Asymptotic F Test in a GMM Framework with Cross Sectional Dependence

Yixiao Sun* Department of Economics University of California, San Diego

> Min Seong Kim[†] Department of Economics Ryerson University

First version: October 22, 2010 This version: June 29, 2012

Abstract

The paper develops an asymptotically valid F test that is robust to spatial autocorrelation in a GMM framework. The test is based on the class of series covariance matrix estimators and fixed-smoothing asymptotics. The fixed-smoothing asymptotics and F approximation are established under mild sufficient conditions for a central limit theorem. These conditions can accommodate a wide range of spatial processes. This is in contrast with the standard arguments, which often impose very restrictive assumptions so that a functional central limit theorem holds. The proposed F test is very easy to implement, as critical values are from a standard F distribution. To a great extent, the asymptotic F test achieves triple robustness: it is asymptotically valid regardless of the spatial autocorrelation, the sampling region, and the limiting behavior of the smoothing parameter. Simulation shows that the F test is more accurate in size than the conventional chi-square tests, and it has the same size accuracy and power property as nonstandard tests that require computationally intensive simulation or bootstrap.

JEL Classification: C12; C14; C18; C31.

Keywords: F distribution, Fixed-smoothing asymptotics, Heteroskedasticity and Autocorrelation Robust, Robust Standard Error, Series Method, Spatial Analysis, Spatial Autocorrelation.

^{*}Correspondence to: yisun@ucsd.edu. Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA. This paper replaces a previous version entitled "Asymptotic F Test in the Presence of Nonparametric Spatial Dependence."

[†]minseong.kim@ryerson.ca. Department of Economics, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada.

1 Introduction

In this paper, we consider spatial data models in a GMM framework. Like time series data, a salient feature of spatial data is that the observations are statistically dependent. To capture general and unspecified dependence structure, we often use heteroskedasticity and autocorrelation robust (HAR) variance estimators. For recent contributions, see, for example, Kelejian and Prucha (2007) and Kim and Sun (2011). Most commonly used HAR variance estimators are formulated using conventional kernel smoothing techniques. Under some rate conditions, the HAR variance estimator is consistent, and we obtain asymptotic normal and chi-square tests. While convenient in empirical implementations, consistent HAR procedures do not capture the randomness of the HAR variance estimator, and the associated test often has large size distortion, especially when the spatial dependence is high.

To address the size distortion problem, Bester, Conley, Hansen and Vogelsang (2011, BCHV hereafter) extend the fixed-*b* asymptotics in the time series setting to the spatial setting. Under this type of asymptotics, the truncated lag in the kernel HAR variance estimator is set equal to a fixed fraction *b* of the sample size *N*. See for example Kiefer and Vogelsang (2005) and Sun, Phillips and Jin (2008). The fixed-*b* asymptotics is in contrast with the conventional asymptotics where $b \rightarrow 0$ as the sample size *N* increases. Like the conventional asymptotic distribution, the fixed-*b* asymptotic distributions of the Wald statistic and *t* statistic are asymptotically nuisance parameter free, albeit nonstandard. BCHV show by simulation that the nonstandard test has better size properties than the conventional normal or chi-square test.

While BCHV make an important contribution in extending the fixed-b asymptotics to spatial settings, several challenging issues remain to be addressed. From a theoretical point of view, the fixed-b asymptotics in BCHV is obtained under a set of assumptions that are very restrictive in the spatial setting. As BCHV use a functional central limit theorem (FCLT) for a scaled moment process, they impose strong assumptions that may not hold if spatial processes are heteroskedastic or have spatially heterogeneous dependence. BCHV also require a "quadrant-wise monotone boundaries" condition on the shape of the sampling region. This condition may be hard to verify in empirical situations. From a practical point of view, implementation of their fixed-b asymptotic test can be computationally intensive. As the fixed-b asymptotic distribution is nonstandard and depends on the sampling region, critical values have to be obtained via simulation or bootstrap.

We confront these challenges by considering the class of series HAR variance estimators, a class of HAR variance estimators that is different from but closely related to the class of kernel HAR variance estimators. Both classes of estimators belong to the larger class of quadratic estimators. From a broad perspective, these two classes are analogous to the respective kernel estimators and series/sieve estimators in nonparametric regressions. In the present setting, the series HAR variance estimator we recommend for practical use is a kernel HAR variance estimator with a special kernel function. In general, there is no clear finite sample advantage of

one estimator over the other one. However, it is advantageous to use the series HAR variance estimator in developing a new and more accurate approximation.

The smoothing parameter underlying the series HAR variance estimator is the number of terms K in the series expansion. Depending on whether K is fixed or grows with the sample size, we obtain the fixed-K asymptotics or the more conventional large-K asymptotics. We establish the fixed-K asymptotic theory without using an FCLT or maintaining restrictive assumptions on the sampling region. Our basic insight is that an FCLT is not necessary for establishing the fixed-K asymptotics. It is sufficient to invoke a CLT which can hold under much weaker conditions and therefore can accommodate a wide range of spatial processes. For example, we can use the CLT developed by Jenish and Prucha (2009, JP hereafter) which allows the spatial process to be nonstationary and even to have asymptotically unbounded moments. In addition, a CLT is much less demanding than an FCLT on the shape of the sampling region. We impose only a mild boundary condition.

In the time series setting and the spatial setting of BCHV, the fixed-b asymptotic distribution is often represented by a functional of Brownian motion or Brownian sheet. We will not use this type of representation. Instead, we introduce the notion of asymptotically equivalent distributions. We show that, under the fixed-K asymptotics, the Wald statistic is asymptotically equivalent to a quadratic form in a standard normal vector with an independent and random weighting matrix. The random weighting matrix captures the estimation uncertainty of the series HAR variance estimator. The asymptotically equivalent distribution is nuisance parameter free. Compared to the representation that involves a Brownian sheet, the asymptotically equivalent distribution has a simpler representation. It is also easier to simulate, as it is a function of only N iid standard normal vectors.

A further innovation of the paper is that we design a sequence of basis functions so that the asymptotically equivalent distribution becomes a standard distribution. For any given basis functions, we first center them and then orthonormalize the centered basis functions via the Gram-Schmidt procedure. We use the transformed basis functions in our series HAR variance estimation. By construction, the transformed basis functions are orthonormal and integrate to zero. These two properties ensure that the random weighting matrix follows a Wishart distribution and is independent of the standard normal vector. So the quadratic form in this standard normal vector follows exactly an F distribution. The transformed basis functions, coupled with the fixed-K asymptotics, give rise to our asymptotic F test. The F test is very convenient to use in practice, as critical values from the F distribution can be obtained from statistical tables or software packages. No computationally intensive simulation or bootstrap is needed.

The next step in using the series HAR variance estimator is to select the number of terms K. We consider the asymptotic mean squared error (AMSE) criterion. The proposed smoothing parameter depends on unknown parameters, which can be estimated by a parametric plugin procedure. We employ the Matérn model as the approximating parametric model. As a widely-used model in spatial analysis, the Matérn model is very flexible in capturing various decaying patterns of spatial dependence. Simulation studies show that the fixed-K asymptotic tests, which include the F test, are more accurate in size than the conventional chi-square test. They are also as accurate in size as the BCHV test with similar power properties.

The fixed-K asymptotics and fixed-b asymptotics may be collectively referred to as the fixed-smoothing asymptotics as they effectively involve smoothing over a fixed number of quantities of interest. On the other hand, the conventional large-K asymptotics where $K \to \infty$ or the small-b asymptotics where $b \rightarrow 0$ may be referred to as the increasing-smoothing asymptotics, as they involves smoothing over an increasing number of quantities. The two specifications can be viewed as different asymptotic devices to obtain approximations to the finite sample distribution. While we recommend using the fixed-smoothing asymptotic approximation, this does not mean that we advocate the use of an inconsistent HAR variance estimator. Instead, we follow conventional practices in constructing the HAR variance estimator and the associated test statistic. Only in the last stage of inference that requires a reference distribution do we use the fixed-smoothing asymptotic approximation. When $K \to \infty$ or $b \to 0$, it can be shown that the fixed-smoothing asymptotic approximation reduces to the conventional chi-square approximation. So critical values from the fixed-smoothing asymptotics are asymptotically valid regardless of whether the amount of smoothing is held fixed or grows with the sample size or not. We can regard the fixed-smoothing asymptotic approximation as a robust approximation.

The rest of the paper is organized as follows. Section 2 describes the problem at hand and introduces the series HAR variance estimator. Section 3 establishes the fixed-K asymptotic theory and F approximation under transformed basis functions. Section 4 develops a bandwidth selection procedure. Section 5 discusses practical issues in implementing our proposed tests. Section 6 presents simulation evidence. The next section applies the proposed testing procedures to Conley and Udry (2010) and reexamines their main results. The last section provides some concluding discussion. Proofs are given in the Appendix.

2 GMM Estimation and HAR Inference

We are interested in a $d \times 1$ vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$. Let θ_0 be the true value and assume that θ_0 is an interior point of the compact parameter space Θ . The moment conditions

$$Ef_{i}(\theta) = 0$$

hold if and only if $\theta = \theta_0$, where $f_j(\theta) = f(Y_j, \theta)$ is a $d_f \times 1$ vector of continuously differentiable functions with $d_f \ge d$, rank $E[\partial f_j(\theta) / \partial \theta'|_{\theta=\theta_0}] = d$, and $Y_j \in \mathbb{R}^{d_Y}$ is a vector of observations at location j. The number of observed locations or the sample size is N. We allow the moments to exhibit general forms of spatial correlation where the strength of the correlation depends on some observable distance measure between two locations. For simplicity, we follow Conley (1999) and assume that it is possible to map the data onto a finite-dimensional integer lattice so that the distance can be expressed in terms of lattice indices. To simplify the presentation further, we consider the two dimensional case so that $j = (j_1, j_2) \in \mathbb{Z}^2$. Extending our results to higher dimensions is straightforward.

We assume that the sampling region can be represented by $B_n = nB \cap \mathbb{Z}^2$ where $n = (n_1, n_2)$, $nB = \{(n_1b_1, n_2b_2) : (b_1, b_2) \in B\}$, and B is a compact set with positive Lebesgue measure. As n_1 and n_2 increase, the sampling region B_n gets "inflated" in each direction and more observations become available. We allow the possibility of nonnested sampling regions, i.e. B_n may not be a subset of $B_{n'}$ for n < n' and the data generating process is understood to be a triangular array data generating process. We consider the increasing-domain asymptotics where $n_1 \to \infty$ and $n_2 \to \infty$ and there is a minimum positive distance between the locations where the observations are taken.

Based on the moment conditions $Ef_j(\theta_0) = 0$, the GMM estimator (Hansen, 1982) of θ_0 is given by

$$\hat{\theta}_N = \arg\min_{\theta\in\Theta} \left[\frac{1}{N} \sum_{j\in B_n} f_j(\theta) \right]' \mathcal{W}_N \left[\frac{1}{N} \sum_{j\in B_n} f_j(\theta) \right],$$

where \mathcal{W}_N is a $d_f \times d_f$ positive definite and symmetric weighting matrix. Let

$$g_N(\theta) = \frac{1}{N} \sum_{j \in B_n} f_j(\theta), \ G_N(\theta) = \frac{1}{N} \sum_{j \in B_n} \frac{\partial f_j(\theta)}{\partial \theta'}.$$

If $\operatorname{plim}_{N\to\infty} G_N(\hat{\theta}_N) = G$ and $\operatorname{plim}_{N\to\infty} \mathcal{W}_N = \mathcal{W}$ for some G and \mathcal{W} with full column rank, then under some regularity conditions, $\hat{\theta}_N$ satisfies

$$\sqrt{N}\left(\hat{\theta}_{N}-\theta_{0}\right) = -\sqrt{N}\left[G'\mathcal{W}G\right]^{-1}G'\mathcal{W}g_{N}\left(\theta_{0}\right) + o_{p}\left(1\right)$$
$$= -\sqrt{N}H^{-1}S_{N}\left(\theta_{0}\right) + o_{p}\left(1\right)$$

where

$$H = G' \mathcal{W}G, \ S_N(\theta) = \frac{1}{N} \sum_{j \in B_n} s_j(\theta)$$

and $s_j(\theta) = G' \mathcal{W} f_j(\theta)$, a linear transformation of the original moment functions. To simplify the notation, we write $S_N = S_N(\theta_0)$ and $s_j = s_j(\theta_0)$. If $\sqrt{N}S_N \Rightarrow N(0,\Omega)$, where

$$\Omega = \lim_{n \to \infty} \frac{1}{N} E\left(\sum_{j \in B_n} s_j\right) \left(\sum_{j \in B_n} s_j\right)',$$

then

$$\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \Rightarrow N(0, H^{-1}\Omega H^{-1}).$$
(1)

In the time series literature, Ω is called the long-run variance. We may refer to Ω as the global variance in the spatial setting, as it is not a variance associated with a single location but rather a variance contributed by all locations.

Since H can be consistently estimated by its sample analog $\hat{H} = G'_N(\hat{\theta}_N) \mathcal{W}_N G_N(\hat{\theta}_N)$, it suffices to estimate Ω in order to conduct inference about θ_0 . In this paper, we employ the series type estimator for Ω . Let

$$\hat{s}_j = s_j\left(\hat{\theta}_N\right) = G'_N(\hat{\theta}_N)\mathcal{W}_N f_j(\hat{\theta}_N).$$

For each $k = (k_1, k_2) \in (\mathbb{Z}^+ \cup \{0\}) \otimes (\mathbb{Z}^+ \cup \{0\})$, define

$$\mathcal{A}_{k} = \mathcal{A}_{k}(\hat{\theta}_{N}) = \frac{1}{\sqrt{N}} \sum_{(j_{1}, j_{2}) \in B_{n}} \Phi_{k_{1}, k_{2}}(\frac{j_{1}}{n_{1}}, \frac{j_{2}}{n_{2}}) \hat{s}_{j_{1}, j_{2}}$$

for some basis function $\Phi_{k_1,k_2}(\cdot,\cdot) = \Phi_k(\cdot,\cdot)$ that may be complex and $\Phi_{(0,0)}(\cdot,\cdot) \equiv 1$. The basis functions are used to directly capture the global variation of the spatial process. For notational simplicity, we denote $j/n = (j_1/n_1, j_2/n_2)$ from now on so that we can write $\Phi_{k_1,k_2}(j_1/n_1, j_2/n_2) = \Phi_k(j/n)$. Construct the direct estimator:

$$\tilde{\Omega}_k = \operatorname{Re}\left(\mathcal{A}_k \mathcal{A}_k^*\right),$$

where \mathcal{A}_k^* is the conjugate transpose of \mathcal{A}_k . Taking a simple average of the direct estimators yields a new estimator:

$$\hat{\Omega} = \frac{1}{K} \sum_{k \in \mathbb{K}} \hat{\Omega}_k$$

where

$$\mathbb{K} = \mathbb{K}_0 \setminus \{(0,0)\}, \mathbb{K}_0 = (0,1,...,K_1) \otimes (0,1,...,K_2),$$

 K_1 and K_2 are smoothing parameters, and $K = K_1K_2 + K_1 + K_2$. The larger K is, the larger the amount of smoothing is.

