# Spatial Unit Roots

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

This Draft: March 2023

**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 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 develop asymptotically valid spatial unit root tests, stationarity tests, and inference methods for the local-to-unity parameter. 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, functional central limit theorem
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 $I(1)$, HAC/HAR inference fails and OLS produces "spurious regressions" (Granger and Newbold (1974)) with estimators and test statistics behaving in non-standard ways (Phillips (1986)). Panel (a) of Figure 1 illustrates this well-known phenomenon: the realization of two independent random walks of length $n = 250$ are strongly correlated in sample, with a corresponding Newey and West (1987) t-statistic that is highly significant.

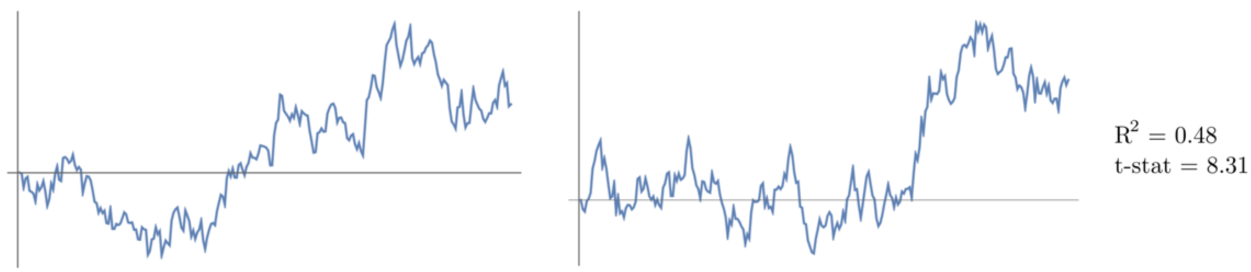
Variables measured over points in space exhibit correlation patterns that in many ways are analogous to serial correlation in time series, and this correlation also complicates inference in spatial regressions. There is a reasonably well-developed literature on HAC/HAR corrections that are required in 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)). Panel (b) of Figure 1 illustrates the issue: the realization of two independent spatial "unit root" processes with values for each of the $n = 722$ commuter zones in the 48 contiguous U.S. states are strongly correlated in sample, and a t-statistic that is clustered by U.S. states is highly significant. This raises several questions. What is a natural spatial analogue of an $I(1)$ time series process, such as the process in Figure 1 (b)? Do such processes systematically induce spuriously significant regression coefficients? 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 these questions.

Throughout the paper we use spatial data and regressions from Chetty, Hendren, Kline, and Saez (2014) to illustrate the issues and methods. These authors construct an index of intergenerational mobility for commuting zones in the United States, and study its relationship to other socioeconomic factors using bivariate regressions with standard errors clustered by U.S. states. As an example, Figure 1 (c) plots their mobility index along with the teenage labor force participation rate. The apparent similarity of these data with the simulated data

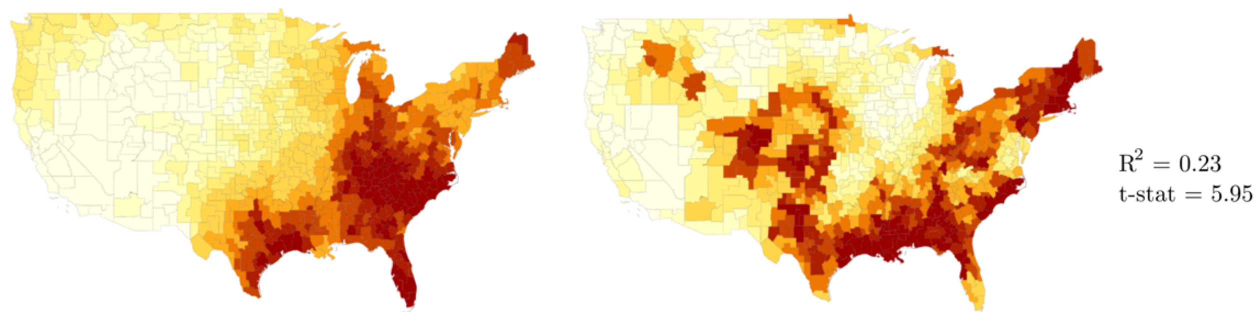
[1] Conley (1999) is a leading example of spatial HAC inference. See Müller and Watson (2022a, 2022b) for a discussion of the post-Conley literature and new suggestions for inference in regression models with weak spatial dependence.

Figure 1: Strongly Dependent Data in Time and Space

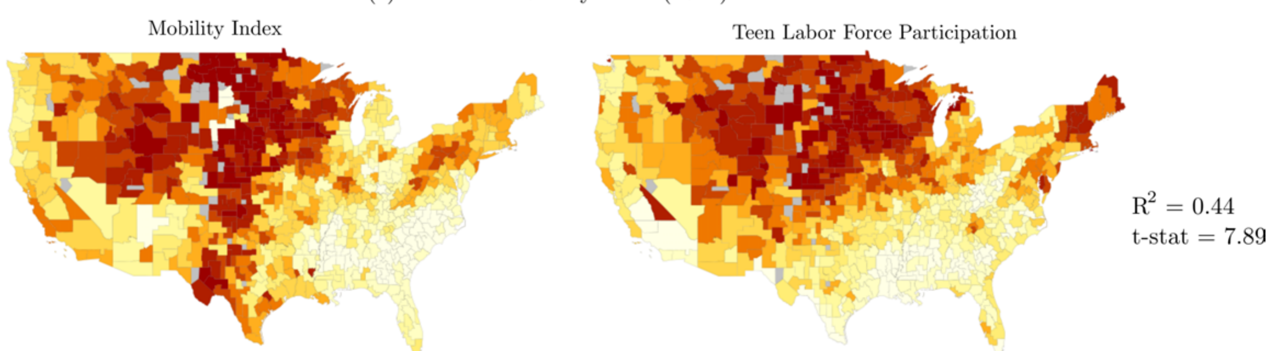
(a) Independent Time Series Random Walks




(b) Independent Spatial Unit Root Processes




(c) Data from Chetty et al. (2014)



of panel (b) suggests that the issues considered in this paper are of empirical relevance.

Much of our analysis parallels the analysis of persistent 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, we begin with a stochastic process $Y(s)$ 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 $Y(s)$ is a continuous time process and where the sample data correspond to realizations of $y_l = Y(s_l)$ observed at potentially irregularly spaced points $s_l \in \mathbb{R}$. More abstractly, as discussed in Conley (1999), "locations" might index an economic characteristic and the "economic distance" between two locations measures the dissimilarity of the characteristic.

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] SARs are simultaneous equation models, typically estimated by GMM/QMLE methods. There is a small literature on unit roots and spurious regression in SAR models, initiated by Fingleton (1999) and summarized in Rossi and Lieberman (2023). SAR models require a spatial weight (or proximity) matrix, which is usually normalized so that its rows sum to unity (Ord (1975)). Under this normalization, the unit root SAR model is not well defined. Lee and Yu (2009, 2013) study asymptotic properties of the row normalized SAR model with a SAR 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. These results are markedly different from our findings based on a continuous parameter model $Y(\cdot)$ of a spatial $I(1)$ process. More in line with our findings Fingleton (1999) generates data from a version of the SAR model that is well-defined with a unit SAR coefficient and presents Monte Carlo results suggesting spurious regression phenomena. Using a related model, Rossi and Lieberman

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

(2023) derive non-standard large-sample distributions for the estimated SAR coefficient for particular specifications of the SAR weight matrix, and suggest that model-specific non-standard results will hold more generally.

With this background in place, the roadmap of the paper is a follows. Section 2 provides our definition of a spatial $I(1)$ process. 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 its properties. In standard time series models, more general $I(1)$ processes can be constructed by replacing the white noise increments of a random walk with a weakly correlated stationary series. For example, stationary ARMA$(p, q)$ noise yields a ARIMA$(p, 1, q)$ process. Section 2 similarly defines the spatial $I(1)$ process by replacing the white noise innovations in the moving average representation of Lévy-Brownian motion with a weakly dependent stationary spatial process.

An important insight from time series analysis is that the large sample distributions of functions of $I(1)$ processes can be approximated by the distributions of corresponding functions of Wiener processes. The functional central limit theorem (FCLT) is the key driver of such approximations, 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 spatial $I(1)$ 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 and clustered versions 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.

In standard time series models, strong persistence also leads to non-standard sampling distributions for estimated autoregressions, such as those suggested by Dickey and Fuller (1979) to test the null hypothesis of a unit root. Section 4 studies a spatial analogue of such autoregressions. In particular, we define an "isotropic differencing" transformation that, for each location $s_l$, computes the weighted averages of $Y(s_\ell) - Y(s_l)$ over neighboring locations $s_\ell$. Regressions of such isotropic differences on $Y(s_l)$ are roughly analogous to a time series regression of $\Delta y_t$ on $y_{t-1}$.

Section 5 takes up the problem of conducting inference about the degree of spatial persistence in a scalar 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). We also consider versions of these tests that can be applied to residuals of one spatial variable on another, yielding spatial analogues of the residual-based cointegration tests in Engle and Granger (1987).

First-differencing an $I(1)$ time series yields an $I(0)$ process, so spurious time series regressions can be avoided by taking first differences of $I(1)$ variables. An analogous transformation for spatial $I(1)$ processes are the isotropic differences introduced in Section 4. Section 6 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 6 compares the coverage and length of feasible confidence intervals from versions of these methods. We find the GLS transformation to be particularly effective.

Section 7 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

As described in the introduction, our analysis requires three ingredients. The first is the spatial sampling region under consideration, denoted by $\mathcal{S}$. The second are the locations, $s_l \in \mathcal{S}$ that are observed. The third is the stochastic process, $Y$, that is defined on $\mathcal{S}$. Taken together, these three ingredients describe the observations

$$y_l = Y(s_l) \text{ for } l = 1, ..., n. \tag{1}$$

We introduce conditions and notation regarding the sampling region and the observed locations in this subsection. The following two subsections discuss the stochastic process $Y$.

We utilize a large-sample framework and assume that the locations $s_l$, $l = 1, \ldots, n$ are non-stochastic (or, equivalently, they are independent of all other random elements), unless stated otherwise. The locations are allowed to depend on $n$ in a double-array fashion, but we do not make this dependence explicit in the notation. We assume the following regularity condition.[3]

**Condition 1.** *(a) The locations $s_l$ are elements of $\mathcal{S}_n = \lambda_n \mathcal{S}^0 = \{s : \lambda_n^{-1} s \in \mathcal{S}^0\}$ for some fixed and compact set $\mathcal{S}^0 \subset \mathbb{R}^d$ and deterministic non-decreasing positive real sequence $\lambda_n$.*

*(b) The empirical cumulative distribution function $G_n$ of $\{\lambda_n^{-1} s_l\}_{l=1}^n \subset \mathcal{S}^0$ converges to $G$, $G_n(s) \to G(s)$ for all $s \in \mathcal{S}^0$, with $G$ an absolutely continuous distribution with support equal to $\mathcal{S}^0$.*

A familiar example helps clarify the sampling framework: consider a regularly spaced time series process observed at time periods $l = 1, ..., n$, so that $s_l = l$. In this example, the sampling region can be represented as $\mathcal{S}_n = [0, n]$, with a domain increasing at the rate $\lambda_n = n$. Thus, $\lambda_n^{-1} s_l = l/n$ and $\mathcal{S}^0 = [0, 1]$. The empirical distribution of the locations is $G_n(s) = n^{-1} \lfloor sn \rfloor \to s$ for $s \in [0, 1]$, so that $G$ is the uniform distribution. Condition 1 extends this familiar example to a general spatial setting with a general prototypical sampling region $\mathcal{S}^0 \subset \mathbb{R}^d$ that grows an arbitrary rate $\lambda_n$.

[3] This coincides with Lahiri's (2003) large-sample framework, except that he replaces Condition 1 with an assumption that the locations are i.i.d. draws from the distribution $G$.

## 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 $I(1)$ time series 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. In the following subsection we discuss 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

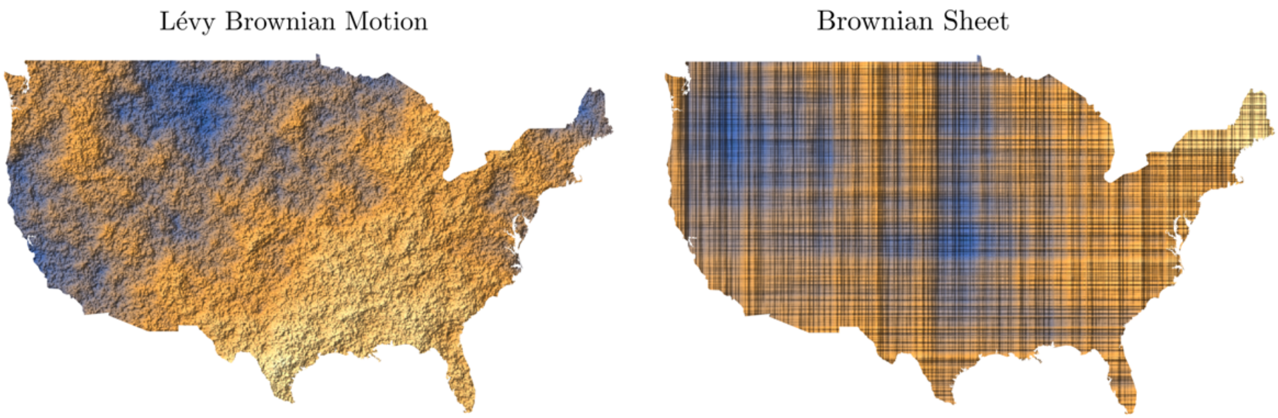
$$\mathbb{E}[L(s)L(r)] = \tfrac{1}{2}(|s| + |r| - |s - r|) \tag{2}$$

with $|x| = \sqrt{x'x}$ for $x \in \mathbb{R}^d$, 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 (2) 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 $O$. Moreover, like the Wiener process, Lévy-Brownian motion is *self-similar*, that is, $L(a\cdot) \sim a^{1/2} L(\cdot)$ for any scalar $a > 0$.

The left panel of Figure 2 plots a realization of $L$ on the sampling region representing the 48 contiguous U.S. States.[4] The right panel shows a realization of another generalization of the Wiener process to $d > 1$, the *Brownian sheet* $\int_{\mathbb{R}^d} \mathbf{1}[0 \leq r \leq s] dW(r)$, $s \geq 0$, where the inequality $0 \leq r \leq s$ is to be understood element by element. The Brownian sheet is not

[4] See Section S.1 in the supplemenatary appendix for details on the generation of Figures 2-4.

