written as a function of two lattice points.

To be concrete, fix a region \( B \subseteq [-1/2, 1/2] \). For a scalar \( c \), define \( cB = \{ cu : u \in B \} \). For each \( n \in \mathbb{N} \), define a set of 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 \( N(n) \equiv \|B_n\| = O(n^2) \), and we will have \( N(n)/n^2 \to \lambda(B) \) where \( \lambda \) is Lesbesgue measure. Associated with each observation \( t \) is a point on the plane \( d(t) = \{i,j\} \) where \( d(t) \in B_n \). For the remainder of the paper, we write \( N \) for the sample size, supressing the dependence on \( n \), and write limits as \( N \to \infty \).
and $n \to \infty$ interchangably. We also note that, while we rely on this representation of the sampling regions for our theoretical results, the bootstrap procedure we propose for practical applications does not require specifying a region $B$.

We can then rewrite the sample moment condition that defines $\hat{\theta}_N$ as

$$\frac{1}{N} \sum_{(i,j) \in B_n} \hat{s}_{i,j} = 0,$$

where $\hat{s}_{i,j} = s_t(\hat{\theta}_N)$ is associated with observation $t$ for which $d(t) = (i, j)$. It is clear that equation (2.2) gives the same estimate of $\theta$ as (2.1). Define the analog to $\hat{s}_{i,j}$ as $s_{i,j} = s_t(\theta_0)$, again where $d(t) = (i, j)$.

Under suitable regularity conditions, $\hat{\theta}_N$ will have an asymptotic normal distribution with variance-covariance matrix of the sandwich form with middle matrix given by

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

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

$$\hat{\Omega} = \frac{1}{N} \sum_{(i_1,j_1) \in B_n} \sum_{(i_2,j_2) \in B_n} K\left( \frac{i_1}{h_1}, \frac{i_2}{h_1}, \frac{j_1}{h_2}, \frac{j_2}{h_2} \right) \hat{s}_{i_1,j_1} \hat{s}_{i_2,j_2}^t,$$

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.

Among other things, this avoids this issue of a researcher having a sample of size $N_0$ and being concerned that potentially none of the sequence of regions $\{B_n\}$ contains exactly $N_0$ data points.

We could allow for the possibility that we may not have an observation at every lattice point in our sample under the assumption that the existence of an observation at a lattice is independent of $s_{i,j}$ and that the number of missing observations is $o(N^{1/2})$. This extension is straightforward but complicates the notation so we do not pursue it here.
2.1. QMB Regions

Part of the contribution of our paper is to explicitly characterize the dependence of the limiting distribution of $\hat{\Omega}$ on the shape of the sampling region as represented by $B$. To achieve this, we restrict the region $B$ to have “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 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 non-monotone.

![Figure 1](image-url)

**Figure 1.** Sample Regions Satisfying (a) and Violating (b) Definition of Quadrant-Wise Monotone Boundary
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 nonincreasing 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

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

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

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 \((2.3)\) is invariant to this rotation.

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. Analytically calculating these critical values would involve potentially complicated integration over set-indexed Brownian motions. To avoid this, we show in Section 3.3 that critical values may be approximated using an easily implemented i.i.d. bootstrap.

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\(^9\)

\[
\hat{\Omega}_T \Rightarrow \Lambda \left[ \int_0^1 \int_0^1 \frac{\partial^2 K(r,s)}{\partial r \partial s} Y_p(r)Y_p(s)drds \right] \Lambda' 
\]

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 analogue 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

\[
\text{Cov}(W(A), W(B)) = \lambda(A \cap B),
\]

\(^9\)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.
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 $B$. The limiting distribution of $\hat{\Omega}$ 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

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

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

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

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

10
\[ P_B(u) = \begin{cases} 
\left( [0,u_1] \times [0,u_2] \right) \cap B & u \in B_1 \\
\left( [u_1,u_1] \times [u_2,0] \right) \cap B & u \in B_2 \\
\left( [u_1,0] \times [u_2,u_2] \right) \cap B & u \in B_4 \\
\left( [u_1,0] \times [0,u_2] \right) \cap B & u \in B_3 
\end{cases} \]

where \( [u_1,u_1] \times [u_2,u_2] \) is the smallest rectangle containing \( B \). Note that \( R_B(u) \) is the intersection of \( B \) with a rectangle with vertices at \( u \) and the origin while \( P_B(v) \) 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

\[
L_{p,I,I}(b_1, b_2, B) = \int_{u \in B} \int_{v \in B} \frac{\partial^4 K_{b_1, b_2}(u, v)}{\partial u_1 \partial u_2 \partial v_1 \partial v_2} W_p(R_B(u), B)W_p(R_B(v), B)' dv du,
\]

where \( K_{b_1, b_2} = K(\frac{u_1}{b_1}, \frac{v_1}{b_1}, \frac{u_2}{b_2}, \frac{v_2}{b_2}) \). These integrals arise as limits of partial sums taken on the interior of the rectangle.\(^{11}\) Unlike the time series case, however, the limit distributions also involve integrals such as

\[
L_{p,I,II}(b_1, b_2, B) = \int_{u \in B} \int_{v \in \partial B} \frac{\partial^3 K_{b_1, b_2}(u, v)}{\partial u_1 \partial u_2 \partial z(v)} W_p(R_B(u), B)W_p(R_B(v), B)' dv du
\]

\[
L_{p,I,III}(b_1, b_2, B) = \int_{u \in B} \int_{v \in \partial B} \frac{\partial^3 K_{b_1, b_2}(u, v)}{\partial u_1 \partial u_2 \partial t(v)} W_p(R_B(u), B)W_p(P_B(v), B)' dv du,
\]

where \( z(v) = 1 (v \in B_2 \cap B_3) v_1 + 1 (v \in B_1 \cap B_4) v_2 \), and \( t(v) \) is a unit length vector tangent to the boundary of the region, \( \partial B \), in a counterclockwise direction. Finally, we will also need integrals weighted against \( W_B \) evaluated over the four subregions \( B_i \), e.g.,

\[
L_{p,I,IV} = \sum_{i=1}^{4} \int_{u \in B} \frac{\partial^2 K_{b_1, b_2}(u, e_i)}{\partial u_1 \partial u_2} W_p(R_B(u), B)W_p(B_i, B)' du
\]

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

where \( e_i \) is the ‘endpoint’ of each region when the boundary \( \partial B \) is traversed in a counterclockwise direction starting at \( (\pi_2, 0) \).\(^{12}\) 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 \( v \).

\(^{11}\)See Lemma 2 in the Appendix.

\(^{12}\)That is, \( e_1 = (0, \pi_1), e_2 = (\pi_2, 0), \text{ etc.}, e_3 = (0, \pi_1), \) and \( e_4 = (\pi_2, 0) \).
As in the time series case (3.1), each term in the limit distribution of \( \hat{\Omega} \) is a random variable whose distribution depends explicitly on both the kernel, \( K \), and 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 \( \hat{\Omega} \). Restrictions on the kernel may simplify these terms. For example, if the kernel is symmetric, \( K(u,v) = K(v,u) \), then any two terms with the same indices are equivalent, e.g., \( L_{p,1,II} = L_{p,II,1} \). 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 \( A \) be a collection of Borel subsets of \([-1/2,1/2]^2\) that satisfies (i) \( \int_0^1 \sqrt{H(A,x)}dx < \infty \) where \( H(A,\epsilon) \) is the logarithm of the smallest number of open disks of radius \( \epsilon \) which form a covering of \( A \) and (ii) \( \lambda(\partial A) = 0 \) for all \( A \in A \) where \( \partial A \) is the boundary of \( A \).

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

**ASSUMPTION 2.** Suppose that \( B \in A \) is a QMB-region such that \( A \subseteq B \) for all \( A \in A \). Assume that \( B \) is inscribed in the rectangle \([u_1, \overline{u}_1] \times [u_2, \overline{u}_2]\) where \( \max_{j=1,2} \{ |u_j|, |\overline{u}_j| \} = 1/2 \).
and $\min_{j=1,2} \{ |u_j|, |\bar{u}_j| \} = c > 0$. Let $B_n = nB$ 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. 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.** $\{s_t(\theta_0)\}_{i \in \mathbb{Z}^d}$ is a strictly stationary field of $\phi$–mixing random variables with mean zero, finite variance, mixing coefficients $\phi_2(k)$ where (i) $E(|s_t(\theta_0)|^4) < \infty$ and $\sum_{k>0} k^3 \phi_2(k) < \infty$ or (ii) for some $b$ in $]2, 4[,$ $E(|s_t|^{2b/(b-2)}) < \infty$ and $\phi_2(k) = O(k^{-b})$. Let $\Lambda = \lim_{N \to \infty} \frac{1}{N} E \left[ \left( \sum_{(i,j) \in B_n} s_{i,j} \right) \left( \sum_{(i,j) \in B_n} s_{i,j} \right)' \right]$.

The scores $s_t$ 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, which is often assumed in the time series case but is restrictive in the spatial setting. In

---

13 We assume that one of the ‘endpoints’ $e_j$ defined in footnote 12 is at distance exactly 1/2 from the origin for convenience. Any $\tilde{B} = cB$ for $0 < c < 1$ would give the same limit distribution.