In the definition of $\hat{\Omega}$, we have explicitly excluded the case $k_1 = k_2 = 0$. We do so because when $\Phi_{(0,0)}(r,s) = 1$, we have $\mathcal{A}_{(0,0)} = 0$ and hence $\hat{\Omega}_{(0,0)} \equiv 0$, using the definition of the estimator $\hat{\theta}_N$.

As an example, consider using the complex exponential:

$$\Phi_k(r,s) = \exp\left[-i\left(2\pi k_1 r + 2\pi k_2 s\right)\right], \ k_1, k_2 \in \mathbb{K},\tag{2}$$

as the basis functions. In this case, \mathcal{A}_k becomes

$$\mathcal{A}_{k} = \frac{1}{\sqrt{N}} \sum_{(j_{1}, j_{2}) \in B_{n}} \exp\left[-i\left(\frac{2\pi k_{1}}{n_{1}}j_{1} + \frac{2\pi k_{2}}{n_{2}}j_{2}\right)\right] s_{j_{1}, j_{2}}\left(\hat{\theta}_{N}\right),$$

which is the finite Fourier transform of the spatial process $s_{j_1,j_2}(\hat{\theta}_N)$. Furthermore, some elementary manipulations show that

$$\hat{\Omega} = \frac{1}{N} \sum_{(j_1, j_2) \in B_n} \sum_{\left(\tilde{j}_1, \tilde{j}_2\right) \in B_n} s_{j_1, j_2} \left(\hat{\theta}_N\right) W\left(\frac{j_1 - \tilde{j}_1}{n_1}, \frac{j_2 - \tilde{j}_2}{n_2}\right) s'_{\tilde{j}_1, \tilde{j}_2} \left(\hat{\theta}_N\right)$$

where

$$W(x_1, x_2) = \frac{1}{K} \sum_{k_1=0}^{K_1} \sum_{k_2=0}^{K_2} \cos\left[2\pi \left(k_1 x_1 + k_2 x_2\right)\right]$$
$$= \frac{1}{K} \cos\pi \left[K_1 x_1 + K_2 x_2\right] \frac{\sin\pi x_1 \left(K_1 + 1\right)}{\sin\pi x_1} \frac{\sin\pi x_2 \left(K_2 + 1\right)}{\sin\pi x_2}$$

So the series HAR variance estimator is a kernel estimator with a special kernel weighting function given above. The above formula also shows that $\hat{\Omega}$ depends on n only through K_1/n_1 and K_2/n_2 , which will be specified in Section 4. See Sun (2011a,b) and references therein for more discussions on the series long-run variance estimation in the time series setting.

Suppose that the null hypothesis of interest is $H_0 : R\theta = \mathfrak{r}$ and the alternative is $H_1 : R\theta \neq \mathfrak{r}$, where R is a $q \times d$ matrix. The F test version of the Wald statistic for testing H_0 against H_1 is given by

$$F_N = \left[\sqrt{N}\left(R\hat{\theta}_N - \mathfrak{r}\right)\right]' \left(R\hat{H}^{-1}\hat{\Omega}\hat{H}^{-1}R'\right)^{-1}\sqrt{N}\left(R\hat{\theta}_N - \mathfrak{r}\right)/q.$$
(3)

When q = 1, we can construct the usual t statistic

$$t_N = \frac{\sqrt{N}(R\hat{\theta}_N - \mathfrak{r})}{\left(R\hat{H}^{-1}\hat{\Omega}\hat{H}^{-1}R'\right)^{1/2}}.$$
(4)

We can also consider nonlinear restrictions; our results remain valid after simple linearization.

3 Asymptotic Properties of Test Statistics

In this section, we establish the asymptotic distribution of the Wald and t statistics under the specification that K_1 and K_2 are fixed. We maintain the following assumptions.

Assumption 1. K_1 and K_2 are fixed as $(n_1, n_2) \to (\infty, \infty)$.

Assumption 2. $\hat{\theta}_N \rightarrow^p \theta_0$.

Assumption 3. (i) For any open set $A \subseteq int(B)$, let $j_L^A = \min\{j_1 : (j_1, j_2) \in nA \cap \mathbb{Z}^2\}$, $j_U^A = \max\{j_1 : (j_1, j_2) \in nA \cap \mathbb{Z}^2\}$ and define $A_n(j_1) = \{j_2 : (j_1, j_2) \in nA \cap \mathbb{Z}^2\}$ for $j_L^A \leq j_1 \leq j_U^A$. The uniform law of large numbers

$$\sup_{\theta \in \mathcal{N}(\theta_0)} \sup_{A \subseteq int(B)} \sup_{j_1 \in [j_L^A, j_U^A]} \left\| \frac{1}{|A_n(j_1)|} \sum_{j_2 \in A_n(j_1)} \frac{\partial s_j(\theta)}{\partial \theta'} - H(\theta) \right\| \to^p 0$$

holds, where $\mathcal{N}(\theta_0)$ is an open neighborhood of θ_0 , $\|\cdot\|$ is the Euclidean norm, and $|A_n(j_1)|$ is the number of elements in $A_n(j_1)$; (ii) $H(\theta)$ is continuous at $\theta = \theta_0$, and $H = H(\theta_0)$ is a nonsingular matrix.

Assumption 4. For $e_j \sim iidN(0, \mathbb{I}_d)$ over $j \in B_n$, we have

$$P\left(\left[\begin{array}{c}\frac{1}{\sqrt{N}}\sum_{j\in B_{n}}\operatorname{Re}\Phi_{k}(\frac{j}{n})s_{j}\\\frac{1}{\sqrt{N}}\sum_{j\in B_{n}}\operatorname{Im}\Phi_{k}(\frac{j}{n})s_{j}\end{array}\right] < x_{k} \text{ for all } k\in\mathbb{K}_{0}\right)$$
$$=P\left(\left[\begin{array}{c}\Lambda\frac{1}{\sqrt{N}}\sum_{j\in B_{n}}\operatorname{Re}\Phi_{k}(\frac{j}{n})e_{j}\\\Lambda\frac{1}{\sqrt{N}}\sum_{j\in B_{n}}\operatorname{Im}\Phi_{k}(\frac{j}{n})e_{j}\end{array}\right] < x_{k} \text{ for all } k\in\mathbb{K}_{0}\right) + o\left(1\right)$$

uniformly over $x_k \in \mathbb{R}^{2d}$ where Λ is the matrix square root of Ω , i.e. $\Lambda \Lambda' = \Omega$, and Ω is positive definite.

Assumption 1 makes it explicit that our asymptotics is taken under fixed K_1 and K_2 and that the sampling region expands in each direction but with possibly different speeds. Assumption 2 is made for convenience. It can be proved under more primitive assumptions and using standard arguments. The ULLN in Assumption 3 is similar to Assumption 6 in BCHV. Similar assumptions are made in the time series setting. For a given set A, $[j_L^A, j_U^A]$ is the range of the first index in the set $nA \cap \mathbb{Z}^2$. The set $\{(j_1, j_2) : j_1 \in [j_L^A, j_U^A], j_2 \in A_n(j_1)\}$ is the same as $nA \cap \mathbb{Z}^2$. Assumption 4 is satisfied if a CLT holds jointly for $N^{-1/2} \sum_{j \in B_n} \operatorname{Re} \Phi_k(j/n)s_j$ and $N^{-1/2} \sum_{j \in B_n} \operatorname{Im} \Phi_k(j/n)s_j$. Some primitive sufficient conditions for Assumption 4 are provided in the Appendix. When Assumption 4 holds, we write

$$\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})s_j \stackrel{a}{\sim} \Lambda \frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})e_j$$

where $\stackrel{a}{\sim}$ signifies that the two sides are asymptotically equivalent in distribution.

Theorem 1 Let Assumptions 1-4 hold. If $\Phi_k(\cdot, \cdot)$ is continuously differentiable, then

$$F_N \stackrel{a}{\sim} F_N^a, \ t_N \stackrel{a}{\sim} t_N^a$$

where

$$F_N^a = \frac{1}{q} \eta' \left(\frac{1}{K} \sum_{k \in \mathbb{K}} \operatorname{Re}\left(\xi_k \xi_k^*\right) \right)^{-1} \eta, \ t_N^a = \frac{\eta}{\left[K^{-1} \sum_{k \in \mathbb{K}} \operatorname{Re}\left(\xi_k \xi_k^*\right) \right]^{1/2}},$$
$$\eta = \frac{1}{\sqrt{N}} \sum_{j \in B_n} \varepsilon_j, \ \xi_k = \frac{1}{\sqrt{N}} \sum_{j \in B_n} \left[\Phi_k(\frac{j}{n}) - \frac{1}{N} \sum_{j \in B_n} \Phi_k(\frac{j}{n}) \right] \varepsilon_j,$$

and $\varepsilon_j \sim iidN(0, \mathbb{I}_q)$.

We prove Theorem 1 under much weaker assumptions than BCHV. As in JP, we allow for very general sampling regions. In contrast, BCHV maintain the very restrictive assumption that the sampling region has a quadrant-wise monotone boundary. In addition, our result is established under the CLT for spatial processes while BCHV require an FCLT indexed by sets. A CLT can hold under very mild conditions to accommodate a wide range of spatial processes. The spatial processes can be nonstationary and can even have asymptotic unbounded moments, see JP. In addition to the conditions given in JP, we need a mild homogeneity condition for Assumption 4 to hold; see Lemma 1 in the Appendix. This condition is weaker than covariance stationarity. In contrast, an FCLT usually requires much stronger assumptions so that the tightness condition can be verified. Furthermore, the asymptotically equivalent distribution in Theorem 1 has a simple representation. When s_j is iid normal, it is exactly equal to the finite sample distribution. This is in contrast to the limiting distribution in BCHV, which is a sum of complicated functionals of set-indexed Brownian sheet processes. In general, the limiting distribution in BCHV is always an approximation to the finite sample distribution.

Since η and ξ_k are normal and $cov(\eta, \xi_k) = N^{-1} \sum_{j \in B_n} \left[\Phi_k(j/n) - N^{-1} \sum_{j \in B_n} \Phi_k(j/n) \right] \mathbb{I}_q = \mathbf{0}_{q \times q}$ for any $k \in \mathbb{K}$, η and ξ_k are independent. So F_N is asymptotically equivalent in distribution to a quadratic form in a standard normal vector with an independent and random weighting matrix. The random weighting matrix captures the estimation uncertainty of the series HAR variance estimator. Like the finite sample distribution, the asymptotically equivalent distribution depends on K, the basis functions, and the sampling region.

When $\Phi_k(\cdot, \cdot)$ is real, ξ_k is normal with mean zero and variance

$$var\left(\xi_{k}\right) = \frac{1}{N} \sum_{j \in B_{n}} \left[\Phi_{k}\left(\frac{j}{n}\right) - \frac{1}{N} \sum_{j \in B_{n}} \Phi_{k}\left(\frac{j}{n}\right)\right]^{2} \mathbb{I}_{q}.$$

The covariance between ξ_{k_1} and ξ_{k_2} is

$$cov\left(\xi_{k_{1}},\xi_{k_{2}}\right) = \frac{1}{N}\sum_{j\in B_{n}}\left[\Phi_{k_{1}}(\frac{j}{n}) - \frac{1}{N}\sum_{j\in B_{n}}\Phi_{k_{1}}(\frac{j}{n})\right]\left[\Phi_{k_{2}}(\frac{j}{n}) - \frac{1}{N}\sum_{j\in B_{n}}\Phi_{k_{2}}(\frac{j}{n})\right]\mathbb{I}_{q}.$$

So the representation of F_N^a enables us to see that the asymptotic distribution of F_N depends

on the sampling region and basis functions via $N^{-1} \sum_{j \in B_n} \Phi_k(j/n)$, $N^{-1} \sum_{j \in B_n} \Phi_k^2(j/n)$, and $N^{-1} \sum_{j \in B_n} \Phi_{k_1}(j/n) \Phi_{k_2}(j/n)$ or their limiting forms.

The following corollary gives the asymptotically equivalent distribution in a special case.

Corollary 2 Let the Assumptions in Theorem 1 hold. If (i) $N^{-1} \sum_{j \in B_n} \Phi_k(j/n) = o(1)$ and (ii) $N^{-1} \sum_{j \in B_n} \Phi_{k_1}(j/n) \Phi_{k_2}(j/n) = 1 \{k_1 = k_2\} (1 + o(1)), \text{ then}$

$$\frac{(K-q+1)}{K}F_N \stackrel{a}{\sim} F_{q,K-q+1} \text{ and } t_N \stackrel{a}{\sim} t_{K-q+1}.$$

Corollary 2 shows that the finite sample distribution of qF_N can be approximated by

$$\frac{K}{K-q+1} \frac{1}{\chi^2_{K-q+1}/(K-q+1)} \chi^2_q.$$

As $K \to \infty$, both $\chi^2_{K-q+1}/(K-q+1)$ and K/(K-q+1) converge to one. As a result, the above limiting distribution reduces to χ^2_q , the conventional asymptotic approximation. A direct implication is that critical values obtained from the F approximation are asymptotically valid under the conventional asymptotics when $K \to \infty$ with the sample size. However, when K is not very large or the number of the restrictions q is large, the F approximation can be very different from the chi-square approximation. Since both the random denominator $\chi^2_{K-q+1}/(K-q+1)$ and the proportional factor K/(K-q+1) shift the probability mass to the right, critical values based on the F approximation are larger than those based on the chi-square approximation.