Figure 2: Sample Realizations of Stochastic Processes for $d = 2$




isotropic, as is clearly visible in the sample realizations in the figure. Because of this, we find Lévy-Brownian motion a more appealing generalization of the Wiener process for most applications and therefore use $y_l = L(s_l)$ as the canonical unit root process for $d > 1$.

*Two Representations of Lévy-Brownian motion*

We will utilize two representations for Lévy-Brownian motion, the Karhunen–Loève expansion and a spatial "moving average" representation. We discuss these in turn.

By Mercer's Theorem, the covariance kernel (2) evaluated at $s, r \in \mathcal{S}^0$ can be represented as

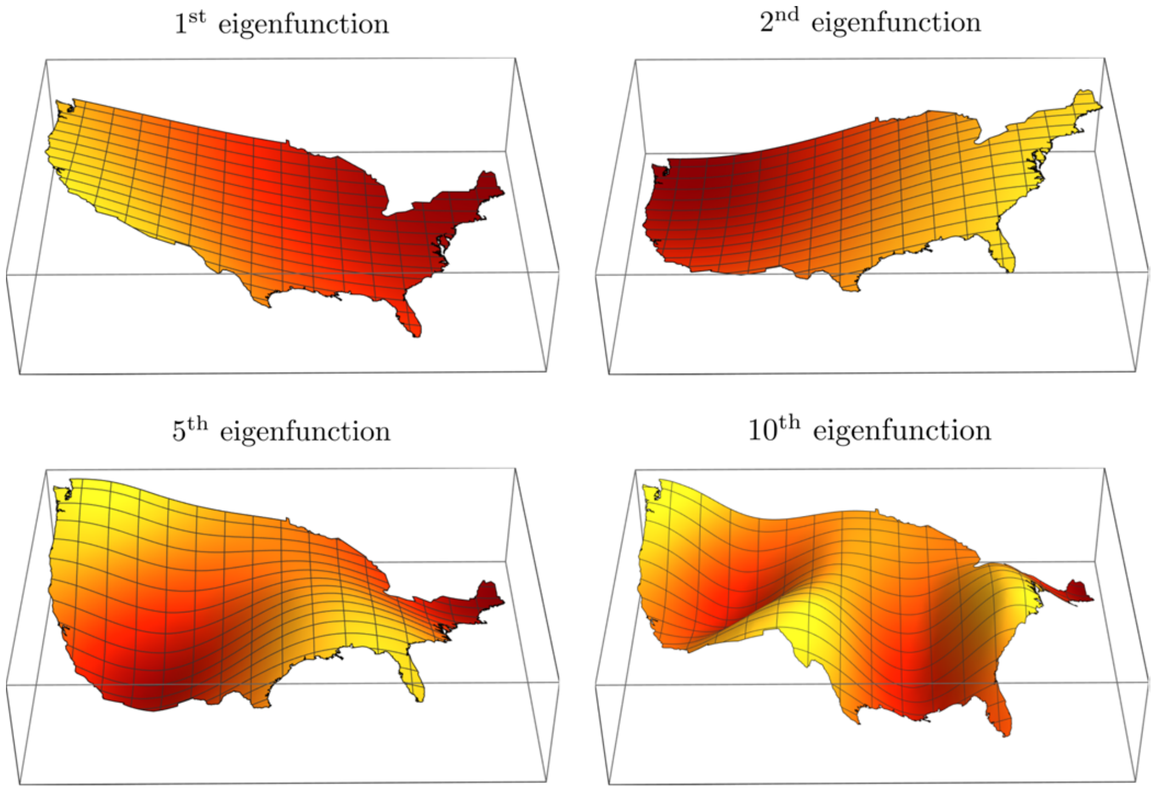
$$\mathbb{E}[L(s)L(r)] = \sum_{j=1}^{\infty} \nu_j \varphi_j(s)\varphi_j(r) \tag{3}$$

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

$$L(s) = \sum_{j=1}^{\infty} \nu_j^{1/2} \varphi_j(s)\xi_j,\ \xi_j \sim iid\mathcal{N}(0, 1) \tag{4}$$

where the right-hand side converges uniformly on $\mathcal{S}^0$ 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. Figure 3 plots some of the eigenfunctions $\varphi_j$ for $\mathcal{S}^0$ the contiguous

Figure 3: Eigenfunctions for Uniform Distribution on Continental U.S.




U.S. and $G$ the uniform distribution.

The spatial moving average representation represents Lévy-Brownian motion as a weighted average of spatial white noise. Recall that a Wiener process can trivially be written as an integral over white noise, $W(s) = \int_0^s dW(r)$. This can be generalized for Lévy-Brownian motion for all $d \geq 1$: from Lindstrøm (1993)

$$L(s) = \int h(r,s)dW(r) = \begin{cases} \int_0^s dW(r) \text{ for } d = 1 \\ \kappa_d \int_{\mathbb{R}^d} (|s-r|^{(1-d)/2} - |r|^{(1-d)/2})dW(r) \text{ for } d > 1 \end{cases} \tag{5}$$

where $\kappa_d > 0$ is a scalar chosen so that $\mathrm{Var}(L(s)) = 1$ when $|s| = 1$.

## 2.3 Spatial $I(1)$ Processes

The commuter-zone data plotted in panel (b) of Figure 1 are realizations of Lévy-Brownian motion evaluated at the zone centers, while the data plotted in panel (c) are variables from

Chetty, Hendren, Kline, and Saez (2014). To the naked eye, the long-range spatial correlation patterns in these figures are similar, suggesting that Lévy-Brownian motion may be a reasonable model for low-frequency correlation in economic and demographic spatial data. That said, the higher-frequency/short-range correlation patterns look different. In this section, we propose a generalization of Lévy-Brownian motion that inherits its long-range properties but allows for more flexible short-range correlation patterns. Following the notation used in time series, we call these (spatial) $I(1)$ processes.

In the standard time series case, $I(1)$ processes are defined as partial sums of a weakly dependent $I(0)$ process, say $u_t$, so that $y_t = \sum_{s=1}^{t} u_s$. Because spatial locations typically do not fall on a regular lattice, this definition does not naturally generalize. Instead, we take advantage of the moving average representation (5) and replace the white noise innovations $dW(r)$ by a weakly dependent random field $B$.

We thus define a *spatial* $I(1)$ *process* on $\mathcal{S}_n$ via

$$Y(s) = \int h(r, s)B(r)dr. \tag{6}$$

Note that if $B$ is isotropic, then so is $Y$.

In general, $B$ does not need to be Gaussian or isotropic, but we impose the following regularity condition.

**Condition 2.** *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 $f : \mathbb{R}^d \mapsto \mathbb{R}$,*

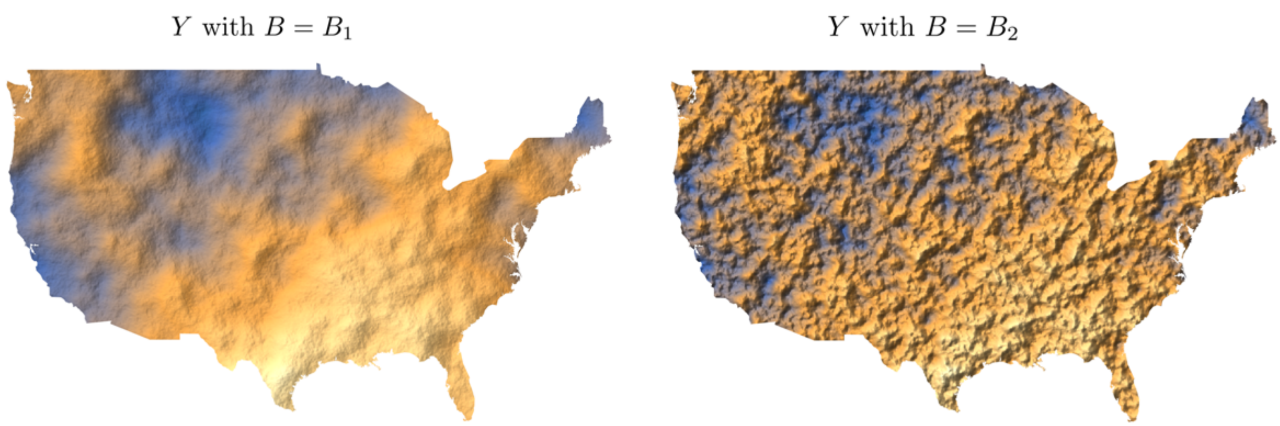
$$\mathbb{E}\left[\left(\int_{\mathbb{R}^d} f(r)B(r)dr\right)^{2m}\right] \leq C_m \left(\int_{\mathbb{R}^d} f(r)^2 dr\right)^{m}.$$

Lemma 1.8.4 of Ivanov and Leonenko (1989) implies that Condition 2 holds for a wide range of covariance stationary mixing random fields $B$.

Note that $\int_{\mathbb{R}^d} |h(r, s)|dr$ does not exist for $d > 1$, so $Y$ in (6) is not defined pathwise for every realization of $B$. However, $\int_{\mathbb{R}^d} h(r, s)^2 dr < \infty$, so under appropriate weak dependence conditions on $B$, the integral that defines $Y$ can be shown to converge in a mean square sense. In particular, we have the following result.

**Lemma 1.** *Under Condition 2, for all $d \geq 1$, $Y(\cdot)$ exists on $\mathcal{S}_n \subset \mathbb{R}^d$ for all $n$ and has*

Figure 4: Sample Realizations of $Y$ with Different Underlying $B$




Notes: $B_1$ and $B_2$ are zero mean Gaussian processes with spectral densities $f_1(\omega) \propto 1/(|\omega|^2+100^2)^{3/2}$ and $f_2(\omega) \propto (|\omega|^2+50^2)^{3/2}/(|\omega|^2+200^2)^3$ for $\omega \in \mathbb{R}^2$, respectively, where the width of the contiguous U.S. is normalized to unity.

*continuous sample paths with probability one.*

Figure 4 plots realization from two such $Y$ processes with $B$ equal to two different isotropic Gaussian processes. These realizations were generated using the same underlying normal variables as the Lévy-Brownian motion plotted in Figure 2. As demonstrated in Figure 4, different $B$ processes can induce quite different local behavior of $Y$, but with the same long-range behavior as Lévy-Brownian motion, a result formalized in the next subsection.

## 2.4 A Functional Central Limit Theorem

In the standard time series case, 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 $Y(\cdot)$ in (6).

The classic time series FCLT involves two rescalings: one that maps time into the unit interval, and one that shrinks the scale of $y_t$ to compensate for its increasing variance. For the spatial $I(1)$ process in (6) we similarly define the process $Y_n^0(\cdot)$ on $\mathcal{S}^0$ via

$$Y_n^0(r) = \lambda_n^{-1/2} Y(\lambda_n r),\ r \in \mathcal{S}^0. \tag{7}$$

We make the following assumption about the process $B$.

**Condition 3.** *For some positive sequence $\zeta_n \to \infty$, let $\mathcal{R}_n = [-\zeta_n, \zeta_n]^d \subset \mathbb{R}^d$, and let $f_n : \mathbb{R}^d \mapsto \mathbb{R}$ be any sequence of functions such that $\limsup_{n\to\infty} \sup_{r\in\mathcal{R}_n} \lambda_n^{d/2}|f_n(r)| < \infty$ and $\mathrm{Var}[\int_{\mathcal{R}_n} f_n(r)B(r)dr] \to \sigma_0^2$. Then $\int_{\mathcal{R}_n} f_n(r)B(r)dr \Rightarrow \mathcal{N}(0, \sigma_0^2)$.*

The central limit theorems in Section 1.7 of Ivanov and Leonenko (1989) provide primitive mixing and moment conditions on $B$ that imply Condition 3.

**Theorem 2.** *Suppose Conditions 2 and 3 hold. If $\lambda_n \to \infty$, then $Y_n^0(\cdot) \Rightarrow \omega L(\cdot)$ on $\mathcal{S}^0$, where $\omega^2 = \int_{\mathbb{R}^d} \sigma_B(r)dr$.*

**Remark 2.1.** Under Condition 1, the rate $\lambda_n$ governs the degree of "infill" versus "outfill" sampling. To see this, note that $\mathcal{S}_n = \lambda_n \mathcal{S}^0$ implies that the volume of $\mathcal{S}_n$ is $\mathrm{vol}(\mathcal{S}_n) = \lambda_n^d \,\mathrm{vol}(\mathcal{S}^0)$ so the average number of observations per unit of volume is proportional to $n/\lambda_n^d$. If $\lambda_n$ is constant this corresponds to pure infill sampling of the fixed sampling domain. The assumption that $\lambda_n \to \infty$ rules out pure infill sampling. In contrast, when $\lambda_n \propto n^{1/d}$, the average number of observations per unit of volume is unchanged as $n$ grows and this corresponds to pure outfill sampling. Finally, when $\lambda_n \to \infty$ with $\lambda_n = o(n^{-1/d})$, the sampling domain is increasing as is the average number of observations per unit of volume, which yields a mix of infill and outfill sampling. The theorem requires $\lambda_n \to \infty$ and thus requires some degree of outfill sampling. Recall that for the standard time series example in Section 2.1, $\mathcal{S}^0 = [0, 1]$, $\lambda_n = n$, and $\mathcal{S}_n = [0, n]$, so $\lambda_n \to \infty$ corresponds to the usual large-$n$ increasing domain requirement.

**Remark 2.2.** In practice, Theorem 2 can be used to argue that the distribution of a functional $\psi(Y_n^0)$ is well approximated by the distribution of $\psi(\omega L)$. Note, however, that this requires $\psi$ 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: For example, the distance between two typical neighboring locations $\lambda_n^{-1}s_l$, $\lambda_n^{-1}s_\ell \in \mathcal{S}^0$ is $O(n^{-1/d})$. For $d > 1$, the difference in the value of $Y_n^0(\cdot)$ evaluated at two such neighboring points $s$ and $s + n^{-1/d}a$, $a \in \mathbb{R}^d$ is given by

$$Y_n^0(s + n^{-1/d}a) - Y_n^0(s) \;=\; \lambda_n^{-1/2} \int_{\mathbb{R}^d} (h(r, \lambda_n(s + n^{-1/d}a)) - h(r, \lambda_n s))B(r)dr$$

$$= \lambda_n^{-1/2} \int_{\mathbb{R}^d} (|\lambda_n n^{-1/d} a - r|^{(1-d)/2} - |r|^{(1-d)/2}) B(\lambda_n s + r) dr. \quad (8)$$

Even under pure outfill sampling, $\lambda_n n^{-1/d}$ does not diverge. The weighting of $B(\lambda_n s + r)$ in (8) thus puts most of its (square integrable) weight on small values $r$, and the (suitably scaled) difference $Y_n^0(s + n^{-1/d}a) - Y_n^0(s) = \lambda_n^{-1/2}(Y(\lambda_n s + \lambda_n n^{-1/d}a) - Y(\lambda_n s))$ *does not* become Gaussian as $n \to \infty$, just as in the time series case.