14 For a probability space $(\Omega, \mathcal{A}, P)$ and two $\sigma$–algebras $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{A}$, define $\phi(\mathcal{U}, \mathcal{V}) = \sup\{\|P(\mathcal{V}|\mathcal{U}) - P(\mathcal{V})\|_\infty, \mathcal{V} \in \mathcal{V} \}$. For a real random field $(\mathbb{R}^d, \mathcal{B}^d, P)$ 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 $\phi_k(m) = \sup\{\phi(\mathcal{F}_1, \mathcal{F}_2), |\Gamma_2| \leq k, d(\Gamma_1, \Gamma_2) \geq m \}$ 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.
particular, it is in general too strong for application to Gibbs random fields. Because we are working directly with the scores, we state consistency of the estimator $\hat{\theta}_N$ defined by (2.1) as a high level assumption.

**ASSUMPTION 4.** $\hat{\theta}_N \xrightarrow{p} \theta_0$ 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 \( \left\{ \frac{1}{n} \sum_{(i,j) \in \mathbb{Z}^2} \lambda (nA \cap R_{(i,j)}) s_{(i,j)} : A \in \mathcal{A} \right\} \) 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 $\hat{\Omega}$ 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 1–3 below would exist, and further that the bootstrap procedure we propose below would remain valid, under much weaker restrictions on the region $B$.

\footnote{See comments and references in Dedecker (2001). A previous version of this paper referenced an FCLT in Goldie and Greenwood (1986), which requires uniform mixing. We thank an anonymous referee for bringing this issue to our attention.}
ASSUMPTION 6.

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

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

Assumption 6 may be interpreted as two requirements, the first of which is a uniform (over both \( \Theta \) and \( A \)) law of large numbers. The second requirement is that the expectation of \( \partial s_t / \partial \theta \) is homogeneous across the sampling region. For example, in the regression context, Assumption 6 requires that \( E(x_t x'_t) \) does not depend on the location, \( t \). 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 \( \hat{\theta}_N \).

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

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

While asymptotic normality of \( \hat{\theta}_N \) seems trivial from Assumptions 5 and 6, note that the FCLT above applies to partial sums of the form \( \lambda(B_n \cap R_t) s_t \), where \( 0 \leq \lambda(B_n \cap R_t) \leq 1 \), where the estimator is defined by sums of \( 1_N(t) s_t \), where \( 1_N(t) = 1 \) if \( t \in B_n \) 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 \( \hat{\Omega} \) 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 $(h_1, h_2) = (b_1n, b_2n)$ 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,

\begin{equation}
\hat{\Omega} \Rightarrow \frac{1}{\lambda(B)} \Lambda Q_p(b_1, b_2, B) \Lambda'
\end{equation}


\begin{equation}
Q_p(\cdot) = L_{p, I, I}(\cdot) + L_{p, I, II}(\cdot) + L_{p, II, II}(\cdot) + \ldots + L_{p, IV, IV}(\cdot),
\end{equation}

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

As for KV in the time series case, the limit of $\hat{\Omega}$ is proportional to $\Lambda \Lambda' \equiv \Omega_0$, but depends on a 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 $Q_p(b_1, b_2, B) = I$, (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 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 an $q$-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

\begin{equation}
\sqrt{N} r(\hat{\theta}_N) \Rightarrow -R(\theta_0)(\lambda(B)^{1/2} \mathcal{J})^{-1} \Lambda W_p(B).
\end{equation}

Note that the right hand side follows a $q$-variate Gaussian distribution.
Let $\hat{J}_{\hat{\theta}N} = \frac{1}{N} \sum_{i,j \in B_n} \frac{\partial s_{i,j}(\hat{\theta}N)}{\partial \theta}$ and define the Wald statistic

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

and the $t$-statistic

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

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

$$F \Rightarrow W_q(B) Q_q(b_1, b_2, B)$$

and

$$t \Rightarrow W_1(B) \frac{Q_1(b_1)}{\sqrt{Q_1(b_1, b_2, B)}}.$$

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. However, the resulting reference distributions are nonstandard and are influenced by the choice of kernel and bandwidth(s) as well as the shape of the sampling region. In the next section, we show that critical values may be obtained via an i.i.d. bootstrap that is both general and simple to carry out in practice.

### 3.3. Critical Values and the Bootstrap

For a given region $B$, the limiting distributions given by Theorem 2 can be simulated given a choice of kernel and bandwidths. In the time series case, KV operationalize these tests by tabulating critical values as a function of the (scalar) bandwidth parameter $b$ for each of several commonly used kernels. Here, unfortunately, such an approach is impractical: The bandwidth parameter is two dimensional, and more importantly, the shape of the sampling region obviously differs depending on the application. In this section, we present a simple
bootstrap procedure that will generate critical values for any choice of kernel, bandwidth, and sampling region. Importantly, this procedure does not require that the researcher specify a particular region $B$; all that is needed is the set of locations in the actual sample.

Note that we use the bootstrap here purely as a simulation device. We use a ‘naive’ bootstrap, where the denominator of our test statistic is calculated using the same formula used with the original data, purely as a Monte Carlo device, as an alternative to simulating from the reference distributions in Theorem 2 numerically. In particular we are not asking the bootstrap to reproduce the dependence structure in the original data, which would require stronger assumptions and a blocking scheme for resampling. While we find the bootstrap procedure appealing due to its simplicity and familiarity to applied researchers, it is obvious that other Monte Carlo procedures could be used.

As in Section 2, let $d(t)$ denote the location of observation $t$ on the integer lattice. For each $d(t) \in B_n$, let $y_t^{**}$ be a value drawn by i.i.d. sampling with replacement from the observed data, $\{y_t\}$. One bootstrap realization consists of the draws $\{y_t^{**}\}$, placed at the same set of locations $B_n$ as the original sample.

Once a bootstrap sample is drawn, the bootstrap statistics are computed as follows. Let $s_{i,j}^{**}(\theta)$ denote the bootstrap score at coordinate $(i, j) \in B_n$. Analogously to (2.2), define the bootstrap estimator $\hat{\theta}^*$ such that

$$\frac{1}{N} \sum_{(i,j) \in B_n} s_{i,j}^{**}(\hat{\theta}^*) = 0.$$
Because of i.i.d. resampling, the conditional mean of the bootstrap score, given a particular realization of the score process \( \{s_{i,j}\} \), is

\[
E^* (s_{i,j}^*(\theta)) = \frac{1}{N} \sum_{(i,j) \in B_n} s_{i,j}(\theta).
\]

Recalling (2.2), it immediately follows that

\[
E^* (s_{i,j}^*(\hat{\theta}_N)) = 0.
\]

Therefore, in the bootstrap world, the population value of \( \theta \) is \( \hat{\theta}_N \) and \( \hat{\theta}^* \) is an estimator of \( \hat{\theta}_N \). We conserve on notation by letting \( \hat{s}_{i,j}^{**} \) denote \( s_{i,j}^*(\hat{\theta}^*) \).

Define the bootstrap versions of the \( t \) and Wald statistics as

\[
F^* = N \left( r(\hat{\theta}^*) - r(\hat{\theta}_N) \right)' \left( R(\hat{\theta}^*) \left( \hat{J}_{\theta}^* \right)^{-1} \hat{\Omega}^* \left( \hat{J}_{\theta}^* \right)^{-1} R(\hat{\theta}^*)' \right)^{-1} \left( r(\hat{\theta}^*) - r(\hat{\theta}_N) \right),
\]

\[
t^* = \frac{\sqrt{N} \left( r(\hat{\theta}^*) - r(\hat{\theta}_N) \right)}{\sqrt{R(\hat{\theta}^*) \left( \hat{J}_{\theta}^* \right)^{-1} \hat{\Omega}^* \left( \hat{J}_{\theta}^* \right)^{-1} R(\hat{\theta}^*)'}}.
\]

where

\[
\hat{\Omega}^* = \frac{1}{N} \sum_{(i_1,j_1) \in B_n} \sum_{(i_2,j_2) \in B_n} K(\frac{i_1}{h_1}, \frac{i_2}{h_1}, \frac{j_1}{h_2}, \frac{j_2}{h_2}) \hat{s}_{i_1,j_1}^{**} \hat{s}_{i_2,j_2}^{**}',
\]

\[
\hat{J}_{\theta}^* = \frac{1}{N} \sum_{(i,j) \in B_n} \frac{\partial s_{i,j}^{**}}{\partial \theta^*}.
\]

Other than the recentering around \( r(\hat{\theta}_N) \), the bootstrap versions of the statistics are computed in the same way that the original statistics are computed.

As shown by Gonçalves and Vogelsang (2011) in a time series regression setting, the i.i.d. bootstrap is first order equivalent to the fixed-\( b \) asymptotic distribution of \( t \) and Wald statistics even when the data has dependence. The i.i.d. bootstrap remains valid when the data is dependent because partial sums of the bootstrap data satisfy a FCLT. Because the
asymptotic variance covariance matrix in the FCLT cancels from the fixed-\(b\) distribution, the bootstrap does not have to mimic the correct variance covariance matrix. The results of Gonçalves and Vogelsang (2011) for the i.i.d. bootstrap naturally extend to the spatial context as the next theorem shows.

To formally establish the result, we make one further high-level assumption.

