Spatial Unit Roots

Ulrich K. Müller and Mark W. Watson
Department of Economics, Princeton University
Princeton, NJ, 08544

This Draft: September 13, 2022

Abstract

This paper proposes a model for, and investigates the consequences of, strong spatial dependence in economic variables. Our approach and findings echo those of the corresponding "unit root" time series literature: We suggest a model for spatial $I(1)$ processes, and establish a Functional Central Limit Theorem that justifies a large sample Gaussian process approximation for such processes. We further generalize the $I(1)$ model to a spatial "local-to-unity" model that exhibits weak mean reversion. We characterize the large sample behavior of regression inference with independent spatial $I(1)$ variables, and establish that spurious regression is as much a problem with spatial $I(1)$ data as it is with time series $I(1)$ data. We also develop asymptotically valid spatial unit root and stationarity tests, as well inference for the local-to-unity parameter. And finally, we consider strategies for valid inference in regressions with persistent ($I(1)$ or local-to-unity) spatial data, such as spatial analogues of first-differencing transformations.

Keywords: spatial correlation, spurious regression, Lévy-Brownian motion
JEL: C12, C20

1 Introduction

Serial correlation complicates inference in time series regressions. When the serial correlation in the regressors and regression errors is weak, that is $I(0)$, inference can proceed as with i.i.d. sampling after using HAC/HAR standard errors that incorporate adjustments for serial correlation. However, when the serial correlation is strong, that is $I(1)$, OLS produces 'spurious regressions' (Granger and Newbold (1974)), HAC/HAR inference fails, and estimators and test statistics behave in non-standard ways (Phillips (1986)).

Variables measured over points in space exhibit correlation patterns that in many ways are analogous to serial correlation in time series, and this correlation complicates inference in spatial regressions. There is a reasonably well-developed literature on the requisite spatial HAC/HAR corrections that are required in spatial regressions with weakly dependent stationary regressors and errors.[1] However, much less is known about the implications of strong spatial correlation despite evidence suggesting its presence in many empirical applications in economics (Kelly (2019, 2020)). There are several natural questions. What is the spatial analogue of an $I(1)$ time series process? What are the properties of OLS estimators and test statistics in spatial regressions when regressors and errors follow these spatial $I(1)$ processes? Does strong spatial persistence produce spurious regressions? How can one test for $I(1)$ spatial persistence? And finally, is there a spatial analogue to the 'first-differencing' transformation in time series that eliminates $I(1)$ persistence? This paper takes up each of these questions.

Much of the analysis parallels the analysis of economic time series, but there is a notable difference worth highlighting at the outset. Time series analysis typically studies observations, say y_t, observed at equidistant points in time, $t = 1, 2, 3, ...$ where t indexes months, quarters, years, etc. Economic variables observed in space are not so neatly arranged. For example, geographical data may be collected at potentially arbitrary locations s_l within a given region $\mathcal{S}$ such as a U.S. state, and each state has its own unique shape. For the analysis to be useful in a wide range of spatial applications, we posit a model that assigns values to all locations that may potentially be observed. Thus, for the general problem with d spatial dimensions,

[1] Conley (1999) is a leading example of spatial HAC inference. See Müller and Watson (2022a, 2022b) for a discussion of the post-Conley literature and new suggestions for inference in regression model with weak spatial dependence.

we begin with a stochastic process $Y_n(s)$ over $s \in \mathbb{R}^d$, where $d = 2$ in the geography example. When $d = 1$, s could index time, so this is a time series model where $Y_n(s)$ is a continuous time process and where the sample data correspond to realizations of $y_l = Y_n(s_l)$ observed at potentially irregularly spaced points $s_l \in \mathbb{R}$.

We thus follow the geostatistical tradition of positing a continuous parameter model of spatial variation, rather than modelling spatial dependence by spatial autoregressive (SAR) models of Cliff and Ord (1974) and Anselin (1988).[2] There is a small literature on unit roots and spurious regression in SAR models, initiated by Fingleton (1999). SAR models require the definition of a spatial weights (or proximity) matrix, which is usually normalized so that its rows sum to unity (Ord (1975)). Under that normalization, the unit root SAR model is not well defined. Fingleton (1999) and Mur and Trívez (2003) modify the weight matrix so that one unit becomes unconnected. In contrast, Beenstock, Feldman, and Felsenstein (2012) abandon the row sum normalization, and Lauridsen and Kosfeld (2006) employ a generalized inverse to solve the model. The only authors that derive asymptotic distribution results are Lee and Yu (2009, 2013), who study the row normalized SAR model with a coefficient that converges to unity. They find that this model does not induce spurious regression effects of the type encountered in time series: OLS coefficients remain asymptotically normal, the regression R^2 converges in probability to zero, and t-statistics do not diverge.

With this background in place, the roadmap of the paper is a follows.

Section 2 defines a (continuous parameter version) of a spatial $I(1)$ process, as well as a functional central limit theorem (FCLT) for approximating the sampling distribution of statistics involving spatial $I(1)$ variables. In time series models ($d = 1$ in our notation), the canonical $I(1)$ process is a Wiener process. Lévy-Brownian motion is a useful generalization of the Wiener process for $d > 1$, and Section 2 begins by reviewing the properties of Lévy-Brownian motion. In time series models, more general $I(1)$ processes can be constructed by replacing the white noise increments of the Wiener process with a weakly correlated stationary series. For example, in a discrete-time time series model, the white noise innovations of a random walk might be replaced with stationary ARMA(p, q) random variables, yielding an ARIMA$(p, 1, q)$ process. Section 2 defines the spatial $I(1)$ process similarly by replacing the Lévy-Brownian motion white noise innovations with a weakly dependent stationary spatial

[2] See Gelfand, Diggle, Guttorp, and Fuentes (2010) and Schabenberger and Gotway (2005) for useful overviews.

process.

An important insight from time series analysis is that the distribution of functions of $I(1)$ processes can be approximated by the distributions of corresponding functions of Wiener processes. The FCLT is the core of such an approximation, and it provides the basis for large-sample inference using statistics constructed from realizations of $I(1)$ processes. Section 2 provides a FCLT that is applicable to $I(1)$ spatial processes. We also show how to appropriately generalize the $I(1)$ model to a spatial "local-to-unity" process, and provide a corresponding FCLT result about its large sample behavior.

Armed with the tools from Section 2, Section 3 studies regressions involving spatial $I(1)$ variables, specifically models where the regressors and dependent variable are independent $I(1)$ processes. The section shows that many of the key results from the spurious time series regression (cf., Phillips (1986)) carry over to the spatial case. For example, OLS regression coefficients and the regression R^2 are not consistent, but have limiting distributions that can be represented by functions of Lévy-Brownian motion. Regression F-statistics—we study HAC in addition to the classical homoskedasticity-only test statistics considered in Phillips (1986)—diverge to infinity. The bottom line is that researchers should be wary of spurious regressions using spatial data, just as they are using time series data.

Section 4 takes up the problem of conducting inference about the degree of persistence in a scalar spatial variable. In particular, we construct spatial analogues of the time series "low-frequency" unit root and stationary tests of Müller and Watson (2008). In addition, we suggest a confidence interval for the mean reversion parameter in the spatial local-to-unity model, analogous to the time series work by Stock (1991).

First-differencing an $I(1)$ time series yields an $I(0)$ process, so spurious time series regressions can be avoided by transforming $I(1)$ variables into first differences. The analogous transformation for spatial $I(1)$ processes is not obvious. Section 5 investigates a transformation that we call 'isotropic differencing' in which weighted averages of neighboring locations are subtracted from each spatial location. The section provides Monte Carlo evidence that regressions using isotropic differences do not suffer from spurious regression problems, and that valid inference can be conducted using the spatial-correlation robust methods developed in Müller and Watson (2022a, 2022b). But there are other intuitively plausible methods that may eliminate or mitigate problems associated with $I(1)$ variables in a spatial regression. These other methods include (i) low-pass and high-pass spectral regressions, (ii) regressions that incorporate small-area fixed effects, (iii) pooling estimates constructed from data in non-overlapping regions, and (iv) employing a GLS transformation based on Lévy-Brownian motion. Section 5 compares the coverage and efficiency of feasible confidence intervals from versions of these methods.

In Sections 4 and 5 we illustrate the methods using spatial data and regressions from Chetty, Hendren, Kline, and Saez (2014). Specifically, we examine the index of intergenerational mobility constructed by the authors for 741 commuting zones in the United States and data on various other socioeconomic factors they examine. Using the methods developed in Section 4, we find that many of these variables exhibit strong spatial persistence. That said, after re-estimating their regressions using the methods developed in Section 5, we find that their substantive conclusions about the correlation of socioeconomic factors and intergenerational mobility are robust.

Section 6 offers some concluding remarks. The appendix contains all proofs.

2 Spatial $I(1)$ Processes and Their Limits

This section is divided into five subsections. The first subsection defines some notation for the environment under study. The second reviews Lévy-Brownian motion, a spatial generalization of the Wiener process. The third subsection provides the definition of a spatial $I(1)$ process, and the fourth provides a corresponding functional central limit theorem. The final subsection presents a spatial generalization of the time-series local-to-unity model.

2.1 Set-up and Notation

The observations are the random variables y_l, $l = 1, \ldots, n$, as well as their associated location $s_l \in \mathbb{R}^d$. We assume that the locations are scaled such that $s_l \in \mathcal{S}$, for some compact $\mathcal{S} \subset \mathbb{R}^d$. Let G_n be the empirical distribution function of the n locations s_l. In our asymptotic analysis, we assume that G_n converges in distribution to G, $G_n(s) \to G(s)$ for all $s \in \mathcal{S}$, with G an absolutely continuous distribution with support equal to $\mathcal{S}$. As mentioned in the introduction, we model y_l as being the value of a stochastic process $Y_n(\cdot)$ on $\mathcal{S}$ evaluated at s_l, that is $y_l = Y_n(s_l)$. In much of our analysis, we treat the locations $\{s_l\}$ as non-stochastic, or equivalently, we condition on the locations and assume that $Y_n(\cdot)$ is independent of $\{s_l\}_{l=1}^n$.

2.2 Lévy-Brownian Motion

Consider the usual time series $I(1)$ process $y_t = \sum_{s=1}^{t} u_s$, $t = 1, \ldots, n$, where u_t is mean zero, covariance stationary and weakly dependent (that is, u_t is $I(0)$). A standard time series FCLT implies that $n^{-1/2} y_{\lfloor \cdot n \rfloor} \Rightarrow \omega W(\cdot)$, where W is a standard Wiener process on the unit interval $[0, 1]$. For this reason, Wiener processes play a key role in the asymptotic analysis of inference involving $I(1)$ time series. Moreover, if $n^{-1/2} y_t = \omega W(t/n)$ holds exactly, then y_t is a Gaussian random walk. Thus, Wiener processes represent the canonical $I(1)$ model, and the FCLT shows that other $I(1)$ processes behave similarly to this canonical model in a well-defined sense.

With this in mind, we begin by defining the generalization of the Wiener process to the spatial case, before discussing more general spatial $I(1)$ processes.

An attractive generalization of the Wiener process to the spatial case is *Lévy-Brownian motion* $L(s)$, $s \in \mathbb{R}^d$ (Lévy (1948)), which will play a corresponding important role in our analysis of $I(1)$ spatial variables. Lévy-Brownian motion is a zero-mean Gaussian process with domain $\mathbb{R}^d$ and covariance function

$$\mathbb{E}[L(s)L(r)] = \tfrac{1}{2}(|s| + |r| - |s - r|) \tag{1}$$

with $|a| = \sqrt{a'a}$ for $a \in \mathbb{R}^d$, so in particular, $\mathrm{Var}(L(s)) = |s|$ and $\mathrm{Var}(L(s) - L(r)) = |s - r|$. When $d = 1$ and $s, r \geq 0$, the covariance function (1) simplifies to $\mathbb{E}[L(s)L(r)] = \min(s, r)$, the covariance function of a Wiener process, so Lévy Brownian motion reduces to a Wiener process. More generally, for any d, the process obtained along a line in $\mathbb{R}^d$, $W_{a,b}(s) = L(a + bs) - L(a)$, $a, b \in \mathbb{R}^d$, $|b| = 1$, $s \in \mathbb{R}$ is a Wiener process. Thus, L is a natural embedding of the canonical time series model of strong persistence to the spatial case. Notice that Lévy-Brownian motion is *isotropic*, that is, $\mathrm{Var}(L(s) - L(r))$ depends on s, r only through $|s - r|$. Thus, Lévy-Brownian motion is invariant to rotations of the spatial axes, $L(Os) \sim L(s)$, for any $d \times d$ rotation matrix O. This attractive invariance property is not shared by a Brownian sheet, another generalization of scalar Brownian motion to higher dimensions.[3]

By Mercer's Theorem, the covariance kernel (1) evaluated at $s, r \in \mathcal{S}$ can be represented as

$$\mathbb{E}[L(s)L(r)] = \sum_{j=1}^{\infty} \nu_j \varphi_j(s) \varphi_j(r) \tag{2}$$

[3] The Brownian sheet, say $X(s)$, for $s \in \mathbb{R}_+^d$ has a covariance function $\mathbb{E}[X(s)X(r)] = \prod_{i=1}^{d} \min(s_i, r_i)$.

where (ν_j, φ_j) are eigenvalue/eigenfunction pairs with $\nu_j \geq \nu_{j+1} \geq 0$ and where $\varphi_j : \mathcal{S} \mapsto \mathbb{R}$ satisfy $\int \varphi_i(s)\varphi_j(s)dG(s) = \mathbf{1}[i = j]$. This spectral decomposition of the covariance kernel leads to a corresponding Karhunen–Loève expansion of L as the infinite sum

$$L(s) = \sum_{j=1}^{\infty} \nu_j^{1/2} \varphi_j(s)\xi_j,\ \xi_j \sim iid\mathcal{N}(0, 1) \tag{3}$$

where the right hand side converges uniformly on $\mathcal{S}$ with probability one (cf. Theorem 3.1.2 of Adler and Taylor (2007)). This result generalizes the corresponding observation in Phillips (1998) about representations of the Wiener process in terms of stochastically weighted averages of deterministic series.

