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

C. Alan Bester, Timothy G. Conley, and Christian B. Hansen
University of Chicago Booth School of Business
Timothy J. Vogelsang
Michigan State University

June 5, 2009
Introduction

Many economic analyses rely on dependent data.

▷ Thinking about dependence in time series applications is common.

▷ Dependence may also exist in cross-sectional and other applications.

We consider inference in **spatially dependent data** – data where dependence is indexed in more than one dimension. Includes

▷ time series - dependence indexed in one dimension (time)

▷ cross-sectional dependence - dependence indexed in m spatial dimensions (e.g., latitude and longitude)

▷ panel - dependence indexed in m+1 dimensions (m spatial dimensions and time)
Usual Approach

- Plug-in HAC estimator for unknown covariance matrix in limiting normal approximation
- Rely on consistency of HAC estimator to justify inference based on Gaussianity
- Ignores estimation uncertainty in HAC estimator, leads to poor finite sample properties

Our Approach

- Use asymptotics where smoothing parameter is proportional to sample size (Kiefer and Vogelsang, 2002, 2005)
- HAC estimator converges to nondegenerate r.v.
- Test statistics asymptotically pivotal, but with nonstandard distributions
Our Contribution

Consider $t$- and Wald statistics using spatial HAC estimator (Conley, 1996, 1999) under ‘fixed-b’ sequence.

Characterize limit distributions of test statistics as functionals of multi-indexed Brownian motion.

Nonstandard limiting distributions imply critical values obtained by simulation. Complicated b/c limit distribution...

- involves integrals against multi-indexed Brownian process.
- depends on kernel and smoothing parameter.
- depends on shape of sampling region.

We show simple iid bootstrap can be used to obtain valid critical values regardless of sampling region.
Simulation Example

Data generated on rectangular integer lattice according to

- \( y_s = \alpha + x_s \beta + \varepsilon_s \)
- \( s \) is a vector \((s_1, s_2)\) indicating lattice point
- \( x_s = \sum_{\|j\| \leq 2} \gamma \|j\| v_{s+j} \) with \( v_s \sim N(0, 1) \)
- \( \varepsilon_s = \sum_{\|j\| \leq 2} \gamma \|j\| u_{s+j} \) with \( u_s \sim N(0, 1) \)
- \( \|j\| = \max(j_1, j_2) \)
- \( \gamma \) controls strength of correlation
- “Full lattice” is 25 × 25 (\( n = 625 \))
- “Sparse lattice” is 625 randomly chosen locations on 36 × 36 lattice
### Table 1. Simulation Results from Cross-Sectional Design

<table>
<thead>
<tr>
<th></th>
<th>Ref. Dist.</th>
<th>Full Lattice</th>
<th>Sparse Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\gamma = 0$</td>
<td>$\gamma = .3$</td>
</tr>
<tr>
<td>IID</td>
<td>N(0,1)</td>
<td>0.051</td>
<td>0.381</td>
</tr>
<tr>
<td>Heteroskedasticity</td>
<td>N(0,1)</td>
<td>0.052</td>
<td>0.385</td>
</tr>
<tr>
<td>Bartlett(2)</td>
<td>N(0,1)</td>
<td>0.056</td>
<td>0.218</td>
</tr>
<tr>
<td>Bartlett(2)</td>
<td>Fixed-b</td>
<td>0.023</td>
<td>0.156</td>
</tr>
<tr>
<td>Gaussian(2)</td>
<td>N(0,1)</td>
<td>0.058</td>
<td>0.182</td>
</tr>
<tr>
<td>Gaussian(2)</td>
<td>Fixed-b</td>
<td>0.023</td>
<td>0.121</td>
</tr>
<tr>
<td>Bartlett(16)</td>
<td>N(0,1)</td>
<td>0.121</td>
<td>0.164</td>
</tr>
<tr>
<td>Bartlett(16)</td>
<td>Fixed-b</td>
<td>0.028</td>
<td>0.050</td>
</tr>
<tr>
<td>Gaussian(16)</td>
<td>N(0,1)</td>
<td>0.169</td>
<td>0.195</td>
</tr>
<tr>
<td>Gaussian(16)</td>
<td>Fixed-b</td>
<td>0.027</td>
<td>0.039</td>
</tr>
</tbody>
</table>
HAC Estimators