An example of Corollary 2 is when B is a rectangle and $\Phi_k(r,s) = \sqrt{2} \cos(2\pi k_1 r + 2\pi k_2 s)$ or $\Phi_k(r,s) = \sqrt{2} \sin(2\pi k_1 r + 2\pi k_2 s)$. It is not hard to show that the assumptions in Corollary 2 hold in this case. So when the sampling region has a square lattice structure, we can use critical values from the F or t distribution to perform the Wald test or the t test.

When $\Phi_k(r, s)$ is complex, we may write

$$\frac{1}{K}\sum_{k\in\mathbb{K}}\operatorname{Re}\left(\xi_{k}\xi_{k}^{*}\right) = \frac{1}{2K}\sum_{k\in\mathbb{K}}\left(\xi_{R,k}\xi_{R,k}^{\prime} + \xi_{I,k}\xi_{I,k}^{\prime}\right)$$

where $\xi_{R,k} = \frac{1}{\sqrt{N}} \sum_{j \in B_n} \tilde{\Phi}_k^R(j/n) \varepsilon_j$, $\xi_{I,k} = \frac{1}{\sqrt{N}} \sum_{j \in B_n} \tilde{\Phi}_k^I(j/n) \varepsilon_j$ and

$$\tilde{\Phi}_{k}^{R}(\frac{j}{n}) = \operatorname{Re}\sqrt{2}\tilde{\Phi}_{k}(\frac{j}{n}), \quad \tilde{\Phi}_{k}^{I}(\frac{j}{n}) = \operatorname{Im}\sqrt{2}\tilde{\Phi}_{k}(\frac{j}{n}),$$
$$\tilde{\Phi}_{k}(\frac{j}{n}) = \Phi_{k}(\frac{j}{n}) - \frac{1}{N}\sum_{j\in B_{n}}\Phi_{k}(\frac{j}{n}).$$

So the complex case is the same as the real case using $\{\sqrt{2} \operatorname{Re} \Phi_k(r,s), \sqrt{2} \operatorname{Im} \Phi_k(r,s), k \in \mathbb{K}\}$ as the basis functions. The number of basis functions is 2K. If these basis functions satisfy the assumptions in Corollary 2, then we immediately have:

$$\frac{(2K-q+1)}{2K}F_N \stackrel{a}{\sim} F_{q,2K-q+1} \text{ and } t_N \stackrel{a}{\sim} t_{2K-q+1}.$$
(5)

When B is a rectangle, a natural choice for complex valued $\Phi_k(\cdot, \cdot)$ is the complex exponential given in (2). It is easy to check that all assumptions in Corollary 2 hold for $\{\sqrt{2} \operatorname{Re} \Phi_k(r,s), \sqrt{2} \operatorname{Im} \Phi_k(r,s), k \in \mathbb{K}\}$.

For a general sampling region B_n and basis functions, the asymptotically equivalent distribution cannot be simplified in general. However, it can be simulated easily. To obtain a draw from this distribution, we only need to generate N iid standard normal vectors. So it is computationally less intensive than simulating the asymptotic distribution represented by functionals of Brownian sheets.

A more interesting and convenient method to deal with a general sampling region is to construct orthonormal basis functions on the region so that Corollary 2 applies. Given the basis function $\Phi_k(\cdot, \cdot)$, we can center it first and then orthonormalize the centered basis functions via the Gram-Schmidt procedure. More specifically, we can follow the three steps below:

(i) Center the basis functions to obtain

$$\tilde{\Phi}_k(\frac{j}{n}) := \tilde{\Phi}_k(\frac{j_1}{n_1}, \frac{j_2}{n_2}) = \exp\left[-2\pi i \left(\frac{k_1 j_1}{n_1} + \frac{k_2 j_2}{n_2}\right)\right] - \frac{1}{N} \sum_{b \in B_n} \exp\left[-2\pi i \left(\frac{k_1 b_1}{n_1} + \frac{k_2 b_2}{n_2}\right)\right]$$

for all $j \in B_n$.

(ii) For each k, let $\tilde{\Phi}_k^R$ be the $N \times 1$ column vector collecting $\left\{ \operatorname{Re} \left[\tilde{\Phi}_k \left(j/n \right) \right], j \in B_n \right\}$ and $\tilde{\Phi}_k^I$ be the corresponding column vector collecting $\left\{ \operatorname{Im} \left[\tilde{\Phi}_k \left(j/n \right) \right], j \in B_n \right\}$. Concatenate $\tilde{\Phi}_k^R$ and $\tilde{\Phi}_k^I$ to form the $N \times 2K$ matrix Φ .

(iii) Let $\mathbf{\Phi} = \mathbf{\Phi}^Q R$ be the QR decomposition of $\mathbf{\Phi}$, where $\mathbf{\Phi}^Q = (\Phi_{m\ell}^Q)$ is the $N \times 2K$ orthonormal matrix. We use the columns of $\mathbf{\Phi}^Q$ as the basis vectors and estimate Ω by

$$\hat{\Omega}_Q = \frac{1}{2K} \sum_{\ell=1}^{2K} \left[\sum_{m=1}^N \Phi^Q_{m\ell} \hat{s}^Q_m \right] \left[\sum_{m=1}^N \Phi^Q_{m\ell} \hat{s}^Q_m \right]'$$
(6)

where \hat{s}_m^Q is the *m*-th transformed moment vector with ordering conformable with the columns of $\boldsymbol{\Phi}$.

Since the columns of $\mathbf{\Phi}^Q$ are just linear combinations of the centered basis functions, they satisfy the conditions in Corollary 2. Therefore, we can use the F distribution or the t distribution as the reference distribution in statistical inference. We call the resulting test the asymptotic F test or the asymptotic t test.

4 Smoothing Parameter Choice

In this section, we select the smoothing parameter based on the mean square error criterion. We focus on the case that $\Phi_k(r,s) = \exp\left[-i\left(2\pi k_1 r + 2\pi k_2 s\right)\right]$ as it provides a complete orthonormal system on $L^2([0,1]^2)$. In addition, for this choice of $\Phi_k(r,s)$, the series HAR variance estimator reduces to a kernel HAR estimator, so we can use the results in Kim and Sun (2011) to facilitate the mean square error calculation.

We consider the case $K_1/n_1 = K_2/n_2 = \kappa$ and select only one smoothing parameter κ . We can allow K_1/n_1 and K_2/n_2 to take different values, but it is often difficult to estimate more than one theoretically-optimal smoothing parameter in finite samples.

We define Ω as the pseudo-estimator that is identical to $\hat{\Omega}$ but is based on the true parameter θ_0 instead of $\hat{\theta}_N$. That is,

$$\tilde{\Omega} = \frac{1}{K} \sum_{k \in \mathbb{K}} \operatorname{Re} \left[\mathcal{A}_{k} \left(\theta_{0} \right) \mathcal{A}_{k}^{*} \left(\theta_{0} \right) \right].$$

Given a $d^2 \times d^2$ weighting matrix V, we define the MSE criterion as

$$MSE(\tilde{\Omega}, V) = E\left\{\left[\operatorname{vec}(\tilde{\Omega} - \Omega)' V \operatorname{vec}(\tilde{\Omega} - \Omega)\right]\right\},\$$

where $vec(\cdot)$ is the column-by-column vectorization function. If necessary, we employ the asymptotic truncated MSE function in Andrews (1991). Under the assumptions in Kim and Sun (2011) and following the same arguments, we can show that

$$MSE(\tilde{\Omega}, V) = \left\{ \kappa^4 \left[vec\left(\mathcal{B}\right) \right]' Vvec(\mathcal{B}) + \frac{1}{2N\kappa^2} tr \left[V\left(\Omega \otimes \Omega\right) \left(\mathbb{I}_{d^2} + \mathcal{K}_{dd} \right) \right] \right\} (1 + o(1)),$$

where $\mathcal{B} = \Omega^{(11)} + \Omega^{(22)} + \Omega^{(12)}$,

$$\Omega^{(11)} = -\frac{2}{3}\pi^2 \frac{1}{N} \sum_{j \in B_n} \sum_{\tilde{j} \in B_n} Es_j (\theta_0) s'_{\tilde{j}} (\theta_0) (j_1 - \tilde{j}_1)^2$$

$$\Omega^{(22)} = -\frac{2}{3}\pi^2 \frac{1}{N} \sum_{j \in B_n} \sum_{\tilde{j} \in B_n} Es_j (\theta_0) s'_{\tilde{j}} (\theta_0) (j_2 - \tilde{j}_2)^2$$

$$\Omega^{(12)} = -\pi^2 \frac{1}{N} \sum_{j \in B_n} \sum_{\tilde{j} \in B_n} Es_j (\theta_0) s'_{\tilde{j}} (\theta_0) (j_1 - \tilde{j}_1) (j_2 - \tilde{j}_2)$$
(7)

and \mathcal{K}_{dd} is the $d^2 \times d^2$ commutation matrix.

According to the MSE criterion, the MSE-optimal κ is

$$\kappa_{opt} = \left(\frac{\operatorname{tr}\left[V\left(\Omega \otimes \Omega\right)\left(\mathbb{I}_{d^{2}} + \mathcal{K}_{dd}\right)\right]}{4\left[\operatorname{vec}\left(\mathcal{B}\right)\right]' \operatorname{Vvec}(\mathcal{B})}\right)^{1/6} \left(\frac{1}{N}\right)^{1/6}$$

When V assigns equal weights only to the diagonal elements of $\tilde{\Omega}$, the above formula simplifies to

$$\kappa_{opt} = \left(\frac{\|diag\left(\Omega\right)\|^2}{2 \|diag(\mathcal{B})\|^2}\right)^{1/6} \left(\frac{1}{N}\right)^{1/6}.$$

When the spatial process has positive dependence such that $diag(Es_j(\theta_0) s'_{\tilde{j}}(\theta_0)) \ge 0$ for all (j, \tilde{j}) , we can show that

$$\begin{split} \|diag(\mathcal{B})\|^{2} &= \left\| diag \left[\Omega^{(11)} + \Omega^{(22)} + \Omega^{(12)} \right] \right\|^{2} \\ &\leq 3 \left(\left\| diag \left[\Omega^{(11)} \right] \right\|^{2} + \left\| diag \left[\Omega^{(22)} \right] \right\|^{2} + \left\| diag \left[\Omega^{(12)} \right] \right\|^{2} \right) \\ &\leq 3 \left(\left\| diag \left[\Omega^{(11)} \right] \right\|^{2} + \left\| diag \left[\Omega^{(22)} \right] \right\|^{2} + \frac{3}{4} \left\| diag \left[\Omega^{(11)} \right] \right\|^{2} + \frac{3}{4} \left\| diag \left[\Omega^{(22)} \right] \right\|^{2} \right) \\ &= \frac{21}{4} \left(\left\| diag \left[\Omega^{(11)} \right] \right\|^{2} + \left\| diag \left[\Omega^{(22)} \right] \right\|^{2} \right). \end{split}$$

In this case, $\kappa_{opt} \geq \kappa^*$ where

$$\kappa^* = \left(\frac{2 \|diag(\Omega)\|^2}{21 \left(\|diag[\Omega^{(11)}]\|^2 + \|diag[\Omega^{(22)}]\|^2\right)}\right)^{1/6} \left(\frac{1}{N}\right)^{1/6}.$$

To control the asymptotic bias more effectively, we suggest employing the lower bound κ^* to choose K_1 and K_2 . That is, we take $K_1^* = \lfloor n_1 \kappa^* \rfloor$ and $K_2^* = \lfloor n_2 \kappa^* \rfloor$ where $\lfloor x \rfloor$ is the integer part of x. This suggestion is line with Sun, Phillips and Jin (2008) who show that the asymptotic bias under some testing-oriented criterion should be of smaller order than the asymptotic bias under the MSE criterion. In addition, using κ^* instead of κ_{opt} avoids estimating $\Omega^{(12)}$, which is often a difficult task.

To implement the data-driven K_1^* and K_2^* , we use an approximating parametric model to capture the spatial dependence. There are two classes of parametric models that are commonly used in the literature. The first is to model the process itself. This approach is based on the work of Cliff and Ord (1981) and requires the use of a weight matrix. Kim and Sun (2011) use this approach. The second approach is to model the covariance structure directly. In this approach, rather than starting with the process and deriving the covariance matrix, a functional form for the covariance structure is assumed. The parameters of this function are then estimated. Here we use the second approach as it does not require specifying a weight matrix.

We employ the flexible class of Matérn models as the approximating covariance model.