A Wiener process can trivially be written as an integral over white noise, $W(s) = \int_0^s dW(u)$. A similar representation exists for Lévy-Brownian motion: from Stoll (1986) and Lindstrøm (1993)

$$L(s) = \begin{cases} \int_0^s dW(u) \text{ for } d = 1 \\ \kappa_d \int_{\mathbb{R}^d} (|s-u|^{(1-d)/2} - |u|^{(1-d)/2}) dW(u) \text{ for } d > 1 \end{cases} \tag{4}$$

where dW is white noise on $\mathbb{R}^d$, and $\kappa_d > 0$ is a scalar chosen so that $\mathrm{Var}(L(s)) = 1$ when $|s| = 1$.

2.3 Spatial $I(1)$ Processes

In the standard time series case, $I(1)$ processes are defined as partial sums over of a weakly dependent $I(0)$ process u_t, $y_t = \sum_{s=1}^{t} u_s$. Because spatial locations typically do not fall on a regular lattice, this definition does not naturally generalize. Instead, it makes sense to take advantage of the representation (4) and replace the white noise innovations $dW(u)$ by a weakly dependent random field B. We impose the following regularity condition on B.

Condition 1. *The mean-zero random field B with domain $\mathbb{R}^d$ is covariance stationary with $\mathbb{E}[B(s)B(r)] = \sigma_B(s-r)$ and $\int_{\mathbb{R}^d} \sigma_B(s)ds < \infty$, and B is such that for some $m > 2d$, $C_m > 0$ and any square integrable function $h : \mathbb{R}^d \mapsto \mathbb{R}$,*

$$\mathbb{E}\left[\left(\int_{\mathbb{R}^d} h(u)B(u)du\right)^{2m}\right] \leq C_m \left(\int_{\mathbb{R}^d} h(u)^2 du\right)^m.$$

Lemma 1.8.4 of Ivanov and Leonenko (1989) implies that Condition 1 holds for a wide range of covariance stationary mixing random fields B.

To begin the discussion of spatial $I(1)$ processes, it is useful to first consider the $d = 1$ case with $\mathcal{S} = [0, 1]$. The continuous time analogue of the usual definition $y_t = \sum_{s=1}^{t} u_s$ is given by $Y_n(s) = \int_0^{sn} B(r)dr$. In fact, with $u_t = \int_{t-1}^{t} B(r)dr$, the definitions coincide in the sense that $y_t = Y_n(t/n)$ for $t = 1, \ldots, n$. Notice that this definition of $Y_n(s)$ is useful for studying y_t generated by the usual 'outfill sampling' in which y_t is observed over longer time spans as n increases. These longer time spans for y_t correspond to the increasing domain of B that determine $Y_n(\cdot)$. In contrast, infill-sampling keeps the relevant domain of B fixed, with y_t sampling this fixed domain more finely as n increases. In this case, the relevant definition of Y_n becomes $Y_n(s) = \int_0^{s\lambda} B(r)dr$, where λ is a fixed constant.

More generally, to accommodate both infill and outfill sampling, consider the process

$$Y_n(s) = \int_0^{s\lambda_n} B(r)dr = \lambda_n \int_0^{s} B(\lambda_n r)dr \tag{5}$$

for some deterministic sequence $\lambda_n > 0$, where $\lambda_n = n$ yields the usual outfill sampling, $\lambda_n = \lambda$ yields infill sampling, and $\lambda_n \to \infty$ with $\lambda_n/n \to 0$ corresponds to a mixture of infill and outfill sampling of B.

The representation (4) leads to a straightforward generalization of $I(1)$ processes for $d > 1$

$$\begin{aligned} Y_n(s) &= \kappa_d \int_{\mathbb{R}^d} (|\lambda_n s - u|^{(1-d)/2} - |u|^{(1-d)/2}) B(u) du \qquad (6) \\ &= \kappa_d \lambda_n^{(d+1)/2} \int_{\mathbb{R}^d} (|s - u|^{(1-d)/2} - |u|^{(1-d)/2}) B(\lambda_n u) du. \end{aligned}$$

Note that $\int_{\mathbb{R}^d} |(|s-u|^{(1-d)/2} - |u|^{(1-d)/2})| du$ does not exist, so $Y_n(s)$ is not defined path-wise for every realization of B. However, $\int_{\mathbb{R}^d} (|s-u|^{(1-d)/2} - |u|^{(1-d)/2})^2 du < \infty$, so under appropriate weak dependence conditions on B, the integral that defines $Y_n(s)$ can be shown to converge in a mean square sense. In particular, we have the following result.

Lemma 1. *Under Condition 1, $Y_n(\cdot)$ exists for any integers $d, n \geq 1$ and has continuous sample paths with probability one.*

Remark 2.1. If the n locations s_l in $\mathcal{S}$ are roughly equidistant, then their distance is of order $n^{-1/d}$. Thus pure outfill sampling of B corresponds to $\lambda_n \propto n^{1/d}$, while $\lambda_n \to \infty$ with $\lambda_n/n^{1/d} \to 0$ corresponds to a mixture of infill and outfill. The scaling of $Y_n(s)$ in (6) is such that $\mathrm{Var}(Y_n(s)) \propto \lambda_n$, just like in the $d = 1$ case (5).

Remark 2.2. If B is isotropic, then $Y_n(s)$ in (6) shares the isotropic properties of Lévy-Brownian motion and is invariant to rotations of the spatial axes.

2.4 A Functional Central Limit Theorem

Recall that for discrete-time time series, a functional central limit theorem (FCLT) yields $n^{-1/2}y_{\lfloor \cdot n\rfloor} = n^{-1/2}\sum_{t=1}^{\lfloor \cdot n\rfloor} u_t \Rightarrow \omega W(\cdot)$ for a covariance stationary and weakly dependent time series u_t, where $\omega^2 = \sum_{k=-\infty}^{\infty}\mathbb{E}[u_t u_{t-k}]$ is the so-called long-run variance of u_t. We now develop a similar result for the spatial $I(1)$ process $Y_n(\cdot)$ in (5) and (6).

We assume the following central limit condition on B.

Condition 2. *For some positive sequence $\lambda_n \to \infty$, let $\mathcal{R}_n = [-\lambda_n, \lambda_n]^d \subset \mathbb{R}^d$, and let $h_n : \mathbb{R}^d \mapsto \mathbb{R}$ be any sequence of functions such that $\limsup_{n\to\infty}\sup_{u\in\mathcal{R}_n}\lambda_n^{d/2}|h_n(u)| < \infty$ and $\mathrm{Var}[\int_{\mathcal{R}_n} h_n(u)B(u)du] \to \sigma_0^2$. Then $\int_{\mathcal{R}_n} h_n(u)B(u)du \Rightarrow \mathcal{N}(0,\sigma_0^2)$.*

The central limit theorems (CLTs) in Section 1.7 of Ivanov and Leonenko (1989) provide primitive mixing and moment conditions on B that imply Condition 2.

Theorem 2. *Suppose Conditions 1 and 2 hold. If $\lambda_n \to \infty$, then $\lambda_n^{-1/2}Y_n(\cdot) \Rightarrow \omega L(\cdot)$ with $\omega^2 = \int_{\mathbb{R}^d}\sigma_B(r)dr$.*

Remark 2.3. In practice, Theorem 2 can be used to argue that the distribution of a functional $\psi(Y_n)$ is well approximated by the distribution of $\psi(\lambda_n^{1/2}\omega L)$. Note, however, that this requires ψ to be sufficiently continuous for the continuous mapping theorem to be applicable. For instance, recall that for y_t a mean-zero zero $I(1)$ time series, a FCLT implies that $n^{-1/2}y_{\lfloor \cdot n\rfloor} \Rightarrow \omega W(\cdot)$, yet $y_t - y_{t-1} = u_t$ does not in general converge to a Gaussian variable. The same holds for our generalization to spatial $I(1)$ processes Y_n: As noted above, a typical distance between two neighboring locations is of order $n^{-1/d}$. For $d > 1$, the difference in the value of $Y_n(\cdot)$ evaluated at two such neighboring points s and $s + n^{-1/d}a$, $a \in \mathbb{R}^d$ is given by

$$\begin{aligned} Y_n(s+n^{-1/d}a) - Y_n(s) &= \int_{\mathbb{R}^d}(|\lambda_n(s+n^{-1/d}a) - u|^{(1-d)/2} - |\lambda_n s - u|^{(1-d)/2})B(u)du \\ &= \int_{\mathbb{R}^d}(|\lambda_n n^{-1/d}a - u|^{(1-d)/2} - |u|^{(1-d)/2})B(\lambda_n s + u)du. \qquad (7) \end{aligned}$$

Even under pure outfill sampling, $\lambda_n n^{-1/d}$ does not diverge. The weighting of $B(\lambda_n s + u)$ in (7) thus puts most of its (square integrable) weight on small values u, and the (suitably scaled) difference $Y_n(s+n^{-1/d}a) - Y_n(s)$ *does not* become Gaussian as $n \to \infty$, just like in the time series case.

2.5 Spatial Local-to-Unity Processes

A large time series literature, initiated by Chan and Wei (1987) and Phillips (1987), concerns a generalization of the $I(1)$ model to the weakly mean reverting local-to-unity model. In this model, y_t satisfies $n^{-1/2}(y_{\lfloor \cdot n \rfloor} - y_1) \Rightarrow \omega(J_c(\cdot) - J_c(0))$, with J_c a stationary Ornstein-Uhlenbeck (OU) process with covariance kernel $\mathbb{E}[J_c(s)J_c(r)] = \exp[-c|s-r|]/(2c)$, $c > 0$. As demonstrated by Elliott (1999), $J_c(\cdot) - J_c(0)$ converges to a Wiener process as $c \to 0$. We now generalize the spatial $I(1)$ process defined above to an analogous local-to-unity spatial model.

In particular, for $d > 1$, define J_c on $\mathbb{R}^d$ as the stationary and isotropic Gaussian process with covariance function $\mathbb{E}[J_c(s)J_c(r)] = \exp[-c|s-r|]/(2c)$, $c > 0$. This is a special case of the Matérn class of covariance functions, with a spectral density $f_c(\omega)$, $\omega \in \mathbb{R}^d$ proportional to $(|\omega|^2 + c^2)^{-(d+1)/2}$. A calculation shows that just like in the univariate case, $J_c(\cdot) - J_c(0)$ converges to $L(\cdot)$ as $c \to 0$ for any integer d.

Furthermore, from equation (2.4.7) in Matérn (1986), J_c has the moving average representation

$$J_c(s) = \kappa_{c,d} \int_{\mathbb{R}^d} |s-u|^{(1-d)/4} K_{(1-d)/4}(c|s-u|) dW(u) \tag{8}$$

for a suitable choice of constant $\kappa_{c,d}$, where K_ν is the modified Bessel function of the second kind, $d \geq 1$.[4] Thus, define the spatial local-to-unity process via

$$Y_n(s) = \kappa_{c,d} \lambda_n^{(d+1)/2} \int_{\mathbb{R}^d} |s-u|^{(1-d)/4} K_{(1-d)/4}(c|s-u|) B(\lambda_n u) du. \tag{9}$$

Proceeding as for the $I(1)$ model, the appendix shows that $Y_n(\cdot)$ in (9) exists under Condition 1 and under the conditions of Theorem 2, $\lambda_n^{-1/2} Y_n(\cdot) \Rightarrow \omega J_c(\cdot)$.

Remark 2.4. These results allow for an extension of the central-limit results in Lahiri (2003) for weighted averages of weakly dependent covariance stationary spatial processes to weighted averages of spatial $I(1)$ and local-to-unity processes.

[4] For $d = 1$, the usual one-sided (causal) representation for a stationary OU process is $J_c(s) = \int_{-\infty}^{s} e^{-c(s-u)} dW(u)$. Equation (8) is an alternative two-sided (non-causal) representation when $d = 1$.

3 Spurious Regressions with Spatial $I(1)$ Variables

As a first application of the results in Section 2, consider the regression model

$$y_l = \alpha + \mathbf{x}_l'\boldsymbol{\beta} + u_l \tag{10}$$

for $l = 1, ..., n$, where $(y_l, \mathbf{x}_l) = (Y_n(s_l), \mathbf{X}_n(s_l)) \in \mathbb{R}^{k+1}$ follow $k+1$ independent spatial $I(1)$ processes. The FCLT in Theorem 2 allows for a straightforward spatial extension of the classic spurious time-series regression results in Phillips (1986).