**ASSUMPTION 7.** Assume that (i) \( \hat{\theta}^* - \hat{\theta}_N \overset{p}{\to} 0 \) and (ii)
\[
\sup_{A \in A, \theta \in \mathcal{N}(\theta_0)} \left\| \frac{1}{N} \sum_{(i,j) \in nA} \frac{\partial s_{i,j}^*(\theta)}{\partial \theta^*} - \frac{\lambda(A)}{\lambda(B)} \mathcal{J}^*(\theta) \right\| P^* \to 0
\]
where \( \mathcal{J}^*(\theta_0) \) is nonsingular and \( \mathcal{N}(\theta_0) \) is an open neighborhood of \( \theta_0 \).\(^{16}\)

The first condition in Assumption 7 that \( \hat{\theta}^* - \hat{\theta}_N \overset{p}{\to} 0 \) holds under essentially the same regularity conditions as does Assumption 4; see Gonçalves and White (2004). A uniform law of large numbers for the derivative of the score about some matrix \( \mathcal{J}^*(\theta) \) is also assumed to hold in the bootstrap samples. Note that because it cancels from the numerator and denominator of the considered test-statistics, \( \mathcal{J}^*(\theta_0) \) need not equal \( \mathcal{J}(\theta_0) \). In the time series case, Gonçalves and Vogelsang (2011) show that this condition holds with \( \mathcal{J}^*(\theta_0) = \mathcal{J}(\theta_0) \) in the linear model under conventional mixing and moment conditions.

**THEOREM 3.** Suppose the i.i.d. bootstrap is used to compute \( F^* \) and \( t^* \). Suppose the conditions of Theorem 1 are satisfied and assume that Assumption 7 is also satisfied. Then
\[
F^* \Rightarrow W_q(B)' Q_q(b_1, b_2, B)^{-1} W_q(B) \quad \text{and} \quad t^* \Rightarrow \frac{W_1(B)}{\sqrt{Q_1(b_1, b_2, B)}}.
\]

\(^{16}\)We use the convention in the bootstrap literature that \( x^* \overset{p}{\to} 0 \) denotes \( \lim_N P^*([x^* > \eta] > \epsilon) = 0 \) for any \( \epsilon > 0 \) and \( \eta > 0 \) where \( P^* \) is the measure induced by the bootstrap.
Theorem 3 establishes that the i.i.d. bootstrap is a valid simulation method for obtaining critical values from the fixed-$b$ distributions given in Theorem 2.

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. Though this approximation leads to nonstandard reference distributions for $t$ and $F$-statistics, we have shown that the i.i.d. bootstrap provides valid critical values. In this section, we provide evidence on the finite-sample inference properties of tests using the proposed procedure through simulation experiments.

4.1. Linear Regression Simulation Results

For our first simulations, we generate data as

$$y_s = \alpha + x_s \beta + \varepsilon_s,$$

where $x_s$ is a scalar, $s$ is a vector $(s_1, s_2)$ indicating which lattice point the observation corresponds to, $\alpha = 0$, and $\beta = 1$. We generate $x_s$ and $\varepsilon_s$ as geometrically decaying spatial moving averages of order 15, i.e.

$$x_s = \sum_{\|j\| \leq 15} \gamma_{\|j\|} u_{s+j},$$

$$\varepsilon_s = \sum_{\|j\| \leq 15} \gamma_{\|j\|} u_{s+j}.$$
with \( \|j\| = \max(j_1, j_2) \), \( u_s \sim N(0, 1) \), \( v_s \sim N(0, 1) \), and \( u_s \) and \( v_s \) are independent for all \( i \) and \( j \). We consider four different values of \( \gamma \), \( \gamma \in \{0, -0.2, 2, 4\} \). \(^{17}\)

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 points wide. See Figure 3. This design of the two regions gives an equal number of observations \( (N = 225) \) within each region.

\(^{17}\)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.
point estimates, i.i.d. standard errors, and heteroskedasticity robust errors across the two regions. Thus, the only differences in testing performance across the two regions in the i.i.d. design come through the differences in denominators due to the use of HAC covariance estimators and the different shapes of the sampling 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.

Table 1 reports rejection rates for 5% level $t$-tests of $H_0 : \beta = 1$ with the OLS estimate $\hat{\beta}$. All results are based on 2000 simulation replications. Row labels indicate which covariance matrix estimator and smoothing parameter are used: i.i.d. and Heteroskedasticity use conventional OLS standard errors and heteroskedasticity robust standard errors, respectively. For all HAC estimators, we use a Gaussian product kernel and smoothing parameter equal to $h$, i.e.,

\[ K(s, \tilde{s}) = \exp\{-0.5[(s_1 - \tilde{s}_1)/(h/2)]^2\} \exp\{-0.5[(s_2 - \tilde{s}_2)/(h/2)]^2\} \]

Column 2 indicates from which reference distribution critical values were obtained, with “fixed-b” indicating that the approximations developed in this paper were used. All fixed-b critical values were obtained via the i.i.d. bootstrap procedure in Section 3.3 with 1000 bootstrap replications. Results for the square shaped region are shown in panel A of the table while the cross-shaped region is shown in panel B.

The results in Table 1 strongly favor the testing procedure we propose in this paper relative to the conventional Gaussian approximation. For a given positive smoothing parameter, the empirical size of a test based on the fixed-$b$ critical value is closer to the nominal 5% than a test using the same HAC estimator and the standard normal approximation across all designs considered. For small values of $h$, the improvement is generally slight. Intuitively,
one may think of the fixed-\(b\) reference distribution as capturing the variance in the HAC estimator and part of the bias. HAC estimators have two sources of bias. One is a fundamental bias that is decreasing in the bandwidth. Neither fixed-\(b\) theory nor a consistency argument captures this bias. The second bias is a mechanical bias induced when the scores are evaluated at the estimated parameters, i.e. the demeaning bias. The demeaning bias is downward and becomes more severe as the bandwidth increases.\(^{18}\) The fixed-\(b\) reference distribution captures the demeaning bias quite well. For small \(h\), the HAC estimator has small variance, small demeaning bias, but large fundamental bias, so use of the fixed-\(b\) critical values is marginally better than standard critical values. As the smoothing parameter increases, the fundamental bias falls but the variance and the demeaning bias of the HAC estimator increases. The variance and the demeaning bias dominate and are captured by the fixed-\(b\) distribution. Because the demeaning bias is downward and severe, the standard approximation exhibits over-rejection. Even with i.i.d. data (i.e., \(\gamma = 0\)), tests based on \(N(0,1)\) critical values are severely size distorted (roughly 25% nominal size when \(h = 16\)). Tests based on our fixed-\(b\) results, by contrast, have approximately correct size for large bandwidths for all but the most severe dependence we consider since the fundamental bias is increasing in the strength of dependence.

Focusing on the \(\gamma = .2\) and \(.4\) cases, we see that the two testing procedures behave very differently when the smoothing parameter is increased. With the Gaussian approximation, increasing \(h\) initially improves empirical size because the fundamental bias decreases faster than the demeaning bias increases. For large values of \(h\), however, the normal approximation to the \(t\)-statistic breaks down because the demeaning bias dominates, and size dramatically

\(^{18}\)See ? for details on the bias and variance formulas for HAC estimators.
worsens. The fixed-\(b\) tests, by contrast, almost uniformly improve as the smoothing parameter increases. Further, in all six designs featuring dependence, the empirical size of the test using the largest value of \(h\) and the fixed-\(b\) critical value is much closer to the nominal 5\% than any of the tests based on the standard normal approximation.

To better understand how our testing procedure works, we report the fixed-\(b\) critical values in Table 2. As in Table 1, row labels correspond to HAC estimators with a given value for the smoothing parameter. Two types of critical values are shown: those from the i.i.d. bootstrap, indicated by ‘boot’ in the second column, and simulated finite sample critical values, labeled FS. In both cases, the critical values are the 95\(^{\text{th}}\) percentile of the absolute value of the \(t\)-statistic and may be compared to 1.96 for the standard normal approximation. First, notice that the bootstrapped critical values are unchanged across values of \(\gamma\) since we are using a ‘naive’ or i.i.d. bootstrap. The point of the i.i.d. bootstrap is not to replicate the dependence structure in the data but to better capture the variability in the HAC estimator, \(\hat{\Omega}\).

Second, note that both the finite sample and fixed-\(b\) critical values differ according to the shape of the sampling region. This difference is extremely small for small \(h\) but is dramatic when \(h = 16\). When dependence is strong, large bandwidths need to be used to control size, and accounting for the shape of the region is crucial. The bootstrap does this. A major goal of this paper is to produce a simple, easily implemented inference procedure that reflects the different configurations in which spatial data may be observed. Our simulations suggest that the methods presented in our paper are promising in that regard.
4.2. Two Stage Least Squares Simulation Results

In this section, we provide additional simulations 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).\(^{19}\) The IM t-statistic is formed by breaking the data into \(G\) groups. A different point estimate of a parameter \(\theta\) is constructed within each group, call it \(\hat{\theta}_g\). Letting \(\bar{\theta}_G\) denote the cross group average of the within-group point estimates, \(\bar{\theta}_G = \frac{1}{G} \sum_g \hat{\theta}_g\), the IM test statistic for the null hypothesis that \(\theta = \theta_0\) is

\[
t_{IM} = \frac{\bar{\theta}_G - \theta_0}{\frac{1}{\sqrt{G}} \sqrt{\frac{1}{G-1} \sum_g (\hat{\theta}_g - \bar{\theta}_G)^2}}.
\]

Under mild regularity conditions, \(t_{IM}\) converges to a limiting t-distribution with \(G - 1\) degrees of freedom.