For each component $s_{j}^{(p)}(\theta_{0})$ of $s_{j}(\theta_{0})$, the covariance function is

$$C_p(h) = Es_j^{(p)}\left(\theta_0\right) \left[s_{\tilde{j}}^{(p)}\left(\theta_0\right)\right]' = \frac{\sigma_p^2}{2^{\nu_p - 1}\Gamma\left(\nu_p\right)} \left(\frac{\|h\|}{\tau_p}\right)^{\nu_p} \mathfrak{K}_{\nu_p}\left(\frac{\|h\|}{\tau_p}\right)$$

where $\nu_p > 0$, $\tau_p > 0$, $h = (j_1 - \tilde{j}_1, j_2 - \tilde{j}_2)$, and \Re_{ν_p} is the modified Bessel function of the second kind (Abramowitz and Stegun, 1965, pp. 374-379). This function has been implemented in standard programming packages. For example, in Matlab[®] the function besselk(nu,z) computes $\Re_{\nu}(z)$. The corresponding variogram is given by

$$\gamma_p(h) = C_p(0) - C_p(h) = \sigma_p^2 \left[1 - \frac{1}{2^{\nu_p - 1} \Gamma(\nu_p)} \left(\frac{\|h\|}{\tau_p} \right)^{\nu_p} \mathfrak{K}_{\nu_p} \left(\frac{\|h\|}{\tau_p} \right) \right]$$

The variogram expression is the same as equation (5.32) in Webster and Oliver (2001, p. 94). For the Matérn class of models, the spectral density for $s_{\ell,m}^{(p)}(\theta_0)$ has the form:

$$f(\omega_1, \omega_2) = \frac{\sigma_p^2 \tau_p^{-2\nu_p}}{\pi \left(\tau_p^{-2} + \omega_1^2 + \omega_2^2\right)^{\nu_p + 1}},$$

which can be obtained by setting $\tau_p = 1/\alpha$, $\sigma_p^2 = 2^{\nu_p - 1} \phi \Gamma(\nu_p + 1)$, and d = 2 in equation (32) of Stein (1999, pp. 48-49). Under the above specification, we have

$$\|diag(\Omega)\|^{2} = \sum_{p=1}^{d} \left(4\pi\sigma_{p}^{2}\tau_{p}^{2}\right)^{2},$$
$$\|diag\left[\Omega^{(11)}\right]\|^{2} + \|diag\left[\Omega^{(22)}\right]\|^{2} = 2\sum_{p=1}^{d} \left(\frac{4}{3}\pi^{2} \times 4\pi\left(\nu_{p}+1\right)\sigma_{p}^{2}\tau_{p}^{4}\right)^{2},$$

and so

$$\kappa_N^* = 0.25501 \left(\frac{\sum_{p=1}^d \sigma_p^4 \tau_p^4}{\sum_{p=1}^d (\nu_p + 1)^2 \sigma_p^4 \tau_p^8} \right)^{1/6} N^{-1/6}.$$
 (8)

In our simulation study below, we set $\nu_p = \nu$ for all p and consider two different values of ν : $\nu = 1/2$, 1, which correspond to the exponential model and the Whittle model. In the former case $\gamma(h) = \sigma^2 \{1 - \exp(-\|h\|/\tau)\}$ while in the latter case $\gamma(h) = \sigma^2 [1 - (\|h\|/\tau) \Re_1(\|h\|/\tau)]$ We estimate the rest of parameters σ_p^2 and τ_p by fitting the theoretical variogram $\gamma_p(h)$ to the empirical estimates $\hat{\gamma}_p(h)$ based on nonlinear least squares (NLS):

$$\left(\hat{\sigma}_{p}^{2},\hat{\tau}_{p}\right) = \arg\min_{\sigma_{p}^{2},\tau_{p}}\sum_{h}\left\{\sigma_{p}^{2}\left[1-\frac{1}{2^{\nu-1}\Gamma\left(\nu\right)}\left(\frac{\|h\|}{\tau_{p}}\right)^{\nu}\mathfrak{K}_{\nu}\left(\frac{\|h\|}{\tau_{p}}\right)\right]-\hat{\gamma}_{p}\left(h\right)\right\}^{2}.$$

In principle, we can estimate σ_p^2 and τ_p using more efficient estimators such as the MLE or weighted NLS estimator. Here we are content with the above ordinary NLS estimator as the

approximating model we specified is not necessarily correct. Plugging $\hat{\sigma}_p^2$ and $\hat{\tau}_p$ into (8) yields the data-driven choices $\hat{\kappa}_N^*$, $K_1^* = \lfloor n_1 \hat{\kappa}_N^* \rfloor$, and $K_2^* = \lfloor n_2 \hat{\kappa}_N^* \rfloor$.

5 Empirical Implementation

In practical situations, observations may not be located on a regular integer lattice. More often than not, the observation locations are characterized by two covariates, say $L_m = (L_{1m}, L_{2m})$ for m = 1, ..., N. To map the observation locations to an integer lattice, we can follow the steps below:

(i) Normalize $L_m = (L_{1m}, L_{2m})$ to obtain $\tilde{L}_m = (\tilde{L}_{1m}, \tilde{L}_{2m})$ where

$$\tilde{L}_{1m} = \frac{L_{1m} - \min_{m} (L_{1m})}{median_{m}(\min_{\breve{m}, L_{1m} \neq L_{1\breve{m}}} | L_{1m} - L_{1\breve{m}} |)},\\ \tilde{L}_{2m} = \frac{L_{2m} - \min_{m} (L_{2m})}{median_{m}(\min_{\breve{m}, L_{2m} \neq L_{2\breve{m}}} | L_{2m} - L_{2\breve{m}} |)}.$$

We employ the normalization to ensure that L_m is invariant to the units of measurement in each coordinate. Other normalizations are possible.

(ii) On the basis of $\{\tilde{L}_m, m = 1, ..., N\}$, compute $\rho_0 = \min_{m \neq \ell} \left\{ \max\left(|\tilde{L}_{1m} - \tilde{L}_{1\ell}|, |\tilde{L}_{2m} - \tilde{L}_{2\ell}| \right) \right\},$

which is the minimum of pairwise distances.

(iii) Compute $j_m = (j_{1m}, j_{2m})$

$$j_{1m} = \left\lceil \frac{\tilde{L}_{1m}}{\rho_0} \right\rceil, j_{2m} = \left\lceil \frac{\tilde{L}_{2m}}{\rho_0} \right\rceil$$

where $\lceil \cdot \rceil$ is the ceiling function.

These steps map each location (L_{1m}, L_{2m}) in the original space into a unique point (j_{1m}, j_{2m}) in the integer lattice $B_n \subset \mathbb{Z}^+ \otimes \mathbb{Z}^+$. The mapping is similar to the one that is used in Conley (1999). Here we have implicitly assumed that $\rho_0 > 0$, a necessary condition for the increasing domain asymptotics.

With the lattice mapping, we can follow the steps below to compute the test statistic:

(i) Set $n_1 = \max\{j_{1m}, m = 1, ..., N\}$ and $n_2 = \max\{j_{2m}, m = 1, ..., N\}$

(ii) For $k_1 = 0, ..., K_1$ and $k_2 = 0, ..., K_2$, compute

$$\mathcal{A}_{k_1,k_2}\left(\hat{\theta}_N\right) = \frac{1}{\sqrt{N}} \sum_{(j_1,j_2)\in B_n} \exp\left[-2\pi i \left(\frac{k_1 j_1}{n_1} + \frac{k_2 j_2}{n_2}\right)\right] s_{j_1,j_2}\left(\hat{\theta}_N\right),$$

and $\hat{\Omega} = K^{-1} \sum_{k_1=0}^{K_1} \sum_{k_2=0}^{K_2} \operatorname{Re} \left[\mathcal{A}_{k_1,k_2} \left(\hat{\theta}_N \right) \mathcal{A}_{k_1,k_2}^* \left(\hat{\theta}_N \right) \right].$

(iii) Plugging $\hat{\Omega}$ into (3) or (4), we obtain the test statistic F_N or t_N .

We can follow a similar procedure to compute $\hat{\Omega}_Q$ and the associated F statistic or t statistic.

6 Simulation Study

This section provides some simulation evidence on the finite sample performance of our proposed nonstandard test and the F test.

The model we consider is a linear regression model:

$$y_j = \alpha + x'_j \beta + \varepsilon_j \tag{9}$$

where x_j and $\beta = (\beta_1, \beta_2)'$ are 2×1 vectors and $\{j = (j_1, j_2)\}$ indicates the lattice points where the observations are located. Without loss of generality, we set $\alpha = 0$. We generate x_j and ε_j according to

$$x_j = \sum_{|a| \le 2} \gamma^{|a|} u_{j+a}^x, \ \varepsilon_j = \sum_{|a| \le 2} \gamma^{|a|} u_{j+a}^\varepsilon, \tag{10}$$

where $|a| = \max(|a_1|, |a_2|)$, $u_j^x \sim iidN(0, \mathbb{I}_{d_x})$, $u_j^{\varepsilon} \sim iidN(0, 1)$, and $\{u_j^x\}$ is independent of $\{u_j^{\varepsilon}\}$. We consider four different values of $\gamma = 0, 0.3, 0.6, 0.9$. We also consider two different sampling regions. The first is a circular lattice given in Figure 1. The sample size is N = 529. The other is a square lattice. For the square lattice, we introduce two different designs: one uses a full lattice and the other uses a sparse lattice. For the full lattice case, we take a regular 25×25 lattice and use the data generated at each location for a total sample size of N = 625. For the sparse lattice case, we generate the data on the full 36×36 lattice but then randomly sample (without replacement) 625 of the potential 1296 locations. We condition on the same set of 625 locations in each of the simulation replications. Our simulation results are not sensitive to the initial sampling of the locations. The sparse square lattice together with the circular lattice are presented in Figure 1.

To explore the size properties, we generate the data under the following null hypothesis:

$$H_{0q}: R_q(\alpha, \beta')' = \mathbf{0}_{q \times 1} \tag{11}$$

where $R_q = (\mathbf{0}_{q \times (d_x+1-q)}, \mathbb{I}_q)$. Thus, for this testing problem, $\mathfrak{r} = \mathbf{0}_{q \times 1}$. When q = 1, the null is $\beta_2 = 0$; when q = 2, the null is $\beta_1 = \beta_2 = 0$. We set the significance level to be 5%, which is also the nominal size. We compute the empirical size based on 5000 simulation replications.

For the circular lattice, we first consider the two tests proposed in this paper. The first uses the complex exponentials as the basis functions and employs simulated critical values from the nonstandard asymptotically equivalent distribution. We refer to this test as the



Figure 1: Circular Lattice and Sparse Square Lattice

series nonstandard test (Series SIM). The other uses the transformed basis matrix via the Gram-Schmidt procedure and employs critical values from the standard F distribution. We refer to this test as the orthonormal series asymptotic F test (OS F). For both tests, we consider the corresponding tests that use critical values from the χ_q^2/q distribution, leading to the series χ^2 test and the OS χ^2 test. For comparative purposes, we also consider two alternative tests: the test based on the Gaussian kernel estimator and the nonstandard fixed-*b* asymptotic distribution (BCHV); and the χ^2 test based on the Gaussian kernel estimator and the same DGP as above but with $\gamma = 0$. The Gaussian kernel we use here is:

$$G(j_1, j_2) = \exp\left(-\frac{2 \|j_1 - j_2\|^2}{d^2}\right),$$

where d is the smoothing parameter. This Gaussian kernel is also considered by BCHV.

We use the data-driven smoothing parameters. For the series variance estimator, the data-driven choice of K_1 and K_2 is given in (8). For the above Gaussian kernel variance estimator, we can derive the MSE-optimal d. Following Kim and Sun (2011), we find that the MSE-optimal d is

$$d_N^{\star} = \left(\frac{16\left[vec\left(\mathcal{B}_G\right)\right]' Vvec(\mathcal{B}_G)}{\pi \operatorname{tr}\left[V(\Omega \otimes \Omega)(\mathbb{I}_{d^2} + \mathcal{K}_{dd})\right]}\right)^{1/6} N^{1/6}$$
(12)

for $\mathcal{B}_G = \Omega^{(11)} + \Omega^{(22)}$, where $\Omega^{(11)}$ and $\Omega^{(22)}$ are defined in (7). Using the same plug-in

procedure as before, we can estimate d_N^{\star} by

$$\hat{d}_N^{\star} = 3.0968 \left(\frac{(\nu+1)^2 \sum_{p=1}^d \hat{\sigma}_p^4 \hat{\tau}_p^8}{\sum_{p=1}^d \hat{\sigma}_p^4 \hat{\tau}_p^4} \right)^{1/6} N^{1/6} = 0.78971 \left(\hat{\kappa}_N^* \right)^{-1},$$

which is a data-driven choice of d when V assigns equal weights only to the diagonal elements of Ω .

Table 1 reports the empirical size of the tests in the circular lattice case. As it is clear from the table, the size distortion tends to increase with spatial dependence and the number of restrictions being jointly tested. Compared to the tests based on the fixed-smoothing asymptotics, the conventional series and kernel based chi-square tests tend to suffer more from over-rejection. For example, when $\gamma = 0.9$ and q = 2, the empirical type I errors of the series χ^2 , OS χ^2 and kernel χ^2 tests with $\nu = 0.5$ are 0.282, 0.268 and 0.354 respectively. In contrast, the fixed-K asymptotic tests and the BCHV test succeed in reducing the size distortion significantly. The empirical type I errors for the Series SIM, OS F, and BCHV tests are 0.095, 0.097 and 0.101 respectively. Overall, tests based on the fixed-smoothing asymptotics have more or less similar size distortion. Results not reported here show that the selected K_1 and K_2 values decrease with the spatial dependence. So our data-driven smoothing parameter selection procedure effectively captures the degree of spatial dependence.

For the square lattice configuration, we consider two sets of testing procedures. The first set consists of the series χ^2 test and the series F test, which employs an F distribution as the reference distribution based on Corollary 2. We do not need to conduct any transformation of basis functions because complex exponentials are orthonormal and integrated to zero on the full square lattice. The second set consists of the BCHV and kernel χ^2 tests. For the square lattice, we consider *a priori* fixed smoothing parameter choice as well as data-driven smoothing parameter choice. In the former case, we set $K_1 = K_2 = 1$ or 2 for the series estimator and d = 6 or 12 for the kernel estimator.

Table 2 gives the empirical size of the tests when the smoothing parameters are fixed a priori. Both full lattice results and sparse lattice results are reported. We see that the conventional chi-square tests can have large size distortion, especially when the amount of smoothing is small (small K or large d). As expected, the series F test and the BCHV test are more accurate in size. Comparing the full lattice results with the sparse lattice results, we find that the size distortion is smaller for the sparse lattice. Having a sparse lattice is analogous to have a weaker spatial dependence. Similar results were found by BCHV.

Table 3 presents the empirical size as in Table 2 except that the smoothing parameters are data-driven. The qualitative observations from Tables 1 and 2 remain valid.

In Tables 1 and 3, all the fixed smoothing asymptotic tests have slightly more accurate size when $\nu = 1/2$ as compared to $\nu = 1$. Simulation results not reported here show that the smaller value of ν delivers smaller amount of smoothing (smaller K and larger d). Table 2 also shows that the smaller the amount of smoothing is, the more accurate in size the series F test and the BCHV test are. Thus, Tables 1-3 reveal that the fixed-smoothing asymptotic tests are most accurate when the amount of smoothing is small. Finally, as before, the size distortion for the sparse lattice case is smaller compared to the full lattice case. The selected amount of smoothing is larger (larger K and smaller d). The larger the amount of smoothing is, the smaller the difference between the fixed-smoothing asymptotics and the increasing-smoothing asymptotics is. So it is not surprising to see that the fixed-smoothing asymptotic tests do not reduce the size distortion of their respective χ^2 tests by a large margin.