Let $\tilde{y}_l = y_l - n^{-1}\sum_{\ell=1}^n y_\ell$ denote the demeaned value of y_l and similarly for $\tilde{\mathbf{x}}_l$. Let $s_{\tilde{y}\tilde{y}} = n^{-1}\sum_{l=1}^n \tilde{y}_l^2$, $\mathbf{S}_{\tilde{x}\tilde{x}} = n^{-1}\sum_{l=1}^n \tilde{\mathbf{x}}_l\tilde{\mathbf{x}}_l'$ and $\mathbf{S}_{\tilde{x}\tilde{y}} = n^{-1}\sum_{l=1}^n \tilde{\mathbf{x}}_l\tilde{y}_l$. The OLS estimator is $\hat{\boldsymbol{\beta}} = \mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\mathbf{S}_{\tilde{x}\tilde{y}}$, the regression $R^2 = \mathbf{S}_{\tilde{x}\tilde{y}}'\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\mathbf{S}_{\tilde{x}\tilde{y}}/s_{\tilde{y}\tilde{y}}$, the OLS estimator for the variance of u_l is $s_u^2 = \frac{n}{n-k-1}(s_{\tilde{y}\tilde{y}} - \mathbf{S}_{\tilde{x}\tilde{y}}'\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\mathbf{S}_{\tilde{x}\tilde{y}})$, and the classical (non-spatial-correlation robust, homoskedastic) F-statistic for testing $H_0 : \mathbf{H}\boldsymbol{\beta} = 0$, where $\mathbf{H}$ is a non-stochastic matrix with $\operatorname{rank}(\mathbf{H}) = m \leq k$, is $F^{\text{Hom}} = \frac{n-k-1}{m}\hat{\boldsymbol{\beta}}'\mathbf{H}'(\mathbf{H}'\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\mathbf{H})^{-1}\mathbf{H}\hat{\boldsymbol{\beta}}/s_u^2$.

Suppose $(y_l, \mathbf{x}_l) = (Y_n(s_l), \mathbf{X}_n(s_l))$ follow spatial $I(1)$ processes with

$$\begin{bmatrix} \lambda_n^{-1/2} Y_n(\cdot) \\ \lambda_n^{-1/2}\mathbf{X}_n(\cdot) \end{bmatrix} \Rightarrow \begin{bmatrix} Y(\cdot) \\ \mathbf{X}(\cdot) \end{bmatrix} \tag{11}$$

where $[Y(\cdot), \mathbf{X}(\cdot)]$ are $k+1$ independent and arbitrarily scaled Lévy-Brownian motions. Let $\tilde{Y}(\cdot) = Y(\cdot) - \int Y(u)dG(u)$ denote the demeaned version of Y using spatial-weighted average demeaning, and define $\tilde{\mathbf{X}}$ analogously.

Theorem 3. *If $G_n \Rightarrow G$ and (11) hold, then*

(i) $\lambda_n^{-1}s_{\tilde{y}\tilde{y}} \Rightarrow \Xi_{\tilde{y}\tilde{y}} = \int_{\mathcal{S}} \tilde{Y}^2(u)dG(u)$, $\lambda_n^{-1}\mathbf{S}_{\tilde{x}\tilde{x}} \Rightarrow \boldsymbol{\Xi}_{\tilde{x}\tilde{x}} = \int_{\mathcal{S}} \tilde{\mathbf{X}}(u)\tilde{\mathbf{X}}(u)'dG(u)$ *and* $\lambda_n^{-1}\mathbf{S}_{\tilde{x}\tilde{y}} \Rightarrow \boldsymbol{\Xi}_{\tilde{x}\tilde{y}} = \int_{\mathcal{S}} \tilde{\mathbf{X}}(u)\tilde{Y}(u)dG(u)$,

(ii) $\hat{\boldsymbol{\beta}} \Rightarrow \boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}}$,

(iii) $R^2 \Rightarrow \boldsymbol{\Xi}_{\tilde{x}\tilde{y}}'\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}}/\Xi_{\tilde{y}\tilde{y}}$,

(iv) $\lambda_n^{-1}s_u^2 \Rightarrow \Xi_{\tilde{y}\tilde{y}} - \boldsymbol{\Xi}_{\tilde{x}\tilde{y}}'\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}}$,

(v) $n^{-1}F^{Hom} \Rightarrow m^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}}'\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\mathbf{H}'(\mathbf{H}\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\mathbf{H}')^{-1}\mathbf{H}\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}}/(\Xi_{\tilde{y}\tilde{y}} - \boldsymbol{\Xi}_{\tilde{x}\tilde{y}}'\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}})$.

Remark 3.1. In the one-dimensional case with $d = 1$, $\mathcal{S}$ the unit interval and G the uniform distribution, these results coincide with the spurious time-series regression limits derived in Phillips (1986). In the general spatial case, the limits are seen to depend on the spatial distribution of locations G and its support $\mathcal{S}$. Section 5 provides numerical results for the behavior of R^2 for a large set of spatial designs with $d = 2$.

Remark 3.2. An implication of part (v) of Theorem 3 is that the classical F-test statistic diverges to infinity so that $\mathbb{P}(F^{\mathrm{Hom}} > \mathrm{cv}) \to 1$ for any $\mathrm{cv} \geq 0$.

A more relevant question in practice is whether the spurious significance of the F-statistic also generalizes to heteroskedasticity and HAC-corrected standard errors. We now establish that it does. In particular, consider the class of correlation-robust-HAC F-statistics

$$F^{\mathrm{HAC}} = \frac{n}{m}\hat{\boldsymbol{\beta}}'\mathbf{H}'(\mathbf{H}\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\hat{\boldsymbol{\Omega}}_n\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\mathbf{H}')^{-1}\mathbf{H}\hat{\boldsymbol{\beta}} \tag{12}$$

where $\hat{\boldsymbol{\Omega}}_n$ is a kernel-based estimator of $\mathrm{Var}\left(n^{-1/2}\sum \tilde{x}_l u_l\right)$ of the form

$$\hat{\boldsymbol{\Omega}}_n = n^{-1}\sum_{l,\ell=1}^{n} \kappa\left(b_n(s_l - s_\ell)\right)\mathbf{e}_l\mathbf{e}_\ell' \tag{13}$$

with $\mathbf{e}_l = \tilde{\mathbf{x}}_l(\tilde{y}_l - \tilde{\mathbf{x}}_l'\hat{\boldsymbol{\beta}})$, b_n a bandwidth (that may depend both on $\{s_l\}$ and the data $\{(y_l, \mathbf{x}_l)\}$) with $b_n^{-1} = o_p(1)$ and $\kappa : \mathbb{R}^d \mapsto \mathbb{R}$ a kernel weighting function satisfying

$$\sup_r |\kappa(r)| = \bar{\kappa} < \infty, \qquad \lim_{\lambda\to\infty}\sup_{|a|=1}|\kappa(\lambda a)| = 0. \tag{14}$$

The assumption of $b_n^{-1} = o_p(1)$ ensures that in large samples, $\hat{\boldsymbol{\Omega}}_n$ in (13) puts negligible weight on pairs of locations with $|s_l - s_\ell| > \varepsilon$, for all positive ε. Since $s_l \in \mathcal{S}$ with $\mathcal{S}$ compact, this is necessary for a kernel estimator to be consistent under weak spatial dependence. These conditions are satisfied, for instance, for the spatial correlation robust estimator suggested in Conley (1999). Heteroskedasticity robust standard errors correspond to $\kappa(r) = \mathbf{1}[r = 0]$, which also satisfies (14).

The following result shows that inference using any such consistent spatial HAC estimator does not avoid spurious significance of the F^{HAC} test statistic.

Theorem 4. *If $G_n \Rightarrow G$ and (11) and (14) hold, then $\mathbb{P}(F^{HAC} > \mathrm{cv}) \to 1$ for any $\mathrm{cv} \geq 0$.*

Remark 3.3. Theorems 3 and 4 also hold for local-to-unity processes, that is, if $[Y(\cdot), \mathbf{X}(\cdot)]$ in (11) are $k+1$ independent processes of the type (8), with arbitrary and potentially different mean-reversion parameters c. This is because the asymptotics that yield convergence to $J_c(\cdot)$ are "pure infill" relative to the degree of mean reversion, and pure infill asymptotics are known to potentially lead to inconsistent parameter estimators (cf. Zhang and Zimmerman

(2005) for references and further discussion).[5] In contrast, no degree of "outfill" (or increasing domain) asymptotics can remedy the spurious regression effect in the $I(1)$ model, again just as in the time series case.

Remark 3.4. It follows from the Karhunen–Loève representation of L in (3) and the FCLT result in Theorem 2 that the coefficients of a regressions of $\lambda_n^{-1/2} y_l$ on the eigenfunctions $[\varphi_1(s_l), \ldots, \varphi_k(s_l)]$ converge to independent $\mathcal{N}(0, \omega^2 \nu_j)$ random variables. This generalizes the "understanding spurious regressions" result in Theorem 3.1 (a) of Phillips (1998) to the spatial case. More generally, the coefficients of a regression of $\lambda_n^{-1/2} y_l$ on smooth deterministic functions of s_l, say $\boldsymbol{\psi}(s_l) \in \mathbb{R}^k$, converge to $\left(\int \boldsymbol{\psi}(s)\boldsymbol{\psi}(s)' dG(s)\right)^{-1} \omega \int \boldsymbol{\psi}(s) L(s) dG(s)$ and are asymptotically significant as measured by a corresponding F^{Hom} or F^{HAC} statistic. Kelly (2019) observes such a phenomenon empirically in a number of applications with spatial data.

4 Inference for Spatial Persistence

The autoregressive representation for a discrete-time $I(1)$ time series process has a unit root in its autoregressive polynomial, making it straightforward to test for $I(1)$ persistence. Spatial $I(1)$ processes as defined in Section 2 do not have an analogous autoregressive representation, so Dickey-Fuller or related unit-roots tests do not generalize.[6] Similarly, popular time series tests of the $I(0)$ null hypothesis, such as Kwiatkowski, Phillips, Schmidt, and Shin (1992), do not directly generalize to spatial processes. An alternative approach pursued in Müller and Watson (2008) is to discriminate $I(1)$ from $I(0)$ time series by studying the properties of q suitably chosen weighted averages, and this approach generalizes fairly directly to spatial settings.

[5] At the same time, fixed-b type spatial HAR inference (Bester, Conley, Hansen, and Vogelsang (2016), Sun and Kim (2012)) does not lead to diverging F-statistics, and the spatial correlation robust inference derived in Müller and Watson (2022a) explicitly accommodates some degree of "strong" persistence of the type exhibited by the spatial local-to-unity model for large enough c. See Section 5 below for corresponding numerical results.

[6] The fact that the spatial processes studied here do not have an autoregressive representation suggests that this paper's title "Spatial Unit Roots" and the "local-to-unity" label for the weakly mean reverting process (9) are misnomers. We retain these terms because they correctly convey a close analogy between the large sample behavior of persistent time series and of the persistent spatial processes considered here.

The intuition underlying this approach is straightforward. The Karhunen–Loève expansion (3) implies that eigenfunction weighted averages of a Lévy-Brownian motion recover independent normal variates with a variance that is proportional to the eigenvalues. Focussing on the q eigenfunctions corresponding to the largest eigenvalues yields a set of independent normal random variables with sharply decaying variance. In contrast, when the data are i.i.d. Gaussian random variables, these weighted averages are also i.i.d. by the orthogonality of the eigenfunctions. This difference in behavior makes it possible to empirically distinguish between these two canonical cases. What is more, the FCLT result in Theorem 2 and the CLT in Lahiri (2003) may be used to generalize these tests to general forms of spatial $I(0)$ and $I(1)$ processes, as well as to the local-to-unity model (9). The remainder of this section expands on this intuition to formally develop tests for the degree of spatial persistence.[7]

4.1 Dimension Reduction by Weighted Averages

Let $\mathbf{Y}_n = (y_1, \ldots, y_n)'$ and let $\boldsymbol{\Sigma}_{n,L}$ be the $n \times n$ covariance matrix induced by Lévy-Brownian motion $y_l = L(s_l)$, conditional on $\{s_l\}_{l=1}^n$. We are interested in tests that are invariant to translation shifts $\mathbf{Y}_n \to \mathbf{Y}_n + a\mathbf{1}$, where $\mathbf{1}$ is a vector of ones. We therefore seek weighted averages of $\mathbf{Y}_n$ that sum to zero. Let $\mathbf{M} = \mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$, and let $\mathbf{R}_n$ be the $n \times q$ matrix of eigenvectors of $\mathbf{M}\boldsymbol{\Sigma}_{n,L}\mathbf{M}$ corresponding to the q largest eigenvalues, where $\mathbf{R}_n$ satisfies $n^{-1}\mathbf{R}_n'\mathbf{R}_n = \mathbf{I}_q$. If $\mathbf{Y}_n \sim (\mathbf{0}, \boldsymbol{\Sigma}_{n,L})$, the columns of $\mathbf{R}_n$ extract the q linear combinations of $\mathbf{Y}_n$ with the largest variance. Let $\mathbf{Z}_n = \mathbf{R}_n'\mathbf{M}\mathbf{Y}_n = \mathbf{R}_n'\mathbf{Y}_n$, a $q \times 1$ random vector, denote the associated weighted averages of the data, where the final equality holds because $\mathbf{R}_n'\mathbf{1} = \mathbf{0}$. As in Müller and Watson (2008), we treat $\mathbf{Z}_n$ as the effective observation, that is, we seek to conduct inference about the persistence properties of $\mathbf{Y}_n$ with a test that is a function of $\mathbf{Z}_n$ only.

Different models for persistence in $\mathbf{Y}_n$ imply different values for $\mathrm{Var}(\mathbf{Z}_n) = \boldsymbol{\Omega}_n$. Consider first the generic problem of testing $H_0 : \boldsymbol{\Omega} = \boldsymbol{\Omega}_0$ versus $H_a : \boldsymbol{\Omega} = \boldsymbol{\Omega}_a$ when $\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. A standard calculation shows that the most powerful scale invariant test rejects for large values

[7] Previous approaches to testing for the presence of spatial correlation, such as Moran's (1950) I statistic or Geary's (1954) c, require the specification of a spatial weight matrix and test the null hypothesis of zero spatial correlation.

of

$$\frac{\mathbf{Z}_n'\mathbf{\Omega}_0^{-1}\mathbf{Z}_n}{\mathbf{Z}_n'\mathbf{\Omega}_a^{-1}\mathbf{Z}_n} \tag{15}$$

with a critical value that is computed under the null distribution $\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_0)$.