**Remark 2.3.** It is well known that suitably scaled partial sums over rectangles of random variables defined on a lattice converge to a Brownian Sheet under suitable mixing and moment conditions; see, for instance, Deo (1975). In contrast, we are not aware of previous results about convergence to Lévy-Brownian motion.

## 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, the time series $y_t$ 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 $\mathbb{E}[J_c(s)J_c(r)] = \exp[-c|s-r|]/(2c)$, $c > 0$. Taking the limit of this covariance kernel shows that $J_c(\cdot) - J_c(0)$ converges to a Wiener process as $c \to 0$ (see Elliott (1999)). 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 proportional to $(|\omega|^2 + c^2)^{-(d+1)/2}$, $\omega \in \mathbb{R}^d$. As in the $d = 1$ model, $J_c(\cdot) - J_c(0)$ converges to $L(\cdot)$ as $c \to 0$ for any integer $d$. Also, along any line $J_c(a + bs)$, $a, b \in \mathbb{R}^d$, $|b| = 1$, $s \in \mathbb{R}$ is a standard OU process.

From equation of (3.2.8) of Matérn (1986), $J_c$ has the moving average representation

$$J_c(s) = \int_{\mathbb{R}^d} h_c(r, s) dW(r) \quad (9)$$

with $h_c(r, s) = \kappa_{c,d}|s - r|^{(1-d)/4} K_{(1-d)/4}(c|s - r|)$ for a suitable choice of constant, where $K_\nu$ is the modified Bessel function of the second kind, $d \geq 1$.[5] We proceed as in the spatial $I(1)$

[5] For $d = 1$, the usual one-sided (causal) representation for a stationary OU process is $J_c(s) = \int_{-\infty}^{s} e^{-c(s-r)} dW(r)$. Equation (9) is an alternative two-sided (non-causal) representation when $d = 1$.

model (6) and replace the white noise term by the weakly dependent random field $B$,

$$Y_c(s) = \int_{\mathbb{R}^d} h_c(r,s)B(r)dr,$$

and define the *spatial local-to-unity process* on $\mathcal{S}_n$ as the sequence of processes $Y_{c/\lambda_n}$. In this definition, the parameter $c/\lambda_n$ is a drifting sequence, generalizing the corresponding local-to-unity time series device. The rate of this drift is such that the overall degree of mean reversion of

$$Y^0_{n,c}(r) = \lambda_n^{-1/2} Y_{c/\lambda_n}(\lambda_n r),\ r \in \mathcal{S}^0 \tag{10}$$

over the fixed set $\mathcal{S}^0$ converges as $n \to \infty$.

The appendix shows that under Condition 2, $Y_{c/\lambda_n}$ exists on $\mathcal{S}_n$ for all $n$. Furthermore, under the conditions of Theorem 2, $Y^0_{n,c}$ in (10) satisfies

$$Y^0_{n,c}(\cdot) \Rightarrow \omega J_c(\cdot). \tag{11}$$

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

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

$$y_l = \alpha + x_l'\beta + u_l \tag{12}$$

for $l = 1, ..., n$, where $(y_l, x_l) = (Y(s_l), X(s_l)) \in \mathbb{R}^{p+1}$ follow $p+1$ 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 $x_l$. Let $s_{\tilde{y}\tilde{y}} = n^{-1}\sum_{l=1}^n \tilde{y}_l^2$, $S_{\tilde{x}\tilde{x}} = n^{-1}\sum_{l=1}^n \tilde{x}_l\tilde{x}_l'$ and $S_{\tilde{x}\tilde{y}} = n^{-1}\sum_{l=1}^n \tilde{x}_l\tilde{y}_l$. The OLS estimator is $\hat{\beta} = S_{\tilde{x}\tilde{x}}^{-1}S_{\tilde{x}\tilde{y}}$, the regression $R^2 = S'_{\tilde{x}\tilde{y}}S_{\tilde{x}\tilde{x}}^{-1}S_{\tilde{x}\tilde{y}}/s_{\tilde{y}\tilde{y}}$, the OLS estimator for the variance of $u_l$ is $s_u^2 = \frac{n}{n-p-1}(s_{\tilde{y}\tilde{y}} - S'_{\tilde{x}\tilde{y}}S_{\tilde{x}\tilde{x}}^{-1}S_{\tilde{x}\tilde{y}})$, and the classical (non-spatial-correlation robust, homoskedastic) F-statistic for testing $H_0 : H\beta = 0$, where $H$ is a non-stochastic matrix with rank$(H) = m \leq p$, is $F^{\mathrm{Hom}} = \frac{n}{m}\hat{\beta}'H'(H'S_{\tilde{x}\tilde{x}}^{-1}H)^{-1}H\hat{\beta}/s_u^2$.

Suppose $(y_l, x_l) = (Y(s_l), X(s_l))$ follow spatial $I(1)$ processes with

$$\begin{bmatrix} Y_n^0(\cdot) \\ X_n^0(\cdot) \end{bmatrix} = \begin{bmatrix} \lambda_n^{-1/2} Y(\lambda_n \cdot) \\ \lambda_n^{-1/2} X(\lambda_n \cdot) \end{bmatrix} \Rightarrow \begin{bmatrix} Y^0(\cdot) \\ X^0(\cdot) \end{bmatrix} \tag{13}$$

on $\mathcal{S}^0$, where $[Y^0(\cdot), X^0(\cdot)']$ are $p+1$ independent and arbitrarily scaled Lévy-Brownian motions. Let $\tilde{Y}(\cdot) = Y^0(\cdot) - \int Y^0(r)dG(r)$ denote the demeaned version of $Y^0$ using spatial-weighted average demeaning, and define $\tilde{X}$ analogously.

**Theorem 3.** *Under Condition 1 and (13)*

*(i)* $\lambda_n^{-1} s_{\tilde{y}\tilde{y}} \Rightarrow \Xi_{\tilde{y}\tilde{y}} = \int \tilde{Y}^2(r)dG(r)$, $\lambda_n^{-1} S_{\tilde{x}\tilde{x}} \Rightarrow \Xi_{\tilde{x}\tilde{x}} = \int \tilde{X}(r)\tilde{X}(r)'dG(r)$ *and* $\lambda_n^{-1} S_{\tilde{x}\tilde{y}} \Rightarrow \Xi_{\tilde{x}\tilde{y}} = \int \tilde{X}(r)\tilde{Y}(r)dG(r)$,

*(ii)* $\hat{\beta} \Rightarrow \Xi_{\tilde{x}\tilde{x}}^{-1}\Xi_{\tilde{x}\tilde{y}}$,

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

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

*(v)* $n^{-1}F^{\text{Hom}} \Rightarrow m^{-1}\Xi_{\tilde{x}\tilde{y}}'\Xi_{\tilde{x}\tilde{x}}^{-1}H'(H\Xi_{\tilde{x}\tilde{x}}^{-1}H')^{-1}H\Xi_{\tilde{x}\tilde{x}}^{-1}\Xi_{\tilde{x}\tilde{y}}/(\Xi_{\tilde{y}\tilde{y}} - \Xi_{\tilde{x}\tilde{y}}'\Xi_{\tilde{x}\tilde{x}}^{-1}\Xi_{\tilde{x}\tilde{y}})$.

**Remark 3.1.** In the one-dimensional case with $d = 1$, $\mathcal{S}^0$ 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 $\mathcal{S}^0$. Section 6 provides numerical results for the behavior of $R^2$ for a range 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 $\mathbb{P}(F^{\text{Hom}} > \text{cv}) \to 1$ for any $\text{cv} \geq 0$.

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

$$F^{\text{HAC}} = \frac{n}{m}\hat{\beta}'H'(HS_{\tilde{x}\tilde{x}}^{-1}\hat{\Omega}_n S_{\tilde{x}\tilde{x}}^{-1}H')^{-1}H\hat{\beta} \tag{14}$$

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

$$\hat{\Omega}_n = n^{-1}\sum_{l,\ell=1}^n \kappa\left(b_n(s_l - s_\ell)\right) e_l e_\ell' \tag{15}$$

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

$$\sup_r |\kappa(r)| = \bar{\kappa} < \infty, \qquad \lim_{\lambda\to\infty} \sup_{|a|=1} |\kappa(\lambda a)| = 0. \tag{16}$$

The assumption of $\lambda_n^{-1}b_n^{-1} = o_p(1)$ ensures that in large samples, $\hat{\Omega}_n$ in (15) puts negligible weight on pairs of locations with $\lambda_n^{-1}|s_l - s_\ell| > \varepsilon$, for all positive $\varepsilon$. Since $\lambda_n^{-1}s_l \in \mathcal{S}^0$ with $\mathcal{S}^0$ 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). As long as all locations are distinct, heteroskedasticity robust standard errors correspond to $\kappa(r) = \mathbf{1}[r = 0]$, which also satisfies (16).

Alternatively, researchers sometimes employ clustered standard errors over larger regions to account for spatial dependence. The corresponding $F^{\text{clust}}$ statistic has the same form as $F^{\text{HAC}}$ in (14) with $\hat{\Omega}_n$ replaced by

$$\hat{\Omega}_n^{\text{clust}} = n^{-1} \sum_{j=1}^{n_C} \left( \sum_{l\in C_j} e_l \right) \left( \sum_{l\in C_j} e_l \right)'$$

where the partitions $C_j$ of $\{1, 2, \ldots, n\}$ indicate membership in cluster $j = 1, \ldots, n_C$. With $|C_j|$ the number of observations in cluster $j$, we assume $\max_{1\le j\le n_C} |C_j|/n \to 0$ as $n \to \infty$. As discussed in Hansen and Lee (2019), page 270, this is necessary for the consistency of cluster robust inference under weak dependence.

The following result shows that inference using $F^{\text{HAC}}$ or $F^{\text{clust}}$ does not avoid spurious significance with independent spatial $I(1)$ variables.

**Theorem 4.** *Under Condition 1, (13) and (16),* $\mathbb{P}(F^{\text{HAC}} > \text{cv}) \to 1$ *and* $\mathbb{P}(F^{\text{clust}} > \text{cv}) \to 1$ *for any* $\text{cv} \ge 0$.

**Remark 3.3.** In contrast to spatial HAC inference, 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 6 below for corresponding numerical results.

**Remark 3.4.** Theorems 3 and 4 also hold for local-to-unity processes, that is, if $[Y^0(\cdot), X^0(\cdot)]$ in (13) are $p+1$ independent processes of the type (9), 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). 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 3.5.** It follows from the Karhunen–Loève representation of $L$ in (4) and the FCLT result in Theorem 2 that the coefficients of a regressions of $\lambda_n^{-1/2} y_l$ on the eigenfunctions $[\varphi_1(\lambda_n^{-1}s_l), \ldots, \varphi_p(\lambda_n^{-1}s_l)]$ 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 $\lambda_n^{-1}s_l$, say $\psi(\lambda_n^{-1}s_l) \in \mathbb{R}^p$, converge to $\left(\int \psi(r)\psi(r)' dG(r)\right)^{-1} \omega \int \psi(r)L(r)dG(r)$ and are asymptotically significant as measured by a corresponding $F^{\text{Hom}}$, $F^{\text{HAC}}$ or $F^{\text{clust}}$ statistic. Kelly (2019) observes such a phenomenon empirically in a number of applications with spatial data.

# 4 Isotropic Differences Regression

A natural approach to learning about the spatial dependence in a univariate data set is via regression analysis. In the standard time series case, this corresponds to autoregressions, where the current value is regressed on past values. For strongly dependent data, this is typically implemented using Dickey and Fuller (1979) regressions, where the dependent variable is the first difference of the series, and the regressor its lagged value.

The spatial case is interestingly different, since for $d > 1$, there is no natural ordering of the locations. Instead, one may consider regressions that are "isotropic" in the sense that all directions are treated symmetrically. In particular, consider the transformation

$$y_l^* = \frac{1}{n} \sum_{\ell \neq l} \kappa_b(\lambda_n^{-1}|s_\ell - s_l|)(y_\ell - y_l) \tag{17}$$

for some bounded weighting function $\kappa_b : \mathbb{R} \mapsto \mathbb{R}$ with $\kappa_b(x) = \kappa_0(x/b)$ for $b > 0$ and $\kappa_0(x) = 0$ for $|x| > 1$. We call this transformation "isotropic differencing," as (17) averages over all differences that are within a ball of radius $\lambda_n b$ around $s_l$, so it is invariant to rotations of the data. The normalization by $\lambda_n^{-1}$ exactly counteracts the growth in the sampling region $\mathcal{S}_n$, so this is a spatial version of a "fixed-$b$" (cf. Kiefer and Vogelsang (2005)) kernel, and the larger the bandwidth $b$, the more averaging is being employed.

Consider a regression of $y_l^*$ on $y_l$—this roughly corresponds to the time series regression of $\Delta y_t$ on $y_{t-1}$, except that the difference is computed symmetrically and over a positive fraction of the sample size. In order to avoid border effects, we only consider locations in the regression of $y_l^*$ on $y_l$ where $\lambda_n^{-1} s_l$ is at least a distance $b$ of the boundary $\partial\mathcal{S}^0$ of $\mathcal{S}^0$. Technically, let $\mathcal{I}_b = \{s \in \mathcal{S}^0 : d(s, \partial\mathcal{S}^0) \geq b\}$ be the corresponding interior of $\mathcal{S}^0$.

**Theorem 5.** *Let $\hat{\gamma}$ be the coefficient of an OLS regression of $y_l^*$ in (17) on $y_l$ for all $l$ such that $\lambda_n^{-1} s_l \in \mathcal{I}_b$. Suppose $y_l$ is a spatial local-to-unity process satisfying (11), $\mathcal{I}_b$ has positive volume, $\kappa_0$ has a finite number of discontinuity points and Condition 1 holds. Then*

$$\hat{\gamma} \Rightarrow \frac{\int_{\mathcal{I}_b} J_c(s) \int \kappa_b(|r-s|)(J_c(r) - J_c(s))dG(r)dG(s)}{\int_{\mathcal{I}_b} J_c(s)^2 dG(s)}. \tag{18}$$

**Remark 4.1.** Due to the long-range nature of the differences (17) with $b$ fixed, the long-run variance $\omega^2$ cancels in the limiting expression for $\hat{\gamma}$. Thus, one could use $\hat{\gamma}$ to learn about the degree of mean reversion $c$, analogous to the suggestion in Stock (1991) for the standard time series case. In order to simulate corresponding critical values, one could replace $G$ by the empirical distribution $G_n$ in the r.h.s. of (18), which is asymptotically justified (see the proof of Theorem 5).