- Parameter of interest: $\theta_0$ satisfies $E(s_i(\theta_0)) = 0$ (Use $s_i(\theta_0) = s_i$)
- Estimator: $\hat{\theta}$ satisfies $\frac{1}{N} \sum_{i=1}^{N} s_i(\hat{\theta}) = 0$

Asymptotic distribution:

$$\hat{\theta} \overset{\Delta}{=} N(\theta_0, J^{-1}\Omega J^{-1}/N)$$

$$\Omega = \lim_{N \to \infty} \text{Var} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} s_i \right), \quad J = E[\partial s_i / \partial \theta]$$

Usual approach: Take consistent estimator $\hat{\Omega}$ and “plug in” to get distribution for $\hat{\theta}$; use this for hypothesis tests, CIs.
HAC Estimators

With dependent data, need to use an estimator of $\Omega$ that is consistent under heteroskedasticity and autocorrelation in $s_i$.

$$\hat{\mathcal{V}}_{HAC} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} K_N(d(i,j)) s_i(\hat{\theta}) s_j(\hat{\theta})'$$

$W_N$ a weight function (e.g. Bartlett, Gaussian, ...), $d$ a **distance metric**.

Applies immediately to cross-sectional dependence, panels,...
HAC Estimators

Example $d$

- time series: $|i - j|$ (i.e. \# of lags)
- census tracts: $d(i, j) = 0$ if $i = j$; $d(i, j) = 1$ if $i$ and $j$ are physically adjacent; ...
- firms: $d(i, j) = ((S_i - S_j)^2 + (B_i - B_j)^2)^{1/2}$ where $S_i$ is firm size and $B_i$ is book-to-market

Lots of other candidate $d$. Can use metric as abstract as required by problem (i.e. need not be physical/temporal distance)
Rectangular Lattice

Focus on case where domain of index is 2-d rectangular lattice.

\[
\hat{\theta} : \frac{1}{LM} \sum_{l=1}^{L} \sum_{m=1}^{M} s_{l,m}(\hat{\theta}) = 0
\]

\[
\hat{\Omega} = \frac{1}{LM} \sum_{l_1=1}^{L} \sum_{m_1=1}^{M} \sum_{l_2=1}^{L} \sum_{m_2=1}^{M} K\left(\frac{l_1}{h_L}, \frac{l_2}{h_L}, \frac{m_1}{h_M}, \frac{m_2}{h_M}\right) \hat{s}_{l_1,m_1} \hat{s}'_{l_2,m_2}
\]

$L$ and $M$ are lengths of each lattice dimension. Identical to previous estimators in case of rectangular lattice.
Brownian Sheets and Related Processes

$W(u, v)$ denotes a 2-d Brownian sheet: A mean zero Gaussian process with covariance function

$$\text{Cov}(W(u_1, v_1), W(u_2, v_2)) = \min(u_1, u_2) \ast \min(v_1, v_2).$$

$W_p$ denotes a $p$–vector of independent Brownian sheets.

“Pinned” Brownian sheet: $B_p(u, v) = W_p(u, v) - uvW_p(1, 1)$. On edges, $B_p(1, v) = W_p(1, v) - vW_p(1, 1)$ behaves like 1-d Brownian bridge but correlated with values of the sheet on the interior.
Brownian Sheets and Related Processes
Brownian Sheets and Related Processes

Will make use of functionals:

\[
Q_p^\emptyset(b_1, b_2) = \int_{[0,1]^4} \frac{\partial^4 K}{\partial x_1 \partial x_2 \partial x_3 \partial x_4} \left( \frac{u_1}{b_1}, \frac{u_2}{b_1}, \frac{v_1}{b_2}, \frac{v_2}{b_2} \right) B_p(u_1, v_1) B_p(u_2, v_2) d(u_1 \times v_1 \times u_2 \times v_2)
\]

\[
Q_p^{\{1\}}(b_1, b_2) = \int_{[0,1]^3} \frac{\partial^3 K}{\partial x_2 \partial x_3 \partial x_4} \left( \frac{1}{b_1}, \frac{u_2}{b_1}, \frac{v_1}{b_2}, \frac{v_2}{b_2} \right) B_p(1, v_1) B_p(u_2, v_2) d(v_1 \times u_2 \times v_2)
\]