Inference of this type depends on q, the number of weighted averages used in the construction of $\mathbf{Z}_n$. The choice of q faces a classic efficiency vs. robustness trade-off: large q increases power, but at the expense of taking the (asymptotic) implications of models of persistence seriously over many weighted averages. In practice, a moderate value of q, say a number around 10-20, as in Müller and Watson (2008), yields a reasonable compromise: it is large enough to yield informative inference and yet does not overly stretch the asymptotic approximations of, say, the FCLT in Theorem 2. We leave a more principled argument that endogenously determines q (potentially along the lines of Dou (2019) and Müller and Watson (2022a)) to future research, and set $q = 15$ in our numerical analysis.

When considering large sample approximations based on the FCLT, it is useful to have a result about the large sample properties of the eigenvectors $\mathbf{R}_n$. Intuitively, these eigenvectors should become close in some sense to the eigenfunctions of the covariance kernel of demeaned Lévy-Brownian motion (1). Thus, define

$$\bar{k}(r,s) = k(r,s) - \int k(u,s)dG(u) - \int k(r,u)dG(u) + \int\int k(u,t)dG(u)dG(t)$$

where $k(r,s) = \frac{1}{2}(|s| + |r| - |s-r|)$. Further, write the spectral decomposition as $\bar{k}(s,r) = \sum_{i=1}^{\infty} \bar{\nu}_i \bar{\varphi}_i(s)\bar{\varphi}_i(r)$, where $\int \bar{\varphi}_i(s)\bar{\varphi}_j(s)dG(s) = \mathbf{1}[i=j]$, $\bar{\nu}_i \geq \bar{\nu}_{i+1} \geq 0$ and the eigenfunctions $\bar{\varphi}_i$ satisfy $\int \bar{k}(\cdot, s)\bar{\varphi}_i(s)dG(s) = \bar{\nu}_i\bar{\varphi}_i(\cdot)$. The sample analogue of $\bar{k}(r,s)$ is

$$\hat{k}_n(r,s) = k(r,s) - n^{-1}\sum_{l=1}^{n} k(s_l, s) - n^{-1}\sum_{\ell=1}^{n} k(r, s_\ell) + n^{-2}\sum_{l=1}^{n}\sum_{\ell=1}^{n} k(s_l, s_\ell)$$

and the $n \times n$ matrix $\hat{\mathbf{K}}_n$ with l, ℓ element equal to $\hat{k}_n(r,s)$ satisfies $\hat{\mathbf{K}}_n = \mathbf{M}\mathbf{\Sigma}_{n,L}\mathbf{M}$. Let $(\mathbf{r}_i, \hat{\nu}_i)$ with $\mathbf{r}_i = (r_{i,1}, \ldots, r_{i,n})'$ be the eigenvector-eigenvalue pairs of $n^{-1}\hat{\mathbf{K}}_n$ with $\hat{\nu}_1 \geq \hat{\nu}_2 \geq \ldots \geq \hat{\nu}_n$ and $n^{-1}\mathbf{r}_i'\mathbf{r}_i = 1$. For all i with $\hat{\nu}_i > 0$ define the $\mathcal{S} \mapsto \mathbb{R}$ functions

$$\hat{\varphi}_i(\cdot) = n^{-1}\hat{\nu}_i^{-1}\sum_{l=1}^{n} r_{i,l}\hat{k}_n(\cdot, s_l). \tag{16}$$

Lemma 6 of Müller and Watson (2022a), building of the work of Rosasco, Belkin, and Vito (2010), shows that under the assumption that s_l is i.i.d. with distribution G and $\bar{\nu}_1 > \bar{\nu}_2 >$

$\ldots > \bar{\nu}_q$, $(\hat{\nu}_i, \hat{\varphi}_i)$ converge to $(\bar{\nu}_i, \bar{\varphi}_i)$, $i = 1, \ldots, q$, and the lemma also provides corresponding convergence rates. The following result does away with the i.i.d. assumption on the generation of the locations s_l, but rather assumes that the nonstochastic sequence of locations $\{s_l\}_{l=1}^n$ is such that $G_n \Rightarrow G$. This condition holds for almost all realizations of $\{s_l\}_{i=1}^n$ if $s_l \sim G$ is i.i.d. by the Glivenko-Cantelli Theorem, so in this sense, the following result is stronger, albeit at the cost of not providing convergence rates.

Lemma 5. *Suppose* $\bar{\nu}_1 > \bar{\nu}_2 > \ldots > \bar{\nu}_q$ *and* $G_n \Rightarrow G$. *Then for any* $q \geq 1$, $\sup_{s \in \mathcal{S}, 1 \leq i \leq q} |\hat{\varphi}_i(s) - \bar{\varphi}(s)| \to 0$ *and* $\max_{1 \leq i \leq q} |\hat{\nu}_i - \bar{\nu}_i| \to 0$.

4.2 Tests of the $I(1)$ Null Hypothesis

With this background, consider the problem of testing the $I(1)$ null hypothesis against the local-to-unity alternative. The canonical form of these models are $y_l = L(s_l)$ and $y_l = J_c(s_l)$. This yields $\mathbf{Y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{n,L})$ and $\mathbf{Y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_n(c))$, respectively, with the l, ℓ element of $\mathbf{\Sigma}_n(c)$ equal to $\exp[-c|s_l - s_\ell|]/(2c)$. Thus, optimal tests in this problem are of the form (15) with $\mathbf{\Omega}_0 = \mathbf{\Omega}_{n,L} = \mathbf{R}_n' \mathbf{\Sigma}_{n,L} \mathbf{R}_n$ and $\mathbf{\Omega}_a = \mathbf{\Omega}_n(c_a) = \mathbf{R}_n' \mathbf{\Sigma}_n(c_a) \mathbf{R}_n$ for some $c_a > 0$. This yields the test statistic

$$\mathrm{LFUR}_n = \frac{\mathbf{Z}_n' \mathbf{\Omega}_{n,L}^{-1} \mathbf{Z}_n}{\mathbf{Z}_n' \mathbf{\Omega}_n^{-1}(c_a) \mathbf{Z}_n}, \tag{17}$$

where the notation emphasizes that this is the spatial analogue of the time series low-frequency unit root test (LFUR) from Müller and Watson (2008). In the Gaussian AR(1) time series model with parameter ρ, the null is $\rho = 1$ and the alternative is $\rho = 1 - c_a/n$. To determine a value of c_a that ensures good power for a wide range of values of c, we follow King (1987) and choose c_a such that a 5% level test has 50% power.

By construction, this test is valid under the canonical H_0 model $\mathbf{Y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{L,n})$. But by the FCLT in Theorem 2, Lemma 5 and the CMT, $\lambda_n^{-1/2} n^{-1} \mathbf{Z}_n \Rightarrow \mathcal{N}(0, \omega^2 \operatorname{diag}(\bar{\nu}_1, \ldots, \bar{\nu}_q))$ for the entire class of $I(1)$ processes in (5) and (6) (including the canonical model). Since the test (15) is scale invariant, the scale parameters $\lambda_n^{-1/2} n^{-1}$ and ω^2 do not matter. Thus, the critical value computed from the canonical model converges to the asymptotically correct critical value for generic $I(1)$ processes, and the test is asymptotically valid.

4.3 Tests of the $I(0)$ Null Hypothesis

Now consider a corresponding spatial stationarity test based on $\mathbf{Z}_n$. Here we seek a test of the null hypothesis that y_l exhibits weak spatial correlation. To operationalize this, one must take a stand on what constitutes "weak" correlation. One useful gauge for the strength of correlation is whether HAR-inference remains valid. Müller and Watson (2022a) derive HAR inference that remains valid in the $\mathbf{\Sigma}_n(c)$ model for (all large enough) values of c that induce an average pairwise correlation

$$\bar{\rho}(c) = \frac{1}{n(n-1)} \sum_{l \neq \ell} \exp[-c|s_l - s_\ell|] \tag{18}$$

of no more than 0.03. Denote the corresponding cut-off value of c by $c_{0.03}$. The canonical version of the testing problem then becomes $H_0 : \mathbf{\Omega} = \mathbf{\Omega}_n(c)$, $c \geq c_{0.03}$ against $H_a : \mathbf{\Omega} = \mathbf{\Omega}_n(c) + g_a^2 \mathbf{\Omega}_{n,L}$, $c \geq c_{0.03}$, $g_a \in \mathbb{R}$. This form of alternative, a sum of a stationary and $I(1)$ process, also motivates the time series stationary tests in Nyblom (1989), Kwiatkowski, Phillips, Schmidt, and Shin (1992), etc. The larger the scale g_a of the Lévy-Brownian motion under the alternative, the easier it is to discriminate the two hypotheses, so g_a can again be chosen using the 50% power rule. The stationarity testing problem is complicated by the presence of the additional nuisance parameter c that indexes the covariance matrix $\mathbf{\Omega}_n(c)$ in both the null and alternative. Here numerical experimentation revealed that in many configurations of locations, picking $c = c_{0.001}$ under both H_0 and H_a works well in the sense of generating a test statistic (15) that has a 95% quantile that is fairly constant as a function of $c \geq c_{0.03}$. Thus, the stationary test rejects if

$$\text{LFST}_n = \frac{\mathbf{Z}_n' \mathbf{\Omega}_n(c_{0.001})^{-1} \mathbf{Z}_n}{\mathbf{Z}_n' [\mathbf{\Omega}_n(c_{0.001}) + g_a^2 \mathbf{\Omega}_{n,L}]^{-1} \mathbf{Z}_n} \tag{19}$$

exceeds the critical value $\text{cv}_n^{\text{LFST}}$, where the critical value is chosen to insure the correct size of the test for all values of $c \geq c_{0.03}$. More precisely, $\text{cv}_n^{\text{LFST}}$ solves $\sup_{c \geq c_{0.03}} \mathbb{P}(\text{LFST}_n \geq \text{cv}_n^{\text{LFST}}) = \alpha$, where α is the size of the test and the probability is computed under under $\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_n(c))$, $c \geq c_{0.03}$. We label the statistic "LFST" because it is the spatial generalization of the low-frequency stationarity test proposed in Müller and Watson (2008).

By the same arguments applied to the LFUR_n test, this stationarity test remains valid in large samples under the general local-to-unity model (9) for $c \geq c_{0.03}$. A more subtle question is whether it also remains valid under generic weak dependence, defined as $y_l = B(\lambda_n s_l)$, with

$\lambda_n \to \infty$ and B a weakly dependent random field as in Section 2. The CLT in Lahiri (2003) shows that under such generic weak dependence (and under the assumption that $s_l \sim G$ is i.i.d.), a suitably scaled version of $\mathbf{Z}_n$ becomes Gaussian, but not necessarily with covariance matrix proportional to $\mathbf{I}_q$. In the spatial case, the effect of weak dependence on the covariance of smoothly weighted averages is generically more subtle than a multiplication by the scalar long-run variance. The LFST_n test still remains valid, since for every n, its critical value is chosen to be valid for all $c \geq c_{0.03}$, so it is also valid under all sequences of $c_n \to \infty$, including those that induce the the different possible limits identified by Lahiri's (2003) CLT.

Theorem 6. *If $y_l = B(\lambda_n s_l)$ and $\lambda_n \to \infty$ with $\lambda_n^d/n \to a \in [0, \infty)$, then under the assumptions of Lahiri's CLT in Theorem 3.2, $\limsup_{n\to\infty} \mathbb{P}(\mathrm{LFST}_n \geq \mathrm{cv}_n^{\mathrm{LFST}}) \leq \alpha$.*

4.4 Confidence Sets for $\bar{\rho}(c)$

A closely related problem is the construction of a confidence set for c, the parameter in the spatial local-to-unity model. As usual, one may obtain a $100(1-\alpha)\%$ confidence set by collecting the values of c_0 for which a family of α-level tests of the form $H_0 : c = c_0$ does not reject. What is more, if this family of tests is optimal against the alternative that c is drawn from some probability distribution Π, then the classic result in Pratt (1961) implies that the resulting confidence interval has the smallest Π-weighted expected length.

The degree of mean reversion implied by a given value of c depends on the scaling of the locations s_l, so it is easier to interpret the scale-invariant average correlation $\bar{\rho}(c)$. With Π such that the implied weighting of $\bar{\rho}$ is uniform on $[0, 1]$, the average length minimizing scale-invariant confidence interval collects the values of $\bar{\rho}_0$ for which the test based on

$$\frac{\int_0^1 \det(\mathbf{\Omega}_n(c_{\bar{r}}))^{-1/2}(\mathbf{Z}_n'\mathbf{\Omega}_n(c_{\bar{r}})^{-1}\mathbf{Z}_n)^{-q/2} d\bar{r}}{(\mathbf{Z}_n'\mathbf{\Omega}_n(c_{\bar{\rho}_0})^{-1}\mathbf{Z}_n)^{-q/2}} \tag{20}$$

does not exceed the $100(1-\alpha)$ percentile of (20) under $\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_n(c_{\bar{\rho}_0}))$. The large sample validity of this confidence set for $\bar{\rho} > 0$ in the general local-to-unity model (9) follows from the same arguments as the large sample validity of the LFUR_n test discussed above.