**Remark 4.2.** If a constant is included in the regression, then the same result holds with $J_c$ replaced by $\tilde{J}_c$ with $\tilde{J}_c(s) = J_c(s) - \int_{\mathcal{I}_b} J_c(r)dG(r)$. Also, for a spatial $I(1)$ process $y_l$, (18) holds with $J_c$ replaced by $L$.

**Remark 4.3.** If the isotropic differences are computed with weights that are normalized to sum to one, $y_l^* = \sum_{\ell \neq l} \kappa_b(\lambda_n^{-1}|s_l - s_\ell|)(y_l - y_\ell) / \sum_{\ell \neq l} \kappa_b(\lambda_n^{-1}|s_l - s_\ell|)$, then (18) holds with $\kappa_b(|r-s|)$ replaced by $\kappa_b(|r-s|)/\int \kappa_b(|u-s|)dG(u)$.

By a change of variables, the numerator in (18) may be rewritten as

$$b^d \int_{\mathcal{I}_b} J_c(s) \int_{|r|\leq 1} \kappa_0(|r|)(J_c(s+br) - J_c(s))g(s+br)drdG(s) \tag{19}$$

where $g$ is the density of the distribution $G$. As the bandwidth $b$ shrinks, (19) becomes smaller for two reasons: the leading terms $b^d$ shrinks and for each $r$, the variance of $J_c(s+br)-J_c(s)$ shrinks. This suggests that the numerator in (18) has little variability as $b \to 0$. The following result formalizes this intuition and establishes the limiting constant.

**Theorem 6.** *Suppose the density $g$ of $G$ admits three bounded derivatives on $\mathcal{S}^0$. Then under the assumptions of Theorem 5, as $b \to 0$,*

$$b^{-d-1}\int_{\mathcal{I}_b} J_c(s)\int \kappa_b(|r-s|)(J_c(r)-J_c(s))dG(r)dG(s) \xrightarrow{p} -\tfrac{1}{2}\int_{|r|\leq 1}|r|\kappa_0(|r|)dr\cdot\int g(s)^2 ds. \tag{20}$$

*Furthermore, (20) continues to hold with $J_c(s)$ replaced by $J_c(s)-\hat{m}$ for any random variable $\hat{m}$ with $\mathbb{E}[\hat{m}^2]<\infty$, and also for $J_c$ replaced by $L$.*

**Remark 4.4.** For intuition about the limit in Theorem 6, consider a time series random walk $y_t=\sum_{s=1}^t \varepsilon_t$ with $\varepsilon_t \sim iid(0,1)$. As is well known, $n^{-1}\sum_{t=2}^{n-1}\Delta y_{t+1}y_t \Rightarrow \int_0^1 W(s)dW(s)$. Also, $n^{-1}\sum_{t=2}^{n-1}(y_{t-1}-y_t)y_t = n^{-1}\sum_{t=2}^{n-1}(-\Delta y_t)(y_{t-1}+\Delta y_t) \Rightarrow \int_0^1 W(s)dW(s)-1$. Thus, $n^{-1}\sum_{t=2}^{n-1}(\Delta y_{t+1}+(y_{t-1}-y_t))y_t \xrightarrow{p} -1$. Treating the forward and backward difference as each receiving unit weight for a total of 2, this result accords with (20). Symmetric time series autoregressive estimators of this type are studied in Pantula, Gonzalez-Farias, and Fuller (1994) and Fuller (1996); see, for instance, Theorems 10.1.7 and 10.1.8 in the latter.

**Remark 4.5.** As $b \to 0$, the denominator in (18) converges to $\int J_c(s)^2 dG(s)$. Suppose we pick a weight function $\kappa_0$ such that $\int_{|r|\leq 1}|r|\kappa_0(|r|)dr\cdot\int g(s)^2ds = 1$. Taking $b \to 0$ limits after $n \to \infty$ limits yields a limiting distribution of $-b^{-d-1}\hat{\gamma}$ equal to $\frac{1}{2}/\int J_c(s)^2dG(s)$. Because the distribution of $\frac{1}{2}/\int J_c(s)^2dG(s)$ depends on the value $c$, this suggests that $-b^{-d-1}\hat{\gamma}$ may be used to estimate the degree of mean reversion $c$ of the local-to-unity spatial process $y_l$. This is analogous to using $n(1-\hat{\rho})$, with $\hat{\rho}$ the estimator of the largest autoregressive root in the standard time series case. The limiting distribution of $-b^{-d-1}\hat{\gamma}$ never takes on negative values, again mirroring the corresponding results for the symmetric autoregressive time series estimator $\hat{\rho}$ of Pantula, Gonzalez-Farias, and Fuller (1994) and Fuller (1996).

# 5 Inference for Spatial Persistence

This section develops inference methods for the degree of spatial persistence.[6] As discussed in the last section, regressions can be used for this purpose. An alternative, non-regression based approach to learn about time series persistence is developed in Müller and Watson (2008). That approach is based on the properties of $q$ suitably chosen weighted averages, and generalizes fairly directly to the spatial setting studied here.

The intuition is as follows: The Karhunen–Loève expansion (4) 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 i.i.d. normal random variables because of the orthogonality of the eigenfunctions. This difference in behavior may be used to empirically distinguish between these two canonical cases. What is more, the FCLT result in Theorem 2 and the CLT in Lahiri (2003) implies that these tests are also asymptotically valid under more general forms of spatial $I(0)$ and $I(1)$ processes. The remainder of this section expands on this intuition to develop tests for the degree of spatial persistence.

## 5.1 Local Alternatives and Rescaling of Locations

In a standard time series unit root test, autoregressive local alternatives are of the local-to-unity type, that is, the alternative autoregressive parameter is a sequence that converges to unity. This device ensures that asymptotic power is non-trivial. For a spatial unit root process with $Y_n^0(\cdot) \Rightarrow \omega L(\cdot)$ on $\mathcal{S}^0$, the same holds true: the natural local alternative is a local-to-unity process satisfying $Y_n^0(\cdot) \Rightarrow \omega J_c(\cdot)$ as introduced in Section 2.5. Since $y_l = \lambda_n^{1/2} Y_n^0(\lambda_n^{-1} s_l)$, the canonical alternative is thus of the form $y_l = \lambda_n^{1/2} J_c(\lambda_n^{-1} s_l)$. By self-similiarity, $\{\lambda_n^{1/2} J_c(\lambda_n^{-1} s_l)\}_{l=1}^n \sim \{J_{c/\lambda_n}(s_l)\}_{l=1}^n$. Instead of characterizing the alternatives in terms of the drifting sequence $c/\lambda_n$, it is more convenient to instead rescale the locations: Let $s_l^0 = \lambda_n^{-1} s_l$, where we suppress the dependence on $n$ for notational convenience. The canonical unit root testing problem then simply becomes testing $y_l = \lambda_n^{1/2} L(s_l^0)$ versus $y_l = \lambda_n^{1/2} J_c(s_l^0)$.

[6] Other 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.

The representation proposed in the preceding paragraph depends on the scale factor $\lambda_n$. The inference methods developed in this section are designed to be invariant to the choice of $\lambda_n$. Notice that the choice of $\lambda_n$ affects the representations in two ways. The first is a simple scaling of $y_l$ by $\lambda_n^{1/2}$. We eliminate this effect by focussing on scale-invariant tests. The second effect is more subtle: the value of $c$ and the scale of $\mathcal{S}^0$ are not separately identified: doubling $\mathcal{S}^0$ and the $s_l^0$'s and halving $c$ yields the same distribution for $\{J_c(s_l^0)\}_{l=1}^n \sim \{J_{c/2}(2s_l^0)\}_{l=1}^n$. Thus, $c$ must be interpreted relative to the scale of $\mathcal{S}^0$. All of our choices for $c$ below are (implicitly) self-normalizing in that sense. Thus, while the scale factor $\lambda_n$ is useful for the asymptotic analysis, its value makes no difference in practice and all tests developed below are invariant to using the scaled locations $\{s_l^0\}_{l=1}^n$ or the original locations $\{s_l\}_{l=1}^n$.

## 5.2 Dimension Reduction by Weighted Averages

Let $\mathbf{Y}_n = (y_1, \ldots, y_n)'$ and let $\mathbf{\Sigma}_{n,L}$ be the $n \times n$ covariance matrix of $\mathbf{Y}_n$ induced by Lévy-Brownian motion $y_l = L(s_l^0)$. 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 $\mathbf{M}\mathbf{\Sigma}_{n,L}\mathbf{M}$ corresponding to the $q$ largest eigenvalues, where $\mathbf{R}_n$ satisfies $n^{-1}\mathbf{R}_n'\mathbf{R}_n = \mathbf{I}_q$. If $\mathbf{Y}_n \sim (\mathbf{0}, \mathbf{\Sigma}_{n,L})$, the columns of $\mathbf{R}_n$ extract the $q$ linear combinations of $\mathbf{M}\mathbf{Y}_n$ 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 $\text{Var}(\mathbf{Z}_n) = \mathbf{\Omega}_n$. Consider first the generic problem of testing $H_0 : \mathbf{\Omega} = \mathbf{\Omega}_0$ versus $H_a : \mathbf{\Omega} = \mathbf{\Omega}_a$ when $\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$. A standard calculation shows that the most powerful level $\alpha$ scale invariant test rejects for large values of

$$\frac{\mathbf{Z}_n'\mathbf{\Omega}_0^{-1}\mathbf{Z}_n}{\mathbf{Z}_n'\mathbf{\Omega}_a^{-1}\mathbf{Z}_n} \tag{21}$$

with a critical value that equals the $1 - \alpha$ quantile of (21) 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 exploiting implications of the specific models of persistence 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 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 to the eigenfunctions of the covariance kernel of demeaned Lévy-Brownian motion (2) given by

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

where $k(r,s) = \frac{1}{2}(|s| + |r| - |s-r|)$. Let the spectral decomposition of $\bar{k}(r,s)$ be $\bar{k}(s,r) = \sum_{i=1}^{\infty} \bar{\nu}_i \bar{\varphi}_i(s)\bar{\varphi}_i(r)$, 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

$$\hat{k}_n(r,s) = k(r,s) - n^{-1}\sum_{l=1}^{n} k(s_l^0, s) - n^{-1}\sum_{\ell=1}^{n} k(r, s_\ell^0) + n^{-2}\sum_{l=1}^{n}\sum_{\ell=1}^{n} k(s_l^0, s_\ell^0)$$

and the $n \times n$ matrix $\hat{\mathbf{K}}_n$ with $l, \ell$ element equal to $\hat{k}_n(s_l^0, s_\ell^0)$ 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 $\mathcal{S}^0 \mapsto \mathbb{R}$ functions

$$\hat{\varphi}_i(\cdot) = n^{-1}\hat{\nu}_i^{-1}\sum_{l=1}^{n} r_{i,l}\hat{k}_n(\cdot, s_l^0). \tag{22}$$

Lemma 6 of Müller and Watson (2022a), building of the work of Rosasco, Belkin, and Vito (2010), shows that if the locations $s_l^0$ are i.i.d. with distribution $G$ and $\bar{\nu}_1 > \bar{\nu}_2 > \ldots > \bar{\nu}_q$, then $(\hat{\nu}_i, \hat{\varphi}_i)$ 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 $s_l^0$, but rather assumes that the non-stochastic sequence of locations $\{s_l^0\}_{l=1}^n$ has an empirical distribution $G_n$ that converges to $G$, as in Condition 1. This assumption holds for almost all realizations of $\{s_l^0\}_{i=1}^n$ if $s_l^0 \sim G$ is i.i.d. by the Glivenko-Cantelli Theorem, so in this sense, the following result is more general, albeit at the cost of not providing convergence rates.

**Lemma 7.** *Suppose $\bar{\nu}_1 > \bar{\nu}_2 > \ldots > \bar{\nu}_q > \bar{\nu}_{q+1}$ and Condition 1 holds. Then for any $q \geq 1$, $\sup_{s\in\mathcal{S}^0, 1\leq i\leq q} |\hat{\varphi}_i(s) - \bar{\varphi}(s)| \to 0$ and $\max_{1\leq i\leq q} |\hat{\nu}_i - \bar{\nu}_i| \to 0$.*

## 5.3 Tests of the $I(1)$ Null Hypothesis

With this background in place, consider the problem of testing the $I(1)$ null hypothesis against the local-to-unity alternative. The canonical forms of these models are $y_l = L(s_l^0)$ and $y_l = J_c(s_l^0)$. 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 $l, \ell$ element of $\mathbf{\Sigma}_n(c)$ equal to $\exp[-c|s_l^0 - s_\ell^0|]/(2c)$. Thus, optimal tests in this problem are of the form (21) 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

$$\text{LFUR}_n = \frac{\mathbf{Z}_n' \mathbf{\Omega}_{n,L}^{-1} \mathbf{Z}_n}{\mathbf{Z}_n' \mathbf{\Omega}_n^{-1}(c_a) \mathbf{Z}_n}, \tag{23}$$

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 $\rho$, 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. This choice of $c_a$ makes the test invariant to the scale factor $\lambda_n$, as previewed in Section 5.1.

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 7 and the continuous mapping theorem (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 that satisfy the conditions of Theorem 2, as well as for the canonical model with $\omega^2 = 1$. Since the test (21) is scale invariant, the scale parameters $\lambda_n^{-1/2} n^{-1}$ and $\omega^2$ vanish, and the critical value computed from the canonical model converges to the asymptotically correct critical value for generic $I(1)$ processes.