\[
Q_p^{\{1,4\}}(b_1, b_2) = \int_{[0,1]^2} \frac{\partial^2 K}{\partial x_2 \partial x_3} \left( \frac{1}{b_1}, \frac{u_2}{b_1}, \frac{v_1}{b_2}, \frac{1}{b_2} \right) B_p(1, v_1) B_p(u_2, 1) d(v_1 \times u_2)
\]

\[
Q_p(b_1, b_2) = Q_p^\emptyset(b_1, b_2) + \sum_{i=1}^{4} Q_p^{\{i\}}(b_1, b_2) + \sum_{E = \{1, 2\}, \{3, 4\}, \{2, 3\}, \{1, 4\}} Q_p^E(b_1, b_2)
\]

(Other functionals, e.g. \(Q_p^{\{2\}}, Q_p^{\{2,3\}}\) defined similarly)
Assumptions

A1. As $N \to \infty$, $L \to \infty$, $M \to \infty$, $\frac{L}{N} \to 0$ and $\frac{M}{N} \to 0$.

A2. $\hat{\theta}_N \xrightarrow{p} \theta_0$.

A3. $\frac{1}{\sqrt{LM}} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} s_{l,m} \Rightarrow \Lambda W(r,s)$ for all $(r, s) \in [0, 1]^2$ with $\Omega = \Lambda \Lambda'$.

A4. $\hat{J}_{\theta}^{[rL],[sM]} = (LM)^{-1} \sum_{l=1}^{[rL]} \sum_{m=1}^{[sM]} \frac{\partial s_{l,m}(\hat{\theta}_N)}{\partial \theta'} \xrightarrow{p} rsJ(\theta)$, where $J(\theta)$ is nonsingular and convergence is uniform over $\theta$. 
‘Fixed-b’ Spatial HAC Limit

Theorem

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,

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

- Result for Bartlett in paper
- $Q_p$ is a random variable
- Limit ‘proportional’ to $\Lambda \Lambda'$
Test Statistic Limits

**Theorem**

Suppose the conditions of Theorem 1 are satisfied, and $t$ and $F$ are the usual $t$- and Wald stats computed using $\hat{\theta}_N$ and $\hat{\Omega}$. Then

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

Similar result for Bartlett kernel.

Test stats asymptotically pivotal (sort of). Could be used to obtain critical values via simulation (for rectangular lattice).

Unfortunately these distributions are a pain to work with in practice.
Bootstrap Procedure

- Draw bootstrap sample by sampling (with replacement) conditional on observed locations
  - i.e. Use locations, distances from data
  - At each observed location, place value drawn by iid sampling with replacement from observed data
- Compute bootstrap estimate $\hat{\theta}_b$ and bootstrap test statistic $W_b$
- Repeat B times
- Find critical value from B realizations of bootstrap distribution

Produces bootstrap distribution that is first order valid for the fixed-$b$ asymptotic distribution. (Regularity conditions in paper. Mild under A1-A4.)
Why IID Bootstrap Works

Similar to Gonçalves and Vogelsang (2006).

Key Intuition: Limit dist. of HAC estimator proportional to $\Omega$, so

- Nuisance parameters $\Lambda$ cancel out of distributions of test statistics
- do not need to replicate true covariance structure
  - only need same FCLTs to apply to simulation samples
- need to mimic domain of process
  - assume observed locations representative of population domain
  - could also assume know shape of domain (see “unconditional bootstrap” in paper for rectangular lattice example)
Panel Simulation

- $y_{st} = \beta x_{st} + \alpha_s + \alpha_t + \varepsilon_{st}$
- $s$ indicates state (48 contiguous states + DC)
- $x_{st} = u_{st} + \gamma \sum_{d(s,r)=1} u_{rt}$ where $u_{st} = \rho u_{s(t-1)} + v_{st}$, $v_{st} \sim N(0,1)$, and $d(s, r)$ is one for adjacent states $s$ and $r$ and zero otherwise.
- $\varepsilon_{st}$ generated same way
- $\gamma \in \{.3, .6\}$ controls strength of correlation, $\rho \in \{0, .3, .6\}$

$N = 49, \ T = 13.$

1000 simulation replications.
Panel Simulation

Use the average latitude and average longitude of each state measured in degrees to measure its spatial location for HAC.