4.5 Spatial Correlation in the Chetty et al. (2014) Data

Chetty, Hendren, Kline, and Saez (2014) use administrative records on the incomes of more than 40 million children and parents to study intergenerational income mobility in the United

Table 1: Spatial Persistence of Variables in Chetty et al. (2014)

Variable	Spatial Persistence Statistics			Regression of the AMI onto Variable $\hat{\beta}$ [95% CI]	
	***p*-Value for Test**		**95% CI for $\bar{\rho}$**		
	I(1) Null	**I(0) Null**		**Level [Cluster]**	**LBM-GLS [CSCPC]**
Absolute Mobility Index	0.08	<0.01	[0.14; 1.00]		
Frac. Black Residents	0.02	0.01	[0.02; 0.71]	-0.58 [-0.71; -0.45]	-0.42 [-0.50; -0.34]
Racial Segregation	0.07	0.02	[0.05; 1.00]	-0.36 [-0.45; -0.27]	-0.24 [-0.28; -0.19]
Segregation of Poverty	0.13	0.04	[0.05; 1.00]	-0.41 [-0.54; -0.28]	-0.21 [-0.25; -0.16]
Frac. < 15 Mins to Work	0.69	<0.01	[0.46; 1.00]	0.61 [0.36; 0.85]	0.37 [0.26; 0.48]
Mean Household Income	0.02	0.18	[0.01; 0.61]	0.05 [-0.09; 0.19]	-0.02 [-0.08; 0.04]
Gini	0.56	<0.01	[0.40; 1.00]	-0.58 [-0.76; -0.40]	-0.21 [-0.29; -0.14]
Top 1 Perc. Inc. Share	0.60	0.03	[0.43; 1.00]	-0.19 [-0.33; -0.05]	-0.06 [-0.11; -0.01]
Student-Teacher Ratio	0.03	0.16	[0.04; 0.87]	-0.33 [-0.52; -0.13]	-0.18 [-0.26; -0.09]
Test Scores (Inc. adjusted)	0.40	0.07	[0.27; 1.00]	0.59 [0.42; 0.76]	0.42 [0.34; 0.51]
High School Dropout	0.63	0.02	[0.40; 1.00]	-0.57 [-0.75; -0.40]	-0.31 [-0.42; -0.20]
Social Capital Index	0.73	<0.01	[0.38; 1.00]	0.64 [0.46; 0.82]	0.28 [0.12; 0.44]
Frac. Religious	0.11	0.03	[0.15; 1.00]	0.52 [0.35; 0.69]	0.32 [0.19; 0.45]
Violent Crime Rate	0.52	0.04	[0.38; 1.00]	-0.38 [-0.67; -0.09]	-0.14 [-0.23; -0.06]
Frac. Single Mothers	0.03	<0.01	[0.05; 0.88]	-0.76 [-0.91; -0.62]	-0.60 [-0.69; -0.51]
Divorce Rate	<0.01	0.21	[0.02; 0.53]	-0.49 [-0.68; -0.29]	-0.38 [-0.49; -0.27]
Frac. Married	0.09	0.07	[0.12; 1.00]	0.57 [0.45; 0.69]	0.36 [0.29; 0.43]
Local Tax Rate	0.01	0.25	[0.01; 0.59]	0.32 [0.19; 0.46]	0.07 [0.01; 0.14]
Colleges per Capita	0.57	0.10	[0.00; 1.00]	0.20 [-0.02; 0.42]	0.02 [-0.08; 0.11]
College Tuition	0.21	<0.01	[0.15; 1.00]	-0.02 [-0.15; 0.11]	0.01 [-0.02; 0.04]
Coll. Grad. Rate (Inc. Adjusted)	0.46	0.01	[0.34; 1.00]	0.15 [0.03; 0.28]	0.08 [0.01; 0.15]
Manufacturing Share	0.04	<0.01	[0.10; 1.00]	-0.26 [-0.44; -0.08]	0.06 [-0.03; 0.16]
Chinese Import Growth	0.02	0.07	[0.01; 0.58]	-0.17 [-0.33; -0.02]	0.03 [0.01; 0.04]
Teenage LFP Rate	0.28	<0.01	[0.20; 1.00]	0.63 [0.46; 0.80]	0.25 [0.14; 0.36]
Migration Inflow	0.06	0.11	[0.00; 1.00]	-0.26 [-0.40; -0.11]	-0.13 [-0.18; -0.08]
Migration Outlflow	0.05	0.02	[0.07; 1.00]	-0.16 [-0.30; -0.03]	-0.09 [-0.15; -0.03]
Frac. Foreign Born	0.44	0.02	[0.35; 1.00]	-0.03 [-0.15; 0.10]	-0.12 [-0.24; -0.00]

Notes: The first two columns show p-values for tests of the $I(1)$ and $I(0)$ null hypotheses using the statistics (17) and (19). The third column shows the 95% confidence for $\bar{\rho}$ constructed by inverting the tests in (20). The final two columns show estimated regression coefficients ($\hat{\beta}$) and nominal 95% confidence intervals from regression of the Absolute Mobility Index (AMI) onto each of the variables in the table, where the first column of results uses the levels of the variables with nominal 95% confidence intervals constructed using standard errors clustered by state, and the second column shows the LBM-GLS estimates and CSPC 95% confidence intervals. The variables are standardized (in levels) to have mean zero and unit standard deviation.

States. They construct an index of mobility for each of the 741 commuter zones in the United States and investigate the relationship between mobility and other factors by regressing their mobility index on variables such as racial segregation, school quality and so forth. They find large and statistically significant correlations between their absolute mobility index (AMI) and many socioeconomic indicators. One might suspect that the variables used in their regressions are spatially correlated, and this raises the question of the robustness of their results to sampling error associated with this spatial persistence. This question is answered in Table 1.[8]

The first three columns in the table apply the tests outlined in this section to gauge the spatial correlation in the variables used by Chetty, Hendren, Kline, and Saez (2014). The results indicate that there is substantial spatial correlation across the United States in the socioeconomic variables. The $I(0)$ null is rejected for most series, the $I(1)$ null is not rejected for several, and the confidence intervals for the implied average correlation $\bar{\rho}$, while wide, suggest a high degree of spatial persistence. The final two columns of the table investigate the robustness of the Chetty, Hendren, Kline, and Saez (2014) conclusions to this spatial correlation. We discuss these columns after introducing additional analysis in the next section.

5 Regressions with Transformed Spatial Variables

To avoid spurious regression effects using $I(1)$ time series data, researchers routinely estimate regressions using first differences of the original variables and rely on HAC/HAR inference methods to account for any remaining $I(0)$ autocorrelation. The analogous approach for regressions involving spatial $I(1)$ variables is not obvious: how exactly does one 'first-difference' a spatial series or otherwise transform the series or regression to eliminate the deleterious effects of $I(1)$ persistence? This section explores a number of possibilities. Using simulated data, we assess whether regressions with such transformed data and spatial HAR corrections yields (approximately) valid inference.

[8]The variables are chosen from Figure VIII in Chetty, Hendren, Kline, and Saez (2014). The data are taken from their comprehensive replication materials.

5.1 Simulation Design

We are interested in inference about β_1 in the linear regression (10), maintaining throughout that $\mathbf{Y}_n$ is independent of $\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$. The simulated data sets have $n = 400$ observations and differ both in their distribution of locations $\{s_l\}$ and the distribution of $(\mathbf{Y}_n, \mathbf{X}_n)$. Spatial locations are drawn from the 48-U.S. States design used in Müller and Watson (2022a, 2022b). Specifically, for each of the 48 contiguous U.S. states, we draw 2 sets of 400 locations at random uniformly within the boundaries of the state. Conditional on each of these 96 location set draws, we consider seven distributions for $(\mathbf{Y}_n, \mathbf{X}_n)$, for $k = 1$ and $k = 5$. In each of those, the $k+1$ columns of $(\mathbf{Y}_n, \mathbf{X}_n)$ are independent and identically distributed. The seven distributions for y_l (which are also used to generate each element of $\mathbf{x}_l$) are:

- DGP1: $y_l = L(s_l)$, Lévy-Brownian Motion;
- DGP2: $y_l \sim I(1)$ as in (6) with $B = J_c$ and $c = c_{0.01}$, so the average pairwise correlation of $\{B(s_l)\}_{l=1}^n$ is 0.01;
- DGP3: $y_l \sim I(1)$ with $B = J_c$ and $c = c_{0.03}$;
- DGP4: $y_l \sim I(1)$ with B a Gaussian process with Matérn covariance function equal to $\mathbb{E}[B(s)B(r)] = (1 + c\Delta + (c\Delta)^2/3)\exp(-c\Delta)$ for $\Delta = |s - r|$ and c such that the average pairwise correlation of $\{B(s_l)\}_{l=1}^n$ is equal to $\bar{\rho} = 0.03$,
- DGP5: $y_l = J_c(s_l)$ with $c = c_{0.03}$;
- DGP6: $y_l = J_c(s_l)$ with $c = c_{0.50}$;
- DGP7: $y_l = X(s_l)$, with X a Brownian sheet as defined in footnote 3.

DGP1-DGP4 feature $I(1)$ processes constructed from different $I(0)$ building blocks: white noise in DGP1; weakly correlated ($\bar{\rho} = 0.01$ and $\bar{\rho} = 0.03$) local-to-unity processes in DGP2 and DGP3; and an alternative Matérn process in DGP4. DGP5 and DGP6 exhibit less than $I(1)$ persistence, much less so in DGP5, and are included to examine the potential effects of 'over-differencing' on inference. The final design, DGP7, generates highly persistent data, but is outside the class of $I(1)$ models introduced in Section 2.

5.2 Data Transformations

We consider inference based on six estimators for β_1.

Levels Regression: This is OLS applied to the 'levels' regression (10). When variables are $I(1)$, this is the spurious regression studied in Section 3.

The next four estimators are OLS estimators using transformed versions the variables. Denote an individual transformed data point $(y_l^*, \mathbf{x}_l^*)$, and stack these in the vector $\mathbf{Y}_n^*$ and matrix $\mathbf{X}_n^*$. In all methods, we use the same transformation for the $k+1$ variables in $(y_l, \mathbf{x}_l)$, and then run a linear regression of $\mathbf{Y}_n^*$ on $\mathbf{X}_n^*$. This regression omits a constant, since all transformations involves a demeaning step.

Isotropic Differences: These mimic the differencing operation in time series and sets

$$y_l^* = \sum_{\ell \neq l} \kappa(|s_l - s_\ell|)(y_l - y_\ell) \tag{21}$$

for some weighting function $\kappa_b : \mathbb{R} \mapsto \mathbb{R}$ with $\kappa_b(x) = x^{-1/2}\mathbf{1}[|x| \leq b]$ for some $b > 0$. We call this transformation 'isotropic differencing', as (21) sums over all differences that are within a ball of radius b around s_l, so it is invariant to rotations of the data. The parameter b is a bandwidth, and the term $x^{-1/2}$ in $\kappa_b(x)$ implies that relatively more weight is put on differences that are computed from nearby observations. We consider this method with $b = 0.03, 0.06, \ldots, 0.15$, where the locations $\{s_l\}$ are scale normalized so that $\max_{l,\ell} |s_l - s_\ell| = 1$.

Cluster Fixed Effects: We partition the sampling region $\mathcal{S}$ into m regions $\mathcal{R}_i$, $i = 1, \ldots, m$ by applying the k-means algorithm to the locations $\{s_l\}_{l=1}^{400}$. This is meant to mimic counties partitioning a state, or states partitioning the U.S., and so forth. We then compute deviations from region means

$$y_l^* = y_l - \frac{\sum_{i=1}^m \mathbf{1}[s_l \in \mathcal{R}_i] \sum_{\ell \neq l} \mathbf{1}[s_\ell \in \mathcal{R}_i] y_\ell}{\sum_{i=1}^m \mathbf{1}[s_l \in \mathcal{R}_i] \sum_{\ell \neq l} \mathbf{1}[s_\ell \in \mathcal{R}_i]}.$$

Including fixed effects for each region effectively induces this transformation for all variables in a regression. This is implemented for $m = 30, 60, 120, 240$.

LBM-GLS: Recall from the last section that $\mathbf{\Sigma}_{n,L}$ is the $n \times n$ covariance matrix of $\mathbf{Y}_n$ induced by a Lévy-Brownian motion. If $\mathbf{Y}_n$ is generated by a Lévy-Brownian motion then $\mathbf{Y}_n \sim \mathcal{N}(\mu\mathbf{1}, \mathbf{\Sigma}_{n,L})$ and

$$\mathbf{Y}_n^* = (\mathbf{M}\mathbf{\Sigma}_{n,L}\mathbf{M})^{-1/2}\mathbf{Y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{M}) \tag{22}$$

where $(\mathbf{M}\mathbf{\Sigma}_{n,L}\mathbf{M})^{-1/2}$ is the Moore-Penrose generalized inverse of $(\mathbf{M}\mathbf{\Sigma}_{n,L}\mathbf{M})^{1/2}$. This GLS transformation converts $\mathbf{Y}_n^*$ into a set of demeaned i.i.d. random variables. In a more general

$I(1)$ model, this is no longer true, but given the FCLT in Theorem 2, it is plausible that this LBM-GLS transformation induces enough stationarity for spatial HAR inference to be reliable.