## 5.4 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. This requires a definition of "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 between $y_l$ and $y_\ell$,

$$\bar{\rho}(c) = \frac{1}{n(n-1)} \sum_{l \neq \ell} \exp[-c|s_l^0 - s_\ell^0|], \tag{24}$$

of no more than 0.03. Denote the corresponding cut-off value of $c$ by $c_{0.03}$, that is, $\bar{\rho}(c_{0.03}) = 0.03$. The canonical version of the testing problem then becomes $H_0 : \mathbf{\Omega} = \mathbf{\Omega}_n(c)$, $c \geq c_{0.03}$ against $H_a : \mathbf{\Omega} = \mathbf{\Omega}_n(c) + g_a^2 \mathbf{\Omega}_{n,L}$, $g_a > 0$. 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 (21) that has a 95% quantile that is fairly constant as a function of $c \geq c_{0.03}$. Thus, the stationary test rejects if

$$\text{LFST}_n = \frac{\mathbf{Z}_n' \mathbf{\Omega}_n(c_{0.001})^{-1} \mathbf{Z}_n}{\mathbf{Z}_n' [\mathbf{\Omega}_n(c_{0.001}) + g_a^2 \mathbf{\Omega}_{n,L}]^{-1} \mathbf{Z}_n} \tag{25}$$

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 $\alpha$ is the size of the test and the probability is computed under $\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_n(c))$. 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 $\text{LFUR}_n$ test, this stationarity test remains valid in large samples under the general local-to-unity model (10) with $c \geq c_{0.03}$. A more subtle question asks whether it also remains valid under generic weak dependence, defined as $y_l = B(s_l) = B(\lambda_n s_l^0)$, 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 $s_l^0 \sim G$ 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 standard deviation. Still, the $\text{LFST}_n$ test 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 different possible limits identified by Lahiri's (2003) CLT. This result is summarized in the following theorem.

**Theorem 8.** *If $y_l = B(\lambda_n s_l^0)$ and $\lambda_n \to \infty$ with $\lambda_n^d/n \to a \in [0, \infty)$, then under the assumptions of Lahiri's CLT in his Theorem 3.2,* $\limsup_{n\to\infty} \mathbb{P}(\text{LFST}_n \geq \text{cv}_n^{\text{LFST}}) \leq \alpha$.

**Remark 5.1.** Suppose the $p \times 1$ vector $x_l$ is spatially cointegrated of order one with cointegrating vector $\beta_0$, that is, $\beta_0' x_l \sim I(0)$, but $\beta' x_l \sim I(1)$ for all $\beta$ that are not proportional to $\beta_0$. An asymptotic level $1 - \alpha$ confidence set for $\beta_0$ can then be formed by collecting those values of $b$ for which the level-$\alpha$ $\text{LFST}_n$ test does not reject the $I(0)$ null hypothesis when applied to the series $b' x_l$. This is the spatial analogue of Wright's (2000) idea for inference about the cointegrating vector in time series; also see Müller and Watson (2013).

## 5.5 Confidence Sets for the Local-To-Unity Parameter

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

An easily interpretable transformation of the parameter $c$ is given by the half-life $h_{1/2}(c) = \ln 2/c$, that is, the distance $\Delta$ at which the correlation $\exp[-c\Delta]$ is equal to $1/2$. With $\Pi$ such that the implied weighting of $h_{1/2}$ is uniform on $[0, \Delta_{\max}]$ with $\Delta_{\max} = \max_{l,\ell} |s_l^0 - s_\ell^0|$, the average length minimizing scale-invariant confidence interval collects the values of $h_{1/2,0}$ for which the test based on

$$\frac{\int_0^{\Delta_{\max}} \det(\mathbf{\Omega}_n(\ln 2/h))^{-1/2} (\mathbf{Z}_n' \mathbf{\Omega}_n(\ln 2/h)^{-1} \mathbf{Z}_n)^{-q/2} dh}{(\mathbf{Z}_n' \mathbf{\Omega}_n(\ln 2/h_{1/2,0})^{-1} \mathbf{Z}_n)^{-q/2}} \tag{26}$$

does not exceed the $1-\alpha$ quantile of (26) under $\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_n(\ln 2/h_{1/2,0}))$. The large sample validity of this confidence set for $h_{1/2,0} > 0$ in the general local-to-unity model (10) follows from the same arguments as the large sample validity of the $\text{LFUR}_n$ test discussed above.

We note that the half-life as a fraction of the maximum distance, $h/\Delta_{\max}$, is invariant to the scaling factor $\lambda_n$, so it is the same using the scaled or unscaled locations.

## 5.6 Residual Based Tests

Consider inference about the persistence properties of the disturbance $u_l$ in a linear regression $y_l = x_l'\beta + u_l$ (where $x_l$ may include a constant). The above results are then not directly applicable, since with $\beta$ unknown, $u_l$ is unobserved.

There is an easy solution to this problem if $\mathbf{u}_n = (u_1, \ldots, u_n)'$ is independent of $\mathbf{X}_n = (x_1, \ldots, x_n)'$. Namely, one can simply base inference on weighted averages of $\mathbf{Y}_n$ with weights that they are orthogonal to $\mathbf{X}_n$. Let $\mathbf{R}_n^X$ be the $n \times q$ matrix of the eigenvectors of $\mathbf{M}_X \mathbf{\Sigma}_{n,L} \mathbf{M}_X$ corresponding to the largest $q$ eigenvalues, where $\mathbf{M}_X = \mathbf{I}_n - \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'$ and $n^{-1}\mathbf{R}_n^{X\prime}\mathbf{R}_n^X = \mathbf{I}_q$. Then by construction, $\mathbf{R}_n^{X\prime}\mathbf{X}_n = \mathbf{0}$, so that $\mathbf{Z}_n^X = \mathbf{R}_n^{X\prime}\mathbf{Y}_n = \mathbf{R}_n^{X\prime}\mathbf{u}_n$. With $\mathbf{u}_n$ independent of $\mathbf{X}_n$, one can simply condition on the realization of $\mathbf{X}_n$, and apply the above tests with $\mathbf{Z}_n^X$ in place of $\mathbf{Z}_n$.

In order to invoke the asymptotic arguments above in such an approach, one needs to generalize the eigenfunction convergence of Lemma 7; see Lemma S.2 in the appendix.

More substantively, the assumption that $\mathbf{u}_n$ is independent of the entire set of regressors $\mathbf{X}_n$ is strong. Without that assumption, there is statistical dependence between the eigenvectors $\mathbf{R}_n^X$ and $\mathbf{u}_n$, invalidating an analysis that conditions on $\mathbf{R}_n^X$. But for large sample validity, it suffices to assume *asymptotic* independence between $\mathbf{X}_n$ and $\mathbf{u}_n$ in the sense that with $(u_l, x_l) = (U(s_l), X(s_l))$,

$$\begin{bmatrix} U_n^0(\cdot) \\ X_n^0(\cdot) \end{bmatrix} = \begin{bmatrix} \lambda_n^{-1/2} U(\lambda_n \cdot) \\ \lambda_n^{-1/2} X(\lambda_n \cdot) \end{bmatrix} \Rightarrow \begin{bmatrix} U^0(\cdot) \\ X^0(\cdot) \end{bmatrix} \tag{27}$$

on $\mathcal{S}^0$ with $X^0$ independent of $U^0$. See Theorem S.3 in the appendix.

This result has a particularly noteworthy implication for the test of the null hypothesis of no cointegration among the $p+1$ variables $(x_l, y_l)$, that is, for the spatial analogue of Engle and Granger's (1987) residual based test of cointegration (see Phillips and Ouliaris (1990) for its asymptotic distribution). Here the assumption is that under the null hypothesis, $(x_l', y_l) = \lambda_n^{1/2}(X_n^0(s_l^0)', Y_n^0(s_l^0))$ with $(X_n^0(\cdot)', Y_n^0(\cdot))' \Rightarrow \Phi L_{X,Y}(\cdot)$, where $L_{X,Y}$ is a vector of $p+1$ independent Lévy-Brownian Motions, and $\Phi$ is an arbitrary full rank $(p+1) \times (p+1)$ matrix. Noting that $OL_{X,Y} \sim L_{X,Y}$ for any $(p+1) \times (p+1)$ rotation matrix $O$, it is without loss of generality to assume that $\Phi$ is lower triangular. Letting $\beta$ be equal to the first $p$ elements in the last row of $\Phi$ then yields that $U_n^0(\cdot) = Y_n^0(\cdot) - X_n^0(\cdot)'\beta$ satisfies (27) with $U^0$ a scalar Lévy-Brownian motion independent of the $p$ dimensional Lévy-Brownian motion $X^0$.

To implement such a level $\alpha$ test of the null hypothesis of no spatial cointegration in practice, one computes the LFUR$_n$ statistic (23) using $\mathbf{Z}_n^X$ in place of $\mathbf{Z}_n$, $\mathbf{\Omega}_0 = \mathbf{R}_n^{X\prime}\mathbf{\Sigma}_{l,n}\mathbf{R}_n^X$ and $\mathbf{\Omega}_1 = \mathbf{R}_n^{X\prime}\mathbf{\Sigma}_n(c_a)\mathbf{R}_n^X$, and compares it to $1-\alpha$ quantile of the statistic under $\mathbf{Y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{n,L})$.

## 5.7 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 States. They construct an index of mobility for each of the 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 and many socioeconomic indicators. One might suspect that the variables used in their regressions are strongly spatially correlated, and in light of the spurious regression results of Section 3, this questions the validity of their inference results. This issue is taken up in Table 1.[7]

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) for the contiguous 48 states, which contain 722 of the 741 commuting zones used by Chetty et al. The results indicate that there is substantial spatial correlation across the United States in the socioeconomic variables. The $I(1)$ null is rejected for only a handful of the variables, the $I(0)$ null is rejected for most, and the confidence intervals for the implied value of the half-life, $h_{1/2}$, while wide, suggest a high degree of spatial persistence. The remaining 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.

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

Table 1: Variables and Regressions from Chetty et al. (2014)

| Variable | Spatial Persistence Statistics | | | Regression of AMI onto Variable | | | | |
|---|---|---|---|---|---|---|---|---|
| | $p$-value for Test of | | Half-life | Levels | | | LBM-GLS | |
| | $I(1)$ Null | $I(0)$ Null | 95% CI | $R^2$ | $\hat{\beta}$ [95% CI] Cluster | $p$-value residual $I(1)$ test | $R^2$ | $\hat{\beta}$ [95% CI] C-SCPC |
| Absolute Mobility Ind. (AMI) | 0.39 | <0.01 | [ 0.10,∞] | NA | NA | NA | NA | NA |
| Frac. Black Residents | 0.11 | 0.01 | [ 0.03,∞] | 0.36 | -0.60 [-0.73,-0.47] | 0.21 | 0.10 | -0.43 [-0.51,-0.35] |
| Racial Segregation | 0.01 | 0.12 | [ 0.00, 0.29] | 0.14 | -0.38 [-0.47,-0.29] | 0.29 | 0.18 | -0.24 [-0.29,-0.18] |
| Segregation of Poverty | 0.29 | 0.03 | [ 0.06,∞] | 0.18 | -0.43 [-0.55,-0.30] | 0.28 | 0.16 | -0.21 [-0.26,-0.16] |
| Frac. < 15 Mins to Work | 0.58 | <0.01 | [ 0.14,∞] | 0.48 | 0.69 [ 0.55, 0.84] | 0.14 | 0.16 | 0.37 [ 0.24, 0.50] |
| Mean Household Income | 0.13 | 0.14 | [ 0.02,∞] | 0.00 | 0.05 [-0.10, 0.19] | 0.39 | 0.00 | -0.02 [-0.09, 0.05] |
| Gini Coefficient | 0.78 | <0.01 | [ 0.25,∞] | 0.37 | -0.60 [-0.78,-0.43] | 0.24 | 0.10 | -0.22 [-0.30,-0.13] |
| Top 1 Perc. Inc. Share | 0.31 | 0.02 | [ 0.07,∞] | 0.04 | -0.21 [-0.36,-0.06] | 0.37 | 0.02 | -0.06 [-0.12,-0.01] |
| Student-Teacher Ratio | 0.22 | 0.13 | [ 0.05,∞] | 0.12 | -0.35 [-0.52,-0.18] | 0.45 | 0.03 | -0.19 [-0.25,-0.12] |
| Test Scores (Inc. adjusted) | 0.29 | 0.06 | [ 0.07,∞] | 0.34 | 0.58 [ 0.41, 0.74] | 0.42 | 0.28 | 0.41 [ 0.30, 0.52] |
| High School Dropout | 0.09 | 0.02 | [ 0.03,∞] | 0.34 | -0.58 [-0.74,-0.42] | 0.50 | 0.22 | -0.29 [-0.56,-0.03] |
| Social Capital Index | 0.72 | <0.01 | [ 0.22,∞] | 0.41 | 0.64 [ 0.46, 0.82] | 0.30 | 0.08 | 0.28 [ 0.12, 0.44] |
| Frac. Religious | 0.27 | 0.04 | [ 0.07,∞] | 0.28 | 0.53 [ 0.36, 0.70] | 0.26 | 0.14 | 0.32 [ 0.13, 0.51] |
| Violent Crime Rate | 0.54 | 0.02 | [ 0.14,∞] | 0.21 | -0.45 [-0.64,-0.26] | 0.35 | 0.04 | -0.15 [-0.18,-0.11] |
| Frac. Single Mothers | 0.18 | <0.01 | [ 0.05,∞] | 0.59 | -0.77 [-0.92,-0.63] | 0.11 | 0.51 | -0.61 [-0.68,-0.53] |
| Divorce Rate | 0.05 | 0.17 | [ 0.02,∞] | 0.27 | -0.52 [-0.70,-0.33] | 0.50 | 0.27 | -0.39 [-0.49,-0.28] |
| Frac. Married | 0.05 | 0.08 | [ 0.01,∞] | 0.31 | 0.56 [ 0.43, 0.68] | 0.22 | 0.31 | 0.35 [ 0.26, 0.44] |
| Local Tax Rate | 0.02 | 0.23 | [ 0.01, 0.51] | 0.12 | 0.35 [ 0.22, 0.48] | 0.40 | 0.01 | 0.08 [ 0.03, 0.13] |
| Colleges per Capita | 0.24 | 0.07 | [ 0.06,∞] | 0.06 | 0.24 [-0.01, 0.48] | 0.29 | 0.00 | 0.02 [-0.23, 0.27] |
| College Tuition | 0.38 | <0.01 | [ 0.09,∞] | 0.00 | -0.02 [-0.15, 0.11] | 0.29 | 0.00 | 0.01 [-0.05, 0.08] |
| Coll. Grad. Rate | 0.04 | 0.03 | [ 0.00, 3.00] | 0.02 | 0.15 [ 0.03, 0.28] | 0.36 | 0.03 | 0.08 [ 0.01, 0.15] |
| Manufacturing Share | 0.21 | 0.00 | [ 0.06,∞] | 0.09 | -0.30 [-0.46,-0.13] | 0.37 | 0.01 | 0.06 [-0.01, 0.13] |
| Chinese Import Growth | 0.02 | 0.07 | [ 0.02, 0.43] | 0.03 | -0.17 [-0.33,-0.02] | 0.39 | 0.00 | 0.03 [ 0.01, 0.04] |
| Teenage LFP Rate | 0.51 | <0.01 | [ 0.12,∞] | 0.44 | 0.66 [ 0.50, 0.82] | 0.29 | 0.04 | 0.25 [ 0.11, 0.40] |
| Migration Inflow | 0.30 | 0.08 | [ 0.00,∞] | 0.07 | -0.27 [-0.42,-0.13] | 0.32 | 0.04 | -0.14 [-0.19,-0.09] |
| Migration Outflow | 0.35 | 0.01 | [ 0.08,∞] | 0.03 | -0.16 [-0.30,-0.03] | 0.37 | 0.02 | -0.09 [-0.16,-0.03] |
| Frac. Foreign Born | 0.55 | 0.04 | [ 0.16,∞] | 0.00 | -0.03 [-0.16, 0.09] | 0.40 | 0.02 | -0.12 [-0.21,-0.04] |

Notes: The first two columns show p-values for tests of the $I(1)$ and $I(0)$ null hypotheses using the statistics (23) and (25). The third column shows the 95% confidence for the half-life, $h_{1/2}$, constructed by inverting the tests in (26). The half-life values shown are relative to the largest distance between commuter zones in the dataset, which is approximately 2800 miles. The results for the levels regressions show the $R^2$ and estimated regression coefficients ($\hat{\beta}$) from regression of the Absolute Mobility Index (AMI) onto each of the variables in the table, with nominal 95% confidence intervals constructed using standard errors clustered by state; also shown are the $p$-values for the $I(1)$ and $I(0)$ tests applied to the residuals. The final two columns show the corresponding regression results using variables transformed using the LBM-GLS transformation, with 95% nominal confidence intervals constructed using C-SCPC. Results are based on commuting zones in the contiguous 48 U.S. states. The variables are standardized (in levels) to have mean zero and unit standard deviation.