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

\(^{19}\)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.
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 non-explosive nonstationarity. This type of non-stationarity 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.\footnote{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.}

We report results from simulating a simple linear model with an endogenous explanatory variable. We generate instruments and errors

\[ z_{s,h} = \sum_{\|j\| \leq 15} \gamma^j w_{s+j,h} \text{ for } h = 1, 2, 3, \]
\[ \varepsilon_s = \sum_{\|j\| \leq 15} \gamma^j u_{s+j,1}, \text{ and} \]
\[ v_s = \sum_{\|j\| \leq 15} \gamma^j u_{s+j,2} \]

where \( u_{s,1} = .8u_{s,2} + .6e_s \) 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 \( x_s = \prod_{h=1}^3 z_{s,h} + \sigma(z_1, z_2, z_3) v_s \) and \( y_s = \beta_0 + \beta_1 x_s + E_s \).
We report the simulation results in Table 3. The main results are for size and size-adjusted power of 5% level tests against the alternatives that \( \beta \) is 0, 0.5, 1.5, or 2.0 based on HAC or IM. The columns labeled “HAC Smoothing Parameter” 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 “Ibragimov and Müller” 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.))\(^{21}\) and the proportion of the total number of potential cross-products that are used in forming the covariance matrix estimator (% Weight).\(^{22}\)

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

---

\(^{21}\)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.

\(^{22}\)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 = \( \frac{1}{225} \sum_{s_1=1}^{15} \sum_{s_2=1}^{15} \sum_{\tilde{s}_1=1}^{15} \sum_{\tilde{s}_2=1}^{15} K(s, \tilde{s}) \). For HAC, \( K(s, \tilde{s}) \) is defined above. For IM, \( K(s, \tilde{s}) = 1 \) whenever \( s \) and \( \tilde{s} \) belong to the same group; and \( K(s, \tilde{s}) = 0 \) otherwise.
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). This difference in bias plays a large role in the relative performance of the two procedures.

In terms of testing properties, the procedures that do the best job controlling size across the table are IM with 2 groups and HAC with a large bandwidth (12 or 16). 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 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 also see that HAC clearly dominates IM as long as one focuses on procedures with approximately correct size. The least powerful HAC procedure with a smoothing parameter of 16 has size-adjusted power comparable to the most powerful IM procedure which uses nine groups, and the HAC procedure controls size considerably better. It is also interesting that HAC performs well in the heteroskedastic design considered, though theory indicates that this is likely not generically the case.

Overall, these simulation results are very favorable to HAC using 'fixed-b' critical values. We see that HAC does well in terms of controlling size relative to IM and also tends to have more power in the designs we have 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 non-degenerate 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 show that the i.i.d. bootstrap provides a simple, feasible simulation procedure that can be used to construct valid critical values.

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 controls size well across
all simulations we consider as long as sufficiently large values of smoothing parameters are used. Simulations also show that appropriate critical values for tests can differ substantially depending on the shape of the sampling region.

We believe the results in this paper will be very useful to researchers using common economic data in which spatial correlation is likely. While the fixed-$b$ asymptotic distributions are complicated, the ability to simply obtain critical values via the i.i.d. bootstrap should make using the approximation accessible to a wide range of researchers. One important question that deserves more attention 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.

**Appendix A. Proofs**

A.1. Lemmas

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

$$\sum_{i,j \in B_n} s_{i,j}(\theta_0) + \Delta_{B_n}(\hat{\theta}_N, \theta_0, \delta_{B_n})(\hat{\theta}_N - \theta_0) = 0,$$
where $\Delta_{B_n}(\hat{\theta}_N, \theta_0, \delta_{B_n})$ is a $p \times p$ matrix with $l^{th}$ row given by the $l^{th}$ row of the matrix
\[
\left( \sum_{i,j \in B_n} \frac{\partial s_{i,j}(\theta)}{\partial \theta^l} \right)_{|_{\theta = \hat{\theta}_B^{(l)}}} = \delta_{B_n}^{(l)} \theta_0 + (1 - \delta_{B_n}^{(l)}) \hat{\theta}_N
\]
where $\delta_{B_n}^{(l)}$ is the $l^{th}$ element of the vector $\delta_{B_n}$ and $0 \leq \delta_{B_n}^{(l)} \leq 1$. It follows that
\[
\frac{1}{n} \sum_{i,j \in B_n} s_{i,j}(\theta_0) = - \left( \frac{1}{n^2} \Delta_{B_n}(\hat{\theta}_N, \theta_0, \delta_{B_n}) \right)^{-1} \frac{1}{n} \sum_{i,j \in B_n} s_{i,j}(\theta_0).
\]

Now note that
\[
\frac{1}{n} \sum_{i,j \in B_n} s_{i,j}(\theta_0) = \frac{1}{n} \sum_{i=-n/2}^{n/2} \sum_{j=-n/2}^{n/2} 1_{i,j} s_{i,j}(\theta_0)
\]
\[
= \frac{1}{n} \sum_{i=-n/2}^{n/2} \sum_{j=-n/2}^{n/2} \lambda_{i,j}^{n} s_{i,j}(\theta_0) + \frac{1}{n} \sum_{i=-n/2}^{n/2} \sum_{j=-n/2}^{n/2} (1_{i,j} - \lambda_{i,j}^{n}) s_{i,j}(\theta_0),
\]
where $1_{i,j}$ is an indicator which equals one when an observation occurs at lattice point $(i, j)$, $\lambda_{i,j}^{n} = \lambda(B_n \cap \{i-1, i\times[j-1, j]\})$, and $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}^2$ as in Dedecker (2001).

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

For the second term, we define $dB^+ = \{(i, j) : [(i, j) \in B_n \text{ and } (i - 1, j) \notin B_n \text{ or } (i, j - 1) \notin B_n \text{ or } (i - 1, j - 1) \notin B_n] \text{ or } [(i, j) \notin B_n \text{ and } (i - 1, j) \in B_n \text{ or } (i, j - 1) \in B_n \text{ or } (i - 1, j - 1) \in B_n] \}$. 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
\[
\frac{1}{n} \sum_{i=-n/2}^{n/2} \sum_{j=-n/2}^{n/2} (1_{i,j} - \lambda_{i,j}^{n}) s_{i,j}(\theta_0) = \frac{1}{n} \sum_{i,j \in dB^+} (1_{i,j} - \lambda_{i,j}^{n}) s_{i,j}(\theta_0).
\]

We then have
\[
E\left\| \frac{1}{n} \sum_{i,j \in dB^+} (1_{i,j} - \lambda_{i,j}^{n}) s_{i,j}(\theta_0) \right\|^2 \leq \frac{1}{n^2} \sum_{i,j \in dB^+} \sum_{k,l \in dB^+} |1_{i,j} - \lambda_{i,j}^{n}| |1_{k,l} - \lambda_{k,l}^{n}| E|s_{i,j}(\theta_0) s_{k,l}(\theta_0)'| |
\]
\[
\leq \frac{1}{n^2} Cn \rightarrow 0
\]
The first equality 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 $d^\text{th}$-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

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

from Assumptions 2 and 6.

Combining these results gives

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

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

$$\sqrt{N}(\hat{\theta}_N - \theta_0) = \sqrt{\frac{N}{n^2}} n(\hat{\theta}_N - \theta_0) \Rightarrow (-\lambda(B)^{1/2} J(\theta_0))^{-1} \Lambda W_p(B). \quad \blacksquare$$

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

**Lemma 2.** Suppose Assumptions 4, 6 hold. Then, as $N \to \infty$,

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

for $A \subseteq B$. 

36
**Proof of Lemma 2:** Note that
\[
\frac{1}{n} \sum_{i,j \in A_n} \hat{s}_{i,j} = \frac{1}{n} \sum_{i,j \in A_n} s_{i,j}(\theta_0) + \frac{1}{n^2} \Delta_{A_n}(\hat{\theta}_N, \theta_0, \delta_{A_n}) n(\hat{\theta}_N - \theta_0),
\]
where \( \Delta_{A_n}(\hat{\theta}_N, \theta_0, \delta_{A_n}) \) is defined similarly to \( \Delta_{B_n}(\hat{\theta}_N, \theta_0, \delta_{B_n}) \). 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 \( \Delta_{B_n}(\hat{\theta}_N, \theta_0, \delta_{B_n}) \), we have
\[
\frac{1}{n^2} \Delta_{A_n}(\hat{\theta}_N, \theta_0, \delta_{A_n}) \overset{p}{\rightarrow} \lambda(A) J(\theta_0).
\]
Plugging these two results in and using Lemma 1 then gives
\[
\frac{1}{n} \sum_{i,j \in A_n} \hat{s}_{i,j} = \frac{1}{n} \sum_{i,j \in A_n} s_{i,j}(\theta_0) + \frac{1}{n^2} \Delta_{A_n}(\hat{\theta}_N, \theta_0, \delta_{A_n}) \sqrt{\frac{n^2}{N}} \sqrt{N}(\hat{\theta}_N - \theta_0)
\Rightarrow \Lambda W_p(A) - \lambda(A) J(\theta_0) \lambda(B)^{-1/2} (\lambda(B)^{1/2} J(\theta_0))^{-1} \Lambda W_p(B)
\equiv \Lambda \left( W_p(A) - \frac{\lambda(A)}{\lambda(B)} W_p(B) \right).
\]

The next lemma provides a representation of sums over QMB regions in terms of partial sums over sub-regions. 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:

\[ A_1 = \{ i, j : 0 \leq i \leq M_1(j + 1) - 1, \ 0 \leq j \leq \bar{N}_1 - 1 \}, \]
\[ A_2 = \{ i, j : i = M_1(j + 1), \ 0 \leq j \leq \bar{N}_1 - 1 \}, \]
\[ A_3 = \{ i, j : 0 \leq i \leq M_1 - 1, j = \bar{N}_1(i) \}, \]
\( A_4 = \{ i, j : i = \bar{M}_1, j = \bar{N}_1(\bar{M}_1) \} \),

\( A_5 = \{ i, j : 0 \leq i \leq \bar{M}_2 - 1, 1 \leq j \leq \bar{N}_2(i + 1) - 1 \} \),