Figures 2-3 present the finite sample power of different testing procedures with $\nu = 0.5$. We use the circular lattice presented above and employ the DGP in (9) and (10). We consider the following local alternative hypothesis:

$$H_{1q}\left(\delta^{2}\right): R_{q}(\alpha, \beta')' = c_{q}/\sqrt{N}$$

$$\tag{13}$$

where $c_q = \left(R_q E(X_j X'_j)^{-1} \left[N^{-1} \sum_{\ell \in B_n} \sum_{m \in B_n} E(\varepsilon_\ell \varepsilon_m X_\ell X'_m)\right] E(X_j X'_j)^{-1} R'_q\right)^{-1/2} \tilde{c}_q$ with $X_\ell = (1, x'_\ell)'$ and \tilde{c}_q is uniformly distributed over a sphere with radius δ , that is, $\tilde{c}_q = \delta \psi / \|\psi\|, \psi \sim N(\mathbf{0}_{q \times 1}, \mathbb{I}_q)$. The scaling matrix before \tilde{c}_q is computed by simulation. Unrestricted parameters are set at their default value of zero.

We compute the power using the 5% empirical critical values under the null and with data-driven smoothing parameter choice. Thus the finite sample power is size-adjusted and power comparison among tests with different HAR variance estimators is meaningful. Of course, the size adjustment is not feasible in practice. We consider the testing procedures with three different HAR variance estimators: series estimator (Series), series estimator with the transformed basis functions by the Gram-Schmidt procedure (OS), and Gaussian kernel estimator (Gauss). The smoothing parameters are data-driven. As illustrated in Figures 2-3, we do not see any significant power difference among the testing procedures.

7 Empirical Application

In this section, we revisit Conley and Udry (2010, CU hereafter) and apply our fixed-K asymptotic tests to their main regression.

CU study the role of social learning in the diffusion of a new agricultural technology in Ghana. More specifically, they investigate the effect of news about fertilizer productivity from a farmer's information neighborhood on the future innovation in his fertilizer use in pineapple production. The base regression specification in their paper is:

$$\Delta x_{i,t} = \beta_1 M_{i,t} + \beta_2 \Gamma_{i,t} + z'_{i,t} \beta_3 + v_{i,t}.$$
(14)

 $\Delta x_{i,t}$ is farmer i's change in input use (the amount of fertilizer) from his previous planting

opportunity at period t. $M_{i,t}$ is the index of input levels associated with good news from farmer *i*'s information neighborhood. $\Gamma_{i,t}$ is defined as a deviation of the average fertilizer use in plots that are proximate to plot *i* at time *t* from plot *i*'s previous fertilizer use. This variable controls movements in $x_{i,t}$ induced by spatially and temporally correlated growing conditions. CU also introduce a regressor that is similar to $\Gamma_{i,t}$ but based on financial neighborhoods instead of geographic neighborhoods. $z_{i,t}$ includes wealth, soil characteristics, indicators for religion, clan, village, round of the planting, and the experience of farmers (novice vs. veteran). The data set is composed of observations on 107 changes in fertilizer use by 47 farmers for a two-year survey (1996–1998). For more details, see sections II and III in CU.

For inference, CU use the Bartlett kernel covariance estimator $\hat{\Omega}_{CU}$ suggested by Conley (1999). The estimator is defined as

$$\hat{\Omega}_{CU} = \frac{1}{N} \sum_{i,j} s_{L_{1i},L_{2i}} \left(\hat{\theta}_{N}\right) W_{CU} \left(\frac{L_{1i} - L_{1j}}{L_{1,cut}}, \frac{L_{2i} - L_{2j}}{L_{2,cut}}\right) s_{L_{1j},L_{2j}}' \left(\hat{\theta}_{N}\right),$$

where

$$W_{CU}\left(\frac{L_{1i} - L_{1j}}{L_{1,cut}}, \frac{L_{2i} - L_{2j}}{L_{2,cut}}\right) = \left(1 - \left|\frac{L_{1i} - L_{1j}}{L_{1,cut}}\right|\right) \mathbf{1} \left\{ \left|\frac{L_{1i} - L_{1j}}{L_{1,cut}}\right| \le 1 \right\}$$
$$\times \left(1 - \left|\frac{L_{2i} - L_{2j}}{L_{2,cut}}\right|\right) \mathbf{1} \left\{ \left|\frac{L_{2i} - L_{2j}}{L_{2,cut}}\right| \le 1 \right\}$$

and (L_{1j}, L_{2j}) are horizontal and vertical coordinates of plot j. They choose $(L_{1,cut}, L_{2,cut}) = (1500m, 1500m)$ to be the truncation lags.

To implement our procedures, we first transform the observation locations into integer lattice points, following the steps presented in Section 5. We then obtain K_1^* and K_2^* using our data driven selection procedure. To compare our results with those in CU, we also consider the values for K_1 and K_2 that are approximately equivalent to the truncation lags used in CU. Following the steps in Section 5, we find the CU's choice of truncation lags corresponds to $(L_{1,cut}^{lattice}, L_{2,cut}^{lattice}) = (115, 150)$ in the transformed lattice. As the effective truncation lags for the series estimator are $n_1/[2(K_1 + 1)]$ and $n_2/[2(K_2 + 1)]$, we set

$$K_1^{CU} = \left\lceil \frac{n_1}{2L_{1,cut}^{lattice}} \right\rceil - 1 \text{ and } K_2^{CU} = \left\lceil \frac{n_2}{2L_{2,cut}^{lattice}} \right\rceil - 1,$$

where $n_1 = 1870, n_2 = 873$. We have $K_1^{CU} = 8$ and $K_2^{CU} = 2$.

Tables 4-6 report the OLS estimates and the p-values for the null hypothesis of no social learning. The p-values are calculated in 5 different ways. The numbers in (·) are p-values using the testing method in CU. The p-values in $\{\cdot\}$ and $\{\cdot\}_{CU}$ are from the series nonstandard test with smooth parameters (K_1^*, K_2^*) and (K_1^{CU}, K_2^{CU}) respectively. Those in [·] and $[\cdot]_{CU}$ are from the asymptotic F test.

Column A in Table 4 presents the results from the base regression model (14). In this column, we find that the coefficient of M_{it} is significantly positive under all the tests considered. This is strong evidence that a farmer adjusts the fertilizer use in response to the news in fertilizer productivity of his information neighbors. Column B in Table 4 estimates the coefficients of M_{it} separately for novice and veteran farmers. While novice farmers' change in fertilizer use is significantly associated with the news in their information neighborhood, this is not the case for veteran farmers.

Columns C-F in Tables 5 and 6 report estimation results from the models which divide information neighborhoods into subgroups according to experience, wealth, farm size and similarity of soil respectively and use them as separate regressors. In column C, we see that while the coefficient of a veteran neighbor's input information is significantly positive, the coefficient of a novice neighbor's input information is not. Column D indicates that information only from neighbors with the same wealth is significantly associated with a farmer's fertilizer input change. In columns E and F, we can see that M_{it} is significantly associated with farmer *i*'s fertilizer input regardless of the neighbor's farm size or similarity of soil. In addition, our tests as well as CU's imply that the spatially and temporally correlated growing conditions are significantly associated with a farmer's input change under any model specification.

Overall, the proposed tests yield very similar results to CU at the 5% level, which reinforces their empirical findings that farmers tend to increase (decrease) the level of fertilizer use when their information neighbors achieve higher than expected profit with more (less) fertilizer than they previously used. However, it is clear that the p-values obtained from the fixed-Kasymptotic tests are larger than those from CU in most cases. In particular, our proposed tests show that the coefficient of "wealth" is not significantly different from zero at 5% level while the test by CU indicates its statistical significance under some model specifications.

8 Conclusion

The paper studies series HAR variance estimation and inference that are robust to spatial autocorrelation. The proposed tests are more accurate in size than the conventional chisquare test because they are based on the fixed-smoothing asymptotics that captures the randomness of the variance estimator. We establish the fixed-smoothing asymptotic theory under very general conditions that accommodate a wide range of spatial data in practice. Among the tests we propose, the F test is asymptotically valid regardless of the sampling region, spatial correlation, and limiting behavior of the smoothing parameter. The F test is especially convenient in empirical applications, as the critical values are from standard F distributions.

In this paper, we focus on the asymptotic MSE criterion, which may not be most suitable for hypothesis testing or CI construction. It is interesting to extend the methods by Sun, Phillips and Jin (2008) on time series HAR estimation to the spatial setting. The idea of using a CLT rather than a more demanding functional CLT in establishing the fixed-smoothing asymptotics can be used for both kernel and series HAR variance estimators in the time series setting and for kernel HAR variance estimators in the spatial setting.

| | | | q = 1 | | | |
|---|--|--|---|--|--|---|
| | Series SIM | | OS | F | Gaussian | (fixed-b) |
| | $\nu = 1/2$ | $\nu = 1$ | $\nu = 1/2$ | $\nu = 1$ | $\nu = 1/2$ | $\nu = 1$ |
| $\gamma = 0$ | 0.080 | 0.080 | 0.050 | 0.050 | 0.082 | 0.082 |
| $\gamma = 0.3$ | 0.108 | 0.135 | 0.094 | 0.113 | 0.097 | 0.117 |
| $\gamma = 0.6$ | 0.070 | 0.104 | 0.085 | 0.101 | 0.082 | 0.098 |
| $\gamma = 0.9$ | 0.082 | 0.118 | 0.094 | 0.113 | 0.087 | 0.107 |
| | Series | χ^2 | OS ; | χ^2 | Gaussia | $n(\chi^2)$ |
| | $\nu = 1/2$ | $\nu = 1$ | $\nu = 1/2$ | $\nu = 1$ | $\nu = 1/2$ | $\nu = 1$ |
| $\gamma = 0$ | 0.052 | 0.053 | 0.052 | 0.052 | 0.052 | 0.052 |
| $\gamma = 0.3$ | 0.115 | 0.122 | 0.115 | 0.125 | 0.140 | 0.122 |
| $\gamma = 0.6$ | 0.152 | 0.137 | 0.148 | 0.135 | 0.208 | 0.188 |
| $\gamma = 0.9$ | 0.157 | 0.147 | 0.153 | 0.145 | 0.212 | 0.193 |
| | | | | | | |
| | | | q = 2 | | | |
| | Series | SIM | q = 2 OS | F | Gaussian | (fixed-b) |
| | Series $\nu = 1/2$ | $\frac{\text{SIM}}{\nu = 1}$ | q = 2 OS $\nu = 1/2$ | F $ u = 1 $ | Gaussian $\nu = 1/2$ | (fixed-b) $\nu = 1$ |
| $\gamma = 0$ | Series $\nu = 1/2$ 0.051 | SIM $\nu = 1$ 0.050 | q = 2 OS $\nu = 1/2$ 0.053 | F $\nu = 1$ 0.055 | Gaussian $\nu = 1/2$ 0.051 | $\begin{array}{l} \text{(fixed-b)} \\ \nu = 1 \\ 0.050 \end{array}$ |
| $\gamma = 0$ $\gamma = 0.3$ | Series $\nu = 1/2$ 0.051 0.092 | SIM $\nu = 1$ 0.050 0.119 | q = 2 OS $\nu = 1/2$ 0.053 0.103 | F $\nu = 1$ 0.055 0.132 | Gaussian $\nu = 1/2$ 0.051 0.089 | (fixed- <i>b</i>) $\nu = 1$ 0.050 0.094 |
| $ \begin{aligned} \gamma &= 0 \\ \gamma &= 0.3 \\ \gamma &= 0.6 \end{aligned} $ | Series $\nu = 1/2$ 0.051 0.092 0.088 | SIM $\nu = 1$ 0.050 0.119 0.103 | q = 2 OS $\nu = 1/2$ 0.053 0.103 0.087 | F $\nu = 1$ 0.055 0.132 0.110 | Gaussian $ \nu = 1/2 $ 0.051 0.089 0.089 | (fixed-b) $\nu = 1$ 0.050 0.094 0.104 |
| $ \begin{aligned} \gamma &= 0 \\ \gamma &= 0.3 \\ \gamma &= 0.6 \\ \gamma &= 0.9 \end{aligned} $ | Series $ \nu = 1/2 $ 0.051 0.092 0.088 0.095 | SIM $\nu = 1$ 0.050 0.119 0.103 0.117 | $q = 2$ OS $\nu = 1/2$ 0.053 0.103 0.087 0.097 | F $\nu = 1$ 0.055 0.132 0.110 0.128 | Gaussian $\nu = 1/2$ 0.051 0.089 0.089 0.101 | (fixed-b) $\nu = 1$ 0.050 0.094 0.104 0.115 |
| $egin{aligned} &\gamma=0\ &\gamma=0.3\ &\gamma=0.6\ &\gamma=0.9 \end{aligned}$ | $Series \\ \nu = 1/2 \\ 0.051 \\ 0.092 \\ 0.088 \\ 0.095 \\ Series$ | SIM $\nu = 1$ 0.050 0.119 0.103 0.117 $5 \chi^2$ | $q = 2$ OS $\nu = 1/2$ 0.053 0.103 0.087 0.097 OS | F = 1 = 1 = 0.055 = 0.132 = 0.110 = 0.128 = | Gaussian $\nu = 1/2$ 0.051 0.089 0.089 0.101 Gaussia | $(fixed-b) \\ \nu = 1 \\ 0.050 \\ 0.094 \\ 0.104 \\ 0.115 \\ m (\chi^2)$ |
| $egin{aligned} &\gamma=0 \ &\gamma=0.3 \ &\gamma=0.6 \ &\gamma=0.9 \end{aligned}$ | Series $ \nu = 1/2 $ 0.051 0.092 0.088 0.095 Series $ \nu = 1/2 $ | SIM $\nu = 1$ 0.050 0.119 0.103 0.117 $\frac{1}{5}\chi^{2}$ $\nu = 1$ | $\begin{array}{c} q = 2 \\ \text{OS} \\ \nu = 1/2 \\ 0.053 \\ 0.103 \\ 0.087 \\ 0.097 \\ \text{OS} \\ \nu = 1/2 \end{array}$ | F $\nu = 1$ 0.055 0.132 0.110 0.128 χ^2 $\nu = 1$ | Gaussian $ \nu = 1/2 $ 0.051 0.089 0.089 0.101 Gaussia $ \nu = 1/2 $ | (fixed-b) $\nu = 1$ 0.050 0.094 0.104 0.115 un (χ^2) $\nu = 1$ |
| $ \begin{array}{c} \gamma = 0 \\ \gamma = 0.3 \\ \gamma = 0.6 \\ \gamma = 0.9 \end{array} \end{array} $ | Series $\nu = 1/2$ 0.051 0.092 0.088 0.095 Series $\nu = 1/2$ 0.056 | SIM $\nu = 1$ 0.050 0.119 0.103 0.117 $\sqrt{2}$ $\nu = 1$ 0.056 | $\begin{array}{c} q = 2 \\ \text{OS} \\ \nu = 1/2 \\ 0.053 \\ 0.103 \\ 0.087 \\ 0.097 \\ \text{OS} \\ \nu = 1/2 \\ 0.057 \end{array}$ | F = 1 = 0.055 = 0.132 = 0.110 = 0.128 = 0.128 = 0.05 | Gaussian $\nu = 1/2$ 0.051 0.089 0.089 0.101 Gaussia $\nu = 1/2$ 0.060 | $\begin{array}{c} \text{(fixed-b)} \\ \nu = 1 \\ 0.050 \\ 0.094 \\ 0.104 \\ 0.115 \\ \text{in } (\chi^2) \\ \nu = 1 \\ 0.059 \end{array}$ |
| $\begin{array}{c} \gamma = 0 \\ \gamma = 0.3 \\ \gamma = 0.6 \\ \gamma = 0.9 \end{array}$ $\begin{array}{c} \gamma = 0 \\ \gamma = 0.3 \end{array}$ | Series $\nu = 1/2$ 0.051 0.092 0.088 0.095 Series $\nu = 1/2$ 0.056 0.166 | $SIM \\ \nu = 1 \\ 0.050 \\ 0.119 \\ 0.103 \\ 0.117 \\ 5 \chi^2 \\ \nu = 1 \\ 0.056 \\ 0.165 \\ 0.165 \\ 0.165 \\ 0.165 \\ 0.056 \\ 0.165 \\ 0.056 \\ 0.165 \\ 0.056 \\ 0.055 \\ 0.056 \\ 0.055 $ | $\begin{array}{c} q = 2 \\ \text{OS} \\ \nu = 1/2 \\ 0.053 \\ 0.103 \\ 0.087 \\ 0.097 \\ \hline \text{OS} \\ \nu = 1/2 \\ 0.057 \\ 0.161 \end{array}$ | $F = 1$ 0.055 0.132 0.110 0.128 χ^{2} $\nu = 1$ 0.058 0.169 | Gaussian $\nu = 1/2$ 0.051 0.089 0.089 0.101 Gaussia $\nu = 1/2$ 0.060 0.229 | $(fixed-b) \nu = 1 0.050 0.094 0.104 0.115 un (\chi^2)\nu = 1 0.059 0.174$ |
| $\begin{array}{c} \gamma = 0 \\ \gamma = 0.3 \\ \gamma = 0.6 \\ \gamma = 0.9 \end{array}$ $\begin{array}{c} \gamma = 0 \\ \gamma = 0.3 \\ \gamma = 0.6 \end{array}$ | $Series \\ \nu = 1/2 \\ 0.051 \\ 0.092 \\ 0.088 \\ 0.095 \\ Series \\ \nu = 1/2 \\ 0.056 \\ 0.166 \\ 0.275 \\ \end{cases}$ | SIM $\nu = 1$ 0.050 0.119 0.103 0.117 χ^2 $\nu = 1$ 0.056 0.165 0.235 | $\begin{array}{c} q = 2 \\ & \text{OS} \\ \nu = 1/2 \\ 0.053 \\ 0.103 \\ 0.087 \\ 0.097 \\ \hline \\ \text{OS} \\ \nu = 1/2 \\ 0.057 \\ 0.161 \\ 0.258 \end{array}$ | $F = 1$ 0.055 0.132 0.110 0.128 χ^{2} $\nu = 1$ 0.058 0.169 0.228 | Gaussian $\nu = 1/2$ 0.051 0.089 0.089 0.101 Gaussia $\nu = 1/2$ 0.060 0.229 0.346 | $(fixed-b) \nu = 1 0.050 0.094 0.104 0.115 on (\chi^2) \nu = 1 0.059 0.174 0.314 $ |

