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

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

August 2008

Abstract. This paper develops a method for performing inference using spatially dependent data. We consider test statistics formed using nonparametric covariance matrix estimators that account for heteroskedasticity and spatial correlation (spatial HAC). We provide distributions of commonly used test statistics under "fixed-b" asymptotics, in which HAC smoothing parameters are proportional to the sample size. Under this sequence, spatial HAC estimators are not consistent but converge to non-degenerate limiting random variables that depend on the HAC smoothing parameters and kernel. We show that the limit distributions of commonly used test statistics are pivotal but non-standard, so critical values must be obtained by simulation. We provide a simple and general simulation procedure based on the i.i.d. bootstrap that can be used to obtain critical values. We illustrate the potential gains of the new approximation through simulations and an empirical example that examines the effect of unjust dismissal doctrine on temporary help services employment.

Keywords: HAC, panel, robust

JEL Codes: C12, C21, C22, C23

*We thank Silvia Gonçalves fpr helpful discussion. We are grateful for support from the Neubauer Family Faculty Fund and from the IBM Corporation Faculty Research Fund at the University of Chicago Graduate School of Business.

†Graduate School of Business, University of Chicago

‡Department of Economics, Michigan State University

1. Introduction

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

An oft neglected feature of cross-sectional and panel data is cross-sectional or spatial correlation: Data or residuals from one observation may be statistically related to data or residuals from other neighboring observations. As is the case with temporal correlation, inference procedures that do not account for such spatial correlation will usually be incorrect when it exists in the data. To account for this problem, spatial HAC procedures have been developed.[3] These spatial HAC estimators allow for processes with dependence indexed in more than one dimension. This setup includes time series (where dependence is indexed in one dimension) and panel data (where dependence is indexed spatially and over time) as special cases in addition to covering the purely cross-sectional case. In this paper, we present a novel way to use spatial HAC estimators. Our approach is based on an alternate

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

[2]See Wooldridge (2003) for a concise review of this literature. See also Liang and Zeger (1986), Arellano (1987), Bertrand, Duflo, and Mullainathan (2004), Hansen (2007), and Bester, Conley, and Hansen (2008).

[3]See Conley (1996, 1999) for early examples.

asymptotic approximation to the behavior of associated test statistics, referred to in time series contexts as "fixed-b" asymptotics following Kiefer and Vogelsang (2002, 2005).

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

In contrast, the time series fixed-b approach of Kiefer and Vogelsang (2002, 2005) (KV) considers the behavior of statistics under an asymptotic sequence in which the HAC smoothing parameter is proportional to the sample size (this fixed proportion is the source of the label "fixed-b"). This contrasts with the conventional sequence where the smoothing parameter grows more slowly than the sample size. Under the KV sequence, the HAC estimator is not consistent but converges in distribution to a non-degenerate random variable that depends on both the kernel and smoothing parameter used in forming the HAC estimator. KV calculate an approximate distribution for commonly-used test statistics accounting for this random limit of the HAC covariance estimator. The resulting limit distribution for t and Wald statistics are pivotal but nonstandard, so critical values are obtained by simulation. However, because they explicitly treat the HAC estimator as random, these approximations account for both sampling uncertainty and bias in the HAC estimator. Through simulations, KV provide convincing evidence that their approximation outperforms the plug-in approach.[4]

In this paper, we extend the analysis of KV to allow for spatially dependent data where dependence is indexed in more than one dimension. This framework is very general and allows for general forms of cross-sectional correlation, time series dependence, and panel or other multilevel data. We consider the behavior of spatial HAC estimators as in Conley (1996, 1999) under asymptotic sequences in which smoothing parameters are proportional to the sample size. We follow KV and refer to this asymptotic sequence as fixed-b asymptotics. As in KV, this analysis produces a random variable limit for the HAC estimator that depends on both the HAC smoothing parameters and kernel. We provide the limiting behavior of t and Wald statistics and show that, as in the time series case, they have pivotal limiting distributions but that these distributions are nonstandard.

Extension of these results from the time series to the spatial context is both important and nontrivial. Nonparametric estimation of covariances can be challenging in the time series case, and is even more so when the data exhibits dependence in more than one dimension. The resulting variance estimates exhibit pronounced sampling variability and small sample biases, both of which can be captured to a large degree through the use of fixed-b asymptotics. However, application of fixed-b asymptotics to spatial HAC is nontrivial due to the

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

importance of boundary effects in spatial data and the fact that any given point has many more neighbors when dependence is indexed in multiple dimensions than when it is indexed on the line (as in a time series). Resulting distributions for test statistics are considerably more complicated than those obtained in the time series case. Critical values depend not only on the kernel and chosen values of smoothing parameters for the HAC estimator, but also on the distribution of observed locations. Therefore they must typically be recalculated for each specific application. Direct calculation requires approximating functions of Brownian sheets with non-trivial boundary conditions and are cumbersome at best. A major contribution of this paper is to provide a simple simulation method to obtain critical values. We show that a simple procedure based on an i.i.d. bootstrap is valid for the fixed-b asymptotic distribution and can be used to obtain critical values quite generally.

To illustrate the potential improvements available from using this approach, we present simulation results in a pure cross-sectional and a panel setting. In both cases, we see that inference procedures that ignore spatial correlation are highly size-distorted when spatial correlation is present. We also see that large size distortions persist when spatial HAC estimators are employed with the usual asymptotic normal approximation. Across the simulations, correctly-sized tests are only obtained using the inference procedure developed in our paper along with HAC estimators using large values of the smoothing parameter.

We also present results from a simple empirical example based on Autor (2003) that estimates the effect of restrictions on firing practices on temporary help services employment. Conventional methods find statistically significant evidence that there is a positive association between judicial recognition of restrictions of firing practices in a state and temporary help services employment in that state. However, when cross-sectional correlation is taken into account via our method, the resulting confidence intervals are quite wide, encompassing economically meaningful negative and positive values. Our findings suggest that inference made in this example without accounting for spatial correlation substantially overstates the true precision of the estimated treatment effects.

There are, of course, other ways to perform inference with spatially dependent cross section or panel data. Kelejian and Prucha (1999, 2001) and Lee (2004, 2007a, 2007b) consider inference based on specific models for the spatial dependence structure. Methods that do not rely on parametric covariance structures are also available. The Fama and MacBeth (1973) procedure consists of partitioning a panel data set into its time periods and estimating the parameters of interest within each cross-section. Inference is then conducted by taking this set of point estimates, one for each time series observation in the sample, and forming a t-statistic. Since this effectively collapses the cross-sectional dimension out of the problem, inference is robust to arbitrary cross-sectional dependence, though the procedure needs adjustment in the presence of intertemporal dependence. A recent paper by Ibragimov and Müller (2006) provides a formal and general treatment of inference procedures based on partitioning data and estimating parameters within each partition, focusing upon properties of t-tests using these sets of point estimates. Bester, Conley, and Hansen (2008) consider a different, but related procedure based on using a clustered covariance estimator with researcher defined clusters. In both Ibragimov and Müller (2006) and Bester, Conley, and Hansen (2008), the key condition required for such tests' validity is that averages of scores within clusters be asymptotically uncorrelated, which may be achieved through appropriate choice of partitions.[5] Both papers are also in the spirit of fixed-b asymptotics, as both

[5]This is trivially achieved with time series data and may also be achieved with more general spatial dependence under conditions given in Bester, Conley, and Hansen (2008).

approaches advocate the use of large groups and provide asymptotic analysis that accounts for uncertainty in estimating the covariance matrix. Driscoll and Kraay (1998) present a different approach that uses a HAC estimator over the within-period averages of OLS scores. Vogelsang (2008) provides a fixed-b asymptotic analysis of Driscoll and Kraay (1998) type standard errors as well as standard errors estimated by averaging across time series HAC covariance estimators formed for each individual series in the panel.

Our paper complements this literature. We obtain inference that is robust to general forms of spatial correlation without functional form assumptions on the dependence structure. Unlike Fama and MacBeth (1973), Driscoll and Kraay (1998), and Vogelsang (2008), the spatial HAC estimator we consider applies to cross-sectional data as well as time series and long and short T panel data. Relative to Ibragimov and Müller (2006) and Bester, Conley, and Hansen (2008), the procedure in this paper allows finer control over assumed dependence structure through the use of downweighting via the kernel function and smoothing parameters. By allowing for smooth decay of covariance across observational units, it is less susceptible to boundary problems than the grouping schemes of Ibragimov and Müller (2006) and Bester, Conley, and Hansen (2008) as these estimators by construction ignore covariances between observations on the boundary of one group with observations in other groups.

The remainder of this paper is organized as follows. In Section 2, we review spatial HAC estimation. We present the main asymptotic results in Section 3, and show how the simple i.i.d. bootstrap can be used as a simulation procedure to obtain critical values. Section 4 contains results from our simulation experiments, and an empirical example is given in Section 5. Section 6 concludes.

2. **Spatial HAC Estimation**

We propose a new approach to inference in models with spatial dependence. The approach is based on the spatial HAC procedure proposed in Conley (1996, 1999) but follows Kiefer and Vogelsang (2005) in that the bandwidth is taken as a constant fraction of the sample size for the asymptotic analysis. We are interested in an estimator, $\widehat{\theta}_N$, of a $p \times 1$ parameter vector θ_0 that satisfies

$$\frac{1}{N}\sum_{i=1}^{N} s_i(\widehat{\theta}_N) = \mathbf{0} \tag{2.1}$$

where N is the sample size, $s_i(\theta)$ is a $p \times 1$ score vector such that

$$E(s_i(\theta_0)) = \mathbf{0}.$$

The score vector is a function of observable data indexed by i. In general we would like to allow the data and score vector to exhibit general forms of spatial correlation where the strength of the correlation depends on some observable distance measure between any two observations.

For simplicity, we follow Conley (1999) and assume that, given the distance measure, it is possible to map the data onto a two-dimensional integer lattice so that the distance between pairs of observations can be expressed in terms of the lattice indices when presenting most of our theoretical results. To be concrete, suppose a location $d(i)$ is given for each observation and that the range of $d(i)$ is $\{1, \ldots, L\} \times \{1, \ldots, M\} \subset \mathbb{Z}_+^2$, i.e., the observations live on a lattice in $\mathbb{Z}_+^2$. We can then rewrite the sample moment condition that defines $\widehat{\theta}_N$ as

$$\frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} \mathbf{1}_{l,m}\widehat{s}_{l,m} = \mathbf{0}, \tag{2.2}$$

where $\widehat{s}_{l,m} = s_i\left(\widehat{\theta}_N\right)$ is associated with observation i for which $d(i) = (l, m)$ and the indicator function $\mathbf{1}_{l,m}$ denotes the presence of an observation at a particular lattice point. Note that we do not assume the presence of an observation at every lattice point, and that equation (2.2) gives the same estimate of θ as (2.1). Define the population analog to $\widehat{s}_{l,m}$ as $s_{l,m} = s_i(\theta_0)$, again where $d(i) = (l, m)$. Because asymptotic analysis is determined by the properties of the random field $\mathbf{1}_{l,m}s_{l,m}$, we conserve on notation by defining

$$s^*_{l,m} = \mathbf{1}_{l,m}s_{l,m} \quad \text{and} \quad \widehat{s}^*_{l,m} = \mathbf{1}_{l,m}\widehat{s}_{l,m}.$$

We assume a rectangular lattice domain for several reasons, and also note that extension to non-rectangular lattices is straightforward. As we shall see below, we present most of our theoretical results using a functional central limit theorem of the form

$$\frac{1}{\sqrt{LM}} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} s^*_{l,m} \Rightarrow \Lambda W(r, s)$$

where $r, s \in (0, 1]$, $[x]$ denotes the integer part of x, and W is a Wiener process. Functional central limit theorems (FCLTs) of this form appear in Deo (1975) and Basu and Dorea (1979), and are simply the spatial analog of the usual time series FCLTs that deliver convergence of partial sums of the score process to Brownian motion on the unit interval. When the domain is rectangular,[6] W is a Brownian sheet, or Brownian motion on the unit rectangle, which we discuss further in Section 3.1. The loss of generality is minimal because a slightly more general FCLT in Goldie and Greenwood (1986) provides convergence of partial sums of data indexed on a domain A to a Wiener process over a domain $A \subseteq [0, 1]^2$. We focus on the rectangular lattice case for the formal theoretical results as it simplifies notation and

[6]Intuitively, this occurs when the limiting sequence of indicators $\mathbf{1}_{l,m}$ in (2.2) is dense in $\mathbb{Z}^2_+$ when it is normalized to the unit rectangle, so that observed locations "fill in" the unit rectangle.

proofs considerably and allows us to explicitly characterize boundary terms in the asymptotic distribution of test statistics that do not arise in time series results but are important in the spatial context. In addition, an abstract discussion of Brownian processes over general domains seems beyond the scope of the paper. Finally and most importantly, we argue in Section 3.4 that the inference procedure we propose in practice, which uses an i.i.d. bootstrap to produce approximate critical values for test statistics, will work the same way on a more general domain as in the rectangular lattice case.

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

$$\Omega = \lim_{L\to\infty, M\to\infty} \frac{1}{LM} E\left[\left(\sum_{l=1}^{L}\sum_{m=1}^{M} s^*_{l,m}\right)\left(\sum_{l=1}^{L}\sum_{m=1}^{M} s^{*\prime}_{l,m}\right)\right].$$

The nonparametric HAC estimator of Ω is given by

$$\widehat{\Omega} = \frac{1}{LM}\sum_{l_l=1}^{L}\sum_{m_1=1}^{M}\sum_{l_2=1}^{L}\sum_{m_2=1}^{M} K(\frac{l_1}{h_L}, \frac{l_2}{h_L}, \frac{m_1}{h_M}, \frac{m_2}{h_M})\widehat{s}^*_{l_1,m_1}\widehat{s}^{*\prime}_{l_2,m_2}, \tag{2.3}$$

where $K(x_1, x_2, x_3, x_4)$ is a kernel weighting function and h_L and h_M are bandwidth tuning parameters where typically $h_L \leq L$ and $h_M \leq M$.