\( A_6 = \{ i, j : i = 0 \leq i \leq \bar{M}_2 - 1, j = \bar{N}_2(i + 1) \} \),

\( A_7 = \{ i, j : i = \bar{M}_2(j), 1 \leq j \leq \bar{N}_2 - 1 \} \),

\( A_8 = \{ i, j : i = \bar{M}_2(\bar{N}_2), j = \bar{N}_2 \} \),

\( A_9 = \{ i, j : 1 \leq i \leq \bar{M}_3 - 1, 0 \leq j \leq \bar{N}_3(i + 1) - 1 \} \),

\( A_{10} = \{ i, j : i = 1 \leq i \leq \bar{M}_3 - 1, j = \bar{N}_3(i + 1) \} \),

\( A_{11} = \{ i, j : i = \bar{M}_3(j), 0 \leq j \leq \bar{N}_3 - 1 \} \),

\( A_{12} = \{ i, j : i = \bar{M}_3(\bar{N}_3), j = \bar{N}_3 \} \),
\[ A_{13} = \{ i, j : 1 \leq i \leq \bar{M}_4(j + 1) - 1, \ 1 \leq j \leq \bar{N}_4 - 1 \} , \]
\[ A_{14} = \{ i, j : i = \bar{M}_4(j + 1), \ 1 \leq j \leq \bar{N}_4 - 1 \} , \]
\[ A_{15} = \{ i, j : 1 \leq i \leq \bar{M}_4 - 1, j = \bar{N}_4(i) \} , \]
\[ A_{16} = \{ i, j : i = \bar{M}_4, j = \bar{N}_4(\bar{M}_4) \} . \]

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

\[
D^{i,j,1}W(i, j, k, l) = (W(i, j, k, l) - W(i, j + 1, k, l))
\]
\[
- (W(i + 1, j, k, l) - W(i + 1, j + 1, k, l))
\]
\[
D^{i,j,2}W(i, j, k, l) = W(\bar{M}_1(j + 1), j, k, l) - W(\bar{M}_1(j + 1), j + 1, k, l)
\]
\[
D^{i,j,3}W(i, j, k, l) = W(i, \bar{N}_1(i), k, l) - W(i + 1, \bar{N}_1(i + 1), k, l)
\]
\[
D^{i,j,4}W(i, j, k, l) = W(\bar{M}_1, \bar{N}_1(\bar{M}_1), k, l)
\]
\[
D^{i,j,5}W(i, j, k, l) = (W(i, -j, k, l) - W(i + 1, -j, k, l))
\]
\[
- (W(i, -(j + 1), k, l) - W(i + 1, -(j + 1), k, l))
\]
\[
D^{i,j,6}W(i, j, k, l) = W(i, -\bar{N}_2(i + 1), k, l) - W(i + 1, -\bar{N}_2(i + 1), k, l)
\]
\[
D^{i,j,7}W(i, j, k, l) = W(\bar{M}_2(j), -j, k, l) - W(\bar{M}_2(j + 1), -(j + 1), k, l)
\]
\[
D^{i,j,8}W(i, j, k, l) = W(\bar{M}_2(\bar{N}_2), -\bar{N}_2, k, l)
\]
\[
D^{i,j,9}W(i, j, k, l) = (W(-i, j, k, l) - W(-i + 1, j, k, l))
\]
\[
- (W(-i, j + 1, k, l) - W(-i + 1, j + 1, k, l))
\]
\[
D^{i,j,10}W(i, j, k, l) = W(-i, \bar{N}_3(i + 1), k, l) - W(-i + 1, \bar{N}_3(i + 1), k, l)
\]
\[
D^{i,j,11}W(i, j, k, l) = W(-\bar{M}_3(j), j, k, l) - W(-\bar{M}_3(j + 1), j + 1, k, l)
\]
\[ D_{i,j,12} W(i, j, k, l) = W(-\bar{M}_3(N_3), \bar{N}_3, k, l) \]

\[ D_{i,j,13} W(i, j, k, l) = (W(-i, -j, k, l) - W(-i, -(j + 1), k, l)) \]

\[ - (W(-(i + 1), -j, k, l) - W(-(i + 1), -(j + 1), k, l)) \]

\[ D_{i,j,14} W(i, j, k, l) = W(-\bar{M}_4(j + 1), -j, k, l) - W(-\bar{M}_4(j + 1), -(j + 1), k, l) \]

\[ D_{i,j,15} W(i, j, k, l) = W(-i, -\bar{N}_4(i), k, l) - W(-(i + 1), -\bar{N}_4(i + 1), k, l) \]

\[ D_{i,j,16} W(i, j, k, l) = W(-\bar{M}_4, -\bar{N}_4(\bar{M}_4), k, l). \]

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:

\[ b^1(i, j) = \sum_{p=0}^{i} \sum_{q=0}^{j} b_{p,q}, \quad b^2(i, j) = \sum_{p=0}^{M_1(j + 1)} \sum_{q=0}^{j} b_{p,q} \]

\[ b^3(i, j) = \sum_{p=0}^{i} \sum_{q=0}^{N_1(p)} b_{p,q}, \quad b^4(i, j) = \sum_{p=0}^{M_1} \sum_{q=0}^{N_1(p)} b_{p,q} \]

\[ b^5(i, j) = \sum_{p=0}^{i} \sum_{q=1}^{j} b_{p,-q}, \quad b^6(i, j) = \sum_{q=1}^{N_2(i+1)} \sum_{p=0}^{i} b_{p,-q} \]

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

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

\[ b^{11}(i, j) = \sum_{q=0}^{\bar{M}_3(q)} \sum_{p=1}^{j} b_{p,q}, \quad b^{12}(i, j) = \sum_{p=0}^{\bar{N}_3} \sum_{q=0}^{\bar{M}_3(q)} b_{p,q} \]

\[ b^{13}(i, j) = \sum_{p=1}^{i} \sum_{q=1}^{j} b_{p,-q}, \quad b^{14}(i, j) = \sum_{p=1}^{\bar{M}_4(j+1)} \sum_{q=1}^{j} b_{p,-q} \]
\[ b^{15}(i, j) = \sum_{p=1}^{i} \sum_{q=1}^{\mathcal{N}_4(p)} b_{p-q, q} \quad b^{16}(i, j) = \sum_{p=1}^{M_4} \sum_{q=1}^{\mathcal{N}_4(p)} b_{p-q, q} \]

**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.

**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
\[
\sum_{i,j \in A} a_{i,j} b_{i,j} = \sum_{i=0}^{L} \sum_{j=0}^{\mathcal{M}(i)} a_{i,j} b_{i,j} \\
= \sum_{j=0}^{\mathcal{M}(i)} \sum_{i=0}^{L-1} [(a_{i,j} - a_{i,j+1}) - (a_{i+1,j} - a_{i+1,j+1})] \sum_{p=0}^{L} \sum_{q=0}^{J} b_{p,q} \\
+ \sum_{j=0}^{L-1} (a_{L(j+1),j} - a_{L(j+1),j+1}) \sum_{p=0}^{L} \sum_{q=0}^{J} b_{p,q} \\
+ \sum_{i=0}^{\mathcal{M}(i)} (a_{i,\mathcal{M}(i)} - a_{i+1,\mathcal{M}(i)+1}) \sum_{p=0}^{\mathcal{M}(p)} \sum_{q=0}^{J} b_{p,q} \\
+ a_{L,\mathcal{M}(i)} \sum_{p=0}^{L} \sum_{q=0}^{J} b_{p,q}.
\]
for any region \( A \) that satisfies \( \sum_{i,j \in A} f_{i,j} = \sum_{i=0}^{L} \sum_{j=0}^{M(i)} f_{i,j} = \sum_{j=0}^{L(j)} \sum_{i=0}^{M(j)} f_{i,j} \). 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 \( \hat{s}^x(i, j) \) for \( x = 1, \ldots, 16 \) be defined as the partial sums in Lemma 3. Then

\[
\frac{1}{n} \hat{s}^x(i, j) = W_p(A^x(u_1, u_2), B)
\]

for

\[
A^1(u_1, u_2) = ([0, u_1] \times [0, u_2]) \cap B_1, \quad A^2(u_1, u_2) = ([0, f_1(u_2)] \times [0, u_2]) \cap B_1
\]

\[
A^3(u_1, u_2) = ([0, u_1] \times [0, g_1]) \cap B_1, \quad A^4(u_1, u_2) = B_1
\]

\[
A^5(u_1, u_2) = ([0, u_1] \times [0, u_2]) \cap B_2, \quad A^6(u_1, u_2) = ([0, u_1] \times [g_2(u_1), 0]) \cap B_2
\]

\[
A^7(u_1, u_2) = ([0, \bar{f}_2] \times [0, u_2]) \cap B_2, \quad A^8(u_1, u_2) = B_2
\]

\[
A^9(u_1, u_2) = ([u_1, 0] \times [0, u_2]) \cap B_3, \quad A^{10}(u_1, u_2) = ([u_1, 0] \times [0, g_3(-u_1)]) \cap B_3
\]

\[
A^{11}(u_1, u_2) = ([u_1, 0] \times [0, u_2]) \cap B_3, \quad A^{12}(u_1, u_2) = B_3
\]

\[
A^{13}(u_1, u_2) = ([u_1, 0] \times [0, u_2]) \cap B_4, \quad A^{14}(u_1, u_2) = ([u_1, 0] \times [0, -u_2, 0]) \cap B_4
\]

\[
A^{15}(u_1, u_2) = ([u_1, 0] \times [0, g_4, 0]) \cap B_4, \quad A^{16}(u_1, u_2) = B_4
\]