# 6 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 best approach for regressions involving spatial $I(1)$ variables is not so obvious, and in this section we explore a number of possibilities. Using simulated data, we assess the coverage and length properties of the corresponding confidence intervals.

## 6.1 Simulation Design

We are interested in inference about $\beta_1$, the first element of $\beta$, in the linear regression (12), maintaining throughout that $\mathbf{Y}_n$ is independent of $\mathbf{X}_n = (x_1, \ldots, x_n)'$. 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 $p = 1$ and $p = 5$. In each of those, the $p + 1$ 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 of the $p$ elements of $x_l$, are:

- DGP1: $y_l = L(s_l)$, Lévy-Brownian motion;
- DGP2: $Y_l = Y(s_l)$ with $Y \sim I(1)$ 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 $\bar{\rho} = 0.01$;
- 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 $\bar{\rho} = 0.03$,
- 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: $y_l = \int_{\mathbb{R}^2} \mathbf{1}[0 \leq r \leq s_l] dW(r)$, so dependence is induced by a Brownian sheet.

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.

## 6.2 Data Transformations

We consider inference based on six estimators for $\beta_1$.

*Levels Regression:* This is OLS applied to the "levels" regression (12). 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 $(y_l^*, x_l^*)$, and stack these in the vector $\mathbf{Y}_n^*$ and matrix $\mathbf{X}_n^*$. In all methods, we use the same transformation for the $p+1$ variables in $(y_l, x_l)$, and then run a linear regression of $\mathbf{Y}_n^*$ on $\mathbf{X}_n^*$. This regression omits a constant, since all transformations involve a demeaning step.

*Isotropic Differences:* This is transformation (17) which we apply for bandwidths $b = 0.03, 0.06, \ldots, 0.15$, where the locations $\{s_l^0\} = \{\lambda_n^{-1} s_l\}$ are scale normalized so that $\max_{l,\ell} |s_l^0 - s_\ell^0| = 1$.

*Cluster Fixed Effects:* We partition the sampling region $\mathcal{S}_n$ 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

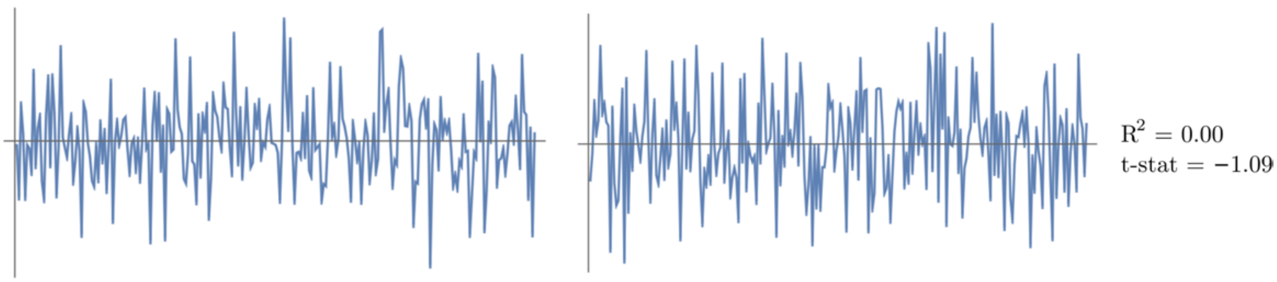
$$y_l^* = y_l - \frac{\sum_{i=1}^m \mathbf{1}[s_l \in \mathcal{R}_i] \sum_{\ell \neq l} \mathbf{1}[s_\ell \in \mathcal{R}_i] y_\ell}{\sum_{i=1}^m \mathbf{1}[s_l \in \mathcal{R}_i] \sum_{\ell \neq l} \mathbf{1}[s_\ell \in \mathcal{R}_i]}.$$

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

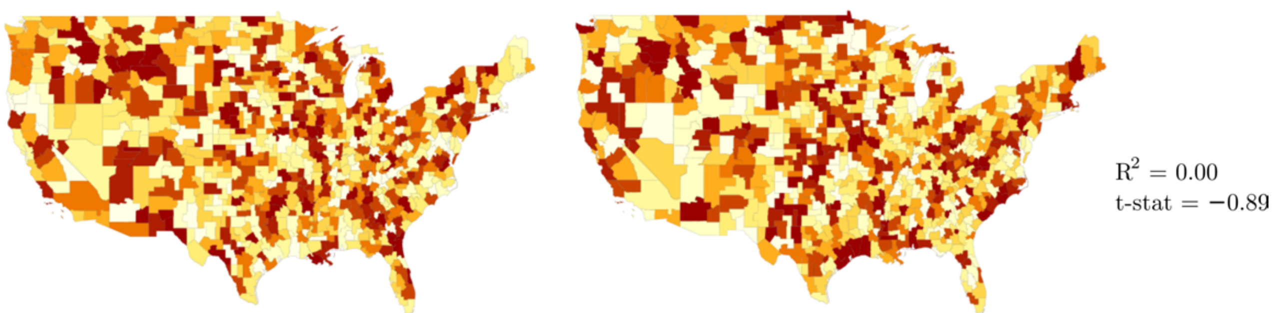
*LBM-GLS:* First differences in a regular time series are a GLS transformation under the canonical random walk model of $I(1)$ persistence. Recall from the last section that $\mathbf{\Sigma}_{n,L}$ is the $n \times n$ covariance matrix of $\mathbf{Y}_n$ induced by a Lévy-Brownian motion, the canonical $I(1)$

Figure 5: Transformed Strongly Dependent Data

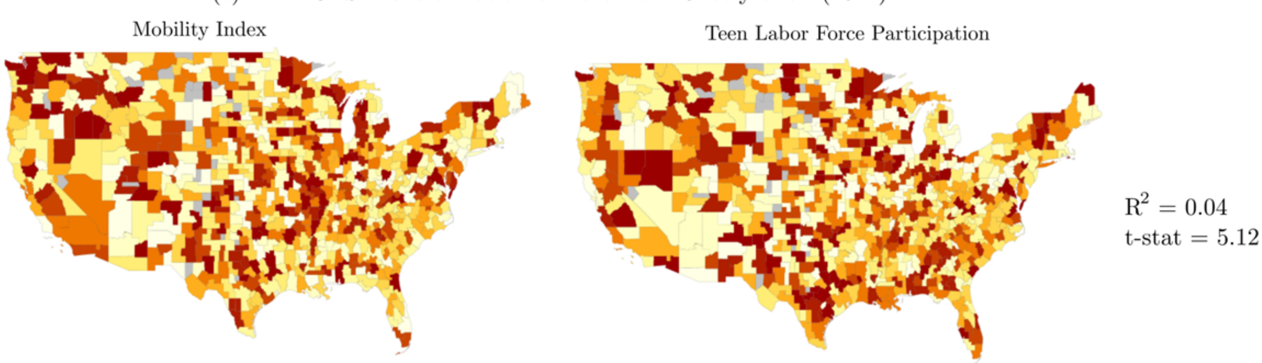
(a) First Difference of Independent Time Series Random Walks




(b) LBM-GLS Transformation of Independent Spatial Unit Root Processes




(c) LBM-GLS Transformation of Data from Chetty et al. (2014)



model of spatial persistence. With $\mathbf{Y}_n \sim \mathcal{N}(\mu\mathbf{1}, \mathbf{\Sigma}_{n,L})$,

$$\mathbf{Y}_n^* = (\mathbf{M}\mathbf{\Sigma}_{n,L}\mathbf{M})^{-1/2}\mathbf{Y}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{M}) \tag{28}$$

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 $\mathbf{Y}_n^*$. 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. Figure 5 illustrates the GLS transformations for the data in Figure 1.

**Remark 6.1.** The LBM-GLS estimator of $\beta$ is closely related to the OLS estimator obtained after controlling for a smooth spline function, say $\eta(s)$, in the regression.[8] 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 $n \times p$, $\boldsymbol{\beta}$ is $p \times 1$, $\boldsymbol{\eta}$ 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 $\min_{\boldsymbol{\beta},\boldsymbol{\eta}}[(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta} - \boldsymbol{\eta})'(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta} - \boldsymbol{\eta}) + \lambda^{-1}\boldsymbol{\eta}'\mathbf{V}^{-1}\boldsymbol{\eta}]$ for an appropriately chosen matrix $\mathbf{V}$ that captures the smoothness properties of $\eta(\cdot)$. The solution to the problem yields the estimator $\hat{\boldsymbol{\beta}} = [\mathbf{X}_n'(\mathbf{I}+\lambda^{-1}\mathbf{V})^{-1}\mathbf{X}_n]^{-1}[\mathbf{X}_n'(\mathbf{I}+\lambda^{-1}\mathbf{V})^{-1}\mathbf{Y}_n]$, which is recognized as the GLS estimator using the error covariance matrix $\mathbf{I} + \lambda^{-1}\mathbf{V}$. 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 $\mathbf{\Sigma}_{n,L}$ for $\mathbf{V}$ imposes a Lévy-Brownian motion smoothness prior of $\eta$; 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 $\mathbf{M}\mathbf{\Sigma}_{n,L}\mathbf{M}$ corresponding to the $q < n$ largest eigenvalues $\hat{\boldsymbol{\nu}}_n$. By construction, if $\mathbf{Y}_n \sim \mathcal{N}(\mu\mathbf{1}, \mathbf{\Sigma}_{n,L})$, then

$$\mathbf{Y}_n^* = \text{diag}(\hat{\boldsymbol{\nu}}_n)^{-1/2}\mathbf{Z}_n \sim \mathcal{N}(\mathbf{0}, n^{-1}\mathbf{I}_q). \tag{29}$$

Here $\mathbf{Y}_n^*$ is $q \times 1$, rather than $n \times 1$, as in the previous transformations. Note that setting

[8] 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)).

$q = n - 1$ amounts to the LBM-GLS transformation (28). 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 (29) has an asymptotic $\mathcal{N}(\mathbf{0}, \omega^2 \mathbf{I}_q)$ 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 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 6.1, the resulting regression controls for smooth spatial function spanned by the columns of $\mathbf{R}_n$. We implement this with $q = 5, 10, 20, 50, 100$.

*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 $\hat{\beta}_i = (\hat{\beta}_{i,1}, \ldots, \hat{\beta}_{i,p})'$ be the corresponding coefficients. The $m$ estimators $\hat{\beta}_{i,j}$ are then treated as independent and Gaussian information about $\beta_j$, 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 $\beta_1$ are used in conjunction with three types of standard errors: We present results using heteroskedasticity robust standard errors for the LBM-GLS. For the cluster fixed effects 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). Their so-called C-SCPC method is calibrated to control size under spatial dependence with an average pairwise 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)).

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

| Method | Lévy-BM | $I(1)_{c_{0.01}}$ | $I(1)_{c_{0.03}}$ | $I(1)_{\text{Matérn}}$ | $J_{c_{0.03}}$ | $J_{c_{0.50}}$ | Br. Sheet |
|---|---|---|---|---|---|---|---|
| | | | | (a) $k = 1$ | | | |
| OLS (C-SCPC) | 0.23 | 0.24 | 0.27 | 0.25 | 0.04 | 0.14 | 0.25 |
| Isotropic difference (C-SCPC) | 0.04 | 0.05 | 0.07 | 0.06 | 0.03 | 0.04 | 0.04 |
| Cluster fixed effects (Cluster) | 0.08 | 0.24 | 0.35 | 0.30 | 0.07 | 0.07 | 0.13 |
| Cluster fixed 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-SCPC) | 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 | 0.05 | 0.07 | 0.08 | 0.10 | 0.05 | 0.05 | 0.10 |
| Ibragimov-Müller | 0.08 | 0.13 | 0.15 | 0.13 | 0.05 | 0.07 | 0.16 |
| Addendum: Avg. $R^2$ | 0.14 | 0.18 | 0.21 | 0.19 | 0.01 | 0.09 | 0.14 |
| | | | | (b) $k = 5$ | | | |
| OLS (C-SCPC) | 0.20 | 0.20 | 0.23 | 0.20 | 0.04 | 0.14 | 0.23 |
| Isotropic difference (C-SCPC) | 0.04 | 0.05 | 0.07 | 0.06 | 0.03 | 0.04 | 0.05 |
| Cluster fixed effects (Cluster) | 0.08 | 0.24 | 0.35 | 0.31 | 0.07 | 0.08 | 0.14 |
| Cluster fixed 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-SCPC) | 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 | 0.05 | 0.07 | 0.08 | 0.09 | 0.05 | 0.05 | 0.10 |
| Ibragimov-Müller | 0.05 | 0.06 | 0.08 | 0.07 | 0.05 | 0.05 | 0.07 |
| Addendum: Avg. $R^2$ | 0.43 | 0.56 | 0.64 | 0.59 | 0.05 | 0.31 | 0.44 |

## 6.3 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. Here we present the key conclusions in two tables.

Table 2 summarizes the null 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 methods that exhibit large size distortions in Table 2.