3. Fixed-b Asymptotics

This section presents asymptotic theory for the spatial HAC covariance estimator described in the previous section under a sequence where the smoothing parameters h_L and h_M grow at a rate proportional to the sample size. Under this sequence, the HAC estimator converges to a random variable whose distribution involves stochastic processes called Brownian sheets, which are essentially Brownian motions indexed in more than one dimension (e.g., space and time). Section 3.1 provides a brief overview and definitions of these processes. In Section 3.2, we provide results on the limit distributions of the HAC estimator and related test statistics.

As mentioned above, the distributions of t and Wald tests are pivotal, but nonstandard, and explicitly reflect both the kernel and chosen values of smoothing parameters. Because of the many bandwidth and kernel choices available in practice and the multiple bandwidths involved, it is impractical to tabulate all relevant critical values. Critical values must therefore typically be calculated by simulation on a case-by-case basis, which unfortunately involves integration of multi-indexed stochastic processes. However, we show in Section 3.3 that critical values may be approximated using an i.i.d. bootstrap, which is easily implemented in a wide variety of applications.

3.1. **Brownian Sheets and related processes**

For $\mathbf{u} \in \mathbf{R}^d$, let $W(\mathbf{u})$ be a Brownian sheet, a continuous Gaussian process with mean zero and covariance function

$$\mathrm{Cov}(W(\mathbf{u}_1), W(\mathbf{u}_2)) = \prod_{k=1}^{d} \min(u_1^{(k)}, u_2^{(k)}).$$

With $d = 1$, this is the usual Brownian motion. Again for simplicity we deal with the $d = 2$ case corresponding to a location space in $\mathbb{R}_+^2$. Below we think of W as being defined on the unit rectangle. Denote by W_p a p-vector of independent Brownian sheets.

Define the "pinned" Brownian Sheet $B_p(u, v) = W_p(u, v) - uvW_p(1, 1)$, see e.g., McKeague and Sun (1996). Note that on the outer edges of the unit rectangle, $B_p(1, v) = W_p(1, v) - vW_p(1, 1)$ behaves analogously to a one-dimensional Brownian bridge but may be correlated with values of the sheet on the interior. Denote by $K(x_1, x_2, x_3, x_4)$ the kernel used above. We define the following functionals of the Brownian sheet:

$$
\begin{aligned}
Q_p^{\emptyset}(b_1,b_2) &= \int_{[0,1]^4} \frac{\partial^4 K}{\partial x_1 \partial x_2 \partial x_3 \partial x_4}\left(\frac{u_1}{b_1},\frac{u_2}{b_1},\frac{v_1}{b_2},\frac{v_2}{b_2}\right) B_p(u_1,v_1)B_p(u_2,v_2)d(u_1\times v_1\times u_2\times v_2)\\
Q_p^{\{1\}}(b_1,b_2) &= \int_{[0,1]^3} \frac{\partial^3 K}{\partial x_2 \partial x_3 \partial x_4}\left(\frac{1}{b_1},\frac{u_2}{b_1},\frac{v_1}{b_2},\frac{v_2}{b_2}\right) B_p(1,v_1)B(u_2,v_2)_p d(v_1\times u_2\times v_2)\\
Q_p^{\{1,4\}}(b_1,b_2) &= \int_{[0,1]^2} \frac{\partial^2 K}{\partial x_2 \partial x_3}\left(\frac{1}{b_1},\frac{u_2}{b_1},\frac{v_1}{b_2},\frac{1}{b_2}\right) B_p(1,v_1)B_p(u_2,1)d(v_1\times u_2).
\end{aligned}
$$

Other functionals, e.g. $Q_p^{\{2\}}$, $Q_p^{\{2,3\}}$ are similarly defined: For Q_p^E, all arguments in the set E are omitted from the derivative, and evaluation in those arguments is at the boundary. Finally, we define

$$
Q_p(b_1,b_2) = Q_p^{\emptyset}(b_1,b_2) + \sum_{i=1}^{4} Q_p^{\{i\}}(b_1,b_2) + \sum_{\substack{E=\{1,2\},\{3,4\},\\ \{2,3\},\{1,4\}}} Q_p^E(b_1,b_2). \tag{3.1}
$$

At this point, we avoid any assumptions about the kernel K beyond the regularity conditions given below. Note, however, that some widely used assumptions about the kernel simplify (3.1) considerably. For example, $K(x_1,x_2,x_3,x_4) = K(x_1-x_2,x_3-x_4)$ would imply $Q_p^{\{1\}} = Q_p^{\{2\}}$, $Q_p^{\{3\}} = Q_p^{\{4\}}$, and $Q_p^{\{1,4\}} = Q_p^{\{2,3\}}$.

3.2. **Main Theorems**

Define Λ such that $\Lambda\Lambda' = \Omega$. The following set of assumptions are sufficient for obtaining the main results of the paper.

ASSUMPTION 1. *As* $N\to\infty$, $L\to\infty, M\to\infty$, $\frac{L}{N}\to 0$ *and* $\frac{M}{N}\to 0$.

ASSUMPTION 2. $\widehat{\theta}_N \xrightarrow{p} \theta_0$.

ASSUMPTION 3. $\frac{1}{\sqrt{LM}}\sum_{l=1}^{[rL]}\sum_{m=1}^{[sM]} s_{l,m}^* \Rightarrow \Lambda W(r,s)$ *for all* $(r,s)\in[0,1]^2$.

ASSUMPTION 4.

$$\widehat{J}_{\widehat{\theta}_N}^{[rL],[sM]} = (LM)^{-1} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \frac{\partial s_{l,m}^{*}(\widehat{\theta}_N)}{\partial \theta'} \xrightarrow{p} rs\mathcal{J},$$

where $\mathcal{J}$ is a nonsingular matrix.

Assumption 1 states that neither dimension is dominant, in which case the problem would reduce to one involving a scalar-indexed process (e.g., a time series). Note that Assumption 3 requires a FCLT hold for the random field $s_{l,m}^{*}$. For simplicity we assume that the locations at which data are observed are dense in $\mathbb{Z}_{+}^{2}$, e.g. as in Deo (1975) or Basu and Dorea (1979), so that the limiting process W is the Brownian sheet discussed in the previous section.[7] Assumption 4 is a fairly standard assumption. Note that Assumption 2 and 4 imply that

$$\widehat{J}_{\overline{\theta}}^{[rL],[sM]} = (LM)^{-1} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \frac{\partial s_{l,m}^{*}(\overline{\theta})}{\partial \theta'} \xrightarrow{p} rs\mathcal{J},$$

where $\overline{\theta}$ is between (in the vector sense) $\widehat{\theta}_N$ and θ_0. When $r = s = 1$ (i.e., the sum is taken over all observations in the sample), we will simply write $\widehat{J}_{\overline{\theta}}^{L,M} = \widehat{J}_{\overline{\theta}}$.

LEMMA 1. *Suppose Assumptions 1-4 hold. Then, as $N \to \infty$,*

$$\sqrt{LM}\left(\widehat{\theta}_T - \theta_0\right) \Rightarrow -\mathcal{J}^{-1}\Lambda W_p(1,1) \sim N(0,V)$$

The following theorem is analogous to Kiefer and Vogelsang (2005) and establishes convergence of $\widehat{\Omega}$ to a functional of pinned Brownian sheets.

[7]As discussed above, this assumption simplifies notation and proofs considerably and avoids an abstract discussion of Wiener processes over more general domains. In Section 3.4, we argue our results extend to the non-rectangular lattice case due to the FCLT in Goldie and Greenwood (1986). Most importantly, the inference procedure we propose in practice works the same way in either case.

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

$$\widehat{\Omega} \Rightarrow \Lambda Q_p(b_1, b_2)\Lambda'.$$

We obtain a typical fixed-b result where the limit of $\widehat{\Omega}$ is proportional to Ω through $\Lambda\Lambda'$ but depends on a the random matrix $Q_p(b_1, b_2)$ which reflects the choice of kernel and bandwidths. $Q_p(b_1, b_2)$ approximates the finite sample randomness in $\widehat{\Omega}$ in contrast to a consistency result for $\widehat{\Omega}$ where $Q_p(b_1, b_2)$ is replaced with an identity matrix in which case randomness in $\widehat{\Omega}$ is ignored.

An important kernel without fourth derivatives not covered by Theorem 1 is the Bartlett kernel. In time series settings the Bartlett kernel is widely used because of its relative simplicity and the fact that the Bartlett kernel guarantees a positive definite HAC estimate. These benefits extend to the spatial context as argued by Conley (1999) and we include it in our analysis. We use a Bartlett kernel in product form defined as

$$K(x_1, x_2, x_3, x_4) = (1 - |x_1 - x_3|)^+ (1 - |x_2 - x_4|)^+$$

where $z^+ = z\mathbf{1}(z > 0)$, which leads to the HAC estimator

$$\widehat{\Omega} = \frac{1}{LM} \sum_{l_l=1}^{L} \sum_{m_1=1}^{M} \sum_{l_2=1}^{L} \sum_{m_2=1}^{M} K_{l_1,l_2}^{h_L} K_{m_1,m_2}^{h_M} \widehat{s}^{*}_{l_1,m_1} \widehat{s}^{*\prime}_{l_2,m_2}, \tag{3.2}$$

where

$$K_{l_1,l_2}^{h_L} = \left(1 - \left|\frac{l_1 - l_2}{h_L}\right|\right)^+, \quad \text{and} \quad K_{m_1,m_2}^{h_M} = \left(1 - \left|\frac{m_1 - m_2}{h_M}\right|\right)^+.$$

The next theorem gives the fixed-b result for the Bartlett kernel. The result is stated in terms of the functional

$$\begin{aligned}\widetilde{Q}_p(u,v,b) &= \frac{2}{b}\int_0^1 B_p(u,s)B_p(v,s)'ds - \frac{1}{b}\int_0^{1-b}\left[B_p(u,s+b)B_p(v,s)' + B_p(u,s)B_p(v,s+b)'\right]ds \\ &\quad -\frac{1}{b}\int_{1-b}^1 \left[B_p(u,s)B_p(v,1)' + B_p(u,1)B_p(v,s)'\right]ds + B_p(u,1)B_p(v,1)'.\end{aligned}$$

Note that $\widetilde{Q}_p(v,u,b)' = \widetilde{Q}_p(u,v,b)$.

THEOREM 2. *Suppose K is the Bartlett kernel given by (3.3) with bandwidths $(h_L, h_M) = (b_1 L, b_2 M)$ where $(b_1, b_2) \in (0,1]^2$. Then, under Assumptions 1-4,*

$$\widehat{\Omega} \Rightarrow \Lambda Q_p^{bart}(b_1, b_2)\Lambda',$$

where

$$\begin{aligned}Q_p^{bart}(b_1,b_2) &= \frac{2}{b_1}\int_0^1 \widetilde{Q}_p(r,r,b_2)dr - \frac{1}{b_1}\int_0^{1-b_1}\left[\widetilde{Q}_p(r+b_1,r,b_2) + \widetilde{Q}_p(r,r+b_1,b_2)\right]dr \\ &\quad -\frac{1}{b_1}\int_{1-b_1}^1 \left[\widetilde{Q}_p(r,1,b_2) + \widetilde{Q}_p(1,r,b_2)\right]dr + \widetilde{Q}_p(1,1,b_2).\end{aligned}$$

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

$$\sqrt{LM}r(\widehat{\theta}_N) \Rightarrow -R(\theta_0)\mathcal{J}^{-1}\Lambda W_p(1,1). \tag{3.3}$$

Note that the right hand side follows a q-variate Gaussian distribution. Define the Wald statistic

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

and when $q = 1$, the t- statistic,

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

The null distributions of F and t are given in the following theorem.

THEOREM 3. *Suppose the conditions of Theorem 1 are satisfied. Then*

$$F \Rightarrow B_q(1,1)Q_q(b_1,b_2)^{-1}B_q(1,1) \qquad \text{and} \qquad t \Rightarrow \frac{B_1(1,1)}{\sqrt{Q_1(b_1,b_2)}}.$$

Suppose the conditions of Theorem 2 are satisfied. Then

$$F \Rightarrow B_q(1,1)Q_q^{bart}(b_1,b_2)^{-1}B_q(1,1) \qquad \text{and} \qquad t \Rightarrow \frac{B_1(1,1)}{\sqrt{Q_1^{bart}(b_1,b_2)}}.$$

As in the time series fixed-b approximations of KV, the t- and Wald tests are asymptotically pivotal. However, the resulting reference distributions are nonstandard and are influenced by the choice of kernel and bandwidth(s), as well as the shape of the sampling region (in this case a rectangular lattice). In the next section, we show that critical values may be obtained via an i.i.d. bootstrap that is both general and simple to carry out in practice.

3.3. **Critical Values and the Bootstrap**

The limiting distributions given by Theorems 2 and 3 can be simulated given a choice of kernel and bandwidths. Because of the myriad of bandwidth and kernel choices available in practice, it is impractical to tabulate all relevant critical values especially if joint hypothesis are of interest. Even focusing on a single kernel, the two dimensional nature of the bandwidth makes tabulation of critical values tedious. In addition, the limiting behavior of test statistics depends on the shape of the sampling region as discussed in Section 3.4 below.

This dependence makes it virtually impossible to tabulate critical values given the myriad of potential shapes that could be encountered in practice. Given this, we present a simple bootstrap procedure that will be generally applicable.