Table 1: Empirical size of different tests with data-driven smoothing parameters and different number of restrictions in the circular lattice case

Note: 'Series SIM' denotes the series nonstandard test and 'OS F' denotes the orthonormal series asymptotic F test. 'Gaussian (fixed-b)' is the test developed in BCHV. 'Series χ^2 ', 'OS χ^2 test' and 'Gaussian (χ^2)' are conventional χ^2 tests.

| | Series F | | Series χ^2 | | Gaussia | n (fixed-b) | Gaussian (χ^2) | |
|----------------|-----------|-----------|---------------------|------------|---------|--------------|---------------------|--------|
| | $K_i = 1$ | $K_i = 2$ | $K_i = 1$ | $K_i = 2$ | d = 6 | d = 12 | d = 6 | d = 12 |
| | | | Reg | gular Latt | tice | | | |
| $\gamma = 0$ | 0.047 | 0.045 | 0.092 | 0.065 | 0.046 | 0.043 | 0.066 | 0.120 |
| $\gamma = 0.3$ | 0.062 | 0.074 | 0.118 | 0.098 | 0.072 | 0.060 | 0.103 | 0.149 |
| $\gamma = 0.6$ | 0.071 | 0.095 | 0.135 | 0.122 | 0.098 | 0.074 | 0.129 | 0.170 |
| $\gamma = 0.9$ | 0.079 | 0.109 | 0.143 | 0.136 | 0.109 | 0.079 | 0.141 | 0.174 |
| | | | Sp | arse Latt | ice | | | |
| $\gamma = 0$ | 0.058 | 0.054 | 0.104 | 0.072 | 0.047 | 0.060 | 0.068 | 0.089 |
| $\gamma = 0.3$ | 0.056 | 0.063 | 0.110 | 0.085 | 0.060 | 0.066 | 0.086 | 0.103 |
| $\gamma = 0.6$ | 0.060 | 0.076 | 0.117 | 0.099 | 0.078 | 0.075 | 0.103 | 0.115 |
| $\gamma=0.9$ | 0.063 | 0.083 | 0.121 | 0.104 | 0.086 | 0.080 | 0.112 | 0.119 |

Table 2: Empirical size of different tests with $a \ priori$ fixed smoothing parameters in the square lattice case

Note: 'Series F' is the test developed in this paper. 'Gaussian (fixed-b)' is the test developed in BCHV. 'Series χ^2 ' and 'Gaussian (χ^2)' are conventional χ^2 tests.

| | Series F | | Series χ^2 | | Gaussian | (fixed-b) | Gaussian (χ^2) | | |
|----------------|----------------|-----------|-----------------|-----------|-------------|-----------|---------------------|-----------|--|
| | $\nu = 1/2$ | $\nu = 1$ | $\nu = 1/2$ | $\nu = 1$ | $\nu = 1/2$ | $\nu = 1$ | $\nu = 1/2$ | $\nu = 1$ | |
| | | | Reg | gular La | ttice | | | | |
| $\gamma = 0$ | 0.045 | 0.045 | 0.047 | 0.047 | 0.040 | 0.039 | 0.045 | 0.045 | |
| $\gamma = 0.3$ | 0.076 | 0.092 | 0.097 | 0.107 | 0.067 | 0.070 | 0.121 | 0.106 | |
| $\gamma = 0.6$ | 0.072 | 0.079 | 0.136 | 0.127 | 0.078 | 0.078 | 0.170 | 0.161 | |
| $\gamma=0.9$ | 0.081 | 0.095 | 0.142 | 0.135 | 0.080 | 0.084 | 0.173 | 0.166 | |
| | Sparse Lattice | | | | | | | | |
| $\gamma = 0$ | 0.053 | 0.052 | 0.054 | 0.053 | 0.050 | 0.051 | 0.055 | 0.054 | |
| $\gamma = 0.3$ | 0.071 | 0.084 | 0.083 | 0.081 | 0.069 | 0.063 | 0.091 | 0.086 | |
| $\gamma = 0.6$ | 0.072 | 0.078 | 0.103 | 0.098 | 0.072 | 0.075 | 0.127 | 0.118 | |
| $\gamma = 0.9$ | 0.080 | 0.086 | 0.108 | 0.105 | 0.073 | 0.080 | 0.130 | 0.122 | |

Table 3: Empirical size of different tests with data-driven smoothing parameters in the square lattice case

| | A | | В | |
|--|-------|--|-------|--|
| Index of Inputs on Successful Experiments $(M_{i,t})$ | 1.05 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.001\}^*_{CU}, [.001]^*_{CU}$ | | |
| $\mathbf{M}_{i,t} \times$ novice farmer | | | 1.07 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.001\}^*_{CU}, [.000]^*_{CU}$ |
| $\mathbf{M}_{i,t} \times$ veteran farmer | | | -0.46 | $(.176), \{.230\}, [.255]$ $\{.182\}_{CU}, [.195]_{CU}$ |
| Novice farmer | | | 3.97 | $(.137), \{.177\}, [.226]$ $\{.267\}_{CU}, [.237]_{CU}$ |
| Avg. dev. of geog. neighbors from previous use $[\Gamma_{i,t}]$ | 0.52 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ | 0.56 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ |
| Avg. dev. of financial neighbors from prev. use | 0.52 | $(.378), \{.462\}, [.492]$ $\{.496\}_{CU}, [.357]_{CU}$ | 0.55 | $(.335), \{.382\}, [.340]$ $\{.360\}_{CU}, [.292]_{CU}$ |
| Village 1 | -7.50 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ | -8.09 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ |
| Village 2 | -0.47 | $(.759), \{.815\}, [.793]$ $\{.816\}_{CU}, [.811]_{CU}$ | -1.91 | $(.356), \{.376\}, [.468]$ $\{.474\}_{CU}, [.482]_{CU}$ |
| Wealth (million cedis) | 0.10 | $(.689), \{.807\}, [.707]$ $\{.770\}_{CU}, [.691]_{CU}$ | 0.41 | $(.016)^*, \{.079\}, [.062]$ $\{.045\}^*_{CU}, [.025]^*_{CU}$ |
| Clan 1 | -2.36 | $(.094), \{.131\}, [.120]$ $\{.076\}_{CU}, [.121]_{CU}$ | -2.44 | $(.051), \{.095\}, [.062]$ $\{.070\}_{CU}, [.056]$ |
| Clan 2 | -0.35 | $(.808), \{.805\}, [.837]$ $\{.831\}_{CU}, [.802]_{CU}$ | 0.00 | $(1.000), \{.995\}, [.999]$ $\{.995\}_{CU}, [.999]_{CU}$ |
| Church 1 | 0.13 | $(.921), \{.932\}, [.948]$ $\{.934\}_{CU}, [.941]_{CU}$ | 0.63 | $(.577), \{.615\}, [.701]$ $\{.604\}_{CU}, [.659]_{CU}$ |

Table 4: Predicting innovations in per plant fertilizer use — differential effects by source of information

Note: The numbers in (\cdot) are p-values obtained from the test by CU. The numbers in $\{\cdot\}$ and $\{\cdot\}_{CU}$ are p-values obtained from the series nonstandard test. The former uses the data driven smoothing parameters (K_1^*, K_2^*) and the latter uses (K_1^{CU}, K_2^{CU}) which are approximately equivalent to the smoothing parameters in CU. The numbers in $[\cdot]$ and $[\cdot]_{CU}$ are p-values obtained from the orthonormal series asymptotic F test with (K_1^*, K_2^*) and (K_1^{CU}, K_2^{CU}) respectively. * denotes statistical significance at the 5% level. For definitions of the variables in the regressions, see CU.

| Table 5: | Predicting | innovations | in per | plant | fertilizer | use — | differential | effects by | ^r source of |
|----------|------------|-------------|--------|-------|------------|-------|--------------|------------|------------------------|
| informat | ion | | | | | | | | |

| | C | | D | |
|------------------------------------|-------|--|-------|---|
| Index of good news input levels | -0.05 | $(.898), \{.944\}, [.954]$ | | |
| by novice farmer | | $\{.908\}_{CU}, [.952]_{CU}$ | | |
| | | () * () + [] * | | |
| Index of good news input levels | 1.05 | $(.000)^*, \{.000\}^*, [.000]^*$ | | |
| by veteran farmer | | $\{.000\}_{CU}^*, [.000]_{CU}^*$ | | |
| | 1 | | 1.00 | (000)*(000)*[000]* |
| Index of good news input levels | | | 1.00 | $(.000)$, $\{.000\}$, $[.000]$ |
| by farmers with same wealth | | | | $\{.000\}_{CU}^{\circ}, [.001]_{CU}$ |
| Index of good news input levels | 1 | | -0.32 | $(317) \{468\} [447]$ |
| by farmers with different wealth | | | 0.02 | $\{291\}_{CH}$ $[420]_{}$ |
| by farmers with different wearth | | | | (.231)CU, $[.120]CU$ |
| Novice farmer | 4.03 | $(.133), \{.162\}, [.225]$ | 4.02 | $(.132), \{.158\}, [.220]$ |
| | | $\{.263\}_{CU}, [.236]_{CU}$ | | $\{.266\}_{CU}, [.232]_{CU}$ |
| | I | | I | |
| Avg. dev. of geog. neighbors | 0.56 | $(.000)^*, \{.000\}^*, [.000]^*$ | 0.57 | $(.000)^*, \{.000\}^*, [.000]^*$ |
| from previous use $[\Gamma_{i,t}]$ | | $\{.000\}_{CU}^*, [.000]_{CU}^*$ | | $\{.000\}_{CU}^{*}, [.000]_{CU}^{*}$ |
| | | | | |
| Avg. dev. of financial | 0.38 | $(.512), \{.561\}, [.529]$ | 0.41 | $(.448), \{.508\}, [.438]$ |
| neighbors from prev. use | | $\{.560\}_{CU}, [.514]_{CU}$ | | $\{.489\}_{CU}, [.426]_{CU}$ |
| 5711 4 | | (000)*(000)*[000]* | 0.10 | (000)*(000)*[000]* |
| Village 1 | -7.97 | $(.000)^*, \{.000\}^*, [.000]^*$ | -8.10 | $(.000)^*, \{.000\}^*, [.000]^*$ |
| | | $\{.000\}_{CU}^{+}, [.000]_{CU}$ | | $\{.000\}_{CU}^{+}, [.000]_{CU}$ |
| Village 2 | _1 94 | $(330) \{354\} [446]$ | _1 98 | (339) { 366} [456] |
| v mage 2 | -1.54 | $\{469\}_{GU}$ $[460]_{H}$ | -1.50 | $\{466\}_{GU}$ $[470]_{HI}$ |
| | | (.402)CU, [.400]CU | | (.400)CU, [.410]CU |
| Wealth (million cedis) | 0.36 | $(.046)^*, \{.174\}, [.113]$ | 0.40 | $(.026)^*, \{.097\}, [.072]$ |
| × , | | $\{.084\}_{CU}, [.048]_{CU}^*$ | | $\{.061\}_{CU}, [.034]_{CU}^*$ |
| | I | | I | |
| Clan 1 | -2.43 | $(.056), \{.095\}, [.063]$ | -2.32 | $(.059), \{.103\}, [.076]$ |
| | | $\{.074\}_{CU}, [.057]_{CU}$ | | $\{.077\}_{CU}, [.070]_{CU}$ |
| - | 1 | | I | |
| Clan 2 | -0.10 | $(.941), \{.939\}, [.949]$ | -0.13 | $(.923), \{.924\}, [.938]$ |
| | | $\{.945\}_{CU}, [.937]_{CU}$ | | $\{.935\}_{CU}, [.926]_{CU}$ |
| Church 1 | 0.49 | (660) [687] [769] | 0.41 | (717) [749] [ONK] |
| | 0.40 | (.000), (.001), [.100] | 0.41 | (····/, (···±), [·000] ∫ 727] ~ [777] |
| | | $1 \cdot 100 \int CU$, $[\cdot 124] CU$ | I | $\bigcup CU, [\cdots]CU$ |
| | | | | |