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 \( \bar{f}_2 = \sup_{u_2 \in [0,1/2]} f_2(-u_2); \) \( -u_2 = -g_2(u_1) \) 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

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

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

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

(A.1)

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

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 provide 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
\[ T_{1,1} = \frac{n^2 N}{N} \sum_{i=0}^{\hat{N}_1-1} \sum_{k=0}^{\hat{N}_1(i+1)-1} \sum_{j=0}^{\hat{N}_1(j+1)-1} \left[ D^{k,l,1} \left( D^{i,j,1} K \left( \frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n} \right) \right) \right] \]

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

Observe that
\[
n^4 \left[ D^{k,l,1} \left( D^{i,j,1} K \left( \frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n} \right) \right) \right] \rightarrow \frac{1}{b_1^2 b_2^2} \frac{\partial K(x_1, x_2, x_3, x_4)}{\partial x_1 \partial x_2 \partial x_3 \partial x_4} \bigg|_{x_1 = u_1/b_1, x_2 = u_2/b_2, x_3 = v_1/b_1, x_4 = v_2/b_2} = K_{b_1, b_2}^4 (u_1, u_2, v_1, v_2). \]
Using Assumption \(2\) Lemma 4, and the continuous mapping theorem, we have that

\[
T_{1,1} = \frac{n^2}{N} \int \sum_{l=0}^{N} \sum_{k=0}^{l} \sum_{j=0}^{k} \sum_{i=0}^{j} n^4 \left[ D^{i,j,1}_1 \left( D^{i,j,1}_1 K \left( \frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n} \right) \right) \right] 
\]

\[
\left( \sum_{p=0}^{i} \sum_{q=0}^{j} \tilde{s}_{p,q}/n \right) \left( \sum_{r=0}^{k} \sum_{t=0}^{l} \tilde{s}_{r,t}/n \right) 
\]

\( \Rightarrow \Lambda Q_{1}^{1}(b_1, b_2, B) \Lambda' \)

where

\[
Q_{1}^{1}(b_1, b_2, B) = \frac{1}{\lambda(B)} \int_{u \in B_1} \int_{v \in B_1} K^{4}_{b_1,b_2}(u,v)W_p(A^{1}(u), B)W_p(A^{1}(v), B)dvdu,
\]

\( u = (u_1, u_2), v = (v_1, v_2), \) and \( \int_{u \in B_1} h(u)du \equiv \int_{0}^{f_1} \int_{0}^{g_1(u_1)} h(u)du du_1. \) 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

\[
T_{1,5} \Rightarrow \Lambda Q_{p}^{1.5}(b_1, b_2, B) \Lambda', \quad T_{1,9} \Rightarrow \Lambda Q_{p}^{1.9}(b_1, b_2, B) \Lambda', \quad \text{and} \quad T_{1,13} \Rightarrow \Lambda Q_{p}^{1.13}(b_1, b_2, B) \Lambda'
\]

where

\[
Q_{p}^{1.5}(b_1, b_2, B) = -\frac{1}{\lambda(B)} \int_{u \in B_1} \int_{v \in B_2} K^{4}_{b_1,b_2}(u,v_1,-v_2)W_p(A^{1}(u), B)W_p(A^{5}(v), B)dvdu
\]

\[
Q_{p}^{1.9}(b_1, b_2, B) = -\frac{1}{\lambda(B)} \int_{u \in B_1} \int_{v \in B_3} K^{4}_{b_1,b_2}(u,-v_1,v_2)W_p(A^{1}(u), B)W_p(A^{9}(v), B)dvdu
\]

\[
Q_{p}^{1.13}(b_1, b_2, B) = \frac{1}{\lambda(B)} \int_{u \in B_1} \int_{v \in B_4} K^{4}_{b_1,b_2}(u,-v)W_p(A^{1}(u), B)W_p(A^{13}(v), B)dvdu
\]

with \( \int_{u \in B_2} h(u)du \equiv \int_{0}^{f_2} \int_{0}^{g_2(u_1)} h(u)du du_1, \) \( \int_{u \in B_3} h(u)du \equiv \int_{0}^{f_3} \int_{0}^{g_3(u_1)} h(u)du du_1, \) and \( \int_{u \in B_4} h(u)du \equiv \int_{0}^{f_4} \int_{0}^{g_4(u_1)} h(u)du du_1. \) 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

\[(A.2) \quad Q_{p,1}^1(b_1, b_2, B) \equiv Q_{p,1}^1(b_1, b_2, B) + Q_{p,5}^1(b_1, b_2, B) + Q_{p,9}^1(b_1, b_2, B) + Q_{p,13}^1(b_1, b_2, B) \equiv \frac{1}{\lambda(B)} \int_{u \in B} \int_{v \in B} K_{b_1,b_2}^4(u,v) W_p(A^1(u), B) W_p(R(v), B) d\nu d\mu \]

where

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

Using similar definitions and arguments, one can define

\[(A.4) \quad Q_{p,2}^1(b_1, b_2, B) \equiv Q_{p,5}^5(b_1, b_2, B) + Q_{p,9}^5(b_1, b_2, B) + Q_{p,13}^5(b_1, b_2, B)\]

\[(A.5) \quad Q_{p,3}^1(b_1, b_2, B) \equiv Q_{p,9}^9(b_1, b_2, B) + Q_{p,9}^9(b_1, b_2, B) + Q_{p,9}^9(b_1, b_2, B) + Q_{p,13}^9(b_1, b_2, B)\]

\[(A.6) \quad Q_{p,4}^1(b_1, b_2, B) \equiv Q_{p,13}^{13}(b_1, b_2, B) + Q_{p,13}^{13}(b_1, b_2, B) + Q_{p,13}^{13}(b_1, b_2, B) + Q_{p,13}^{13}(b_1, b_2, B).\]

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

\[(A.7) \quad L_1(b_1, b_2, B) = \frac{1}{\lambda(B)} \int_{u \in B} \int_{v \in B} K_{b_1,b_2}^4(u,v) W_p(R(u), B) W_p(R(v), B) d\nu d\mu.\]

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 \(x = 3\):

\[
T_{4,3} = \frac{n^2}{N} \sum_{i=0}^{N_1-1} \left[ D^{k,l,4} \left( D^{i,j,3} K \left( \frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n} \right) \right) \right] \\
\left( \sum_{p=0}^{N_1} \sum_{q=0}^{N_1} \tilde{s}_{p,q} \right) \left( \sum_{r=0}^{N_1} \sum_{t=0}^{N_1} \tilde{s}_{r,t} \right)
\]

46
where

\[
D_{k,l,4} \left( D_{i,j,3} K \left( \frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n} \right) \right) \\
= K \left( \frac{i}{b_1 n}, \frac{N_1(i)}{b_2 n}, \frac{M_1}{b_1 n}, \frac{N_1(M_1)}{b_2 n} \right) - K \left( \frac{i+1}{b_1 n}, \frac{N_1(i+1)}{b_2 n}, \frac{M_1}{b_1 n}, \frac{N_1(M_1)}{b_2 n} \right).
\]

Once again, observe that

\[
nD_{k,l,4} \left( D_{i,j,3} K \left( \frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n} \right) \right)
\rightarrow - \frac{1}{b_1} \frac{\partial K}{\partial x_1} \left( \frac{u_1}{b_1}, \frac{g_1(u_1)}{b_2}, \frac{f_1}{b_1}, 0 \right) - \frac{1}{b_2} \frac{\partial K}{\partial x_2} \left( \frac{u_1}{b_1}, \frac{g_1(u_1)}{b_2}, \frac{f_1}{b_1}, 0 \right) \frac{\partial g_1(z)}{\partial z} \bigg|_{z=\tilde{u}_1}
\]

\[
\equiv - \frac{\partial K}{\partial t_{b_1,b_2}(u)} \left( \frac{u_1}{b_1}, \frac{g_1(u_1)}{b_2}, \frac{f_1}{b_1}, 0 \right)
\]

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

\[
T_{4,3} = n^2 \frac{1}{N} \frac{1}{n} \sum_{i=0}^{N_1-1} n \left[ D_{k,l,4} \left( D_{i,j,3} K \left( \frac{i}{b_1 n}, \frac{j}{b_2 n}, \frac{k}{b_1 n}, \frac{l}{b_2 n} \right) \right) \right]
\]

\[
\left( \sum_{p=0}^{M_1} \sum_{q=0}^{N_1(p)} \frac{\tilde{s}_{p,q}}{n} \right) \left( \sum_{r=0}^{M_1} \sum_{t=0}^{N_1(r)} \frac{\tilde{s}'_{r,t}}{n} \right)
\]

\[
\Rightarrow \Lambda Q_{p,T}^{4,3}(b_1, b_2, B) \Lambda'
\]

where

\[
Q_{p,T}^{4,3}(b_1, b_2, B) = - \frac{1}{\lambda(B)} \int_0^{f_1} \frac{\partial K}{\partial t_{b_1,b_2}(u)} \left( \frac{u_1}{b_1}, \frac{g_1(u_1)}{b_2}, \frac{f_1}{b_1}, 0 \right) W_p(A^3(u_1, u_2), B) W_p(B_1, B) du_1
\]