Since the fixed-b limiting behavior of test statistics depends of the shape of the sampling region, it is important that one uses critical values that capture this shape. We consider two simple bootstrap procedures that do this in different ways. In the first, one assumes that the shape of the locations observed in the sample is representative of the shape of locations in the population and performs a bootstrap conditional on this shape. In the second, one assumes a hypothetical shape for the population region where locations in the hypothetical population region without data in the sample are assumed to be missing at random and/or data would fill the hypothetical region asymptotically. The first procedure will be valid when the observed locations are representative of the population locations, and the second will be valid when the hypothesized region has the same shape as the population region. We present formal results for the latter on a rectangular lattice since it is notationally more convenient, but results generalize to the bootstrap conditional on locations as discussed in Section 3.4.

We consider a bootstrap approach that is first order valid with respect to the fixed-b asymptotic distribution. Consider the following resampling scheme. Create a bootstrap realization of the sample by i.i.d. resampling with replacement using the observed locations. That is, at each observed location, place a value drawn by i.i.d. sampling with replacement from the observed data. This procedure conditions on the locations in the observed sample and will be valid when the sampled locations mimic the population locations. We refer to this bootstrap resampling scheme as the i.i.d. "conditional on locations" bootstrap.

A slightly different variant of the i.i.d. bootstrap for a rectangular lattice structure with potentially randomly missing locations is to resample from all locations of the lattice. At each point in the rectangle, an i.i.d. draw with replacement is made from the full set of lattice locations unconditional on whether a location has an observation. If the draw is a location with a data point, that data point is placed on the bootstrap lattice. If the draw is a location without a data point, that point on the bootstrap lattice is left empty and is treated as missing. This bootstrap is appropriate for rectangular lattices with missing data if the missing locations are viewed as randomly assigned. We refer to this variant of the bootstrap as the i.i.d. “unconditional on locations” bootstrap.

It is worth noting that validity of this bootstrap procedure is very much related to how we are using the bootstrap. We use a ‘naive’ bootstrap, where the denominator of our test statistic is calculated using the same formula used with the original data, purely as a Monte Carlo device, as an alternative to simulating from the reference distributions in Theorem 3 numerically. In particular we are *not* asking the bootstrap to reproduce the dependence structure in the original data, which would almost certainly require stronger assumptions and a blocking scheme for resampling.

Once a bootstrap sample is drawn using either resampling scheme, the bootstrap statistics are computed as follows. Let $s^{**}_{l,m}(\theta)$ denote the bootstrap score at coordinate (l, m). Define the bootstrap estimator $\widehat{\theta}^*$ such that

$$\frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} s^{**}_{l,m}(\widehat{\theta}^*) = \mathbf{0}.$$

Because of i.i.d. resampling, the mean of the bootstrap score is given by

$$E^*(s^{**}_{l,m}(\theta)) = \frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} s^{*}_{l,m}(\theta).$$

Recalling (2.2), it immediately follows that

$$E^*(s^{**}_{l,m}(\widehat{\theta}_N)) = 0.$$

Therefore, in the bootstrap world, the population value of θ is $\widehat{\theta}_N$ and $\widehat{\theta}^*$ is an estimator of $\widehat{\theta}_N$. We adopt the following short hand notation that mimics the notation used for the true data. Let $\widehat{s}^{**}_{l,m}$ denote $s^{**}_{l,m}(\widehat{\theta}^*)$.

Define the bootstrap versions of the t and Wald statistics as

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

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

where

$$\widehat{\Omega}^* = \frac{1}{LM}\sum_{l_l=1}^{L}\sum_{m_1=1}^{M}\sum_{l_2=1}^{L}\sum_{m_2=1}^{M}K(\frac{l_1}{h_L},\frac{l_2}{h_L},\frac{m_1}{h_M},\frac{m_2}{h_M})\widehat{s}^{**}_{l_1,m_1}\widehat{s}^{**\prime}_{l_2,m_2},$$

$$\widehat{J}^*_{\widehat{\theta}^*} = (LM)^{-1}\sum_{l=1}^{L}\sum_{m=1}^{M}\frac{\partial s^{**}_{l,m}(\widehat{\theta}^*)}{\partial\theta'}.$$

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

As shown by Gonçalves and Vogelsang (2006) in a time series regression setting, the i.i.d. bootstrap is first order equivalent to the fixed-b asymptotic distribution of t and Wald statistics even when the data has dependence. The i.i.d. bootstrap remains valid when the data is dependent because partial sums of the bootstrap data satisfy a FCLT. Because the asymptotic variance covariance matrix in the FCLT cancels from the fixed-b distribution, the bootstrap does not have to mimic the correct variance covariance matrix. The results of Gonçalves and Vogelsang (2006) for the i.i.d. bootstrap naturally extend to the spatial context as the next theorem shows.

THEOREM 4. *Suppose the i.i.d. "unconditional on locations" bootstrap is used to compute* F^* *and* t^**. Suppose the conditions of Theorem 1 are satisfied and assume that* $\widehat{\theta}^* - \widehat{\theta}_N \xrightarrow{p} 0$*. Then*

$$F^* \Rightarrow B_q(1,1)Q_q(b_1,b_2)^{-1}B_q(1,1) \qquad \text{and} \qquad t^* \Rightarrow \frac{B_1(1,1)}{\sqrt{Q_1(b_1,b_2)}}.$$

Suppose the conditions of Theorem 2 are satisfied. Then

$$F^* \Rightarrow B_q(1,1)Q_q^{bart}(b_1,b_2)^{-1}B_q(1,1) \qquad \text{and} \qquad t^* \Rightarrow \frac{B_1(1,1)}{\sqrt{Q_1^{bart}(b_1,b_2)}}.$$

Theorem 4 establishes that i.i.d. "unconditional on locations" bootstrap critical values for spatial HAC robust tests are first order asymptotic equivalent to fixed-b critical values on either a full rectangular lattice or a rectangular lattice with randomly assigned missing locations. Note that the assumption that $\widehat{\theta}^* - \widehat{\theta}_N \xrightarrow{p} 0$ holds under essentially the same regularity conditions as does Assumption 2.[8]

3.4. **Non-Rectangular Lattice Domains**

In this section, we sketch an extension of our main theorems to non-rectangular domains and lattice domains with missing locations where inference is carried out conditional on observed locations. Most importantly, we argue that the "conditional bootstrap" procedure discussed in the previous section can be used regardless of the shape of the sampling region.

We make use of the rectangular lattice assumption at three points, the first of which is stating that the limiting process $W(r,s)$ in Assumption 3 is the Brownian Sheet described

[8] See Gonçalves and White (2004).

in Section 3.1. In this case Assumption 3 could be justified by FCLTs for random fields, e.g. Deo (1975) and Basu and Dorea (1979). The rectangular domain also simplifies notation and the proofs of Theorems 1 and 2. Finally, we are able to explicitly characterize 'boundary terms' $Q_p^{\{\cdot\}}$ in the rectangular lattice case as integrals of the pinned Brownian sheet on the boundaries of the unit rectangle. Although practitioners can use the i.i.d. bootstrap procedure in the previous section to avoid the need to simulate each boundary term individually, we believe it instructive to point out that these boundary terms are relevant (unlike in the time series case) and that other Monte Carlo approaches are possible.

We briefly sketch the extension to more general domains, which is based on an FCLT in Goldie and Greenwood (1986). First fix a subset of the unit rectangle, $A \subset [0,1]^2$. Let $\mathcal{B}(A)$ be the σ-algebra of Borel subsets of A. Now for a given $C \in \mathcal{B}(A)$ and positive integers L and M, let $\mathbf{n} = (L, M)$, and define the function $\mathbf{1}_{l,m}^{\mathbf{n}} : \mathcal{B}(A) \times \{1, \ldots, L\} \times \{1, \ldots, M\} \to [0,1]$ as

$$\mathbf{1}_{l,m}^{\mathbf{n}}(C) = \left| C \cap \left(\frac{l-1}{L}\right] \times \left(\frac{m-1}{M}\right] \right|,$$

where $|\cdot|$ denotes Lesbesgue measure. Let $\{s_{l,m}^{\mathbf{n}}\}$, with $\mathbf{n} = (L, M)$, be a sequence of random fields, i.e. for each $\mathbf{n} = (L, M)$ a collection of random variables indexed by $l = 1, \ldots, L$, $m = 1, \ldots, M$, each of which satisfies $\mathrm{E}(s_{l,m}^{\mathbf{n}}) = 0$ and $\mathrm{Var}(s_{l,m}^{\mathbf{n}}) = 1$. Theorem 1.1 in Goldie and Greenwood (1986) states that, under mild regularity conditions on the sequence $\{s_{l,m}^{\mathbf{n}}\}$ and the set A, and with $L, M \to \infty$ as in our Assumption 1, the partial sum process defined over $\mathcal{B}(A)$, converges weakly[9] to a standard Wiener process on the domain A,

$$Z_{\mathbf{n}}(C) = \sqrt{\frac{|C|}{LM}} \sum_{l=1}^{L} \sum_{m=1}^{M} \mathbf{1}_{l,m}^{\mathbf{n}}(C)\, s_{l,m}^{\mathbf{n}} \quad \Rightarrow \quad W(C).$$

[9] For a precise definition of this convergence, see Goldie and Greenwood (1986), p.818-20.

Though a full discussion of Wiener processes over arbitrary domains is beyond the scope of this paper, we note that the limiting process W has the usual properties, i.e., $W(C)$ is a Gaussian random variable with $\mathrm{E}[W(C)] = 0$ and $\mathrm{E}[W(B)W(C)] = |B \cap C|$. As such, versions of Theorems 1 and 2 remain true under a modified version of Assumption 3, where the limiting process W is a Wiener process with domain $A \subseteq [0,1]^2$, and an appropriately modified version of Assumption 4, which guarantees that partial sums of the derivatives of the scores over any $C \in \mathcal{B}(A)$ converge in probability to an additive function $\mathcal{J}(C)$. The chief difference, intuitively, is that boundary terms take the form of integrals of the appropriate 'pinned' Wiener process along the boundary of A. The first order equivalence of the i.i.d. "conditional on locations" bootstrap and the fixed-b asymptotic distribution will continue to hold as long as the bootstrapped scores satisfy an FCLT in the same form as the original data (Assumption 3) with possibly a different variance matrix, and that the bootstrapped Hessian $\widehat{J}$ satisfy a ULLN, again in the same form as the original data (Assumption 4) with possibly a different probability limit. We therefore claim that Theorem 4 will hold under the modified versions of Assumptions 3 and 4 mentioned in the previous paragraph.

Whether one uses a rectangular lattice-based FCLT or the more general one discussed here depends upon whether one believes that observed locations will "fill in" the rectangular integer lattice as the size of the sample increases. This matters in practice because the reference distributions for obtaining critical values will be different in the two cases. Because the computation of the t and F statistics is the same in either case, different critical values will lead to potentially different outcome for tests and different confident intervals. Fortunately, the i.i.d. bootstrap will deliver valid critical values but the practitioner must use the version of the i.i.d. bootstrap appropriate for the relevant reference asymptotic distribution. If the rectangular lattice structure is reasonable and missing locations (if there are any) are randomly assigned, the i.i.d. "unconditional on locations" bootstrap is appropriate. If missing locations on a rectangular lattice are not randomly assigned or if the rectangular structure is unrealistic to begin with, the i.i.d. "conditional of locations" bootstrap is appropriate. We note that in all of our simulation results, we use the "conditional on location" bootstrap.

4. Simulation Examples

In the previous sections, we have provided approximations to the behavior of test-statistics based on spatial HAC covariance estimators under a sequence which keeps the HAC bandwidths proportional to the sample size. Under this sequence, estimation uncertainty associated with the covariance matrix is accounted for, unlike in the usual approximations which ignore this source of uncertainty. We expect that, in many cases, this will lead to improved finite sample performance of our approximation relative to the usual normal approximation. We have also shown that a relatively simple i.i.d. bootstrap procedure can be used very generally to obtain critical values for test statistics. In this section, we provide evidence on the finite-sample inference properties of tests using our new approximation through simulation experiments first in purely cross-sectional data and then in a panel. In all of our simulations, we consider inference about a slope coefficient from a linear regression model.

4.1. Cross-Sectional Example

We present simulation results obtained in cross-sectional data with data generated on a rectangular lattice. All of our results are based on data generated using a spatial moving average model. Specifically, we consider data generated on an $M \times M$ integer lattice with

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

where x_s is a scalar, s is a vector (s_1, s_2) indicating which lattice point the observation corresponds to, $\alpha = 0$, and $\beta = 1$. We generate x_s and ε_s as

$$x_s = \sum_{\|j\| \leq 2} \gamma^{\|j\|} v_{s+j},$$

$$\varepsilon_s = \sum_{\|j\| \leq 2} \gamma^{\|j\|} u_{s+j}$$

with $\|j\| = max(j_1, j_2)$ in this expression, $u_s \sim N(0,1)$, and $v_s \sim N(0,1)$ for all i and j. We consider three different values of γ, $\gamma \in \{0, .3, .6\}$. We consider two different designs, one of which uses a full lattice and one of which uses a sparse lattice. For the full lattice results, we set $M = 25$ and use the data generated at each location for a total sample size of $N = 625$.[10] For the sparse lattice results, we set $M = 36$ and generate 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.

Table 1 reports rejection rates for 5% level tests from this simulation experiment. All results are based on 1000 simulation replications. Row labels indicate which covariance matrix estimator and smoothing parameter are used. I.i.d. and Heteroskedasticity use conventional OLS standard errors and heteroskedasticity robust standard errors respectively. For all HAC estimators, we use product kernels[11] and equal bandwidths in both coordinate

[10]We draw u_s and v_s on a 29×29 lattice to generate the 25×25 lattice of x_s and ε_s.