Remark 5.1. The LBM-GLS estimator of β is closely related to the OLS estimator obtained after controlling for a smooth spline function, say $\eta(s)$, in the regression.[9] To see this, write the stacked regression as $\mathbf{Y}_n = \mathbf{X}_n\boldsymbol{\beta} + \boldsymbol{\eta} + \mathbf{e}$, where $\mathbf{Y}_n$ is $n \times 1$, $\mathbf{X}_n$ is $n \times k$, $\boldsymbol{\beta}$ is $k \times 1$, $\boldsymbol{\eta}$ is $n \times 1$ with $\eta_l = \eta(s_l)$, and $\mathbf{e}$ is a vector of errors. Estimation of $\boldsymbol{\beta}$ subject to a smoothness constraint on $\eta(\cdot)$ can be accomplished by solving the penalized least squares problem $\min_{\boldsymbol{\beta},\boldsymbol{\eta}}[(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta} - \boldsymbol{\eta})'(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta} - \boldsymbol{\eta}) + \lambda^{-1}\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}]$ for an appropriately chosen matrix $\boldsymbol{\Omega}$ that captures the smoothness properties of $\eta(\cdot)$. The solution to the problem yields the estimator $\hat{\boldsymbol{\beta}} = [\mathbf{X}_n'(\mathbf{I}+\lambda^{-1}\boldsymbol{\Omega})^{-1}\mathbf{X}_n]^{-1}[\mathbf{X}_n'(\mathbf{I}+\lambda^{-1}\boldsymbol{\Omega})^{-1}\mathbf{Y}_n]$, which is recognized as the GLS estimator using the error covariance matrix $\mathbf{I} + \lambda^{-1}\boldsymbol{\Omega}$. The location-invariant GLS estimator uses $(\mathbf{M}\mathbf{Y}_n, \mathbf{M}\mathbf{X}_n)$ in place of $(\mathbf{Y}_n, \mathbf{X}_n)$. The use of $\boldsymbol{\Sigma}_{n,L}$ for $\boldsymbol{\Omega}$ imposes a Levy-Brownian motion smoothness prior of η; this is a spatial generalization of the Wiener process smoothing prior that yields a quadratic smoothing spline when $d = 1$. For our designs, we found that inference using this estimator and spatial HAR standard errors performed best using $\lambda = 0$, which coincides with the LBM-GLS method

Low-pass Eigenvector Transformation: Recall that the $q \times 1$ vector $\mathbf{Z}_n$ in the previous section was defined as $\mathbf{Z}_n = \mathbf{R}_n'\mathbf{Y}_n$, where $\mathbf{R}_n$ collects the eigenvectors of $\mathbf{M}\boldsymbol{\Sigma}_{n,L}\mathbf{M}$ corresponding to the $q < n$ largest eigenvalues $\hat{\boldsymbol{\nu}}_n$. By construction, if $\mathbf{Y}_n \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{\Sigma}_{n,L})$, then

$$\mathbf{Y}_n^* = \operatorname{diag}(\hat{\boldsymbol{\nu}}_n)^{-1/2}\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, n^{-1}\mathbf{I}_q). \tag{23}$$

Here $\mathbf{Y}_n^*$ is $q \times 1$, rather than $n \times 1$, as in the previous transformations. Note that setting $q = n - 1$ amounts to the LBM-GLS transformation (22). The potential advantage of using a smaller, fixed q is that the CMT and the FCLT imply that the (suitably scaled) $\mathbf{Y}_n^*$ in (23) has an asymptotic $\mathcal{N}(\mathbf{0}, \mathbf{I}_q)$ distribution in the general $I(1)$ model. Thus, classical Gaussian small sample inference in the regression of $\mathbf{Y}_n^*$ on $\mathbf{X}_n^*$ is asymptotically justified with this

[9] Kelly (2022) proposes a spline-augmented OLS estimator for inference in spatial regressions and presents simulation evidence showing that it improves size control compared to using spatial HAC standard errors and OLS with untransformed regressors. This remark reiterates well-known results on the relationship between smoothing splines, Bayes smoothness priors, and maximum likelihood estimators (cf. Engle and Watson (1988)).

transformation in the general $I(1)$ model. This approach is analogous to what is suggested by Müller and Watson (2017) for persistent time series. We implement this with $q = 10, 20, 50$.

High-pass Eigenvector Transformation: An alternative to using the first q principal components in $\mathbf{Z}_n$ is to use the remaining $n-1-q$ principal components, say $\mathbf{Y}_n^* = \tilde{\mathbf{R}}_n'\mathbf{Y}_n$, where $\tilde{\mathbf{R}}_n$ collects the eigenvectors of $\mathbf{M}\boldsymbol{\Sigma}_{n,L}\mathbf{M}$ corresponding to the $n-1-q$ smallest eigenvalues. The rationale for this approach is that eliminating the first q principal components purges the data of the large-variance components associated with spatial $I(1)$ persistence. Alternatively, in the context of Remark 5.1, the resulting regression controls for smooth spatial function spanned by the columns of $\mathbf{R}_n$. We implement this with $q = 5, 10, 20, 50$.

Ibragimov-Müller: In addition, we consider the approach suggested in Ibragimov and Müller (2010). They suggest dealing with weak spatial dependence by running m independent regressions of y_l on x_l and a constant in each region $\mathcal{R}_i$, $i = 1, \ldots, m$, for some reasonably small m. Let $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{i,1}, \ldots, \hat{\beta}_{i,k})'$ be the corresponding coefficients. The m estimators $\hat{\beta}_{i,j}$ are then treated as independent and Gaussian information about β_j, so inference is conducted using a corresponding t-statistic with a Student-t critical value with $m-1$ degrees of freedom. We form the m regions by applying the k-means algorithm to the locations $\{s_l\}_{l=1}^{400}$, and consider $m = 10, 20, 50$.

These estimators of β_1 are used in conjunction with three types of standard errors: We present results using heteroskedasticity robust standard errors for the LBM-GLS and cluster fixed effects estimators. For the latter estimator, we also consider clustered standard errors. Finally, for all but the last two methods, we use spatial HAR standard errors and critical values suggested by Müller and Watson (2022b). This so-called C-SCPC-method is calibrated to control size under spatial dependence with an average correlation of no more than 0.03 (and, by "conditioning" on the regressor, it is by construction more conservative than the method developed in Müller and Watson (2022a)).

5.3 Simulation Results

The experiments involve 96 different spatial designs and six different estimators, five of which are implemented for several values of a bandwidth or related parameters (b for isotropic differencing, m for clustered fixed effect and Ibragimov-Müller and q for the eigenvector transforms). A detailed summary of the results is provided in the Supplementary Material.

Table 2: Rejection Frequency for Nominal 5% Tests (median over 96 spatial designs)

Method	Lévy-BM	$I(1)_{c_{0.01}}$	$I(1)_{c_{0.03}}$	$I(1)_{Matérn}$	$J_{c_{0.03}}$	$J_{c_{0.50}}$	Br. Sheet
				(a) k = 1			
OLS (C-SCPC)	0.29	0.32	0.35	0.33	0.04	0.20	0.33
Isotropic difference (C-SCPC)	0.04	0.05	0.07	0.06	0.03	0.04	0.04
Cluster fxed-effects (cluster)	0.08	0.24	0.35	0.30	0.07	0.07	0.13
Cluster fxed-effects (C-SCPC)	0.05	0.08	0.12	0.10	0.04	0.05	0.08
LBM-GLS	0.05	0.26	0.39	0.38	0.06	0.05	0.23
LBM-GLS (C-SPC)	0.03	0.05	0.07	0.06	0.03	0.03	0.09
Low-pass Eigenvector	0.05	0.05	0.05	0.05	0.08	0.05	0.13
High-pass Eigenvector (C-SCPC)	0.05	0.10	0.13	0.14	0.05	0.05	0.15
Ibragimov-Müller	0.08	0.13	0.15	0.13	0.05	0.07	0.16
Addendum: Avg. R^2	0.16	0.20	0.24	0.22	0.01	0.10	0.16
				(b) k = 5			
OLS (C-SCPC)	0.29	0.32	0.36	0.33	0.05	0.22	0.32
Isotropic difference (C-SCPC)	0.04	0.05	0.07	0.06	0.03	0.04	0.05
Cluster fxed-effects (cluster)	0.08	0.24	0.35	0.31	0.07	0.08	0.14
Cluster fxed-effects (C-SCPC)	0.05	0.08	0.11	0.09	0.05	0.05	0.08
LBM-GLS	0.05	0.26	0.39	0.38	0.06	0.05	0.23
LBM-GLS (C-SPC)	0.03	0.05	0.07	0.06	0.03	0.03	0.09
Low-pass Eigenvector	0.05	0.05	0.05	0.05	0.08	0.05	0.10
High-pass Eigenvector (C-SCPC)	0.06	0.11	0.14	0.15	0.05	0.06	0.15
Ibragimov-Müller	0.05	0.06	0.08	0.07	0.05	0.05	0.07
Addendum: Avg. R^2	0.47	0.59	0.66	0.62	0.05	0.35	0.47

Notes: Entries are the median rejection frequency across the 96 spatial designs described in the text. For methods that depend on a bandwidth or other parameter, results are shown for the parameter value with the smallest size distortion. "(C-SCPC)" indicates spatial-robust HAR inference from Müller and Watson (2022b). "(cluster)" indicates that clustered standard errors are used.

Here we present the key conclusions in two tables.

Table 2 summarizes the rejection frequency of nominal 5% level tests for each method. It reports median rejection frequencies across the 96 spatial designs and, where a method depends on a parameter, chooses the parameter that yields the rejection frequency closest to the nominal value of 0.05, thus providing a lower bound on the method's size distortion.

Table 3 summarizes the expected length of the resulting (non-size corrected) 95% confidence intervals. It leaves out the levels-OLS estimator, clustered-standard error fixed effects and high-pass eigenvector transform methods because of their significant size distortions.

There are two key takeaways from the tables. First, isotropic differences and LBM-GLS implemented with HAR standard errors have reasonably good size properties in all designs, as does the eigenvalue transformation. Second, LBM-GLS (with HAR standard errors) produces confidence intervals with the smallest average length. These results, along with the

Table 3: Expected Length of Nominal 95% Confidence Intervals (median across the 96 spatial designs

Method	Lévy-BM	$I(1)_{c_{0.01}}$	$I(1)_{c_{0.03}}$	$I(1)_{Matérn}$	$J_{c_{0.03}}$	$J_{c_{0.50}}$	Br. Sheet
	(a) k = 1						
Isotropic difference (C-SCPC)	0.53	0.70	0.73	0.73	0.44	0.52	0.54
Cluster-FE (CSCPC)	0.55	0.81	0.91	0.88	0.43	0.39	0.50
LBM-GLS (C-SPC)	0.25	0.42	0.54	0.55	0.26	0.26	0.33
Low-pass Eigenvector	1.51	1.51	1.51	1.51	0.57	0.57	1.51
Ibragimov-Müller	0.42	0.87	1.00	0.96	0.37	0.41	0.43
	(b) k = 5						
Isotropic difference (C-SCPC)	0.51	0.69	0.77	0.78	0.43	0.51	0.51
Cluster fxed-effects (C-SCPC)	0.35	0.42	0.48	0.47	0.34	0.35	0.37
LBM-GLS (C-SPC)	0.26	0.42	0.54	0.55	0.27	0.26	0.33
Low-pass Eigenvector	2.30	2.30	2.30	2.30	0.60	0.60	2.30
Ibragimov-Müller	0.46	0.65	0.67	0.78	0.36	0.48	0.49

Notes: Entries are the median average length across the 96 spatial designs described in the text. See Table 2 for additional comment notes.

observation that LBM-GLS does not require the choice of a bandwidth or other parameter, suggests that it dominates the other methods considered here.

Remark 5.2. The final rows in Table 2 show the average value of the R^2 in the levels-regression (10). (This is the median value across the 96 spatial designs.) These R^2 values are large for the $I(1)$ models, consistent with the implications of Theorem 3, and for the local-to-unity model with $c = c_{0.50}$, consistent with the discussion in Remark 3.3.

5.4 Regressions in Chetty et al. (2014)

We now return to the results in Table 1. As noted in Section 4.5 the first three columns of the table suggest substantial spatial correlation in many of the variables. The final two columns summarize results from the regression of the Absolute Mobility Index (the first variable in the table) onto each of the other variables. These regressions were reported in Figure VII of Chetty, Hendren, Kline, and Saez (2014). The penultimate column in the Table 1 (labeled "Level(Cluster)") reports the levels-OLS estimate of the regression coefficient with a nominal 95% confidence interval computed using standard errors clustered at the state level. These results are reported in Chetty, Hendren, Kline, and Saez (2014). The final column in the table shows results using LBM-GLS with a C-SCPC 95% confidence interval.

Two results stand out from a comparison of the levels- and LBM-GLS results. First, the LBM-GLS estimates of β tend to be smaller in magnitude than the OLS estimates; the estimated correlations are more muted than those reported in Chetty, Hendren, Kline, and Saez (2014). Second, the LBM-GLS C-CSPC confidence intervals are narrower and, based on the experiments reported earlier, provide approximately valid inference for the correlation between each variable and the mobility index. Our reading of these results is that the substantive conclusions made in Chetty, Hendren, Kline, and Saez (2014) about the correlation of the various socioeconomic factors with intergenerational income mobility largely continue to hold after accounting for the strong spatial correlation in the variables.

6 Concluding Remarks

Applied researchers are well aware of the pitfalls of conducting inference with persistent time series data. Variables are routinely tested for the presence of a unit root, and often differenced to stationarity to avoid spurious regression effects.

This paper demonstrates that inference with highly persistent spatial data is equally fraught: HAC corrections for spatial dependence fail in the presence of strong correlations, leading to spurious significance between independent spatial variables. We have provided tools to detect such strong spatial persistence, akin to time series unit root and stationarity tests.