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,

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

| Method | Lévy-BM | $I(1)_{c_{0.01}}$ | $I(1)_{c_{0.03}}$ | $I(1)_{\text{Matérn}}$ | $J_{c_{0.03}}$ | $J_{c_{0.50}}$ | Br. Sheet |
|---|---|---|---|---|---|---|---|
| | | | | (a) $k = 1$ | | | |
| Isotropic difference (C-SCPC) | 0.53 | 0.70 | 0.73 | 0.73 | 0.44 | 0.52 | 0.54 |
| Cluster fixed effects (C-SCPC) | 0.55 | 0.81 | 0.91 | 0.88 | 0.43 | 0.39 | 0.50 |
| LBM-GLS (C-SCPC) | 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 |
| High-pass Eigenvector | 0.33 | 0.43 | 0.45 | 0.50 | 0.42 | 0.33 | 0.41 |
| 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 fixed effects (C-SCPC) | 0.44 | 0.79 | 0.90 | 0.88 | 0.43 | 0.42 | 0.48 |
| LBM-GLS (C-SCPC) | 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 |
| High-pass Eigenvector | 0.33 | 0.43 | 0.45 | 0.50 | 0.41 | 0.33 | 0.40 |
| Ibragimov-Müller | 0.46 | 0.65 | 0.67 | 0.78 | 0.36 | 0.48 | 0.49 |

as does the low-pass eigenvector transformation. Second, LBM-GLS (with HAR standard errors) produces confidence intervals with the smallest average length. These results, along with the 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 6.2.** The final rows in Table 2 show the average value of the $R^2$ in the levels-regression (12). (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 3.4.

## 6.4 Regressions in Chetty et al. (2014)

We now return to Table 1. As noted in Section 5.7 the first three columns of the table suggest substantial spatial correlation in many of the variables. The final 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 et al. (2014).[9] The first set of results are for regressions using the levels of the variables, and the second set uses the LBM-GLS transformed variables.

We highlight three results from the table. First, the residuals from the levels regressions

[9] The results in Table 1 differ slightly from the results in Chetty et al. because Table 1 only uses data from the 48 contiguous U.S. states.

are highly spatially correlated: the $I(1)$ null is not rejected at the 10% level for any of the regressions. Second, the LBM-GLS estimates of $\beta$ and $R^2$ tend to be smaller in magnitude than the levels-regression estimates. Third, the LBM-GLS C-CSPC confidence intervals are narrower than in the levels regression 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 et al. (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.

# 7 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 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 unit root model as a baseline—we do not yet have a good theoretical understanding of its properties, and future research is required to establish 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 $m > 2d$, $\mathbb{E}\left[(Y(s) - Y(r))^{2m}\right] \leq C|s-r|^m$ for some $C$. Let $m > 2d$ and apply Condition 2 to obtain

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

where the second equality follows from the representation (5) of $L$.

For the corresponding result about spatial local-to-unity processes, we similarly have with $Y_c(s) = \int_{\mathbb{R}^d} h_c(r,s)B(r)dr$

$$\begin{aligned}\mathbb{E}[(Y_c(s)-Y_c(r))^{2m}] &\leq C_m\left(\int_{\mathbb{R}^d}(h_c(u,s)-h_c(u,r))^2du\right)^m\\ &= C_m\mathbb{E}[(J_c(s)-J_c(r))^2]^m \qquad (30)\end{aligned}$$

where the last equality follows from the representation (9) of $J_c$, and

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

□

**Proof of Theorem 2:** Consider first the claim for the convergence for the LTU process (10). From

$$\int_{\mathbb{R}^d} h_c(r,0)^2dr = (2c)^{-1} = \lambda^d\int_{\mathbb{R}^d} h_c(\lambda r,0)^2dr = \lambda^{(1+d)/2}\frac{\kappa_{c,d}^2}{\kappa_{\lambda c,d}^2}\int_{\mathbb{R}^d} h_{\lambda c}(r,0)^2dr = \lambda^{(1+d)/2}\frac{\kappa_{c,d}^2}{\kappa_{\lambda c,d}^2}(2c\lambda)^{-1}$$

for all $\lambda > 0$ it follows that $\kappa_{\lambda c,d} = \lambda^{(d-1)/4}\kappa_{c,d}$. Thus, the LTU process can be written as

$$Y_n^0(s) = \lambda_n^{-d/2}\int_{\mathbb{R}^d} h_c(\lambda_n^{-1}r,s)B(r)dr, \quad s\in\mathcal{S}^0. \qquad (31)$$

We show convergence of finite dimensional distributions and tightness of the process $Y_n^0$. The latter follows by Theorem 23.7 of Kallenberg (2021) from (30) and

$$\mathbb{E}[Y_c^0(0)^2] \leq C_2\lambda_n^{-d}\int_{\mathbb{R}^d} h_c(\lambda_n^{-1}r,s)^2dr = C_2\int_{\mathbb{R}^d} h_c(r,s)^2dr = C_2(2c)^{-1}$$

where the inequality invokes Condition 2. For the former, consider for $t_1, \ldots, t_p \in \mathcal{S}^0$, the $p \times 1$ vector $(Y_n^0(t_1), \ldots, Y_n^0(t_p))$. By the Cramér-Wold device, it suffices to establish the convergence $X_n = \sum_{j=1}^p v_j Y_n^0(t_j) \Rightarrow \sum_{j=1}^p v_j \omega J_c(t_j)$ for $(v_1, \ldots, v_p) \in \mathbb{R}^p$. Let $f_v(r) = \sum_{j=1}^p v_j h_c(r, t_j)$, so that from (9), $\sum_{j=1}^p v_j J_c(t_j) \sim \mathcal{N}(0, \int_{\mathbb{R}^d} f_v(r)^2 dr)$ and from (31)

$$X_n = \lambda_n^{-d/2} \int_{\mathbb{R}^d} f_v(\lambda_n^{-1} r) B(r) dr.$$

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

$$X_n^\varepsilon = \lambda_n^{-d/2} \int_{\mathbb{R}^d} f_v^\varepsilon(\lambda_n^{-1} r) B(r) dr.$$

From Condition 2 we find

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

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

Furthermore

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

by dominated convergence, since by Cauchy-Schwarz, $|\int_{\mathbb{R}^d} f_v^\varepsilon(r) f_v^\varepsilon(r + \lambda_n^{-1} s) dr| \leq \int_{\mathbb{R}^d} f_v^\varepsilon(r)^2 dr < \infty$ and $\int_{\mathbb{R}^d} |\sigma_B(s)| ds < \infty$.

Finally, note that $f_v^\varepsilon$ is bounded and $f_v^\varepsilon(\lambda_n^{-1} r) = 0$ for $|r| > \lambda_n / \varepsilon$. Thus, using Condition 3,

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

The result for the LTU process (10) now follows since mean square convergence implies convergence in distribution, and $\varepsilon > 0$ was arbitrary.

For the convergence in Theorem 2, note that from (7), $Y_n^0(s) = \lambda_n^{-d/2} \int_{\mathbb{R}^d} h(\lambda_n^{-1} r, s) B(r) dr$, and $Y_n^0(0) = 0$, so the result follows from the same steps. $\square$

**Proof of Theorem 3:** The results follow straightforwardly from the CMT if we can show that $\lambda_n^{-1/2} n^{-1} \sum_{l=1}^n y_l \Rightarrow \int Y^0(s) dG(s)$, $\lambda_n^{-1/2} n^{-1} \sum_{l=1}^n x_l \Rightarrow \int X^0(s) dG(s)$, $\lambda_n^{-1} n^{-1} \sum_{l=1}^n x_l y_l \Rightarrow \int X^0(s) Y^0(s) dG(s)$ and $\lambda_n^{-1} n^{-1} \sum_{l=1}^n x_l x_l' \Rightarrow \int X^0(s) X^0(s)' dG(s)$.

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

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

so it suffices to show the claim for $\int X^*(s) Y^*(s) dG_n(s)$. Now almost all realizations of the $\mathbb{R}^p \mapsto \mathbb{R}$ function $s \mapsto X^*(s) Y^*(s)$ on $\mathcal{S}^0$ are continuous and bounded. For any such realization, $\int X^*(s) Y^*(s) dG_n(s) \to \int X^*(s) Y^*(s) dG(s)$ by the definition of convergence in distribution. Thus $\int X^*(s) Y^*(s) dG_n(s) \overset{a.s.}{\to} \int X^*(s) Y^*(s) dG(s)$. But almost sure convergence implies convergence in distribution, so the desired result follows. The argument for the other terms is analogous. $\square$

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

**Lemma 9.** *Let $\mathcal{B}_\delta(r) = \{s : |s - r| \leq \delta\} \subset \mathbb{R}^d$ be a ball of radius $\delta$ with center $r$. Under Condition 1, for any $\delta > 0$, $\limsup_{n \to \infty} \sup_{r \in \mathcal{S}^0} G_n(\mathcal{B}_\delta(r)) \leq \sup_{r \in \mathcal{S}^0} G(\mathcal{B}_\delta(r))$, where $G_n(A)$ and $G(A)$ are the measures that are assigned to the Borel set $A \subset \mathbb{R}^d$ by the distributions $G_n$ and $G$, respectively.*

*Proof.* Suppose otherwise. Then there exists $\varepsilon > 0$ and a sequence $r_n$ such that

$$\limsup_{n \to \infty} \sup_{r \in \mathcal{S}^0} G_n(\mathcal{B}_\delta(r)) = \lim_{n \to \infty} G_n(\mathcal{B}_\delta(r_n)) \geq \sup_{r \in \mathcal{S}^0} G(\mathcal{B}_\delta(r)) + \varepsilon.$$

Since $G$ is a continuous distribution, there exists $\delta' > \delta$ such that $\sup_{r \in \mathcal{S}^0} G(\mathcal{B}_{\delta'}(r)) \leq \sup_{r \in \mathcal{S}^0} G(\mathcal{B}_\delta(r)) + \varepsilon/2$. Since $\mathcal{S}^0$ 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

$$G_n(\mathcal{B}_\delta(r_n)) \leq G_n(\mathcal{B}_{\delta'}(r_0)) \to G(\mathcal{B}_{\delta'}(r_0)) \leq \sup_{r \in \mathcal{S}^0} G(\mathcal{B}_\delta(r)) + \varepsilon/2$$

yielding the desired contradiction. $\square$

**Proof of Theorem 4:** From Theorem 3 and the CMT, $H\hat{\beta} \Rightarrow H\boldsymbol{\Xi}_{\tilde{x}\tilde{x}}^{-1}\Xi_{\tilde{x}\tilde{y}}$ with the r.h.s. non-zero with probability one. Thus $H\hat{\beta} = O_p(1)$ (and not $H\hat{\beta} = o_p(1)$). The result hence follows if we can show that $||S_{\tilde{x}\tilde{x}}^{-1}\hat{\Omega}_n S_{\tilde{x}\tilde{x}}^{-1}|| = o_p(n)$ (since this implies that the smallest eigenvalue of $n(HS_{\tilde{x}\tilde{x}}^{-1}\hat{\Omega}_n S_{\tilde{x}\tilde{x}}^{-1}H')^{-1}$ diverges).

Since $\lambda_n^{-1}S_{\tilde{x}\tilde{x}} \Rightarrow \Xi_{\tilde{x}\tilde{x}}$ and $\Xi_{\tilde{x}\tilde{x}}$ is full rank with probability one, it suffices to show that $n^{-1}\lambda_n^{-2}||\hat{\Omega}_n|| \xrightarrow{p} 0$.

Let $\tilde{Y}_n^0(\cdot) = Y_n^0(\cdot) - \int Y_n^0(s)dG_n(s)$, $\tilde{X}_n^0(\cdot) = \tilde{X}_n^0(\cdot) - \int X_n^0(s)dG_n(s)$ and $e_n^0(\cdot) = (\tilde{Y}_n^0(\cdot) - \hat{\beta}\tilde{X}_n^0(\cdot))\tilde{X}_n^0(\cdot)$, so that $e_l = \lambda_n e_n^0(\lambda_n^{-1}s_l)$. By (13), Theorem 3 and the CMT, $e_n^0(\cdot) \Rightarrow e^0(\cdot) = (\tilde{Y}(\cdot) - \tilde{X}(\cdot)'\Xi_{\tilde{x}\tilde{x}}^{-1}\Xi_{\tilde{x}\tilde{y}})\tilde{X}(\cdot)$, so that $\sup_l |e_n^0(s_l)| \Rightarrow \sup_{s\in\mathcal{S}^0}|e^0(s)|$, and therefore $\lambda_n^{-1}\sup_{l\leq n}|e_l| = O_p(1)$. Consider first the HAC estimator. We have

$$\lambda_n^{-2}n^{-2}\left\|\sum_{l,\ell=1}^{n}\kappa(b_n(s_l-s_\ell))e_l e_\ell'\right\| \leq \lambda_n^{-2}(\sup_{l\leq n}|e_l|)^2\cdot n^{-2}\sum_{l,\ell=1}^{n}|\kappa(b_n(s_l-s_\ell))|$$

and with $b_n^0 = \lambda_n b_n$ and $s_l^0 = \lambda_n^{-1}s_l$

$$\begin{aligned}\sum_{l,\ell=1}^{n}|\kappa(b_n(s_l-s_\ell))| &= \sum_{l,\ell=1}^{n}|\kappa(b_n^0(s_l^0-s_\ell^0))| \\ &\leq \bar{\kappa}\sum_{l,\ell=1}^{n}\mathbf{1}[|s_l^0-s_\ell^0|\leq(b_n^0)^{-1/2}] + \sum_{l,\ell=1}^{n}\mathbf{1}[|s_l^0-s_\ell^0|>(b_n^0)^{-1/2}]|\kappa(b_n^0(s_l^0-s_\ell^0))|.\end{aligned}$$

Now

$$n^{-2}\sum_{l,\ell=1}^{n}\mathbf{1}[|s_l^0-s_\ell^0|>(b_n^0)^{-1/2}]|\kappa(b_n^0(s_l^0-s_\ell^0))| \leq \sup_{|a|\geq\sqrt{b_n^0}}|\kappa(a)| \xrightarrow{p} 0$$

by (16) and $1/b_n^0 = o_p(1)$. Furthermore, since $G$ is continuous, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\sup_{r\in\mathcal{S}^0} G(\mathcal{B}_\delta(r)) \leq \varepsilon$ in the notation of Lemma 9. Note that

$$\begin{aligned}n^{-2}\sum_{l,\ell=1}^{n}\mathbf{1}[|s_l^0-s_\ell^0|\leq(b_n^0)^{-1/2}] &\leq \sup_{r\in\mathcal{S}^0} G_n(\mathcal{B}_{(b_n^0)^{-1/2}}(r)) \\ &\leq \sup_{r\in\mathcal{S}^0} G_n(\mathcal{B}_\delta(r)) + \mathbb{P}((b_n^0)^{-1/2} > \delta).\end{aligned}$$

Since by assumption, $1/b_n^0 = o_p(1)$, we have $\mathbb{P}((b_n^0)^{-1/2} > \delta) \to 0$, and by Lemma 9, $\limsup_{n\to\infty}\sup_{r\in\mathcal{S}^0} G_n(\mathcal{B}_\delta(r)) \leq \varepsilon$. But $\varepsilon > 0$ was arbitrary, and the result follows.