[11]For $s, \tilde{s} \in \mathbb{Z}_+^2$, we refer to kernels of the form $K(s, \tilde{s}) = k_1(s_1, \tilde{s}_1)k_2(s_2, \tilde{s}_2)$, where $k_1, k_2 : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}_+$, as product kernels. We use this type of kernel throughout our simulations and empirical example.

directions. Rows labeled Bartlett(h) use HAC estimators with a Bartlett kernel and smoothing parameter equal to h, $K(s, \tilde{s}) = (1 - |s_1 - \tilde{s}_1|/h)^+(1 - |s_2 - \tilde{s}_2|/h)^+$; and rows labeled Gaussian(h) use HAC estimators with a Gaussian kernel and smoothing parameter equal to h, $K(s, \tilde{s}) = \exp\{-.5[(s_1 - \tilde{s}_1)/(h/2)]^2\} \exp\{-.5[(s_2 - \tilde{s}_2)/(h/2)]^2\}$. Column 2 indicates from which reference distribution critical values were obtained with "Fixed-b" indicating that the approximations developed in this paper were used. All fixed-b critical values were obtained via the i.i.d. bootstrap with 200 bootstrap replications. Columns labeled "Full Lattice" and "Sparse Lattice" use the corresponding simulation designs as discussed above.

Some general patterns emerge from the simulation results. The sparsity of the lattice does not adversely affect the properties of the fixed-b approximations and appears to simply correspond to a general weakening of the spatial dependence in the data. In all cases, we see large size distortions in testing procedures that ignore spatial correlation (i.e., 'IID' and 'Heteroskedasticity') when spatial correlation is present. Tests based on HAC estimators also have substantial size distortions when critical values based on the standard asymptotic normal approximation are used. These size distortions occur across all bandwidths when there is spatial correlation in the data, and also occur for large bandwidths even when the data are i.i.d.

When there is spatial correlation in the data, the only testing procedures with size close to the nominal level are those that use a large value for the smoothing parameter and the procedure developed in this paper. This result may not be surprising: When there are modest amounts of dependence in data, one needs to use a smoothing parameter that is large enough to allow this dependence to be captured. However, the large smoothing parameter induces substantial sampling variability and downward bias in the HAC estimator, which is ignored in the usual asymptotic normal approximation. Our fixed-b approximation, like Kiefer and Vogelsang (2002, 2005) in the time series case, uses critical values from a distributional approximation that accounts for sampling uncertainty in the HAC estimator. The results strongly suggest that one should use the approach of this paper or something similar when faced with data in which there may be spatial dependence.

4.2. **Panel Example**

In this section, we present results from a simulated panel with both cross-sectional and intertemporal dependence. Specifically, we generate data from the model

$$y_{st} = \beta x_{st} + \alpha_s + \alpha_t + \varepsilon_{st}$$

where y_{st} is the outcome for observation s at time t, α_s and α_t are unobserved cross-sectional unit and time effects, ε_{st} is the error term, and x_{st} is the variable of interest. In all of the simulations, we set $\beta = 1$, $\alpha_s = 0$ for all $s = 1, ..., N$, and $\alpha_t = 0$ for all $t = 1, ..., T$ when generating the data but include a full set of individual and time dummies when estimating the model. To examine the behavior of our procedure in a relevant, reasonably small sample setting, we set $N = 49$ and use the locations of the 48 contiguous United States plus Washington D.C. to generate spatial dependence. In particular, we generate x_{st} as

$$x_{st} = u_{st} + \gamma \sum_{d(s,r)=1} u_{rt} \quad \text{where}$$

$$u_{st} = \rho u_{s(t-1)} + v_{st},$$

$$v_{st} \sim N(0, 1),$$

$d(s, r)$ is one for adjacent states s and r and zero otherwise, and ρ and γ respectively control the strength of the intertemporal and spatial correlation and are varied in the simulations.

We generate ε_{st} using the same model. We considered simulations for all combinations of ρ and γ with $\rho \in \{0, .3, .6\}$ and $\gamma \in \{.3, .6\}$.[12] Many models are estimated with state level panels with T between 10 and 20, and we arbitrarily chose to set $T = 13$.

We report rejection rates for 5% level tests from this simulation experiment in Table 2. In this case, our HAC estimator needs to account for dependence both in space and over time. In implementing the HAC estimators, we choose to use the average latitude and average longitude of each state measured in degrees to measure its spatial location and treat them as Cartesian coordinates. As in many spatial applications, here there are many possible location metrics, and the metric on which we base our HAC estimators (latitude, longitude) is intentionally different from the data generating process. As in the simulation results reported in the pure cross-sectional case, we consider either Bartlett or Gaussian kernels of the product form. We also allow the smoothing parameter in each dimension (latitude, longitude, and time) to be different. When reporting the results, we refer to tests based on a given HAC estimator by the kernel and the three smoothing parameters. Thus, results in a row labeled Bartlett$(h_{lat}, h_{long}, h_{time})$ correspond to using a kernel $K(st, r\tau) = (1 - |latitude_s - latitude_r|/h_{lat})^+(1 - |longitude_s - longitude_r|/h_{long})^+(1 - |t - \tau|/h_{time})^+$, and results in a row labeled Gaussian$(h_{lat}, h_{long}, h_{time})$ correspond to using a kernel $K(st, r\tau) = \exp\{-.5[(latitude_s - latitude_r)/(h_{lat}/2)]^2\} \exp\{-.5[(longitude_s - longitude_r)/(h_{long}/2)]^2\} \exp\{-.5[(t-\tau)/(h_{time}/2)]^2\}$. We also report results using a clustered

[12] We also considered $\gamma = 0$ but do not report these results for brevity. When $\rho = \gamma = 0$, the results are similar to those reported in the pure cross-section case above for $\gamma = 0$. When $\gamma = 0$ otherwise, simply clustering by state and using critical values from a t distribution with 48 (N-1) degrees of freedom as in Hansen (2007) works very well.

covariance estimator where we cluster by state or time[13] and results based on conventional OLS standard errors (IID) and heteroskedasticity robust standard errors (Heteroskedasticity).

The rejection rates summarized in Table 2 show more or less the same patterns as the results from the pure cross-section case discussed above. Large size distortions occur when either conventional OLS or heteroskedasticity consistent standard errors are used to do inference. We also see large size distortions in tests that use a HAC estimator together with standard normal critical values; this is true with either kernel and any of the bandwidths we consider (including bandwidths corresponding to clustering by state or time). In this example, we see dramatic improvements in size of tests by employing critical values corresponding to the approximation developed in this paper relative to the usual asymptotic approximation. It is interesting that only Gaussian(20,40,∞) controls size uniformly across the strengths of dependence considered, though many of the kernel and bandwidth combinations considered do quite well across each $\{\gamma, \rho\}$ combination.

Even when there is only modest spatial and temporal correlation, one needs to use rather large values for smoothing parameters in order to control the size of tests. As in the pure cross-section case, this involves using many cross-products, and the simulation results clearly demonstrate that the usual approximation which ignores uncertainty in estimating the covariance matrix is inadequate in this scenario. In addition to demonstrating the potential

[13] Note that cluster by state corresponds to using a HAC estimator with any kernel and bandwidths $h_{lat} = 0$, $h_{long} = 0$, and $h_{time} = \infty$, e.g. Bartlett(0,0,∞) or Gaussian(0,0,∞), and cluster by time corresponds to using a HAC estimator with any kernel and bandwidths $h_{lat} = \infty$, $h_{long} = \infty$, and $h_{time} = 0$, e.g. Bartlett(∞,∞,0) or Gaussian(∞,∞,0). In reporting results for these estimators, we use critical values obtained using the same bootstrap we use with the other HAC estimators.

large value added by the procedure in this paper, the simulation results also clearly illustrate that inference procedures that do not adequately account for both spatial and temporal dependence may suffer from severe size distortions.

5. **Empirical Example: Temporary Help Employment**

To illustrate that the methods developed in this paper can be important in practice, we present a brief empirical example. Following Autor (2003), we examine the effect of exceptions to the common law doctrine of employment at will[14] that were recognized in many states during the late 1970s and 1980s on temporary help services employment. In particular, as discussed in Autor (2003), 46 state courts recognized exceptions to employment at will between 1973 and 1995 that limited employers' ability to dismiss workers and opened them to potentially costly litigation in the case of "unjust dismissal."[15] Autor (2003) argues that the recognition of unjust dismissal doctrine should have a positive effect on temporary help employment for a variety of reasons and examines the effect of this unjust dismissal doctrine on the growth of temporary help services employment.

For our empirical example, we use essentially the same data as Autor (2003) which consists of an annual state-level panel on temporary help services employment from 1979 to

[14]Employment at will was recognized throughout the U.S. by 1953 and held that employees and employers had unlimited discretion in terminating an employment relationship unless explicitly contracted against.

[15]See Autor (2003) for a thorough review of the relevant history of employment at will and unjust dismissal doctrine as well as a discussion of other relevant exceptions to employment at will.

1995.[16] In our example, we focus on estimating the baseline specification of Autor (2003):

$$\log(THS_{st}) = \beta(CommonLawExceptions_{st}) + \gamma \log(NonfarmEmployment_{st}) + \alpha_s + \alpha_t + \lambda_s t + \varepsilon_{st}$$

where THS_{st} is total temporary help services employment in state s at time t, $NonfarmEmployment_{st}$ is total employment in the nonfarm sector in state s at time t, $CommonLawExceptions_{st}$ is a vector of dummy variables where a dummy variable is 1 if the state has recognized the given exception,[17] α_s and α_t are respectively state and year effects, and $\lambda_s t$ allows for a state-specific time trend. Details of data construction along with descriptive statistics and detailed descriptions of each variable are in Autor (2003). Autor (2003) also provides a discussion of the strengths and weakness of his strategy for identifying the effect of common law exceptions on temporary help services employment. This discussion is largely tangential to the point of the exercise in this paper, which is to illustrate the potential impact of neglected spatial correlation on the precision of an estimator and the conclusions one might draw in an empirical setting.

In his paper, Autor (2003) notes that there are obvious signs of temporal correlation in the data and accounts for this when performing inference by using standard errors estimated by clustering by state. This approach allows for essentially arbitrary serial dependence within each state over time, but imposes that there is no cross-sectional correlation between states

[16] Our data set differs from that in Autor (2003) in that we exclude Alaska and Hawaii.

[17] As in Autor (2003), we use three dummy variables: one for unjust dismissal doctrine, one for public policy exceptions, and one for good faith exceptions. We choose to focus on estimates for the unjust dismissal doctrine as this is what Autor (2003) focuses on since the estimated effects of the other policies are small and very imprecise even before spatial correlation is taken into account.

in a given time period or across different states in different time periods. Given that there is a strong possibility of spatial correlation in the data, inference obtained by solely accounting for within state intertemporal correlation is potentially highly misleading.

Figures 1 and 2 provide evidence of possible spatial correlation in the residuals. Figure 1 plots the residuals from the regression of $\log(THS_{st})$ on state and year effects and state specific trends, and Figure 2 plots the residuals from the regression of the unjust dismissal dummy (the covariate of interest) on state and year effects and state specific trends. For both the outcome and treatment, we present results for 1979, 1987, and 1995. Looking at these figures, we see obvious signs of spatial correlation even after accounting for the state effects, year effects, and state specific trends. For example, looking at Figure 1 for the outcome, we see that there are large chunks of geographically close states that have very similar outcome values (i.e. the same shading on the map). We also see, with few exceptions, that physically close states are similarly shaded and that transitions between light and dark shaded states (high and low values of the outcome) appear to occur fairly smoothly. The maps for the treatment in Figures 2 display similar behavior. While clearly not conclusive, these figures suggest moderate spatial correlation in the data which, if not properly accounted for, can cause standard inference procedures to produce misleading results. These figures also suggest that physical distance between states will provide a reasonable metric to describe the relevant spatial correlation patterns.

To account for this potential correlation, we employ a HAC estimator and the approximation to the behavior of test statistics developed in this paper. We index observations by their physical location as measured by average latitude and longitude and by their temporal location. Using the panel data simulations from the previous section as a rough guide, we employ a Gaussian kernel when forming the HAC estimator. Thus, the kernel is

$$K(st, r\tau) = \exp\{-.5[(latitude_s - latitude_r)/(h_{lat}/2)]^2\}$$
$$\times \exp\{-.5[(longitude_s - longitude_r)/(h_{long}/2)]^2\}$$
$$\times \exp\{-.5[(t-\tau)/(h_{time}/2)]^2\}$$

which depends on three smoothing parameters: h_{lat} for the latitude dimension, h_{long} for the longitude dimension, and h_{time} for the time dimension. We choose $h_{lat} = 20$, $h_{long} = 40$, and $h_{time} = \infty$ again based on the panel simulations as these are the setting that control size best in the simulations and we anticipate a moderate degree of spatial correlation and a high degree of intertemporal correlation. These bandwidths allow for essentially unconstrained within state intertemporal correlation and also allow for a high degree of spatial correlation across physically close states within a given year as well as a high degree of correlation across different but physically close states in different years. For comparison, we report results from clustering by state as in Autor (2003), which corresponds to the most common method used to estimate standard errors in this context, as well as results where we set smoothing parameters $h_{lat} = 10$, $h_{long} = 20$, and $h_{time} = \infty$, which would be appropriate if we believed spatial correlation decayed fairly rapidly.