Table 6: Predicting innovations in per plant fertilizer use — differential effects by source of information

| | E | | F | |
|--|-------|--|-------|--|
| Index of good news input levels on big farms | 1.17 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ | | |
| Index of good news input levels on small farms | 0.92 | $(.000)^*, \{.000\}^*, [.001]^*$ $\{.001\}^*_{CU}, [.002]^*_{CU}$ | | |
| Index of good news input levels, farmers with same soil | | | 1.08 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.001]^*_{CU}$ |
| Index of good news input levels, farmers with different soil | | | 0.93 | $(.000)^*, \{.000\}^*, [.001]^*$ $\{.001\}^*_{CU}, [.002]^*_{CU}$ |
| Novice farmer | 3.96 | $(.141), \{.172\}, [.237]$ $\{.266\}_{CU}, [.249]_{CU}$ | 3.94 | $(.155), \{.187\}, [.252]$ $\{.285\}_{CU}, [.265]_{CU}$ |
| Avg. dev. of geog. neighbors from previous use $[\Gamma_{i,t}]$ | 0.56 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ | 0.57 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ |
| Avg. dev. of financial neighbors from prev. use | 0.23 | $(.711), \{.739\}, [.716]$ $\{.760\}_{CU}, [.708]_{CU}$ | 0.24 | $(.694), \{.720\}, [.704]$ $\{.732\}_{CU}, [.696]_{CU}$ |
| Village 1 | -7.68 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ | -7.79 | $(.000)^*, \{.000\}^*, [.000]^*$ $\{.000\}^*_{CU}, [.000]^*_{CU}$ |
| Village 2 | -1.60 | $(.421), \{.463\}, [.554]$ $\{.531\}_{CU}, [.561]_{CU}$ | -1.59 | $(.434), \{.465\}, [.563]$ $\{.537\}_{CU}, [.570]_{CU}$ |
| Wealth (million cedis) | 0.24 | $(.253), \{.391\}, [.309]$ $\{.326\}_{CU}, [.306]_{CU}$ | 0.26 | $(.216), \{.365\}, [.264]$ $\{.285\}_{CU}, [.267]_{CU}$ |
| Clan 1 | -2.24 | $(.080), \{.117\}, [.091]$ $\{.110\}_{CU}, [.082]_{CU}$ | -2.33 | $(.073), \{.109\}, [.080]$ $\{.097\}_{CU}, [.073]_{CU}$ |
| Clan 2 | -0.26 | $(.843), \{.840\}, [.876]$ $\{.865\}_{CU}, [.849]_{CU}$ | -0.24 | $(.855), \{.858\}, [.885]$ $\{.875\}_{CU}, [.860]_{CU}$ |
| Church 1 | 0.69 | $(.549), \{.593\}, [.678]$ $\{.592\}_{CU}, [.625]_{CU}$ | 0.74 | $(.520), \{.562\}, [.659]$ $\{.569\}^{CU}, [.606]_{CU}$ |



Figure 2: Size-adjusted power of different testing procedures with q = 1 and $\nu = 1/2$ in the circular lattice case



Figure 3: Size-adjusted power of different testing procedures with q = 2 and $\nu = 1/2$ in the circular lattice case

9 Appendix

Primitive Conditions for Assumption 4

In this section, we use the CLT by JP to provide primitive sufficient conditions for Assumption 4. We consider the case when s_j is a scalar and the distribution of $N^{-1/2} \sum_{j \in B_n} \Phi_k(j/n) s_j$ for a given k. The vector case and the joint asymptotic equivalence over $k \in \mathbb{K}$ can be dealt with using the Cramér-Wold device.

Definitions 1 and 2 below are from JP who provide motivations and detailed discussions of these definitions.

Definition 1 For $U \subseteq B_n$ and $V \subseteq B_n$, let $\sigma_n(U) = \sigma(X_j : j \in U)$, $\alpha_n(U, V) = \alpha(\sigma_n(U), \sigma_n(V))$ and $\phi_n(U, V) = \phi(\sigma_n(U), \sigma_n(V))$, where $\alpha(\cdot, \cdot)$ and $\phi(\cdot, \cdot)$ are α -mixing and ϕ -mixing coefficients respectively. Then, the generalized α -mixing and ϕ -mixing coefficients for the random field $\{X_j : j \in B_n\}$ are defined as follows:

$$\begin{aligned} \alpha_{l_{1},l_{2},n}\left(r\right) &= \sup\left(\alpha_{n}\left(U,V\right), |U| \leq l_{1}, |V| \leq l_{2}, \rho\left(U,V\right) \geq r\right), \\ \phi_{l_{1},l_{2},n}\left(r\right) &= \sup\left(\phi_{n}\left(U,V\right), |U| \leq l_{1}, |V| \leq l_{2}, \rho\left(U,V\right) \geq r\right), \end{aligned}$$

where

$$\rho(U,V) = \inf \left\{ \rho(i,j) : i \in U \text{ and } j \in V \right\}.$$

and for $i = (i_1, i_2)$ and $j = (j_1, j_2)$, $\rho(i, j) = \max\{|i_1 - j_1|, |i_2 - j_2|\}$. We further define

$$\bar{\alpha}_{l_{1},l_{2}}(r) = \sup_{n} \alpha_{l_{1},l_{2},n}(r), \qquad \bar{\phi}_{l_{1},l_{2}}(r) = \sup_{n} \phi_{l_{1},l_{2},n}(r),$$

with $l_1, l_2, r, n \in \mathbb{Z}^+$.

Definition 2 (i) The "upper-tail" quantile function $Q_X : (0,1) \to [0,\infty)$ is defined as

$$Q_X(u) = \inf \left\{ t : P(X > t) \le u \right\}.$$

(ii) For the non-increasing sequence of $\{\bar{\alpha}_{1,1}(r)\}_{r=1}^{\infty}$, set $\bar{\alpha}_{1,1}(0) = 1$ and define its "inverse" function $\alpha_{inv}(u): (0,1) \to \mathbb{Z}^+ \cup \{0\}$ as

$$\alpha_{inv}\left(u\right) = \max\left\{r \ge 0 : \bar{\alpha}_{1,1}\left(r\right) > u\right\}.$$

Definition 3 Let $B_n^{int} = \left\{ j \in B_n : \lim_{n \to \infty} \sum_{\ell \in B_n} E\left(s_j s'_\ell\right) = \Omega \right\}$ and $N^{int} = \left| B_n^{int} \right|$.

Lemma 1 Assume (i) there exists an array of positive real constants $\{c_{j,n}\}$ such that

$$\lim_{C \to \infty} \sup_{n} \sup_{j \in B_n} E\left[\left| \Phi_k(\frac{j}{n}) s_j / c_{j,n} \right|^2 \mathbf{1} \left(\left| \Phi_k(\frac{j}{n}) s_j / c_{j,n} \right| > C \right) \right] = 0,$$

where $\mathbf{1}(\cdot)$ is the indicator function, (ii) $\liminf_{n\to\infty} N^{-1}M_n^{-1}\sigma_n^2 > 0$, where $M_n = \max_{j\in B_n} c_{j,n}$ and $\sigma_n^2 = var\left(N^{-1}\sum_{j\in B_n} \Phi_k(j/n)s_j\right)$, (iii) $\{\Phi_k(j/n)s_j : j\in B_n\}$ is either α -mixing satisfying

- (a) $\lim_{C \to \infty} \lim_{n \to \infty} \sup_{j \in B_n} \int_0^1 \alpha_{inv}^2(u) \left(Q_{|\Phi_k(j/n)s_j| 1(|\Phi_k(j/n)s_j| > C)} \right)^2 du = 0,$
- (b) $\sum_{r=1}^{\infty} r \bar{\alpha}_{l_1, l_2}(r) < \infty \text{ for } l_1 + l_2 \le 4,$
- (c) $\bar{\alpha}_{1,\infty}(r) = O\left(r^{-2-\varepsilon}\right)$ for some $\varepsilon > 0$,

or ϕ -mixing satisfying

(a)
$$\sum_{r=1}^{\infty} r \bar{\phi}_{1,1}^{1/2}(r) < \infty$$

- (b) $\sum_{r=1}^{\infty} r \bar{\phi}_{l_1, l_2}(r) < \infty \text{ for } l_1 + l_2 \le 4,$
- (c) $\bar{\phi}_{1,\infty}(r) = O\left(r^{-2-\varepsilon}\right)$ for some $\varepsilon > 0$.

If $\Phi_k(\cdot, \cdot)$ is continuously differentiable with respect to ρ , $N^{-1} \sum_{\ell \in B_n} \sum_{m \in B_n} \rho(\ell, m) \| E(s_\ell s_m) \| = O(1)$, and $N^{int}/N \to 1$, then

$$\frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_k(\frac{j}{n}) s_j \to^d N\left(0, \lim_{n \to \infty} \frac{n_1 n_2}{N} \Omega \int_{x \in B} \Phi_k^2(x) dx\right).$$

Proof of Lemma 1. Without loss of generality, we assume that $\Phi_k(\cdot)$ is real. As the asymptotic normality of $N^{-1/2} \sum_{j \in B_n} \Phi_k(j/n) s_j$ holds by JP (Theorem 1), it suffices to show that

$$\lim_{n \to \infty} var\left(\frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_k(\frac{j}{n}) s_j\right) = \lim_{n \to \infty} \frac{n_1 n_2}{N} \Omega \int_{x \in B} \Phi_k^2(x) dx.$$

Let $\Gamma_{\ell,m} = E(s_{\ell}s_m)$, then

$$\begin{aligned} var\left(\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})s_j\right) \\ &= \frac{1}{N}\sum_{\ell\in B_n}\sum_{m\in B_n}\Phi_k(\frac{\ell}{n})\Phi_k(\frac{m}{n})\Gamma_{\ell,m} \\ &= \frac{1}{N}\sum_{\ell\in B_n}\sum_{m\in B_n}\left(\Phi_k(\frac{\ell}{n})\Phi_k(\frac{m}{n}) - \Phi_k^2(\frac{\ell}{n})\right)\Gamma_{\ell,m} + \frac{1}{N}\sum_{\ell\in B_n}\sum_{m\in B_n}\Phi_k^2(\frac{\ell}{n})\Gamma_{\ell,m} \\ &: = \mathcal{I}_1 + \mathcal{I}_2.\end{aligned}$$

Let $n_{\min} = \min(n_1, n_2)$. First, $\mathcal{I}_1 = o(1)$ because there exists a positive constant C such that

$$\begin{aligned} |\mathcal{I}_{1}| &\leq \frac{1}{N} \sum_{\ell \in B_{n}} \sum_{m \in B_{n}} \left| \Phi_{k}\left(\frac{\ell}{n}\right) \right| \left| \Phi_{k}\left(\frac{m}{n}\right) - \Phi_{k}\left(\frac{\ell}{n}\right) \right| \left| \Gamma_{\ell,m} \right| \\ &\leq \frac{C}{N} \sum_{\ell \in B_{n}} \sum_{m \in B_{n}} \rho\left(\frac{\ell}{n}, \frac{m}{n}\right) \left| \Gamma_{\ell,m} \right| \\ &= \frac{C}{N} \sum_{\ell \in B_{n}} \sum_{m \in B_{n}} \max\left(\frac{\left|\ell_{1} - m_{1}\right|}{n_{1}}, \frac{\left|\ell_{2} - m_{2}\right|}{n_{2}}\right) \left| \Gamma_{\ell,m} \right| \\ &\leq \frac{C}{n_{\min}N} \sum_{\ell \in B_{n}} \sum_{m \in B_{n}} \rho\left(\ell, m\right) \left| \Gamma_{\ell,m} \right| = o\left(1\right), \end{aligned}$$

as *n* increases. Second, for \mathcal{I}_2 , we have

$$\mathcal{I}_2 = \mathcal{I}_{21} + \mathcal{I}_{22}$$

where

$$\mathcal{I}_{21} = \frac{1}{N} \sum_{\ell \in B_n^{int}} \Phi_k^2(\frac{\ell}{n}) \sum_{m \in B_n} \Gamma_{\ell,m}$$
$$\mathcal{I}_{22} = \frac{1}{N} \sum_{\ell \notin B_n^{int}} \sum_{m \in B_n} \Phi_k^2(\frac{\ell}{n}) \Gamma_{\ell,m}.$$