Consider two different kernels with smoothing parameters $h_{lat}$, $h_{long}$, and $h_{time}$:

- **Bartlett:** $K(st, r\tau) = (1 - |latitude_s - latitude_r|/h_{lat})^+(1 - |longitude_s - longitude_r|/h_{long})^+(1 - |t - \tau|/h_{time})^+$

- **Gaussian:** $K(st, r\tau) = \exp\{-0.5[(latitude_s - latitude_r)/(h_{lat}/2)]^2\} \exp\{-0.5[(longitude_s - longitude_r)/(h_{long}/2)]^2\} \exp\{-0.5[(t - \tau)/(h_{time}/2)]^2\}$
<table>
<thead>
<tr>
<th></th>
<th>Ref. Dist.</th>
<th>(\rho = 0)</th>
<th>(\rho = .3)</th>
<th>(\rho = .6)</th>
<th>(\rho = 0)</th>
<th>(\rho = .3)</th>
<th>(\rho = .6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IID</td>
<td>N(0,1)</td>
<td>0.184</td>
<td>0.218</td>
<td>0.330</td>
<td>0.299</td>
<td>0.313</td>
<td>0.407</td>
</tr>
<tr>
<td>Heteroskedasticity</td>
<td>N(0,1)</td>
<td>0.189</td>
<td>0.217</td>
<td>0.335</td>
<td>0.294</td>
<td>0.300</td>
<td>0.400</td>
</tr>
<tr>
<td>Cluster by State</td>
<td>N(0,1)</td>
<td>0.211</td>
<td>0.214</td>
<td>0.224</td>
<td>0.320</td>
<td>0.308</td>
<td>0.291</td>
</tr>
<tr>
<td>Cluster by State</td>
<td>Fixed b</td>
<td>0.126</td>
<td>0.109</td>
<td>0.137</td>
<td>0.227</td>
<td>0.190</td>
<td>0.194</td>
</tr>
<tr>
<td>Cluster by Time</td>
<td>N(0,1)</td>
<td>0.085</td>
<td>0.120</td>
<td>0.238</td>
<td>0.111</td>
<td>0.105</td>
<td>0.220</td>
</tr>
<tr>
<td>Cluster by Time</td>
<td>Fixed b</td>
<td>0.025</td>
<td>0.042</td>
<td>0.097</td>
<td>0.037</td>
<td>0.030</td>
<td>0.093</td>
</tr>
<tr>
<td>Bartlett(20,40,(\infty))</td>
<td>N(0,1)</td>
<td>0.279</td>
<td>0.278</td>
<td>0.300</td>
<td>0.330</td>
<td>0.319</td>
<td>0.315</td>
</tr>
<tr>
<td>Bartlett(20,40,(\infty))</td>
<td>Fixed b</td>
<td>0.054</td>
<td>0.059</td>
<td>0.058</td>
<td>0.087</td>
<td>0.077</td>
<td>0.076</td>
</tr>
<tr>
<td>Gaussian(20,40,(\infty))</td>
<td>N(0,1)</td>
<td>0.332</td>
<td>0.351</td>
<td>0.356</td>
<td>0.377</td>
<td>0.381</td>
<td>0.367</td>
</tr>
<tr>
<td>Gaussian(20,40,(\infty))</td>
<td>Fixed b</td>
<td>0.029</td>
<td>0.043</td>
<td>0.043</td>
<td>0.058</td>
<td>0.049</td>
<td>0.053</td>
</tr>
</tbody>
</table>
Also in the Paper...

- Non-rectangular domains (via FCLT in Dedecker, 2001)
- Simulations with endogeneity and weak instruments
- Comparison to ‘grouping’ schemes: Bester, Conley, and Hansen (2008) and Ibragimov and Muller (2007)
- Empirical example following Autor (2003)
Conclusions

- extend KV to spatial HAC
- approach is tractable via iid bootstrap
- performs well in simulations
- accounting for spatial dependence may be economically relevant

Bandwidth selection as problematic as ever. Practical suggestions:
- choose big bandwidths (i.e. low bias, high variance in HAC estimator)
- account for uncertainty in covariance matrix with appropriate reference dist