Results from this exercise are reported in Table 3. The point estimate for the effect of the presence of unjust dismissal doctrine on the natural log of temporary help services employment is 0.140 and the clustered standard error is 0.059.[18] In the table, we also report standard errors using the spatial HAC estimator with our preferred smoothing parameter

[18]These numbers are very similar to those reported in Autor (2003). We get identical numbers to those in Autor (2003) when we include Alaska and Hawaii. The clustered standard error we report is identical to that reported in Stata which includes a bias adjustment factor as in Hansen (2007).

settings. For each estimated standard error, we report 95% confidence intervals formed using the usual asymptotic normal approximation (95% CI - Usual) and the fixed-b asymptotic approximation (95% CI - Fixed-b).[19] For the fixed-b confidence intervals using the HAC estimators, we obtain critical values using the i.i.d. bootstrap as advocated in this paper with 25,000 bootstrap replications.

When looking at the results, note that HAC estimators with large values for the smoothing parameters, in addition to being highly variable, may also be substantially biased towards zero.[20] Thus, the cluster by state standard error estimate, which includes a bias adjustment term, is not strictly comparable to the HAC estimators which do not.[21] This point also helps underscore the importance of using appropriate critical values, as the usual asymptotic normal critical values ignore both the variability and bias that HAC estimators exhibit in finite samples. The fixed-b approximation we develop accommodates both of these factors and so the critical values appropriately adjust for both. Evidence for this was provided in the simulations in the previous section.

[19]For the cluster estimator, we use critical values from a t-distribution with 47 degrees of freedom as in Hansen (2007) which is obtained under an approximating sequence for a clustering estimator that is analogous to the fixed-b limit for more general HAC estimators.

[20]With no truncation, HAC estimators are identically equal to 0, and as the value of the smoothing parameters increase, one approaches the no truncation point.

[21]For clustering estimators, the form of the bias is simple, depending only on the number of clusters, and so is readily adjusted for in obtaining estimates of the standard errors. Then, when obtaining critical values, one only needs to account for the variability of the clustered covariance estimator. This differs in the HAC case where the bias is a nontrivial function of both the chosen kernel and values of the smoothing parameters making it more convenient to use critical values that simultaneously adjust for the bias and variance.

Looking at the confidence intervals, we first note that the standard errors from HAC estimators, which allow for correlation in the residuals across states both within and between time periods, are larger than the standard error from clustering by state. In all three cases, however, confidence intervals based on the standard normal approximation exclude zero. Using the standard normal approximation at the 95% level, one would find significant evidence that unjust dismissal doctrines were associated with the large increases in temporary health services employment over this time period after controlling for state and year effects, state specific time trends, and other relevant exceptions to employment at will.

Confidence intervals based on our fixed-b approximation and bootstrap procedure are wider, substantially so in the case of the HAC estimators. More importantly, the difference in the inference one would draw after accounting for cross-sectional correlation is economically meaningful. Both of the HAC estimator intervals include large positive and economically meaningful negative values for the treatment effect. This indicates that, after accounting for correlation in labor market outcomes across states and time periods, there is insufficient evidence in the data to conclude that unjust dismissal doctrines are statistically associated with the rise in temporary help services employment seen during the 1980s and early 1990s. Note that this result is driven more by larger critical values, which correctly account for bias and variability in nonparametric standard error estimates which take spatial-temporal correlation into account, than by the actual point estimates of the standard error.

Overall, this example suggests that accounting for spatial correlation may be important in applied work. It also illustrates the importance not only of thinking about correlation when obtaining standard error estimates but also in using appropriate reference distributions. We believe this example argues that empirical researchers interested in obtaining valid inference should think carefully about how to account for dependence when performing inference, both in calculation of standard errors and use of appropriate reference distributions. The simulation evidence presented in the previous section suggests that our approach offers one way for empirical researchers interested in obtaining correct inference to do this. We also believe the practicality of the i.i.d. bootstrap approach to obtaining critical values should make it appealing to conscientious researchers.

6. Conclusion

In this paper, we have provided asymptotic approximation results for inference procedures based on spatial HAC covariance estimators under asymptotics in which HAC smoothing parameters are proportional to the sample size. Under this asymptotic sequence, the spatial HAC estimator is not consistent but instead converges in distribution to a non-degenerate random variable that depends on the HAC kernel and smoothing parameters. This random variable has expectation that is proportional to the true covariance matrix that is being estimated with bias that depends on the HAC kernel function and smoothing parameters. Asymptotic distributions of test statistics based on the spatial HAC estimator under this sequence thus better account for sampling properties of HAC estimators than the usual approximations that treat the HAC estimator as if it is equal to the true covariance matrix. The approximate distributions for test statistics that we obtain are pivotal but nonstandard, so critical values must be obtained via simulation. We show that the i.i.d. bootstrap provides a simple, feasible simulation procedure that can be used to construct valid critical values.

We illustrate the use of spatial HAC estimators and our approximation through brief simulation examples and an empirical example. In the simulations, we find that spatial

HAC estimators and inference based on the usual asymptotic approximation tends to be badly size distorted. We also find that inference based on spatial HAC estimators and the approximate distributions we develop on this paper controls size very well across all simulations we consider as long as sufficiently large values of smoothing parameters are used. In the empirical example, we consider the impact of the acceptance of exceptions to employment at will at the state level on state level temporary help services employment in a panel. Our findings suggest that the most commonly employed inference procedure in this type of data, using standard errors clustered by state which ignores spatial correlation, substantially overstates the precision with which the effects of interest are estimated relative to inference based on spatial HAC which accounts for the potential spatial correlation.

We believe the results in this paper will be very useful to researchers using common economic data in which spatial correlation is likely. While the asymptotic distributions are complicated, the ability to simply obtain critical values via the i.i.d. bootstrap should make using our approximation accessible to a wide range of researchers. One important question that deserves more attention is smoothing parameter selection. Our current results are helpful for smoothing parameter choice only to the extent that our simulation design matches the dependence structure researchers believe is present in their data. Our simulation results do suggest that one needs to use quite large values for smoothing parameters to control the size of tests based on HAC estimators. This is consistent with Sun, Phillips, and Jin (2008), who consider smoothing parameter selection in the context of Gaussian location model in a time series and show that the rate of increase for the optimal smoothing parameter chosen by trading off size and power under fixed-b asymptotics is much faster than the rate for minimizing MSE of the variance estimator. An interesting direction for future research would be to adapt the arguments of Sun, Phillips, and Jin (2008) to the present context.

7. Proofs

Throughout we adopt Einstein summation conventions such that for a sequence $b_{l,m}$,

$$b^{i,j} = \sum_{l=1}^{i}\sum_{m=1}^{j} b_{l,m}, \qquad b^{i}_{\cdot,m} = \sum_{l=1}^{i} b_{l,m}, \qquad b^{j}_{l,\cdot} = \sum_{m=1}^{j} b_{l,m}$$

where convenient.

Proof of Lemma 1: Using Assumption 2 and a mean-value expansion it follows that

$$\widehat{s}^*_{l,m} = s^*_{l,m} + \frac{\partial s^*_{l,m}(\overline{\theta})}{\partial \theta'}\left(\widehat{\theta}_N - \theta_0\right),$$

Summing over l and m up to $[rL]$ and $[sM]$ gives

$$\widehat{s}^{*[rL],[sM]} = s^{*[rL],[sM]} + \sum_{l=1}^{L}\sum_{m=1}^{M} \frac{\partial s^*_{l,m}(\overline{\theta})}{\partial \theta'}\left(\widehat{\theta}_N - \theta_0\right). \tag{7.1}$$

When $r = 1$ and $s = 1$, the left hand side of (7.1) becomes zero (see (2.2)) and this leads to

$$0 = \sum_{l=1}^{L}\sum_{m=1}^{M} s^*_{l,m}(\theta_0) + \sum_{l=1}^{L}\sum_{m=1}^{M} \frac{\partial s^*_{l,m}(\overline{\theta})}{\partial \theta'}\left(\widehat{\theta}_N - \theta_0\right). \tag{7.2}$$

Rearrangement of (7.2) and scaling gives

$$\begin{aligned}
\sqrt{LM}\left(\widehat{\theta}_N - \theta_0\right) &= -\left((LM)^{-1}\sum_{l=1}^{L}\sum_{m=1}^{M} \frac{\partial s^*_{l,m}(\overline{\theta})}{\partial \theta'}\right)^{-1} (LM)^{-1/2}\sum_{l=1}^{L}\sum_{m=1}^{\bar{M}} s^*_{l,m}(\theta_0) \\
&\Rightarrow -\mathcal{J}^{-1}\Lambda W_p(1,1),
\end{aligned}$$

using Assumptions 3 and 4.

Before proving the theorems, it is convenient to establish the limiting behavior of $\widehat{s}^{*[rL],[sM]}$. Scaling (7.1) by $(LM)^{-1/2}$ gives

$$\begin{aligned}(LM)^{-1/2}\widehat{s}^{*[rL],[sM]} &= (LM)^{-1/2}s^{*[rL],[sM]} + (LM)^{-1}\sum_{l=1}^{L}\sum_{m=1}^{M}\frac{\partial \mathbf{1}_{l,m}s_{l,m}(\overline{\theta})}{\partial\theta'}\sqrt{LM}\left(\widehat{\theta}_N-\theta_0\right),\\ &= (LM)^{-1/2}s^{*[rL],[sM]} + \widehat{J}_{\overline{\theta}}^{[rL],[sM]}\sqrt{LM}\left(\widehat{\theta}_N-\theta_0\right)\\ &\Rightarrow \Lambda W_p(r,s) - (rs\mathcal{J})\mathcal{J}^{-1}\Lambda W_p(1,1),\\ &\equiv \Lambda\left(W_p(r,s)-rsW_p(1,1)\right) = \Lambda B_p(r,s), \qquad (7.3)\end{aligned}$$

where line three of (7.3) follows from Lemma 1 and Assumptions 3 and 4.

Proof of Theorem 1: As stated above, K is a 2-d kernel defined on $U\times U$ with $U\subseteq\mathbb{R}^2$. Define the following:

$$\begin{aligned}K^*_{l_1,l_2,m_1,m_2} &= K\left(\frac{l_1}{b_1L},\frac{l_2}{b_1L},\frac{m_1}{b_2M},\frac{m_2}{b_2M}\right)\\ -\Delta_{l_1}K^*_{l_1,l_2,m_1,m_2} &= \left(K^*_{l_1,l_2,m_1,m_2}-K^*_{l_1+1,l_2,m_1,m_2}\right)\\ -\Delta^2_{m_1,l_1}K^*_{l_1,l_2,m_1,m_2} &= -(\Delta_{l_1}\circ\Delta_{m_1})K^*_{l_1,l_2,m_1,m_2}\\ &= \left(K^*_{l_1,l_2,m_1,m_2}-K^*_{l_1+1,l_2,m_1,m_2}\right)-\left(K^*_{l_1,l_2,m_1,m_2}-K^*_{l_1+1,l_2,m_{1+1},m_2}\right)\end{aligned}$$

Kiefer and Vogelsang (2005) use the identity

$$\sum_{l=1}^{L}a_lb_l=\sum_{l=1}^{L-1}(a_l-a_{l+1})\left(\sum_{i=1}^{l}b_i\right)+a_L\sum_{l=1}^{L}b_l.$$

We use a similar identity for double sums,

$$\sum_{l=1}^{L}\sum_{m=1}^{M}a_{l,m}b_{l,m}=\sum_{l=1}^{L-1}\sum_{m=1}^{M-1}-\Delta^2_{m,l}a_{l,m}b^{l,m}+\sum_{l=1}^{L-1}-\Delta_la_{l,M}b^{l,M}+\sum_{m=1}^{M-1}-\Delta_ma_{L,m}b^{L,m}+a_{L,M}b^{L,M},$$

which may be derived by applying the original identity to the left hand side, first to the sum over m, then again to the sum over l for each of the resulting terms. We apply this identity to (2.3), first to the sums over l_1 and m_1, then to the sums over l_2 and m_2 for each of the resulting terms. The resulting expression for $\widehat{\Omega}$ is

$$
\begin{aligned}
\widehat{\Omega} = \frac{1}{LM}\Bigg\{ & \sum_{l_1,l_2,m_1,m_2}^{\substack{L-1,L-1,\\ M-1,M-1}} -\Delta^4_{m_2,l_2,m_1,l_1} K^*_{l_1,l_2,m_1,m_2} \widehat{s}^{*l_1,m_1} \widehat{s}^{*l_2,m_2\prime} \\
& + \sum_{l_2,m_1,m_2}^{\substack{L-1,\\ M-1,M-1}} -\Delta^3_{m_2,l_2,m_1} K^*_{L,l_2,m_1,m_2} \widehat{s}^{*L,m_1} \widehat{s}^{*l_2,m_2\prime} \\
& + \sum_{l_1,m_1,m_2}^{\substack{L-1,\\ M-1,M-1}} -\Delta^3_{m_2,m_1,l_1} K^*_{l_1,L,m_1,m_2} \widehat{s}^{*l_1,m_1} \widehat{s}^{*L,m_2\prime} \\
& + \sum_{l_1,l_2,m_2}^{\substack{L-1,L-1,\\ M-1}} -\Delta^3_{l_2,m_2,l_1} K^*_{l_1,l_2,M,m_2} \widehat{s}^{*l_1,M} \widehat{s}^{*l_2,m_2\prime} \\
& + \sum_{l_1,l_2,m_1}^{\substack{L-1,L-1,\\ M-1}} -\Delta^3_{l_2,m_1,l_1} K^*_{l_1,l_2,m_1,M} \widehat{s}^{*l_1,m_1} \widehat{s}^{*l_2,M\prime} \\
& + \sum_{l_1,l_2}^{L-1,L-1,} -\Delta^2_{l_2,l_1} K^*_{l_1,l_2,M,M} \widehat{s}^{*l_1,M} \widehat{s}^{*l_2,M\prime} + \sum_{l_1,m_2}^{L-1,M-1} -\Delta^2_{m_2,l_1} K^*_{l_1,L,M,m_2} \widehat{s}^{*l_1,M} \widehat{s}^{*L,m_2\prime} \\
& + \sum_{l_2,m_1}^{L-1,M-1} -\Delta^2_{l_2,m_1} K^*_{L,l_2,m_1,M} \widehat{s}^{*L,m_1} \widehat{s}^{*l_2,M\prime} \Bigg\}
\end{aligned} \tag{7.4}
$$