Using Assumption 2, Lemma 4, and the continuous mapping theorem.
Define the set

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

Also, let

\[(A.9)\]
\[\int_{v \in \partial B^-} h(v) dv = \int_{v \in B_1 : v_1 \geq 0, v_2 = g_1(v_1)} h(v) dv + \int_{v \in B_2 : v_1 = f_2(-v_2), v_1 \leq 0} h(v) dv + \int_{v \in B_3 : v_1 = -f_3(v_2), v_2 \geq 0} h(v) dv + \int_{v \in B_4 : v_1 \leq 0, v_2 = -g_4(-v_1)} h(v) dv.\]

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 yields

\[(A.10)\]
\[Q_p^{1,\text{Type}5}(b_1, b_2, B) \equiv Q_p^{4,3}(b_1, b_2, B) + Q_p^{4,7}(b_1, b_2, B) + Q_p^{4,11}(b_1, b_2, B) + Q_p^{4,13}(b_1, b_2, B) \equiv -\frac{1}{\lambda(B)} \int_{u \in \partial B^-} \frac{\partial K}{\partial t_{b_1, b_2}(u)} \left( \frac{u_1}{b_1}, \frac{u_2}{b_2}, \frac{\bar{f}_1}{b_1}, 0 \right) W_p(P(u_1, u_2), B) W_p(B_1, B)^d u.\]

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 \((A.9):\)

\[(A.11)\]
\[\int_{v \in \partial B^+} h(v) dv = \int_{v \in B_1 : v_1 = f_1(v_2), v_2 \geq 0} h(v) dv + \int_{v \in B_2 : v_1 \geq 0, v_2 = -g_2(v_1)} h(v) dv + \int_{v \in B_3 : v_1 \leq 0, v_2 = g_3(-v_2)} h(v) dv + \int_{v \in B_4 : v_1 = -f_4(-v_2), v_2 \leq 0} h(v) dv,\]
and
\[(A.12) \quad \vec{v}_1 = (\bar{f}_1, 0), \quad \vec{v}_2 = (0, -\bar{g}_2), \quad \vec{v}_3 = (0, \bar{g}_3), \quad \vec{v}_4 = (-\bar{f}_4, 0).\]

Given these objects and following the steps above, we have that \(\hat{\Omega} \Rightarrow \lambda(\Lambda Q_p(b_1, b_2, B)\Lambda)\) where \(Q_p(b_1, b_2, B) = L_{p, I, I}(b_1, b_2, B) + L_{p, I, II}(b_2, b_2, B) + \ldots + L_{p, IV, IV}(b_1, b_2, B)\) where \(L_{p, -}\) 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 \(W_p(A, B), \Lambda, J\) 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 \(\Lambda, J\) and \(R(\theta_0)\) cancelling from the representations.

**A.4. Proof of Theorem 3.**

The first step of the proof is to show that a FCLT holds for the i.i.d. bootstrap. To establish a FCLT for the bootstrap, recall that \(s_{i,j}^{**} = s_{i,j}^{**}(\hat{\theta}_N)\) and \(E^* (s_{i,j}^{**}(\hat{\theta}_N)) = 0\). Because of i.i.d. resampling, the variance covariance matrix of \(s_{i,j}^{**}(\hat{\theta}_N)\) is
\[
E^*[s_{i,j}^{**}(\hat{\theta}_N)s_{i,j}^{**}(\hat{\theta}_N)'] = N^{-1} \sum_{i,j \in B_n} s_{i,j}^{**}(\hat{\theta}_N)s_{i,j}^{**}(\hat{\theta}_N) = \hat{\Gamma}_0^*.
\]
Define the matrix \(\hat{\lambda}_0^*\) as the matrix square root of \(\hat{\Gamma}_0^*\), i.e. \(\hat{\Gamma}_0^* = \hat{\lambda}_0^*\hat{\lambda}_0^{*'}\). Then the random field \(\hat{\lambda}_0^{-1}s_{i,j}^{**}(\hat{\theta}_N)\) is stationary and mean zero with an identity variance covariance matrix in.
which case Assumption 3 holds and \( \hat{\lambda}_0 s_{i,j}^* (\hat{\theta}_N) \) satisfies a FCLT of the form

\[
\frac{1}{n} \sum_{(i,j) \in \mathbb{Z}^2} \lambda(A_n \cap R_{(i,j)}) \hat{\lambda}_0^{-1} s_{i,j}^{**}(\hat{\theta}_N) \Rightarrow W_p(A)
\]

where \( W_p \) is a set-indexed Brownian motion indexed by \( A \) and \( R_{(i,j)} = [i-1, i] \times [j-1, j] \) is the unit square with upper corner at \((i, j)\). This FCLT holds for subsets of \( B \) because the bootstrap preserves the region \( B \) and satisfies Assumptions 1–2. Note that \( \hat{\Gamma}_0^* \) is the sample variance of \( s_{i,j}^* (\theta_0) \). Given that \( \hat{\theta}_N \) is assumed to be a consistent estimator and given that FCLTs for random fields require stationarity of the underlying random field, it follows that \( \hat{\Gamma}_0^* = \lambda_0 \lambda_0' = \Gamma_0 \)

where \( \Gamma_0 \) is the variance matrix of \( s_{i,j}^* (\theta_0) \). Therefore we have

(A.13) \[
\frac{1}{n} \sum_{i,j \in A_n} s_{i,j}^{**}(\hat{\theta}_N) = \hat{\lambda}_0 s_{i,j}^{**}(\hat{\theta}_N) \Rightarrow \Lambda^* W_p(A)
\]

where \( \Lambda^* = \lambda_0 \).

The next step is to derive the limit of \( \sqrt{N} \left( \hat{\theta}^* - \hat{\theta}_N \right) \). Using the mean value theorem we have

\[
\sum_{i,j \in B_n} s_{i,j}^{**}(\hat{\theta}_N) + \Delta_{B_n}(\hat{\theta}^*, \hat{\theta}_N, B_n)(\hat{\theta}^* - \hat{\theta}_N) = 0,
\]

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

\[
n(\hat{\theta}^* - \hat{\theta}_N) = -\left( \frac{1}{n^2} \Delta_{B_n}(\hat{\theta}^*, \hat{\theta}_N, B_n) \right)^{-1} \frac{1}{n} \sum_{i,j \in B_n} s_{i,j}^{**}(\hat{\theta}_N).
\]
Under Assumption 7 (i) $\hat{\theta}^* - \hat{\theta}_N \overset{p}{\rightarrow} 0$ and (ii)

$$\sup_{A \in A, \theta \in \mathcal{N}(\theta_0)} \left\| \frac{1}{N} \sum_{(i,j) \in \pi A} \frac{\partial s_{i,j}^{**}(\theta)}{\partial \theta^v} - \frac{\lambda(A)}{\lambda(B)} \mathcal{J}^*(\theta) \right\| \overset{p}{\rightarrow} 0,$$

(A.14) $\frac{1}{n^2} \Delta_{B_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{B_n}^*) \overset{p}{\rightarrow} \lambda(B)\mathcal{J}^*(\theta_0).$

The limit of $\sqrt{N} \left( \hat{\theta}^* - \hat{\theta}_N \right)$ now easily follows using (A.13) and (A.14):

$$\sqrt{N} \left( \hat{\theta}^* - \hat{\theta}_N \right) = \sqrt{\frac{N}{n^2} n^2} \left( \hat{\theta}^* - \hat{\theta}_N \right) = -\sqrt{\frac{N}{n^2}} \left( \frac{1}{n^2} \Delta_{B_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{B_n}^*) \right) - \frac{1}{n} \sum_{i,j \in B_n} s_{i,j}^{**}(\hat{\theta}_N)$$

(A.15) $\Rightarrow \lambda(B)^{1/2} (\mathcal{J}^*(\theta_0) \lambda(B))^{-1} \Lambda^* W_p(B) \equiv (\lambda(B)^{1/2} \mathcal{J}^*(\theta_0))^{-1} \Lambda^* W_p(B).$

The third step of the proof is to derive the limit of $n^{-1} \sum_{i,j \in A_n} s_{i,j}^{**}(\hat{\theta}^*)$. Using a mean value theorem we can write

$$n^{-1} \sum_{i,j \in A_n} s_{i,j}^{**}(\hat{\theta}^*) = n^{-1} \sum_{i,j \in A_n} s_{i,j}^{**}(\hat{\theta}_N) + n^{-1} \Delta_{A_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{A_n}^*) \left( \hat{\theta}^* - \hat{\theta}_N \right),$$

where $\Delta_{A_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{A_n}^*)$ is defined analogously to $\Delta_{B_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{B_n}^*)$. Again under Assumption 7

(A.16) $\frac{1}{n^2} \Delta_{A_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{A_n}^*) = \frac{N}{n^2} \frac{1}{N} \Delta_{A_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{A_n}^*) \overset{p}{\rightarrow} \lambda(A)\mathcal{J}^*(\theta_0).$

Using (A.13), (A.15) and (A.16) the limit follows as

$$n^{-1} \sum_{i,j \in A_n} s_{i,j}^{**}(\hat{\theta}^*) = n^{-1} \sum_{i,j \in A_n} s_{i,j}^{**}(\hat{\theta}_N) + \frac{1}{n^2} \Delta_{A_n}(\hat{\theta}^*, \hat{\theta}_N, \delta_{A_n}^*) \sqrt{\frac{n^2}{N}} \sqrt{N} \left( \hat{\theta}^* - \hat{\theta}_N \right)$$

$$\Rightarrow \Lambda^* W_p(A) - \lambda(A) \mathcal{J}^*(\theta_0) \lambda(B)^{-1/2} \left( \lambda(B)^{1/2} \mathcal{J}^*(\theta_0) \right)^{-1} \Lambda^* W_p(B)$$

$$= \Lambda^* W_p(A) - \frac{\lambda(A)}{\lambda(B)} \Lambda^* W_p(B) \equiv \Lambda^* W_p(A, B).$$
The remainder of the proof follows closely the proofs of Theorems 1-2 and details are omitted. Note that we do not require $\Lambda^* = \Lambda$ because $\Lambda^*$ cancels from the asymptotic distribution of the bootstrap statistic in the same way that $\Lambda$ cancels from the asymptotic distribution of the actual statistic. This is the sense in which we do not require the bootstrap to capture the dependence in the data.