We have also suggested ways of restoring valid inference by suitably transforming the spatial variables, combined with spatial HAR corrections to accommodate any residual weak correlations. The theory here is less complete, however: For the most promising of these transformations—the FGLS approach using the canonical spatial I(1) model as a baseline—we do not yet have a good theoretical understanding of its properties, and future research is required to fully understand the conditions under which this approach yields valid inference.

A Proofs

Proof of Lemma 1:

By the Corollary on page 48 of Adler (2010), the result holds if for some $m > 2d$, $\mathbb{E}\left[(Y_n(s) - Y_n(r))^{2m}\right] \leq C|s-r|^m$ for some C. Let $m > 2d$ and apply Condition 1 to obtain

$$\begin{aligned}
\mathbb{E}[\lambda_n^{-m}(Y_n(s) - Y_n(r))^{2m}] &\leq \lambda_n^{-m} C_m \left(\kappa_d^2 \int_{\mathbb{R}^d} (|\lambda_n s - u|^{(1-d)/2} - |\lambda_n r - u|^{(1-d)/2})^2 du\right)^m \\
&= C_m \left(\kappa_d^2 \int_{\mathbb{R}^d} (|s-u|^{(1-d)/2} - |r-u|^{(1-d)/2})^2 du\right)^m \\
&= C_m \mathbb{E}[(L(s) - L(r))^2]^m \\
&= C_m |s-r|^m
\end{aligned}$$

where the second equality follows from the representation (4) of L.

For the corresponding result about spatial local-to-unity processes, note that with $h(s,u) = \kappa_{d,c}(|s-u|^{(1-d)/4} K_{(1-d)/4}(c|s-u|))$

$$\begin{aligned}
\mathbb{E}[\lambda_n^{-m}(Y_n(s) - Y_n(r))^{2m}] &\leq \lambda_n^{-m} C_m \left(\lambda_n^{(1-d)} \int_{\mathbb{R}^d} (h(s, \lambda_n^{-1}u) - h(r, \lambda_n^{-1}u))^2 du\right)^m \\
&= C_m \left(\int_{\mathbb{R}^d} (h(s,u) - h(r,u))^2 du\right)^m \\
&= C_m \mathbb{E}[(J_c(s) - J_c(r))^2]^m \qquad (24)
\end{aligned}$$

where the last equality follows from the representation (8) of J_c, and

$$\mathbb{E}[(J_c(s) - J_c(r))^2] = \frac{2 - 2\exp(-c|s-r|)}{2c} \leq |s-r|$$

from a first-order Taylor expansion of the function $x \mapsto \exp(-cx)$, and the result again follows. □

Proof of Theorem 2:

We focus on the convergence for the LTU process (9); the proof for the convergence in Theorem 2 is analogous and is omitted for brevity.

We show convergence of finite dimensional distributions and tightness of the process $\lambda_n^{-1/2} Y_n$. The latter follows by Theorem 23.7 of Kallenberg (2021) from (24) and $J_c(0) \sim \mathcal{N}(0, (2c)^{-1})$. For the former, consider for $t_1, \ldots, t_k \in \mathcal{S}$, the $k \times 1$ vector $\lambda_n^{-1/2}(Y_n(t_1), \ldots, Y_n(t_k))$. By the Cramér-Wold device, it suffices to establish the convergence $X_n = \lambda_n^{-1/2}\sum_{j=1}^k v_j Y_n(t_j) \Rightarrow \sum_{j=1}^k v_j \omega J_c(t_j)$ for $(v_1,\ldots,v_k) \in \mathbb{R}^k$. Let $h(u) = \sum_{j=1}^k v_j \kappa_{d,c}(|t_j - u|^{(1-d)/4}K_{(1-d)/4}(c|t_j - u|))$, so that from (8), $\sum_{j=1}^k v_j J_c(t_j) \sim \mathcal{N}(0, \int_{\mathbb{R}^d} h(u)^2 du)$ and from (9)

$$X_n = \lambda_n^{-d/2}\int_{\mathbb{R}^d} h(\lambda_n^{-1}u)B(u)du.$$

For $\varepsilon > 0$, define $h^\varepsilon(u) = h(u)\mathbf{1}[|u| < 1/\varepsilon]\prod_{j=1}^k \mathbf{1}[|t_j - u| > \varepsilon]$ and let

$$X_n^\varepsilon = \lambda_n^{-d/2}\int_{\mathbb{R}^d} h^\varepsilon(\lambda_n^{-1}u)B(u)du.$$

From Condition 1 we find

$$\begin{aligned}
\mathbb{E}[(X_n^\varepsilon - X_n)^2] &= \lambda_n^{-d}\mathbb{E}\left[\left(\int_{\mathbb{R}^d}(h(\lambda_n^{-1}u) - h^\varepsilon(\lambda_n^{-1}u))B(u)du\right)^2\right] \\
&\leq C_2\lambda_n^{-d}\int_{\mathbb{R}^d}(h(\lambda_n^{-1}u) - h^\varepsilon(\lambda_n^{-1}u))^2 du \\
&= C_2\int_{\mathbb{R}^d}(h(u) - h^\varepsilon(u))^2 du.
\end{aligned}$$

Since $\int_{\mathbb{R}^d}(h(u)-h^\varepsilon(u))^2 du \leq 2\int_{\mathbb{R}^d} h(u)^2 du < \infty$, and $h^\varepsilon(u) \leq h(u)$ for all u, it follows from the dominated convergence theorem that this quantity can be made arbitrarily small by picking ε small enough.

Furthermore

$$\begin{aligned}
\mathbb{E}[(X_n^\varepsilon)^2] &= \lambda_n^{-d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h^\varepsilon(\lambda_n^{-1}u)h^\varepsilon(\lambda_n^{-1}s)\sigma_B(s-u)duds \\
&= \int_{\mathbb{R}^d}\sigma_B(s)\int_{\mathbb{R}^d} h^\varepsilon(u)h^\varepsilon(u + \lambda_n^{-1}s)duds \\
&\to \int_{\mathbb{R}^d}\sigma_B(s)ds\int_{\mathbb{R}^d} h^\varepsilon(u)^2 du
\end{aligned}$$

by dominated convergence, since by Cauchy-Schwarz, $\left[\int_{\mathbb{R}^d} h^\varepsilon(u)h^\varepsilon(u+\lambda_n^{-1}s)du\right]^2 \leq \left[\int_{\mathbb{R}^d} h^\varepsilon(u)^2 du\right]^2$ and $\int_{\mathbb{R}^d}|\sigma_B(s)|ds < \infty$.

Finally, note that h^ε is bounded and $h^\varepsilon(\lambda_n^{-1}u) = 0$ for $|u| > \lambda_n/\varepsilon$. Thus, using Condition 2,

$$X_n^\varepsilon \Rightarrow \mathcal{N}\left(0, \int_{\mathbb{R}^d}\sigma_B(s)ds\int_{\mathbb{R}^d} h^\varepsilon(u)^2 du\right).$$

The result now follows since mean square convergence implies convergence in distribution, and $\varepsilon > 0$ was arbitrary. $\square$

Proof of Theorem 3:

The results follow straightforwardly from the CMT if we can show that $\lambda_n^{-1/2} n^{-1} \sum_{l=1}^n y_l \Rightarrow \int Y(s)dG(s)$, $\lambda_n^{-1/2} n^{-1} \sum_{l=1}^n \mathbf{x}_l \Rightarrow \int \mathbf{X}(s)dG(s)$, $\lambda_n^{-1} n^{-1} \sum_{l=1}^n \mathbf{x}_l y_l \Rightarrow \int \mathbf{X}(s)Y(s)dG(s)$ and $\lambda_n^{-1} n^{-1} \sum_{l=1}^n \mathbf{x}_l \mathbf{x}_l' \Rightarrow \int \mathbf{X}(s)\mathbf{X}(s)'dG(s)$.

Consider the convergence $\lambda_n^{-1} n^{-1} \sum_{l=1}^n \mathbf{x}_l y_l \Rightarrow \int \mathbf{X}(s)Y(s)dG(s)$. By the Skorohod almost sure representation theorem (see, for instance,Theorem 11.7.2 of Dudley (2002)), there exist random elements $(Y_n^*(\cdot), \mathbf{X}_n^*(\cdot))$ such that $\sup_{s\in\mathcal{S}} |(Y_n^*(s) - Y^*(s), \mathbf{X}_n^*(s) - \mathbf{X}^*(s))| \overset{a.s.}{\to} 0$, $(Y^*(\cdot), \mathbf{X}^*(\cdot)) \sim (Y(\cdot), \mathbf{X}(\cdot))$ and $\lambda_n^{-1/2}(Y_n(\cdot), \mathbf{X}_n(\cdot)) \sim (Y_n^*(\cdot), \mathbf{X}_n^*(\cdot))$ for $n = 1, 2, \ldots$. Thus it suffices to show the claim for $\int \mathbf{X}_n^*(s)Y_n^*(s)dG_n(s) = n^{-1}\sum_{l=1}^n \mathbf{X}_n^*(s_l)Y_n^*(s_l) \sim \lambda_n^{-1} n^{-1} \sum_{l=1}^n \mathbf{X}_n(s_l)Y_n(s_l)$. We have

$$\left| \int (\mathbf{X}_n^*(s)Y_n^*(s) - \mathbf{X}^*(s)Y^*(s))dG_n(s) \right| \leq \sup_{s\in\mathcal{S}} |(Y_n^*(s) - Y^*(s), \mathbf{X}_n^*(s) - \mathbf{X}^*(s))| \overset{a.s.}{\to} 0$$

so it suffices to show the claim for $\int \mathbf{X}^*(s)Y^*(s)dG_n(s)$. Now almost all realizations of the $\mathbb{R}^k \mapsto \mathbb{R}$ function $s \mapsto \mathbf{X}^*(s)Y^*(s)$ on $\mathcal{S}$ are continuous and bounded. For any such realization, $\int \mathbf{X}^*(s)Y^*(s)dG_n(s) \to \int \mathbf{X}^*(s)Y^*(s)dG(s)$ by the definition of convergence in distribution. Thus $\int \mathbf{X}^*(s)Y^*(s)dG_n(s) \overset{a.s.}{\to} \int \mathbf{X}^*(s)Y^*(s)dG(s)$. But almost sure convergence implies convergence in distribution, so the desired result follows. The other terms are dealt with in the same manner. □

The following Lemma is used in the proof of Theorem 4.

Lemma 7. *For any $\delta > 0$, $\limsup_{n\to\infty} \sup_{r\in\mathcal{S}} G_n(\{s : |s - r| \leq \delta\}) \leq \sup_{r\in\mathcal{S}} G(\{s : |s - r| \leq \delta\})$, where $G_n(A)$ and $G(A)$ is the measure that is assigned to the Borel set $A \subset \mathbb{R}^d$ by the distributions G_n and G, respectively.*

Proof. Let $a = \sup_{r\in\mathcal{S}} G(\{s : |s - r| \leq \delta\})$. Suppose otherwise. Then there exists $\varepsilon > 0$ and a sequence r_n such that

$$\limsup_{n\to\infty} \sup_{r\in\mathcal{S}} G_n(\{s : |s - r| \leq \delta\}) = \lim_{n\to\infty} G_n(\{s : |s - r_n| \leq \delta\}) \geq a + \varepsilon.$$

Let $\delta' > \delta$ be such that $\sup_{r\in\mathcal{S}} G(\{s : |s - r| \leq \delta'\}) \leq a + \varepsilon/2$. Since $\mathcal{S}$ is compact, $r_n \to r_0$ along some subsequence. Along that subsequence, for all n large enough so that $|r_n - r_0| < \delta' - \delta$, we have

$$G_n(\{s : |s - r_n| \leq \delta\}) \leq G_n(\{s : |s - r_0| \leq \delta'\}) \to G(\{s : |s - r_0| \leq \delta'\}) \leq a + \varepsilon/2$$

yielding the desired contradiction. □

Proof of Theorem 4:

From Theorem 3 and the CMT, $\mathbf{H}\hat{\boldsymbol{\beta}} \Rightarrow \mathbf{H}\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}}$ with the r.h.s. non-zero with probability one. Thus $\mathbf{H}\hat{\boldsymbol{\beta}} = O_p(1)$ (and not $\mathbf{H}\hat{\boldsymbol{\beta}} = o_p(1)$). The result thus follows if we can show that $||\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\hat{\boldsymbol{\Omega}}_n\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}|| = o_p(n)$ (since this implies that the smallest eigenvalue of $n(\mathbf{H}\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\hat{\boldsymbol{\Omega}}_n\mathbf{S}_{\tilde{x}\tilde{x}}^{-1}\mathbf{H}')^{-1}$ diverges).

Now since $\lambda_n^{-1}\mathbf{S}_{\tilde{x}\tilde{x}} \Rightarrow \boldsymbol{\Xi}_{\tilde{x}\tilde{x}}$ and $\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}$ is full rank with probability one, it suffices to show that $n^{-1}\lambda_n^{-2}||\hat{\boldsymbol{\Omega}}_n|| \xrightarrow{p} 0$.