We now need to establish convergence of the RHS terms in (2.3) to the appropriate functionals of the Brownian sheet, i.e.

$$
\widehat{\Omega}^{\emptyset} = \frac{1}{LM} \sum_{l_1,l_2,m_1,m_2}^{\substack{L-1,L-1,\\ M-1,M-1}} -\Delta^4_{m_2,l_2,m_1,l_1} K^*_{l_1,l_2,m_1,m_2} \widehat{s}^{*l_1,m_1} \widehat{s}^{*l_2,m_2\prime} \;\;\Rightarrow\;\; \Lambda Q^{\emptyset}_p(b_1,b_2)\Lambda'
$$

$$
\widehat{\Omega}^{\{1\}} = \frac{1}{LM} \sum_{l_2,m_1,m_2}^{\substack{L-1,\\ M-1,M-1}} -\Delta^3_{m_2,l_2,m_1} K^*_{L,l_2,m_1,m_2} \widehat{s}^{*L,m_1} \widehat{s}^{*l_2,m_2\prime} \;\;\Rightarrow\;\; \Lambda Q^{\{1\}}_p(b_1,b_2)\Lambda'
$$

$$\widehat{\Omega}^{\{1,2\}} = \frac{1}{LM} \sum_{m_1,m_2}^{M-1,M-1} -\Delta^2_{m_2,m_1} K^*_{L,L,m_1,m_2} \widehat{s}^{*L,m_1} \widehat{s}^{*\bar{L},m_2\prime} \;\Rightarrow\; \Lambda Q_p^{\{1,2\}}(b_1,b_2)\Lambda'$$

and so on for each of the terms in (7.4).

Starting with the first term, we can write

$$\widehat{\Omega}^{\emptyset} = \frac{1}{LM} \sum_{l_1=1,m_1=1}^{L,M} \frac{1}{LM} \sum_{l_2=1,m_2=1}^{L,M} -L^2M^2\Delta^4_{m_2,l_2,m_1,l_1} K^*_{l_1,l_2,m_1,m_2} \left(\frac{1}{\sqrt{LM}}\widehat{s}^{*l_1,m_1}\right)\left(\frac{1}{\sqrt{LM}}\widehat{s}^{*l_2,m_2\prime}\right)$$

Similarly to Kiefer and Vogelsang (2002), we define a process $Y_{L,M}(r,s)$ such that

$$Y_{L,M}(r,s) = (LM)^{-1/2}\widehat{s}^{*[rL],[sM]} \Rightarrow \Lambda B_p(r,s),$$

where the limit follows from (7.3).

Observing that as $L, M \to \infty$, $-L^2M^2\Delta^4_{m_2,l_2,m_1,l_1} K^*_{l_1,l_2,m_1,m_2} \to \frac{\partial^4 K^*}{\partial x_1 \partial x_2 \partial x_3 \partial x_4}\Big|_{l_1,l_2,m_1,m_2}$, and using Assumption 3 and the continuous mapping theorem, we have that

$$\begin{aligned}
\widehat{\Omega}^{\emptyset} &= \int_{\substack{\mathbf{u}\in[0,1]^2\\ \mathbf{v}\in[0,1]^2}} -L^2M^2\left(\Delta^4_{m_2,l_2,m_1,l_1} K^*_{[\mathbf{u}L,\mathbf{v}M]}\right) Y_{L,M}(\mathbf{u})Y_{L,M}(\mathbf{v})'d\left(\mathbf{u}\times\mathbf{v}\right)\\
&\Rightarrow \Lambda\left[\int_{\substack{\mathbf{u}\in[0,1]^2\\ \mathbf{v}\in[0,1]^2}} -\frac{\partial^4 K^*}{\partial x_1\partial x_2\partial x_3\partial x_4}(\mathbf{u},\mathbf{v}) B_p(\mathbf{u})B_p(\mathbf{v})'d\left(\mathbf{u}\times\mathbf{v}\right)\right]\Lambda'\\
&= \Lambda Q_p^{\emptyset}(b_1,b_2)\Lambda'.
\end{aligned}$$

This argument is identical to Kiefer and Vogelsang (2005) except that the convergence is to functionals involving Brownian sheets rather than Brownian motion. Note that this requires a more general FCLT (Assumption 3).

The other terms in (7.4) are boundary terms in the sense that they arise from the covariance between scores on the boundary with those on the interior. The argument for these terms is similar. Write $\widehat{\Omega}^{\{1\}}$ as

$$\widehat{\Omega}^{\{1\}} = \left(\frac{1}{M} \sum_{m_1=1}^{M} \frac{1}{LM} \sum_{l_2=1,m_2=1}^{L,M} -LM^2 \Delta^3_{m_2,l_2,m_1} K^*_{L,l_2,m_1,m_2} \left(\frac{1}{\sqrt{LM}} \widehat{s}^{*L,m_1} \right) \left(\frac{1}{\sqrt{LM}} \widehat{s}^{*l_2,m_2\prime} \right) \right).$$

The far right term involving $\widehat{s}^{*l_2,m_2}$ follows the process $Y_{L,M}(\frac{l_2}{L}, \frac{m_2}{M})$ given above. We also have that

$$(LM)^{-1/2} \widehat{s}^{*L,[sM]} = (LM)^{-1/2} S^{L,[sM]} - \widehat{J}^{L,[sM]} \left(\widehat{J}^{L,M} \right)^{-1} (LM)^{-1/2} s^{*L,M} = Y_{L,M}(1,s).$$

Therefore, by Assumption 3 and the continuous mapping theorem, we have

$$\begin{aligned}
\widehat{\Omega}^{\{1\}} &= \int_{\substack{u\in[0,1] \\ \mathbf{v}\in[0,1]^2}} -LM^2 \left(\Delta^3_{m_2,l_2,m_1} K^*_{[L,uL,\mathbf{v}M]} \right) Y_{L,M}(1,u) Y_{L,M}(\mathbf{v})' d\,(u \times \mathbf{v}) \\
&\Rightarrow \Lambda \left[\int_{\substack{u\in[0,1] \\ \mathbf{v}\in[0,1]^2}} - \frac{\partial^3 K^*}{\partial x_2 \partial x_3 \partial x_4} (1,u,\mathbf{v}) B_p(1,u) B_p(\mathbf{v})' d\,(u \times \mathbf{v}) \right] \Lambda' \\
&= \Lambda Q_p^{\{1\}}(b_1,b_2) \Lambda'
\end{aligned}$$

Proof of Theorem 2: As a preliminary calculation, consider the quadratic form

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K_{ij} a_i b_j,$$

where a_i and b_j are generic sequences and $K_{ij} = \left(1 - \left|\frac{i-j}{h}\right|\right) \mathbf{1}\left(\left|\frac{i-j}{h}\right| \le 1\right)$ with $1 < h \le n$. This quadratic form becomes a standard time series Bartlett HAC estimator when scaled by n^{-1} and the a_i and b_j sequences are the same sequence, e.g. residual times regressors in a regression model. Using algebraic results from Hashimzade and Vogelsang (2008) the quadratic form can be rewritten as

(7.5)
$$\sum_{i=1}^{n} \sum_{j=1}^{n} K_{ij} a_i b_j' = \frac{2}{h} \sum_{i=1}^{n-1} a^i b^{i\prime} - \frac{1}{h} \sum_{i=1}^{n-h-1} \left(a^{i+h} b^{i\prime} + a^i b^{i+h\prime} \right) - \frac{1}{h} \sum_{i=n-h}^{n-1} \left(a^n b^{i\prime} + a^i b^{n\prime} \right) + a^n b^{n\prime},$$

where Einstein summation conventions are used. Take (3.2) and rearrange the order of the sums to give

$$\begin{aligned}
\widehat{\Omega} &= \frac{1}{LM}\sum_{l_l=1}^{L}\sum_{m_1=1}^{M}\sum_{l_2=1}^{L}\sum_{m_2=1}^{M} K_{l_1,l_2}^{h_L} K_{m_1,m_2}^{h_M} \widehat{s}_{l_1,m_1}^{*} \widehat{s}_{l_2,m_2}^{*\prime} \\
&= \frac{1}{LM}\sum_{m_1=1}^{M}\sum_{m_2=1}^{M} K_{m_1,m_2}^{h_M} \sum_{l_l=1}^{L}\sum_{l_2=1}^{L} K_{l_1,l_2}^{h_L} \widehat{s}_{l_1,m_1}^{*} \widehat{s}_{l_2,m_2}^{*\prime}.
\end{aligned}$$

Applying (7.5) to the second double sum gives

$$\begin{aligned}
\widehat{\Omega} &= \frac{1}{LM}\sum_{m_1=1}^{M}\sum_{m_2=1}^{M} K_{m_1,m_2}^{h_M} \left(\frac{2}{h_L}\sum_{l=1}^{L-1} \widehat{s}_{\cdot,m_1}^{*l} \widehat{s}_{\cdot,m_2}^{*l\prime} - \frac{1}{h_L}\sum_{l=1}^{L-h_L-1} \left(\widehat{s}_{\cdot,m_1}^{*l+h_L} \widehat{s}_{\cdot,m_2}^{*l\prime} + \widehat{s}_{\cdot,m_1}^{*l} \widehat{s}_{\cdot,m_2}^{*l+h_L\prime} \right) \right. \\
&\left. - \frac{1}{h_L}\sum_{l=L-h_L}^{L-1} \left(\widehat{s}_{\cdot,m_1}^{*L} \widehat{s}_{\cdot,m_2}^{*l\prime} + \widehat{s}_{\cdot,m_1}^{*l} \widehat{s}_{\cdot,m_2}^{*L\prime} \right) + \widehat{s}_{\cdot,m_1}^{*L} \widehat{s}_{\cdot,m_2}^{*L\prime} \right).
\end{aligned}$$

Rearranging sums gives

$$\begin{aligned}
\widehat{\Omega} &= \frac{1}{LM}\left\{ \frac{2}{h_L}\sum_{l=1}^{L-1}\sum_{m_1=1}^{M}\sum_{m_2=1}^{M} K_{m_1,m_2}^{h_M} \widehat{s}_{\cdot,m_1}^{*l} \widehat{s}_{\cdot,m_2}^{*l\prime} - \frac{1}{h_L}\sum_{l=1}^{L-h_L-1}\sum_{m_1=1}^{M}\sum_{m_2=1}^{M} K_{m_1,m_2}^{h_M} \left(\widehat{s}_{\cdot,m_1}^{*l+h_L} \widehat{s}_{\cdot,m_2}^{*l\prime} + \widehat{s}_{\cdot,m_1}^{*l} \widehat{s}_{\cdot,m_2}^{*l+h_L\prime} \right) \right. \\
&\left. - \frac{1}{h_L}\sum_{l=L-h_L}^{L-1}\sum_{m_1=1}^{M}\sum_{m_2=1}^{M} K_{m_1,m_2}^{h_M} \left(\widehat{s}_{\cdot,m_1}^{*L} \widehat{s}_{\cdot,m_2}^{*l\prime} + \widehat{s}_{\cdot,m_1}^{*l} \widehat{s}_{\cdot,m_2}^{*L\prime} \right) + \sum_{m_1=1}^{M}\sum_{m_2=1}^{M} K_{m_1,m_2}^{h_M} \widehat{s}_{\cdot,m_1}^{*L} \widehat{s}_{\cdot,m_2}^{*L\prime} \right\}.
\end{aligned}$$

Expanding the double sums over m_1 and m_2 using (7.5) gives

$$\widehat{\Omega} = \frac{1}{LM}\left\{ \frac{2}{h_L}\sum_{l=1}^{L-1} A_{l,l} - \frac{1}{h_L}\sum_{l=1}^{L-h_L-1} \left(A_{l+h_L,l} + A_{l,l+h_L} \right) - \frac{1}{h_L}\sum_{l=L-h_L}^{L-1} \left(A_{L,l} + A_{l,L} \right) + A_{L,L} \right\},$$

where

$$\begin{aligned}
A_{l_1,l_2} &= \frac{2}{h_M}\sum_{m=1}^{M-1} \widehat{s}^{*l_1,m} \widehat{s}^{*l_2,m\prime} - \frac{1}{h_M}\sum_{m=1}^{M-h_M-1} \left(\widehat{s}^{*l_1,m+h_M} \widehat{s}^{*l_2,m\prime} + \widehat{s}^{*l_1,m} \widehat{s}^{*l_2,m+h_M\prime} \right) \\
&- \frac{1}{h_M}\sum_{m=M-h_M}^{M-1} \left(\widehat{s}^{*l_1,m} \widehat{s}^{*l_2,M\prime} + \widehat{s}^{*l_1,\bar{M}} \widehat{s}^{*l_2,m\prime} \right) + \widehat{s}^{*l_1,M} \widehat{s}^{*l_2,M\prime}.
\end{aligned}$$