Now for some constant C,

$$\left|\mathcal{I}_{22}\right| \leq \frac{C}{N} \sum_{m \notin B_n^{int}} \sum_{\ell \in B_n} \left|\Gamma_{\ell,m}\right| = o\left(1\right).$$

because $(N - N^{int})/N = o(1)$. Using Fubini's Theorem, we have

$$\begin{split} \lim_{n \to \infty} \mathcal{I}_{21} &= \lim_{n \to \infty} \frac{1}{N} \sum_{\ell \in B_n^{int}} \Phi_k^2(\frac{\ell}{n}) \lim_{n \to \infty} \left(\sum_{m \in B_n} \Gamma_{\ell,m} \right) = \left[\lim_{n \to \infty} \frac{1}{N} \sum_{\ell \in B_n^{int}} \Phi_k^2(\frac{\ell}{n}) \right] \Omega \\ &= \lim_{n \to \infty} \frac{1}{N} \left(\sum_{\ell \in B_n} \Phi_k^2(\frac{\ell}{n}) - \sum_{\ell \notin B_n^{int}} \Phi_k^2(\frac{\ell}{n}) \right) \Omega \\ &= \lim_{n \to \infty} \frac{n_1 n_2}{N} \Omega \left(\frac{1}{n_1 n_2} \sum_{\ell \in B_n} \Phi_k^2(\frac{\ell}{n}) \right) - \lim_{n \to \infty} \frac{N - N^{int}}{N} \frac{\Omega}{N - N^{int}} \sum_{\ell \notin B_n^{int}} \Phi_k^2(\frac{\ell}{n}) \\ &= \lim_{n \to \infty} \frac{n_1 n_2}{N} \Omega \int_B \Phi_k^2(x) \, dx. \end{split}$$

Combining the above analyses yields the desired result. \blacksquare

Proofs of Main Results

Proof of Theorem 1. Under Assumption 3, we have

$$\begin{split} &\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})\hat{s}_j \\ &= (1+o_p(1))\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})G'\mathcal{W}f_j(\hat{\theta}_N) \\ &= (1+o_p(1))\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})G'\mathcal{W}\left\{f_j(\theta_0) + \frac{\partial f_j(\tilde{\theta}_N)}{\partial \theta'}\left[\hat{\theta}_N - \theta_0\right]\right\} \\ &= (1+o_p(1))\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})G'\mathcal{W}\left\{f_j(\theta_0) + \frac{\partial f_j(\theta_0)}{\partial \theta'}\left[\hat{\theta}_N - \theta_0\right]\right\} \\ &= (1+o_p(1))\left\{\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})s_j - \frac{1}{N}\sum_{j\in B_n}\Phi_k(\frac{j}{n})G'\mathcal{W}\frac{\partial f_j(\theta_0)}{\partial \theta'}H^{-1}\left[\frac{1}{\sqrt{N}}\sum_{\tilde{j}\in B_n}s_{\tilde{j}}\right]\right\} \\ &= (1+o_p(1))\left\{\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})s_j - \frac{1}{N}\sum_{j\in B_n}\Phi_k(\frac{j}{n})\frac{\partial s_j(\theta_0)}{\partial \theta'}H^{-1}\left[\frac{1}{\sqrt{N}}\sum_{\tilde{j}\in B_n}s_{\tilde{j}}\right]\right\}. \end{split}$$

Let $j_{\min} = \min \{j_1 : (j_1, j_2) \in B_n\}$ and $j_{\max} = \max \{j_1 : (j_1, j_2) \in B_n\}$. For each $j_1 \in [j_{\min}, j_{\max}]$, we partition $\{(j_1, j_2) : (j_1, j_2) \in B_n\}$ into maximal subsets such that each set has consecutive second coordinates j_2 . That is,

$$\{(j_1, j_2) : (j_1, j_2) \in B_n\} = \cup_{j_1=j_{\min}}^{j_{\max}} \cup_{\ell=1}^{L_{j_1}} \left\{ (j_1, j_2) : j_2 \in \left[J_{\min}^{\ell, j_1}, J_{\max}^{\ell, j_1}\right] \right\}$$

for some integers $L_{j_1}, J_{\min}^{\ell,j_1}$ and J_{\max}^{ℓ,j_1} that depend on j_1 . We ensure that $J_{\min}^{\ell+1,j_1} \neq J_{\max}^{\ell,j_1}$ for all j_1 and $\ell = 1, ..., L_{j_1} - 1$. If this happens, we combine the two adjacent sets $[J_{\min}^{\ell,j}, J_{\max}^{\ell,j_1}]$ and $[J_{\min}^{\ell+1,j}, J_{\max}^{\ell+1,j_1}]$ into one larger set.

Now for each j_1 and ℓ , the number of elements in $\left\{j_2: j_2 \in \left[J_{\min}^{\ell, j_1}, J_{\max}^{\ell, j_1}\right]\right\}$ is either finite or proportional to n_2 . Let

$$S(\ell, j_1, j_2) = \sum_{m=J_{\min}^{\ell, j_1}}^{j_2} \frac{\partial s_{(j_1, m)}(\theta_0)}{\partial \theta'} \text{ for } j_2 = J_{\min}^{\ell, j_1}, \dots, J_{\max}^{\ell, j_1} \text{ and } S\left(\ell, j_1, J_{\min}^{\ell, j_1} - 1\right) = 0.$$

By Assumption 3, $S(\ell, j_1, j_2) = O(1)$ or $S(\ell, j_1, j_2) = (j_2 - J_{\min}^{\ell, j_1} + 1) H(1 + o_p(1))$ uniformly

over ℓ, j_1 and j_2 . Using this observation, we have

$$\begin{split} &\frac{1}{N}\sum_{j\in B_{n}}\Phi_{k}(\frac{j}{n})\frac{\partial s_{j}(\theta_{0})}{\partial \theta'} \\ &= \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}}\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})\frac{\partial s_{(j_{1},j_{2})}(\theta_{0})}{\partial \theta'} \\ &= \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}}\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})\left[S\left(\ell,j_{1},j_{2}\right)-S\left(\ell,j_{1},j_{2}-1\right)\right] \\ &= \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}}\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})S\left(\ell,j_{1},j_{2}\right) - \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}-1}^{J_{\max}^{\ell,j_{1}}}\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})S\left(\ell,j_{1},j_{2}\right) - \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}-1}^{J_{\max}^{\ell,j_{1}}}\Phi_{k}(\frac{j_{1}}{n},\frac{j_{2}}{n_{2}})S\left(\ell,j_{1},j_{2}\right) - \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}}\left[\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})-\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}+1}{n_{2}})\right]S\left(\ell,j_{1},j_{2}\right) \\ &+ \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}}}\left[\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})-\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}+1}{n_{2}})\right]H\left(j_{2}-J_{\min}^{\ell,j_{1}}+1\right)\left(1+o_{p}\left(1\right)\right) \\ &+ \frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{J_{\max}^{\ell,j_{1}}}{n_{2}})H\left(J_{\max}^{\ell,j_{1}}-J_{\min}^{\ell,j_{1}}+1\right)\left(1+o_{p}\left(1\right)\right) \end{split}$$

But

$$\frac{1}{N} \sum_{j_{1}=j_{\min}}^{j_{\max}} \sum_{\ell=1}^{L_{j_{1}}} \sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}-1} \left[\left| \Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}}) - \Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}+1}{n_{2}}) \right| \right] H\left(j_{2} - J_{\min}^{\ell,j_{1}} + 1\right) \\ \leq \frac{C}{N} \sum_{j_{1}=j_{\min}}^{j_{\max}} \sum_{\ell=1}^{L_{j_{1}}} \sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}-1} \frac{\left(J_{\max}^{\ell,j_{1}} - J_{\min}^{\ell,j_{1}} + 1\right)}{n_{2}} = O\left(\frac{|B_{n}|}{N}\right) = O\left(1\right)$$

and

$$\frac{1}{N} \sum_{j_1=j_{\min}}^{j_{\max}} \sum_{\ell=1}^{L_{j_1}} \left| \Phi_k(\frac{j_1}{n_1}, \frac{J_{\max}^{\ell, j_1}}{n_2}) \right| H\left(J_{\max}^{\ell, j_1} - J_{\min}^{\ell, j_1} + 1\right) = O\left(1\right),$$

$$\begin{split} &\frac{1}{N}\sum_{j\in B_{n}}\Phi_{k}(\frac{j}{n})\frac{\partial s_{j}(\theta_{0})}{\partial\theta'} \\ &=\frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}-1} \left[\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})-\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}+1}{n_{2}})\right]H\left(j_{2}-J_{\min}^{\ell,j_{1}}+1\right) \\ &+\frac{1}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{J_{\max}^{\ell,j_{1}}}{n_{2}})H\left(J_{\max}^{\ell,j_{1}}-J_{\min}^{\ell,j_{1}}+1\right)+o_{p}\left(1\right) \\ &=\frac{H}{N}\sum_{j_{1}=j_{\min}}^{j_{\max}}\sum_{\ell=1}^{L_{j_{1}}}\sum_{j_{2}=J_{\min}^{\ell,j_{1}}}^{J_{\max}^{\ell,j_{1}}}\Phi_{k}(\frac{j_{1}}{n_{1}},\frac{j_{2}}{n_{2}})+o_{p}\left(1\right) \\ &=\frac{1}{N}\sum_{j\in B_{n}}\Phi_{k}(\frac{j}{n})H+o_{p}\left(1\right). \end{split}$$

Therefore

$$\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})\hat{s}_j = (1+o_p(1))\left\{\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})s_j - \frac{1}{N}\sum_{j\in B_n}\Phi_k(\frac{j}{n})\left[\frac{1}{\sqrt{N}}\sum_{\tilde{j}\in B_n}s_{\tilde{j}}\right]\right\}$$
$$= (1+o_p(1))\frac{1}{\sqrt{N}}\sum_{j\in B_n}\Phi_k(\frac{j}{n})\left(s_j - \frac{1}{N}\sum_{\tilde{j}\in B_n}s_{\tilde{j}}\right).$$

It now follows from Assumption 4 that

$$\begin{split} F_N &\stackrel{a}{\sim} \left[R\hat{H}^{-1}\Lambda \frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_{(0,0)}(\frac{j}{n}) e_j \right]' \left\{ R\hat{H}^{-1}\Lambda \frac{1}{K} \sum_{k \in \mathbb{K}} \operatorname{Re} \left[\frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_k(\frac{j}{n}) \left(e_j - \frac{1}{N} \sum_{j \in B_n} e_j \right) \right] \right\} \\ & \times \frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_k^*(\frac{j}{n}) \left(e_j - \frac{1}{N} \sum_{j \in B_n} e_j \right)' \right] \Lambda' \hat{H}^{-1} R' \\ = ^d \left[\frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_{(0,0)}(\frac{j}{n}) \varepsilon_j \right]' \left\{ \frac{1}{K} \sum_{k \in \mathbb{K}} \operatorname{Re} \left[\frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_k(\frac{j}{n}) \left(\varepsilon_j - \frac{1}{N} \sum_{j \in B_n} \varepsilon_j \right) \right] \right\}^{-1} \left[RH^{-1}\Lambda \frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_{(0,0)}(\frac{j}{n}) e_j \right] / q \\ & \times \frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_k^*(\frac{j}{n}) \left(\varepsilon_j - \frac{1}{N} \sum_{j \in B_n} \varepsilon_j \right)' \right] \\ & \times \frac{1}{\sqrt{N}} \sum_{j \in B_n} \Phi_k^*(\frac{j}{n}) \left(\varepsilon_j - \frac{1}{N} \sum_{j \in B_n} \varepsilon_j \right)' \right] \\ & = \eta' \left(\frac{1}{K} \sum_{k \in \mathbb{K}} \operatorname{Re} \left(\xi_k \xi_k^* \right) \right)^{-1} \eta / q. \end{split}$$

The proof for t_N is similar and is omitted here.

32

 \mathbf{SO}

Proof of Corollary 2. Under assumptions (i) and (ii) in the corollary, $\xi_k \stackrel{a}{\sim} iidN(0, I_q)$ and $\sum_{k \in \mathbb{K}} \xi_k \xi'_k \stackrel{a}{\sim} \mathbb{W}(\mathbb{I}_q, K)$ a Wishart distribution. So F_N is asymptotically equivalent in distribution to Hotelling's T-squared distribution (Hotelling, 1931):

$$qF_N \stackrel{a}{\sim} T^2(q, K).$$

Using the well-known relationship between the T-squared distribution and the F distribution, we have

$$\frac{(K-q+1)}{K}F_N \stackrel{a}{\sim} F_{q,K-q+1}.$$

The proof for the t statistic is similar. \blacksquare

References

- [1] Abramowitz, M. and Stegun, I. (1965): *Handbook of Mathematical Functions*, Dover Publications, New York.
- [2] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica*, 59(3), 817–858.
- [3] Bester, A., Conley, T., Hansen, C., and Vogelsang, T. J. (2011): "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators." Working paper, Michigan State University.
- [4] Cliff, A. D., and Ord, J. K. (1981): Spatial Processes: Models and Applications. London: Pion.
- [5] Conley, T. G. (1999): "GMM Estimation with Cross Sectional Dependence." Journal of Econometrics, 92, 1–45.
- [6] Conley, T. G. and Udry, C. R. (2010): "Learning about a New Technology: Pineapple in Ghana." American Economic Review, 100(1), 35–69.
- [7] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50, 1029–1054.
- [8] Hotelling, H. (1931): "The Generalization of Student's Ratio." Annals of Mathematical Statistics, 2, 360–378.
- [9] Jenish, N. and Prucha, I. R. (2009): "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields." *Journal of Econometrics*, 150, 86–98
- [10] Kelejian, H. and Prucha, I. R. (2007): "HAC Estimation in a Spatial Framework." Journal of Econometrics, 140(1):131–154.
- [11] Kiefer, N. M., and Vogelsang, T. J. (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory*, 21, 1130–1164.
- [12] Kim, M. and Sun, Y. (2011): "Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix." *Journal of Econometrics*, 160(2), 349–371

- [13] Stein, M. L. (1999): Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.
- [14] Sun, Y., Phillips, P. C. B. and Jin, S. (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing." *Econometrica*, 76, 175–194.
- [15] Sun, Y. (2011a): "Robust Trend Inference with Series Variance Estimator and Testingoptimal Smoothing Parameter." *Journal of Econometrics*, 164(2), 345–366.
- [16] Sun, Y. (2011b): "Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator." Working paper, Department of Economics, UC San Diego.
- [17] Webster, R. and Oliver, M. (2001): Geostatistics for Environmental Scientists. Wiley & Sons.