By the FCLT and CMT, $\lambda_n^{-1}\mathbf{e}_n(\cdot) \Rightarrow \mathbf{e}_0(\cdot) = (\tilde{Y}(\cdot) - \tilde{\mathbf{X}}(\cdot)'\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\boldsymbol{\Xi}_{\tilde{x}\tilde{y}})\tilde{\mathbf{X}}(\cdot)$, so that $\lambda_n^{-1}\sup_l |\mathbf{e}_n(s_l)| \Rightarrow \sup_{s\in\mathcal{S}} |\mathbf{e}_0(s)|$, and $\lambda_n^{-1}\sup_l |\mathbf{e}_n(s_l)| = O_p(1)$. We have

$$\lambda_n^{-2}n^{-2}\left\|\sum_{l,\ell=1}^{n}\kappa(b_n(s_l - s_\ell))\mathbf{e}_n(s_l)\mathbf{e}_n(s_\ell)'\right\| \leq \lambda_n^{-2}\sup_l |\mathbf{e}_n(s_l)|^2 \cdot n^{-2}\sum_{l,\ell=1}^{n}|\kappa(b_n(s_l - s_\ell))|$$

and

$$\sum_{l,\ell=1}^{n}|\kappa(b_n(s_l - s_\ell))| \leq \bar{\kappa}\sum_{l,\ell=1}^{n}\mathbf{1}[|s_l - s_\ell| \leq b_n^{-1/2}] + \sum_{l,\ell=1}^{n}\mathbf{1}[|s_l - s_\ell| > b_n^{-1/2}]|\kappa(b_n(s_l - s_\ell))|.$$

Now

$$n^{-2}\sum_{l,\ell=1}^{n}\mathbf{1}[|s_l - s_\ell| > b_n^{-1/2}]|\kappa(b_n(s_l - s_\ell))| \leq \sup_{|a|=1}|\kappa(b_n^{1/2}a)| = o_p(1)$$

by (14) and $b_n^{-1} = o_p(1)$. Furthermore, since G is continuous, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\sup_{r\in\mathcal{S}} G(\{s : |s - r| \leq \delta\}) \leq \varepsilon$. Note that

$$\begin{aligned} n^{-2}\sum_{l,\ell=1}^{n}\mathbf{1}[|s_l - s_\ell| \leq b_n^{-1/2}] &\leq \sup_{r\in\mathcal{S}} G_n(\{s : |s - r| \leq b_n^{-1/2}\}) \\ &\leq \sup_{r\in\mathcal{S}} G_n(\{s : |s - r| \leq \delta\}) + \mathbb{P}(b_n^{-1/2} > \delta). \end{aligned}$$

Since by assumption, $b_n^{-1} = o_p(1)$, we have $\mathbb{P}(b_n^{-1/2} > \delta) \to 0$, and by Lemma 7, $\limsup_{n\to\infty}\sup_{r\in\mathcal{S}} G_n(\{s : |s - r| \leq \delta\}) \leq \varepsilon$. But $\varepsilon > 0$ was arbitrary, and the result follows. $\square$

Proof of Lemma 5:

The proof follows from the same steps as the proof of Lemma 6 in Müller and Watson (2022a). In that proof the assumption of i.i.d. s_l is used on two occasions. Once to argue that $||\int k_0(\cdot, s)(dG_n(s) - dG(s))||_{\mathcal{H}} = O_p(n^{-1/2})$, where $||\cdot||_{\mathcal{H}}$ is the norm of a reproducing kernel Hilbert space $\mathcal{H}$ with kernel $k_0(r,s) = \bar{k}(r,s) + 1$ and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ satisfying $\langle f, k_0(s,\cdot) \rangle_{\mathcal{H}} = f(s)$ for all $f \in \mathcal{H}$. This may be replaced by

$$\begin{aligned}\left\| \int k_0(\cdot,s)(dG_n(s) - dG(s)) \right\|_{\mathcal{H}}^2 &= \int\int k_0(s,r)(dG_n(s) - dG(s))(dG_n(r) - dG(r)) \\ &= \mathbb{E}[k_0(S_n,R_n) - k_0(S,R_n) - k_0(S_n,R) + k_0(S,R)] \to 0\end{aligned}$$

where for each n, S_n, R_n are independent random variables both with distribution G_n, and S, R are two independent random variables with distribution G. The convergence then follows since the $\mathbb{R}^{2d} \mapsto \mathbb{R}$ function $(s,r) \mapsto k_0(s,r)$ is continuous and bounded.

The second occasion is to show that, in the notation of the proof of Lemma 6 in Müller and Watson (2022a), $||\bar{L}_n - \bar{L}||_{HS} = ||M(L_n - L)M||_{HS} \leq ||M||^2||L_n - L||_{HS} = O_p(n^{-1/2})$ since $||M|| < \infty$ and $||L_n - L||_{HS} = O_p(n^{-1/2})$ as in Theorem 7 of Rosasco, Belkin, and Vito (2010). That argument may be replaced by

$$\begin{aligned}||L_n - L||_{HS}^2 &= \sum_{j\geq 0} \left\langle \int e_j(s)k_0(s,\cdot)(dG_n(s) - dG(s)), \int e_j(s)k_0(s,\cdot)(dG_n(s) - dG(s)) \right\rangle_{\mathcal{H}} \\ &= \int\int \left(\sum_{j\geq 0} e_j(s)e_j(r) \right) k_0(s,r)(dG_n(s) - dG(s))(dG_n(r) - dG(r)) \\ &= \int\int k_0(s,r)^2(dG_n(r) - dG(r))(dG_n(r) - dG(r)) \\ &= \mathbb{E}[k_0(S_n,R_n)^2 - k_0(S,R_n)^2 - k_0(S_n,R)^2 + k_0(S,R)^2] \to 0\end{aligned}$$

where $e_0 = 1$ and $e_j = \sqrt{\bar{\nu}_j}\bar{\varphi}_j$, $j = 1, 2, \ldots$ form a basis of $\mathcal{H}$ (see the discussion in the proof of Lemma 6 in Müller and Watson (2022a)), the third equality follows from $\bar{k}(s,r) = \sum_{i=1}^{\infty} \bar{\nu}_i\bar{\varphi}_i(s)\bar{\varphi}_i(r)$, the change of the order of integration and summation is justified by Fubini's Theorem, and the convergence follows, since the $\mathbb{R}^{2d} \mapsto \mathbb{R}$ function $(s,r) \mapsto k_0(s,r)^2$ is bounded and continuous. □

Proof of Theorem 6:

By Lemmas 3 and 12 in Müller and Watson (2022a), we have

$$\lambda_n^{d/2} n^{-1}\mathbf{Z}_n \Rightarrow \mathcal{N}\left(\mathbf{0}, a\sigma_B(0)\int \bar{\boldsymbol{\varphi}}(s)\bar{\boldsymbol{\varphi}}(s)'dG(s) + \omega^2 \int \bar{\boldsymbol{\varphi}}(s)\bar{\boldsymbol{\varphi}}(s)'g(s)dG(s)\right) \tag{25}$$

where $\bar{\boldsymbol{\varphi}} = (\bar{\varphi}_1, \ldots, \bar{\varphi}_q)$, $\omega^2 = \int_{\mathbb{R}^d} \sigma_B(s)ds$ and g is the density of the distribution G. Since the LFST statistic is scale invariant, its limiting distribution under (25) only depends on the properties of B through the ratio $\chi = a\sigma_B(0)/\omega^2 \in [0,\infty)$. We need to show that $\liminf_{n\to\infty} \mathrm{cv}_n^{\mathrm{LFST}}$ is at least as large as the $1-\alpha$ quantile, say $\mathrm{cv}_\chi^{\mathrm{LFST}}$, of the (continuous) asymptotic distribution of LFST for this value of χ.

Note that for $B = J_c$, $\sigma_B(0)/\omega^2 = K_d c^{1+d}$ for some $K_d > 0$. For $a > 0$, let c_* be such $K_d c_*^{1+d} = \chi/a$, and let $c_* = 1$ otherwise. For all n sufficiently large so that $\lambda_n c_* \geq c_{0.03}$, $\mathrm{cv}_n^{\mathrm{LFST}}$ is such that the LFST test controls size under $B = J_{c^*}$. But since $B = J_{c_*}$ satisfies the assumptions of Lahiri (2003), this model induces the same limit (25), so its $1-a$ quantile converges to $\mathrm{cv}_\chi^{\mathrm{LFST}}$, and the result follows. □

References

Adler, R. J. (2010): *The Geometry of Random Fields*, Classics in Applied Mathematics. SIAM, Philadelphia.

Adler, R. J., and J. E. Taylor (2007): *Random Fields and Geometry*, Springer Monographs in Mathematics. Springer, New York.

Anselin, L. (1988): *Spatial Econometrics: Methods and Models.* Kluwer.

Beenstock, M., D. Feldman, and D. Felsenstein (2012): "Testing for Unit Roots and Cointegration in Spatial Cross-Section Data," *Spatial Economic Analysis*, 7, 203–222.

Bester, C., T. Conley, C. Hansen, and T. Vogelsang (2016): "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators," *Econometric Theory*, 32, 154–186.

Chan, N. H., and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014): "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *," *Quarterly Journal of Economics*, 129, 1553–1623.

Cliff, A. D., and J. K. Ord (1974): *Spatial Autocorrelation.* Pion, London.

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

Dou, L. (2019): "Optimal HAR Inference," *Working Paper, Princeton University.*

Dudley, R. M. (2002): *Real Analysis and Probability*. Cambridge University Press, Camebridge, UK.

Elliott, G. (1999): "Efficient Tests for a Unit Root When the Initial Observation is Drawn From its Unconditional Distribution," *International Economic Review*, 40, 767–783.

Engle, R. F., and M. W. Watson (1988): "The Kalman Filter: Applications to Forecasting and Rational Expectations Models," in *Advances in Econometrics, Fifth World Congress of the Econometric Society*, ed. by T. Bewley. Cambridge University Press.

Fingleton, B. (1999): "Spurious Spatial Regression: Some Monte Carlo Results with a Spatial Unit Root and Spatial Cointegration," *Journal of Regional Science*, 39, 1–19.

Geary, R. C. (1954): "The Contiguity Ratio and Statistical Mapping," *The Incorporated Statistician*, 5, 115–145.

Gelfand, A. E., P. Diggle, P. Guttorp, and M. Fuentes (eds.) (2010): *Handbook of Spatial Statistics*CRC Press.

Granger, C. W. J., and P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

Ibragimov, R., and U. K. Müller (2010): "T-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, 453–468.

Ivanov, A. V., and N. N. Leonenko (1989): *Statistical Analysis of Random Fields*. Kluwer Academic Publishers, Dordrecht.

Kallenberg, O. (2021): *Foundations of Modern Probability*. Springer, third edition edn.

Kelly, M. (2019): "The standard errors of persistence," *University College Dublin WP19/13*.

——— (2020): "Understanding Persistence," *CPER Discussion Paper DP15246*.

——— (2022): "Improved Causal Inference on Spatial Observations: A Smoothing Spline Approach," *Working paper, University College Dublin*.

King, M. L. (1987): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159–178.

Lahiri, S. (2003): "Central Limit Theorems for Weighted Sums of a Spatial Process under a Class of Stochastic and Fixed Designs," *Sankhya*, 65(2), 356–388.

Lauridsen, J., and R. Kosfeld (2006): "A test strategy for spurious spatial regression, spatial nonstationarity, and spatial cointegration," *Papers in Regional Science*, 85, 363–377.

Lee, L.-F., and J. Yu (2009): "Spatial Nonstationarity and Spurious Regression: the Case with a Row-normalized Spatial Weights Matrix," *Spatial Economic Analysis*, 4, 301–327.

——— (2013): "Near Unit Root in the Spatial Autoregressive Model," *Spatial Economic Analysis*, 8, 314–351.

Lindstrøm, T. (1993): "Fractional Brownian Fields as Integrals of White Noise," *Bulletin of the London Mathematical Society*, 25, 83–88.

Lévy, P. (1948): *Processus stochastiques et mouvement brownien.* Gauthier-Vilars.

Matérn, B. (1986): *Spatial Variation*, Lecture Notes in Statistics 36. Springer, Berlin.

Moran, P. A. P. (1950): "Notes on Continuous Stochastic Phenomena," *Biometrika*, 37, 17–23.

Müller, U. K., and M. W. Watson (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016.

——— (2017): "Low-Frequency Econometrics," in *Advances in Economics: Eleventh World Congress of the Econometric Society*, ed. by B. Honoré, and L. Samuelson, vol. II, pp. 63–94. Cambridge University Press.

——— (2022a): "Spatial Correlation Robust Inference," *forthcoming in Econometrica.*

——— (2022b): "Spatial Correlation Robust Inference in Linear Regression and Panel Models," *Manuscript.*

Mur, J., and F. J. Trívez (2003): "Unit Roots and Deterministic Trends in Spatial Econometric Models," *International Regional Science Review*, 26, 289–312.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Ord, K. (1975): "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126.

Phillips, P. C. B. (1986): "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311–340.

——— (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (1998): "New Tools for Understanding Spurious Regression," *Econometrica*, 66, 1299–1325.

Pratt, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 549–567.

Rosasco, L., M. Belkin, and E. D. Vito (2010): "On Learning with Integral Operators," *Journal of Machine Learning Research*, 11(30), 905–934.

Schabenberger, O., and C. A. Gotway (2005): *Statistical Methods for Spatial Data Analysis*. Chapman and Hall, Boca Raton.

Stock, J. H. (1991): "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459.

Stoll, A. (1986): "A nonstandard construction of Lévy Brownian motion," *Probability Theorey and Related Fields*, 71, 321–334.

Sun, Y., and M. Kim (2012): "Asymptotic F-Test in a GMM Framework with Cross-Sectional Dependence," *Review of Economics and Statistics*, 91(1), 210–233.

Zhang, H., and D. L. Zimmerman (2005): "Towards reconciling two asymptotic frameworks in spatial statistics," *Biometrika*, 92, 921–936.