Plugging in b_1L for h_L and b_2M for h_M and rearranging gives

$$\widehat{\Omega} = \frac{2}{b_1}\frac{1}{L}\sum_{l=1}^{L-1}\frac{1}{LM}A_{l,l} - \frac{1}{b_1}\frac{1}{L}\sum_{l=1}^{L-b_1L-1}\frac{1}{LM}\left(A_{l+b_1L,l} + A_{l,l+b_1L}\right) - \frac{1}{b_1}\frac{1}{L}\sum_{l=L-b_1L}^{L-1}\frac{1}{LM}\left(A_{L,l} + A_{l,L}\right) + \frac{1}{LM}A_{L,L}. \tag{7.6}$$

Consider the first term in (7.6):

$$\frac{2}{b_1}\frac{1}{L}\sum_{l=1}^{L-1}\frac{1}{LM}A_{l,l} = \frac{2}{b_1}\frac{1}{L}\sum_{l=1}^{L-1}\Bigg\{\frac{2}{b_2}\frac{1}{M}\sum_{m=1}^{M-1}(LM)^{-1/2}\widehat{S}^{*l,m}(LM)^{-1/2}\widehat{S}^{*l,m\prime}$$

$$-\frac{1}{b_2}\frac{1}{M}\sum_{m=1}^{M-b_2M-1}\left((LM)^{-1/2}\widehat{S}^{*l,m+b_2M}(LM)^{-1/2}\widehat{S}^{*l,m\prime} + (LM)^{-1/2}\widehat{S}^{*l,m}(LM)^{-1/2}\widehat{S}^{*l,m+b_2M\prime}\right)$$

$$-\frac{1}{b_2}\frac{1}{M}\sum_{m=M-b_2M}^{M-1}\left((LM)^{-1/2}\widehat{S}^{*l,m}(LM)^{-1/2}\widehat{S}^{*l,M\prime} + (LM)^{-1/2}\widehat{S}^{*l,M}(LM)^{-1/2}\widehat{S}^{*l,m\prime}\right)$$

$$+(LM)^{-1/2}\widehat{S}^{*l,M}(LM)^{-1/2}\widehat{S}^{*l,M\prime}\Bigg\}$$

$$= \frac{2}{b_1}\int_0^1\Bigg\{\frac{2}{b_2}\int_0^1 Y_{L,M}(r,s)Y_{L,M}(r,s)'ds$$

$$-\frac{1}{b_2}\int_0^{1-b_2}\left(Y_{L,M}(r,s+b_2)Y_{L,M}(r,s)' + Y_{L,M}(r,s)Y_{L,M}(r,s+b_2)'\right)ds$$

$$-\frac{1}{b_2}\int_{1-b_2}^1\left(Y_{L,M}(r,s)Y_{L,M}(r,1)' + Y_{L,M}(r,1)Y_{L,M}(r,s)'\right)ds + Y_{L,M}(r,1)Y_{L,M}(r,1)'\Bigg\}dr$$

$$\Rightarrow \Lambda\frac{2}{b_1}\int_0^1\Bigg\{\frac{2}{b_2}\int_0^1 B_p(r,s)B_p(r,s)'ds$$

$$-\frac{1}{b_2}\int_0^{1-b_2}\left(B_p(r,s+b_2)B_p(r,s)' + B_p(r,s)B_p(r,s+b_2)'\right)ds$$

$$-\frac{1}{b_2}\int_{1-b_2}^1\left(B_p(r,s)B_p(r,1)' + B_p(r,1)B_p(r,s)'\right)ds + B_p(r,1)B_p(r,1)'\Bigg\}dr\Lambda'$$

$$= \Lambda\frac{2}{b_1}\int_0^1\widetilde{Q}_p(r,r,b_2)dr\Lambda'.$$

Note that $\int_0^1 \widetilde{Q}_p(r, r, b_2)dr$ is the first term in the limiting expression of $Q_p^{bart}(b_1, b_2)$ as required. The limits of the remaining terms in (7.6) are similar and are omitted.

Proof of Theorem 3: Given (3.3) and Theorems 1 and 2, it is easy to show that F and t converge to functionals of $B_p(r, s)$, Λ, $\mathcal{J}$ and $R(\theta_0)$ using the continuous mapping theorem. Straightforward extensions of arguments used by Kiefer and Vogelsang (2005) in the proof of their Theorem 3 show that the limiting functionals can be rewritten in terms of $B_q(r, s)$ with Λ, $\mathcal{J}$ and $R(\theta_0)$ cancelling from the representations.

Proof of Theorem 4: The first steps of the proof are to show that versions of Assumptions 3 and 4 hold for the i.i.d. "unconditional on locations" bootstrap. To establish a FCLT for the bootstrap, recall that $s_{l,m}^{**} = s_{l,m}^{**}(\widehat{\theta}_N)$ and $E^*(s_{l,m}^{**}(\widehat{\theta}_N)) = \mathbf{0}$. Because of i.i.d. resampling, the variance covariance matrix of $s_{l,m}^{**}(\widehat{\theta}_N)$ is

$$E^*[s_{l,m}^{**}(\widehat{\theta}_N)s_{l,m}^{**}(\widehat{\theta}_N)'] = (LM)^{-1}\sum_{l=1}^{L}\sum_{m=1}^{M} s_{l,m}^{**}(\widehat{\theta}_N)s_{l,m}^{**}(\widehat{\theta}_N)' \equiv \widehat{\Gamma}_0^*.$$

Define the matrix $\widehat{\lambda}_0^*$ as the matrix square root of $\widehat{\Gamma}_0^*$, i.e. $\widehat{\Gamma}_0^* = \widehat{\lambda}_0^*\widehat{\lambda}_0^{*\prime}$. Then the random field $\widehat{\lambda}_0^{*-1}s_{l,m}^{**}(\widehat{\theta}_N)$ is stationary and mean zero with an identity variance covariance matrix in which case $\widehat{\lambda}_0^{*-1}s_{l,m}^{**}(\widehat{\theta}_N)$ satisfies a FCLT of the form

$$(LM)^{-1/2}\widehat{\lambda}_0^{*-1}s^{**[rL],[sM]}(\widehat{\theta}_N) \Rightarrow W_p(r, s).$$

Note that $\widehat{\Gamma}_0^*$ is the sample variance of $s_{l,m}^{**}(\theta_0)$. Given that $\widehat{\theta}_N$ is assumed to be a consistent estimator and given that FCLTs for random fields require stationarity of the underlying random field, it follows that

$$\widehat{\Gamma}_0^* = \widehat{\lambda}_0^*\widehat{\lambda}_0^{*\prime} \xrightarrow{p} \lambda_0\lambda_0' = \Gamma_0$$

where Γ_0 is the variance matrix of $s^*_{l,m}(\theta_0)$. Therefore we have

$$(LM)^{-1/2} s^{**[rL],[sM]}(\widehat{\theta}_N) = \widehat{\lambda}^*_0 (LM)^{-1/2} \widehat{\lambda}^{*-1}_0 s^{**[rL],[sM]}(\widehat{\theta}_N) \Rightarrow \Lambda^* W_p(r,s) \tag{7.7}$$

where $\Lambda^* = \lambda_0$. Now consider

$$\frac{1}{LM} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'}$$

where $\overline{\theta}^*$ is between $\widehat{\theta}^*$ and $\widehat{\theta}_N$ in the vector sense. Because of i.i.d. resampling it follows that

$$E^* \left[\frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'} \right] = \frac{1}{LM} \sum_{l=1}^{L} \sum_{m=1}^{M} \frac{\partial s^*_{l,m}(\overline{\theta}^*)}{\partial \theta'} = \widehat{J}_{\overline{\theta}^*} \xrightarrow{p} \mathcal{J}$$

where the converge follows from Assumption 4 and the assumption $\widehat{\theta}^* - \widehat{\theta}_N \xrightarrow{p} 0$. Some algebra gives

$$\frac{1}{LM} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'} = \widehat{J}_{\overline{\theta}^*} \frac{1}{LM} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \widehat{J}^{-1}_{\overline{\theta}^*} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'} \xrightarrow{p} \mathcal{J} r s I_p = rs\mathcal{J}. \tag{7.8}$$

Note that the converge follows because

$$E^* \left[\widehat{J}^{-1}_{\overline{\theta}^*} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'} \right] = I_p$$

which in turn gives

$$\frac{1}{LM} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \widehat{J}^{-1}_{\overline{\theta}^*} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'} \xrightarrow{p} rsI_p.$$

Using similar arguments it follows that

$$\frac{1}{LM} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \frac{\partial s^{**}_{l,m}(\widehat{\theta}^*)}{\partial \theta'} \xrightarrow{p} rs\mathcal{J}. \tag{7.9}$$

The next steps in the proof are to establish the limits of $\sqrt{LM}\left(\widehat{\theta}^* - \widehat{\theta}_N\right)$ and $(LM)^{-1/2} s^{**[rL],[sM]}(\widehat{\theta}^*)$. Recall that $\widehat{\theta}_N$ is the population value of θ in the bootstrap world. Using the assumption that $\widehat{\theta}^* - \widehat{\theta}_N \xrightarrow{p} 0$, we can use the mean-value theorem to expand $s^{**}_{l,m}(\widehat{\theta}^*)$ around $\widehat{\theta}_N$ giving

$$(7.10) \qquad s^{**}_{l,m}(\widehat{\theta}^*) = s^{**}_{l,m}(\widehat{\theta}_N) + \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'} \left(\widehat{\theta}^* - \widehat{\theta}_N\right),$$

where $\overline{\theta}^*$ is between $\widehat{\theta}^*$ and $\widehat{\theta}_N$ in the vector sense. Summing over l and m gives

$$\mathbf{0} = \sum_{l=1}^{L}\sum_{m=1}^{M} s^{**}_{l,m}(\widehat{\theta}_N) + \sum_{l=1}^{L}\sum_{m=1}^{M} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'} \left(\widehat{\theta}^* - \widehat{\theta}_N\right),$$

where the left hand side follows from the first order conditions for $\widehat{\theta}^*$. After the usual algebraic manipulations and scaling it follows that

(7.11)

$$\sqrt{LM}\left(\widehat{\theta}^* - \widehat{\theta}_N\right) = -\left[\frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'}\right]^{-1} (LM)^{-1/2} s^{**L,M}(\widehat{\theta}_N) \Rightarrow -\mathcal{J}^{-1}\Lambda^* W_p(1,1).$$

using (7.7) and (7.8). Double partial summing (7.10) and scaling by $(LM)^{-1/2}$) gives

$$\begin{aligned}(LM)^{-1/2})\sum_{l=1}^{[rL]}\sum_{m=1}^{[sM]} s^{**}_{l,m}(\widehat{\theta}^*) &= (LM)^{-1/2})\sum_{l=1}^{[rL]}\sum_{m=1}^{[sM]} s^{**}_{l,m}(\widehat{\theta}_N) + \frac{1}{LM}\sum_{l=1}^{[rL]}\sum_{m=1}^{[sM]} \frac{\partial s^{**}_{l,m}(\overline{\theta}^*)}{\partial \theta'}\sqrt{LM}\left(\widehat{\theta}^* - \widehat{\theta}_N\right) \\ &\Rightarrow \Lambda^* W_p(r,s) - rs\mathcal{J}\mathcal{J}^{-1}\Lambda^* W_p(1,1) \equiv \Lambda^* B_p(r,s),\end{aligned}$$

where the convergence follows from (7.7), (7.8) and (7.11). The remainder of the proof follows closely the proofs of Lemma 1 and Theorems 1-3 and details are omitted. Note that we do not require $\Lambda^* = \Lambda$ because Λ^* cancels from the asymptotic distribution of the bootstrap statistic in the same way that Λ cancels from the asymptotic distribution of the actual statistic. This is the sense in which we do not require the bootstrap to capture the dependence in the data.

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Arellano, M. (1987): "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Economics and Statistics*, 49(4), 431–434.

Autor, D. H. (2003): "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing," *Journal of Labor Economics*, 21(1), 1–42.

Basu, A. K., and C. C. Y. Dorea (1979): "On Functional Central Limit Theorem for Stationary Martingale Random Fields," *Acta Mathematica Academia Scientiarum Hungaricae*, 33, 307–316.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 119, 249–275.

Bester, C. A., T. G. Conley, and C. B. Hansen (2008): "Inference for Dependent Data Using Cluster Covariance Estimators," working paper, Chicago GSB.

Conley, T. G. (1996): *Econometric Modelling of Cross-Sectional Dependence*. Ph.D. Dissertation, University of Chicago.

Conley, T. G. (1999): "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92, 1–45.

Deo, C. (1975): "A Functional Central Limit Theorem for Stationary Random Fields," *The Annals of Probability*, 3(4), 708–715.

Driscoll, J., and A. Kraay (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *Review of Economics and Statistics*, 80(4), 549–560.

Fama, E. F., and J. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Goldie, C. M., and P. E. Greenwood (1986): "Variance of Set-Indexed Sums of Mixing Random Variables and Weak Convergence of Set-Indexed Processes," *The Annals of Probability*, 14(3), 817–839.

Gonçalves, S., and T. J. Vogelsang (2006): "Block Bootstrap HAC Robust Tests: The Sophistication of the Naive Bootstrap," Working Paper, Department of Economics, Michigan State University.

Gonçalves, S., and H. White (2004): "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," *Journal of Econometrics*, 119(1), 199–219.

Hansen, C. B. (2007): "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large," *Journal of Econometrics*, 141, 597–620.

Hashimzade, N., and T. J. Vogelsang (2008): "Fixed-b Asymptotic Approximation of the Sampling Behavior of Nonparametric Spectral Density Estimators," *Journal of Time Series Analysis*, 29, 142–162.

Ibragimov, R., and U. K. Müller (2006): "t-statistic Based Correlation and Heterogeneity Robust Inference," Mimeo.

Jansson, M. (2004): "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72(3), 937–946.

Kelejian, H. H., and I. Prucha (1999): "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40, 509–533.