References


Table 1. Empirical Size (%) of Nominal 95% Level $t$-Tests

<table>
<thead>
<tr>
<th>Ref. Dist.</th>
<th>$\Upsilon = 0$</th>
<th>$\Upsilon = -0.2$</th>
<th>$\Upsilon = 0.2$</th>
<th>$\Upsilon = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Square</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IID</td>
<td>N(0,1)</td>
<td>4.8</td>
<td>12.9</td>
<td>22.4</td>
</tr>
<tr>
<td>Hetero</td>
<td>N(0,1)</td>
<td>5.3</td>
<td>13.2</td>
<td>23.5</td>
</tr>
<tr>
<td>Gauss(2)</td>
<td>N(0,1)</td>
<td>5.8</td>
<td>10.9</td>
<td>14.5</td>
</tr>
<tr>
<td>Gauss(2)</td>
<td>Fixed-b</td>
<td>4.7</td>
<td>9.4</td>
<td>12.5</td>
</tr>
<tr>
<td>Gauss(4)</td>
<td>N(0,1)</td>
<td>7.9</td>
<td>10.1</td>
<td>11.9</td>
</tr>
<tr>
<td>Gauss(4)</td>
<td>Fixed-b</td>
<td>5.0</td>
<td>7.0</td>
<td>9.1</td>
</tr>
<tr>
<td>Gauss(8)</td>
<td>N(0,1)</td>
<td>14.4</td>
<td>15.4</td>
<td>17.0</td>
</tr>
<tr>
<td>Gauss(8)</td>
<td>Fixed-b</td>
<td>5.1</td>
<td>6.3</td>
<td>7.3</td>
</tr>
<tr>
<td>Gauss(16)</td>
<td>N(0,1)</td>
<td>28.7</td>
<td>31.8</td>
<td>32.9</td>
</tr>
<tr>
<td>Gauss(16)</td>
<td>Fixed-b</td>
<td>5.3</td>
<td>5.1</td>
<td>6.5</td>
</tr>
<tr>
<td><strong>Panel B: Cross</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IID</td>
<td>N(0,1)</td>
<td>4.8</td>
<td>12.1</td>
<td>22.6</td>
</tr>
<tr>
<td>Hetero</td>
<td>N(0,1)</td>
<td>5.3</td>
<td>12.7</td>
<td>22.9</td>
</tr>
<tr>
<td>Gauss(2)</td>
<td>N(0,1)</td>
<td>5.6</td>
<td>9.9</td>
<td>14.8</td>
</tr>
<tr>
<td>Gauss(2)</td>
<td>Fixed-b</td>
<td>5.0</td>
<td>8.8</td>
<td>13.5</td>
</tr>
<tr>
<td>Gauss(4)</td>
<td>N(0,1)</td>
<td>7.4</td>
<td>9.3</td>
<td>12.7</td>
</tr>
<tr>
<td>Gauss(4)</td>
<td>Fixed-b</td>
<td>4.7</td>
<td>6.2</td>
<td>9.1</td>
</tr>
<tr>
<td>Gauss(8)</td>
<td>N(0,1)</td>
<td>12.2</td>
<td>12.3</td>
<td>15.7</td>
</tr>
<tr>
<td>Gauss(8)</td>
<td>Fixed-b</td>
<td>4.8</td>
<td>5.5</td>
<td>7.4</td>
</tr>
<tr>
<td>Gauss(16)</td>
<td>N(0,1)</td>
<td>24.3</td>
<td>26.0</td>
<td>27.4</td>
</tr>
<tr>
<td>Gauss(16)</td>
<td>Fixed-b</td>
<td>5.3</td>
<td>5.3</td>
<td>6.8</td>
</tr>
</tbody>
</table>

This table reports rejection rates for 5% level $t$-tests of the null hypothesis that the slope coefficient in a linear regression equals its population value (1). Data are generated on square (Panel A) and cross-shaped (Panel B) regions. In both cases results are for $N=225$ data points, 2000 simulation replications, and 1000 bootstrap draws per replication. Row labels indicate kernel and smoothing used, with IID denoting conventional OLS standard errors and Hetero denoting Huber-Eicker-White SEs. Column 2 indicates which critical values were used, with "Fixed-b" indicating that critical values were obtained via the i.i.d. bootstrap procedure proposed in this paper. Both regressor and error term follow the spatial moving average process described in the text, with $\Upsilon$ determining the strength of spatial dependence.
Table 2. Critical Values for 95% Level $t$-Tests

<table>
<thead>
<tr>
<th>Type</th>
<th>$\gamma = 0$</th>
<th>$\gamma = -0.2$</th>
<th>$\gamma = 0.2$</th>
<th>$\gamma = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss(2) FS</td>
<td>2.01</td>
<td>2.41</td>
<td>2.71</td>
<td>3.87</td>
</tr>
<tr>
<td>Gauss(2) boot</td>
<td>2.05</td>
<td>2.05</td>
<td>2.05</td>
<td>2.04</td>
</tr>
<tr>
<td>Gauss(4) FS</td>
<td>2.19</td>
<td>2.42</td>
<td>2.57</td>
<td>3.10</td>
</tr>
<tr>
<td>Gauss(4) boot</td>
<td>2.21</td>
<td>2.21</td>
<td>2.21</td>
<td>2.20</td>
</tr>
<tr>
<td>Gauss(8) FS</td>
<td>2.76</td>
<td>2.90</td>
<td>3.16</td>
<td>3.57</td>
</tr>
<tr>
<td>Gauss(8) boot</td>
<td>2.76</td>
<td>2.76</td>
<td>2.76</td>
<td>2.76</td>
</tr>
<tr>
<td>Gauss(16) FS</td>
<td>4.72</td>
<td>4.60</td>
<td>5.00</td>
<td>5.71</td>
</tr>
<tr>
<td>Gauss(16) boot</td>
<td>4.54</td>
<td>4.53</td>
<td>4.55</td>
<td>4.53</td>
</tr>
<tr>
<td>Gauss(2) FS</td>
<td>2.02</td>
<td>2.37</td>
<td>2.63</td>
<td>3.74</td>
</tr>
<tr>
<td>Gauss(2) boot</td>
<td>2.05</td>
<td>2.05</td>
<td>2.02</td>
<td>2.04</td>
</tr>
<tr>
<td>Gauss(4) FS</td>
<td>2.15</td>
<td>2.32</td>
<td>2.50</td>
<td>3.03</td>
</tr>
<tr>
<td>Gauss(4) boot</td>
<td>2.18</td>
<td>2.18</td>
<td>2.12</td>
<td>2.17</td>
</tr>
<tr>
<td>Gauss(8) FS</td>
<td>2.56</td>
<td>2.68</td>
<td>2.85</td>
<td>3.23</td>
</tr>
<tr>
<td>Gauss(8) boot</td>
<td>2.58</td>
<td>2.59</td>
<td>2.42</td>
<td>2.48</td>
</tr>
<tr>
<td>Gauss(16) FS</td>
<td>3.81</td>
<td>3.79</td>
<td>4.18</td>
<td>4.68</td>
</tr>
<tr>
<td>Gauss(16) boot</td>
<td>3.79</td>
<td>3.79</td>
<td>3.32</td>
<td>3.77</td>
</tr>
</tbody>
</table>

This table shows critical values for the same tests shown in Table 1. Row labels indicate kernel and smoothing parameter. In the second column, 'FS' denotes the (infeasible) simulated finite sample critical value, while 'boot' denotes the "Fixed-b" critical value obtained via the iid bootstrap procedure proposed in this paper. For the 'FS' values, we report the 95th percentile of the absolute value of the $t$-statistic across 2000 simulation replications. For the 'boot' values, we compute this percentile for each replication using 1000 bootstrap draws, then average the resulting CVs across the 2000 replications (taking percentiles of the 2 million total simulated values of the bootstrapped $t$-statistic gives the same answer to 2 decimal places).
Table 3. 2SLS Simulation and Comparison with Ibragimov and Muller (2007)

<table>
<thead>
<tr>
<th>HAC Smoothing Parameter</th>
<th>Ibragimov and Muller Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Weight</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
</tr>
<tr>
<td>Power(0.0)</td>
<td></td>
</tr>
<tr>
<td>Power(0.5)</td>
<td></td>
</tr>
<tr>
<td>Power(1.5)</td>
<td></td>
</tr>
<tr>
<td>Power(2.0)</td>
<td></td>
</tr>
<tr>
<td>Bias(b)</td>
<td></td>
</tr>
<tr>
<td>Std(s.e.)</td>
<td></td>
</tr>
</tbody>
</table>

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 is the frequency of rejection of the hypothesis that the parameter is equal to the true value ($\beta = 1$). 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 2000 simulation replications, and 1000 bootstrap replications are used to generate critical values for the spatial HAC results. Column labels indicate the values of smoothing parameters. Columns under HAC Smoothing Parameters use a Gaussian product kernel and with equal smoothing parameter in both directions equal to the number given in the column. Columns under Ibragimov and Muller Groups 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 $\gamma = .2$ and three instruments.