——— (2001): "On the Asymptotic Distribution of the Moran I Test Statistic with Applications," *Journal of Econometrics*, 104, 219–257.

Kiefer, N. M., and T. J. Vogelsang (2002): "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory*, 18, 1350–1366.

——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164.

Lee, L.-f. (2004): "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Econometric Models," *Econometrica*, 72, 1899–1926.

——— (2007a): "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models," *Journal of Econometrics*, 137, 489–514.

——— (2007b): "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics*, 140, 333–374.

Liang, K.-Y., and S. Zeger (1986): "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73(1), 13–22.

McKeague, I. W., and Y. Sun (1996): "Transformations of Gaussian random fields to Brownian sheet and nonparametric change-point tests," *Statistics & Probability Letters*, 28, 311–319.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Sun, Y., P. C. B. Phillips, and S. Jin (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76(1), 175–194.

Vogelsang, T. J. (2008): "Panel Regression Inference Robust to Heteroskedasticity, Serial Correlation, and General Forms of Spatial Correlation," Working Paper, Department of Economics, Michigan State University.

Wooldridge, J. M. (2003): "Cluster-Sample Methods in Applied Econometrics," *American Economic Review*, 93(2), 133–188.

Table 1. Simulation Results from Cross-Sectional Design

		Full Lattice			Sparse Lattice		
	Ref. Dist.	$\gamma = 0$	$\gamma = .3$	$\gamma = .6$	$\gamma = 0$	$\gamma = .3$	$\gamma = .6$
IID	N(0,1)	0.051	0.381	0.490	0.056	0.053	0.395
Heteroskdasticity	N(0,1)	0.052	0.385	0.497	0.055	0.050	0.403
Bartlett(2)	N(0,1)	0.056	0.218	0.302	0.054	0.053	0.250
Bartlett(2)	Fixed-b	0.023	0.156	0.232	0.030	0.026	0.181
Gaussian(2)	N(0,1)	0.058	0.182	0.253	0.055	0.053	0.212
Gaussian(2)	Fixed-b	0.023	0.121	0.171	0.030	0.026	0.142
Bartlett(4)	N(0,1)	0.060	0.140	0.169	0.064	0.056	0.133
Bartlett(4)	Fixed-b	0.024	0.085	0.112	0.032	0.025	0.086
Gaussian(4)	N(0,1)	0.063	0.107	0.139	0.067	0.063	0.103
Gaussian(4)	Fixed-b	0.022	0.064	0.081	0.029	0.025	0.062
Bartlett(8)	N(0,1)	0.072	0.133	0.137	0.079	0.075	0.110
Bartlett(8)	Fixed-b	0.023	0.067	0.073	0.028	0.029	0.051
Gaussian(8)	N(0,1)	0.085	0.122	0.122	0.093	0.092	0.092
Gaussian(8)	Fixed-b	0.020	0.045	0.050	0.030	0.031	0.037
Bartlett(16)	N(0,1)	0.121	0.164	0.173	0.127	0.126	0.121
Bartlett(16)	Fixed-b	0.028	0.050	0.058	0.032	0.028	0.047
Gaussian(16)	N(0,1)	0.169	0.195	0.192	0.181	0.173	0.123
Gaussian(16)	Fixed-b	0.027	0.039	0.040	0.035	0.033	0.036

Note: This table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment with data generated on a two-dimensional lattice. All results are based on 1000 simulation replications. Row labels indicate which covariance matrix estimator and smoothing parameter are used. IID and Heteroskedasticity use conventional OLS standard errors and heteroskedasticity robust standard errors respectively. Rows labeled Bartlett(h) use HAC estimators with a Bartlett kernel and smoothing parameter equal to h, and rows labeled Gaussian(h) use HAC estimators with a Gaussian kernel and smoothing parameter equal to h. In all cases we use product kernels and equal bandwidths in both coordinate directions. Column 2 indicates from which reference distribution critical values were obtained with "Fixed-b" indicating that the approximations developed in this paper were used. All Fixed-b critical values were obtained via the i.i.d. bootstrap with 200 bootstrap replications. Columns labeled "Full Lattice" use all data points where data are generated on a regular 25 x 25 lattice. Columns labeled "Sparse Lattice" use 625 randomly sampled (without replacement) observations where data were generated on a regular 36 x 36 lattice. All data were generated by a spatial moving average model as described in the text and γ provides a measure of the strength of the cross-sectional dependence.

Table 2. Simulation Results from Panel Design

		$\gamma = .3$			$\gamma = .6$		
	Ref. Dist.	$\rho = 0$	$\rho = .3$	$\rho = .6$	$\rho = 0$	$\rho = .3$	$\rho = .6$
IID	N(0,1)	0.184	0.218	0.330	0.299	0.313	0.407
Heteroskdasticity	N(0,1)	0.189	0.217	0.335	0.294	0.300	0.400
Bartlett(10,20,∞)	N(0,1)	0.195	0.189	0.209	0.251	0.224	0.243
Bartlett(10,20,∞)	Fixed b	0.061	0.062	0.064	0.096	0.089	0.088
Gaussian(10,20,∞)	N(0,1)	0.213	0.203	0.217	0.255	0.235	0.242
Gaussian(10,20,∞)	Fixed b	0.052	0.050	0.049	0.076	0.068	0.071
Bartlett(20,40,∞)	N(0,1)	0.279	0.278	0.300	0.330	0.319	0.315
Bartlett(20,40,∞)	Fixed b	0.054	0.059	0.058	0.087	0.077	0.076
Gaussian(20,40,∞)	N(0,1)	0.332	0.351	0.356	0.377	0.381	0.367
Gaussian(20,40,∞)	Fixed b	0.029	0.043	0.043	0.058	0.049	0.053
Bartlett(10,20,6)	N(0,1)	0.136	0.156	0.226	0.186	0.177	0.238
Bartlett(10,20,6)	Fixed b	0.061	0.075	0.111	0.089	0.084	0.141
Gaussian(10,20,6)	N(0,1)	0.140	0.156	0.222	0.176	0.175	0.224
Gaussian(10,20,6)	Fixed b	0.054	0.062	0.088	0.064	0.072	0.104
Cluster by State	N(0,1)	0.211	0.214	0.224	0.320	0.308	0.291
Cluster by State	Fixed b	0.126	0.109	0.137	0.227	0.190	0.194
Cluster by Time	N(0,1)	0.085	0.120	0.238	0.111	0.105	0.220
Cluster by Time	Fixed b	0.025	0.042	0.097	0.037	0.030	0.093

Note: This table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment with data generated as a state level panel. All results are based on 1000 simulation replications. Row labels indicate which covariance matrix estimator and smoothing parameters are used. IID and Heteroskedasticity use conventional OLS standard errors and heteroskedasticity robust standard errors respectively. Rows labeled Bartlett(h_{lat},h_{long},h_{time}) use HAC estimators with a Bartlett kernel and smoothing parameters equal to h_{lat} in the latitude direction, h_{long} in the longitude dimension, and h_{time} in the time dimension; and rows labeled Gaussian(h_{lat},h_{long},h_{time}) are defined similarly but use a Gaussian kernel. Column 2 indicates from which reference distribution critical values were obtained with "Fixed-b" indicating that the approximations developed in this paper were used. All Fixed-b critical values were obtained via the i.i.d. bootstrap with 200 bootstrap replications. ρ and γ respectively correspond to the strength of intertemporal and cross-sectional correlation and are defined in the text. All data were generated using locations as given by the 48 contiguous United States plus Washington D.C. with the spatial model described in the text, and the number of time periods in the panel is 13.

Table 3. Estimated Standard Errors and Confidence Intervals from Temporary Help Services Example

		HAC	
	Cluster by State	$h_{lat} = 10$, $h_{long} = 20$, $h_{time} = \infty$	$h_{lat} = 20$, $h_{long} = 40$, $h_{time} = \infty$
Estimated s.e.	0.059	0.068	0.062
95% CI - Usual	(0.025,0.255)	(0.007,0.273)	(0.018,0.261)
95% CI - Fixed-b	(0.022,0.258)	(-0.091,0.371)	(-0.201,0.480)

Note: This table reports estimated standard errors and 95% confidence intervals for the effect of the unjust dismissal doctrine on the natural log of temporary help services employment. The coefficient estimate is 0.140. We report standard errors estimated using a clustering estimator where we cluster by state and two other spatial HAC estimators where values of smoothing parameters are given as column labels. In forming the 95% confidence intervals, we report both the interval obtained using the usual asymptotic normal approximation (95% CI - Usual). For the row labeled "95% CI - Fixed-b", we use the asymptotic approximation developed in this paper for the two HAC estimators and obtain critical values through the bootstrap procedure. For the clustered covariance estimator, we report results using critical values from a t-distribution with 47 degrees of freedom as suggested in Hansen (2007) in the "95% CI - Fixed b" column.

Figure 1a. Temporary Help Services Employment Residual 1979

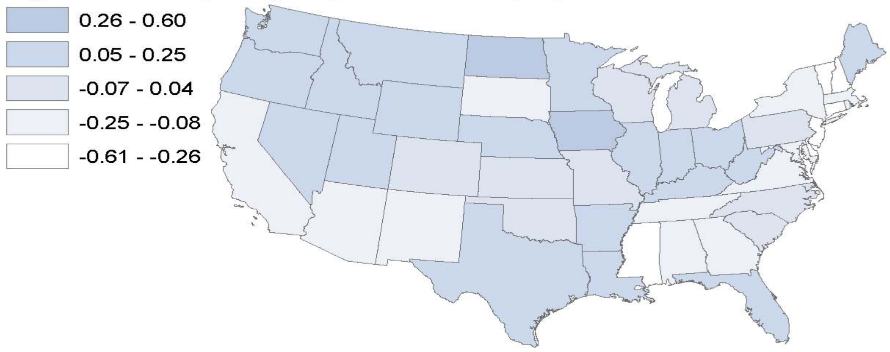

Figure 1b. Temporary Help Services Employment Residual 1987

Figure 1c. Temporary Help Services Employment Residual 1995

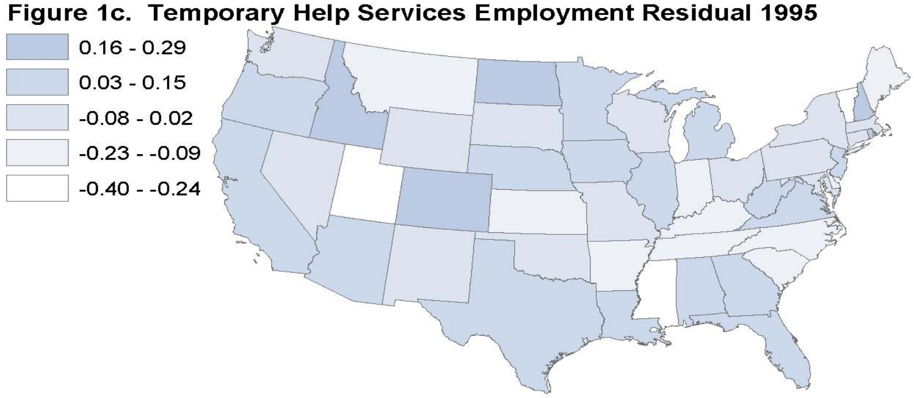

Figure 2a. Unjust Dismissal Residuals 1979

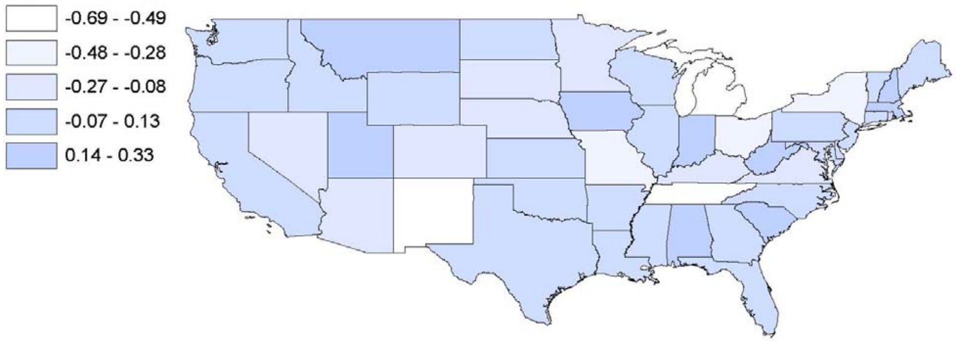

Figure 2b. Unjust Dismissal Residual 1987

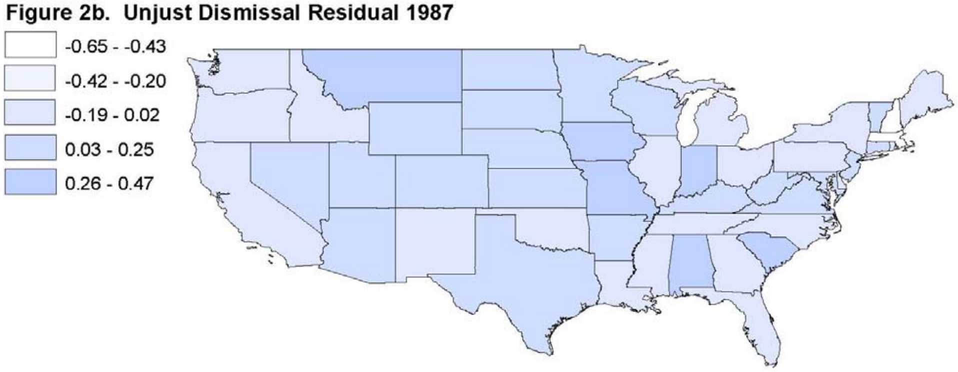

Figure 2c. Unjust Dismissal Residual 1995