For the cluster estimator, we similarly have

$$\lambda_n^{-2} n^{-2} \left\| \sum_{j=1}^{n_C} \left( \sum_{l \in C_j}^{n} e_l \right) \left( \sum_{l \in C_j}^{n} e_l \right)' \right\| \le \lambda_n^{-2} (\sup_{l \le n} |e_l|)^2 \cdot n^{-2} \sum_{j=1}^{n_C} |C_j|^2$$

and $\max_{1 \le j \le n_C} |C_j|/n \to 0$ implies $n^{-2} \sum_{j=1}^{n_C} |C_j|^2 \to 0$, as shown in equation (4) of Hansen and Lee (2019). □

# 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.

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.

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

Dickey, D. A., and W. A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

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 C. W. J. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

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.

Fuller, W. A. (1996): *Introduction to Statistical Time Series.* John Wiley, New York, second edn.

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.

Hansen, B., and S. Lee (2019): "Asymptotic Theory for Clustered Samples," *Journal of Econometrics*, 210, 268–290.

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.

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," *CEPR Discussion Paper DP17429*.

Kiefer, N., and T. J. Vogelsang (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164.

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.

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.

——— (2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81.

——— (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," *Econometrica*, 90, 2901–2935.

——— (2022b): "Spatial Correlation Robust Inference in Linear Regression and Panel Models," *forthcoming, Journal of Business and Economic Statistics.*

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

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.

Pantula, S. G., G. Gonzalez-Farias, and W. A. Fuller (1994): "A Comparison of Unit-Root Test Criteria," *Journal of Business & Economic Statistics*, 12, 449–459.

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.

Phillips, P. C. B., and S. Ouliaris (1990): "Asymptotic Properties of Residual Based Test for Cointegration," *Econometrica*, 58, 165–193.

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.

Rossi, F., and O. Lieberman (2023): "Spatial autoregressions with an extended parameter space and similarity-based weights," *Forthcoming, Journal of Econometrics.*

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.

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.

WRIGHT, J. H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business and Economic Statistics*, 18, 211–222.

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

Supplementary Appendix to

# Spatial Unit Roots

by Ulrich K. Müller and Mark W. Watson

This appendix provides supplemental material. Section S.1 provides details on the technique used to generate Figures 2 and 4. Section S.2 contains proofs of all formal results in Sections 4-5. Section

## S.1 Generation of Figures 2-4

For the left panel of Figure 2 and Figure 4, we approximate the non-stationary processes by stationary ones with a very small degree of mean reversion. In particular, with $f_0(\omega)=1$, let $\tilde{f}_i(\omega)=f_i(\omega)/(c^2+|\omega|^2)^{3/2}$ with $c=0.1$ for the three processes $Y_i$, $i=0,1,2$ of Figures 2 and 4. These spectral densities are isotropic, so the covariance functions satisfy $E[Y_i(r)Y_i(s)]=\sigma_i(|r-s|)$ with

$$\sigma_i(x)=\int_0^\infty J_0(\omega x)f_i(\omega)d\omega$$

where $J_0$ is the Bessel function function of the first kind with zero parameter (cf. equation (1.2.6) in Ivanov and Leonenko (1989)). We approximate $\sigma_i(\cdot)$ numerically on the interval $[0,1]$, and then use Stein's (2002) technique to generate the figures via the fast Fourier transform on a grid of $700\times 700$ points.

The eigenfunctions of Figure 3 are approximated via (22) using 1000 locations $\{s_l^0\}_{l=1}^{1000}$ drawn at random within the contiguous U.S.

## S.2 Proofs of Results from Sections 4 and 5

**Proof of Theorem 5:** Clearly,

$$\hat{\gamma}=\frac{\int_{\mathcal{I}_b}\int Y_n^0(s)\kappa_b(|s-r|)(Y_n^0(r)-Y_n^0(s))dG_n(r)dG_n(s)}{\int_{\mathcal{I}_b}Y_n^0(s)^2dG_n(s)} \tag{S.1}$$

and proceeding as in the proof of Theorem 3 shows that it suffices to show the claim with $Y_n^0(s)$ replaced by $Y^*(s)=\omega J_c(s)$ in (S.1). Denote the resulting expression by $\hat{\gamma}^*$, we have

$$\hat{\gamma}^* \quad = \quad \frac{\mathbb{E}\left[\mathbf{1}[S_n\in\mathcal{I}_b]Y^*(S_n)\kappa_b(|S_n-R_n|)(Y^*(R_n)-Y^*(S_n))|Y^*\right]}{\mathbb{E}\left[\mathbf{1}[S_n\in\mathcal{I}_b]Y^*(S_n)^2|Y^*\right]}$$

$$\begin{aligned}
&\overset{a.s.}{\rightarrow} \frac{\mathbb{E}\left[\mathbf{1}[S\in\mathcal{I}_b]Y^*(S)\kappa_b(|S-R|)(Y^*(R)-Y^*(S))|Y^*\right]}{\mathbb{E}\left[\mathbf{1}[S\in\mathcal{I}_b]Y^*(S)^2|Y^*\right]}\\
= &\frac{\int_{\mathcal{I}_b}\int J_c(s)\kappa_b(|s-r|)(J_c(r)-J_c(s))dG(r)dG(s)}{\int_{\mathcal{I}_b} J_c(s)^2 dG(s)}
\end{aligned}$$

where $(S_n, R_n)$ is a sequence of $\mathbb{R}^{2d}$ random variables with distribution $G_n \times G_n$ converging to $(S, R)$ with distribution $G \times G$, and the convergence follows, since for almost all realizations of $Y^*$, the $\mathbb{R}^{2d} \mapsto \mathbb{R}$ function $(s,r) \mapsto \mathbf{1}[s \in \mathcal{I}_b]Y^*(s)\kappa_b(|s-r|)(Y^*(r) - Y^*(s))$ and the $\mathbb{R}^d \mapsto \mathbb{R}$ function $s \mapsto \mathbf{1}[s \in \mathcal{I}_b]Y^*(s)^2$ is bounded with a discontinuity set of Lebesgue measure zero. $\square$

**Proof of Theorem 6:** We first show the result for $L$ in place of $J_c$. In the proof, $C$ denotes a sufficiently large constant, not necessarily the same in each instance it is used.

As a first step, we show that replacing $L(s)$ by $L(s) - \hat{m}$ induces a $o_p(1)$ difference, where the convergences throughout the proof are with respect to $b \to 0$. By Cauchy-Schwarz, the second moment of the difference is bounded above by

$$\begin{aligned}
&\mathbb{E}\left[\left(b^{-d-1}\hat{m}\int_{\mathcal{I}_b}\int \kappa_b(|s-r|)(L(r)-L(s))dG(r)dG(s)\right)^2\right]\\
\leq\ &\mathbb{E}[\hat{m}^2]\mathbb{E}\left[\left(b^{-d-1}\int_{\mathcal{I}_b}\int \kappa_b(|s-r|)(L(r)-L(s))dG(r)dG(s)\right)^2\right].
\end{aligned}$$

Consider first $d = 1$. The support $\mathcal{S}^0$ of $G$ then consists of a countable number of disjoint intervals, and it suffices to show that the integral over each of those intervals is $o_p(1)$. Take one such interval $[l, u] \subset \mathbb{R}$. We have

$$\int_{l+b}^{u-b}\int_l^u \kappa_b(|s-r|)(L(r)-L(s))dG(r)dG(s) = \int_l^u h_b(r)L(r)dG(r)$$

with $h_b(r) = \int_l^u(\mathbf{1}[l+b \leq s \leq u-b]\kappa_b(|s-r|) - \mathbf{1}[l+b \leq r \leq u-b]\kappa_b(|s-r|))dG(s)$. By inspection, for all small enough $b$, $h_b(r) = 0$ for $r \in [l+2b, u-2b]$, $\sup_{r\in[l,u]}|h_b(r)| \leq Cb$, $\int_l^{l+2b} h_b(r)dr = \int_{u-2b}^u h_b(r)dr = 0$, so that $\int_l^{l+2b} h_b(r)g(r)dr = b\int_0^2 h_b(br)g(l+br)dr = O(b^3)$ from a first order Taylor expansion of $g(\cdot)$ around $g(l)$, and similarly, $\int_{u-2b}^u h_b(r)dG(r) = O(b^3)$. Thus

$$\begin{aligned}
\mathbb{E}\left[\left(\int_l^u h_b(r)L(r)dG(r)\right)^2\right] &= \int_l^u\int_l^u h_b(r)h_b(s)\min(r,s)dG(r)dG(s)\\
&= \int_l^{l+2b}\int_l^{l+2b} h_b(r)h_b(s)(\min(r,s)-l)dG(r)dG(s)\\
&\quad + \int_{u-2b}^u\int_{u-2b}^u h_b(r)h_b(s)(\min(r,s)-u)dG(r)dG(s) + O(b^6)
\end{aligned}$$

$$= O(b^5)$$

so the desired result follows.

For $d > 1$,

$$\begin{aligned} D_b^2 &= \mathbb{E}\left[\left(b^{-d-1}\int_{\mathcal{I}_b}\int \kappa_b(|s-r|)(L(r)-L(s))dG(r)dG(s)\right)^2\right] \\ &= \mathbb{E}\left[\left(b^{-1}\int_{\mathcal{I}_b}\int \kappa_0(|r|)(L(s+br)-L(s))g(s+br)drdG(s)\right)^2\right] \\ &= \int_{\mathcal{I}_b}\int_{\mathcal{I}_b}\int\int b^{-2}\kappa_0(|r|)\kappa_0(|u|)\zeta_b(s,r,t,u)g(s+br)g(t+bu)dr\cdot du\cdot dG(s)dG(t) \end{aligned}$$

with

$$\begin{aligned} 2\zeta_b(s,r,t,u) &= 2\mathbb{E}[(L(s+br)-L(s))(L(t+bu)-L(t))] \\ &= |br+s-t|+|bu+s-t|-|br+bu+s-t|-|s-t|. \end{aligned}$$

Now split the integral over $dG(s)$ and $dG(t)$ into a piece $\mathcal{R}_b^0 = \{s,t:\ s,t\in\mathcal{I}_b,\ |s-t|<2b\}\subset \mathcal{I}_b\times\mathcal{I}_b$ and $\mathcal{R}_b^1 = (\mathcal{I}_b\times\mathcal{I}_b)\backslash\mathcal{R}_b^0$. For the integral over $\mathcal{R}_b^0$, note that for $|s-t|<2b$, $|\zeta_b(s,r,t,u)|<Cb$. At the same time, the area of integration for $\mathcal{R}_b^0$ is of order $b^d$. So with $g$ and $\kappa_0$ bounded, the integral over $\mathcal{R}_b^0$ is of order $b^{d-1}\to 0$, and makes a vanishing contribution to $D_b^2$.

For any $\omega, v\in\mathbb{R}^d$ and $x\in\mathbb{R}$ such that $\omega+xv\neq 0$, we have

$$\begin{aligned} \frac{\partial}{\partial x}|\omega+xv| &= \frac{(\omega+xv)'v}{|\omega+xv|} \\ \frac{\partial^2}{\partial x^2}|\omega+xv| &= -\frac{((\omega+xv)'v)^2}{|\omega+xv|^3}+\frac{v'v}{|\omega+xv|} \\ \frac{\partial^3}{\partial x^3}|\omega+xv| &= 3\frac{((\omega+xv)'v)^3}{|\omega+xv|^5}-3\frac{((\omega+xv)'v)v'v}{|\omega+xv|^3}. \end{aligned}$$

For the integral over $\mathcal{R}_b^1$ where $|s-t|\geq 2b$, apply a second order Taylor expansion to $\zeta_b(s,r,t,u)g(s+br)g(t+bu)$ around $b=0$. Since $\zeta_0(s,r,t,u)=\partial\zeta_b(s,r,t,u)/\partial b|_{b=0}=0$, we find

$$\zeta_b(s,r,t,u)g(s+br)g(t+bu)=\tfrac{1}{2}b^2g(s)g(t)\left(\frac{(s-t)'r(s-t)'u}{|s-t|^3}-\frac{r'u}{|s-t|}\right)+\frac{b^3}{|s-t|^2}\Psi_b(s,r,t,u)$$

where here and below $\Psi_b$ denote uniformly bounded functions, that is,

$\sup_{b>0,s,t\in\mathcal{I}_b,|u|\leq 1,|r|\leq 1}|\Psi_b(s,r,t,u)|<\infty$. By symmetry, for all $|s-t|>2b$

$$\int\int\kappa_0(|r|)\kappa_0(|u|)\left(\frac{(s-t)'r(s-t)'u}{|s-t|^3}-\frac{r'u}{|s-t|}\right)dudr=0.$$

Furthermore,

$$\begin{aligned}\int_{\mathcal{I}_b}\int_{\mathcal{I}_b}\min\left(\frac{b^3}{|s-t|^2},\tfrac{1}{2}b\right)dG(s)dG(t) &\leq C\int_{|s|<C}\min\left(\frac{b^3}{|s|^2},b\right)ds\\ &= C\int_0^C x^{d-1}\min\left(\frac{b^3}{x^2},b\right)dx=O(b^3\log(b)) \qquad \text{(S.2)}\end{aligned}$$

so $D_b^2\to 0$.

Given this first result, it is without loss of generality to assume that $\mathcal{S}^0$ does not contain the origin. Let $Q_b = b^{-1}\int_{\mathcal{I}_b}\int\kappa_0(|r|)(L(s+br)-L(s))g(s+br)drdG(s)$. We will show that $Q_b$ converges in mean square. We have

$$\mathbb{E}\left[Q_b\right]=\tfrac{1}{2}b^{-1}\int_{\mathcal{I}_b}\int\kappa_0(|r|)(|s+br|-|s|-b|r|)g(s+br)drdG(s).$$

By a fist order Taylor expansion, for $|s|\geq 2b$,

$$(|s+br|-|s|-b|r|)g(s+br)=bg(s)\left(\frac{s'r}{|s|}-|r|\right)+b^2\Psi_b(s,r)$$

and $\mathbb{E}\left[Q_b\right]\to -\frac{1}{2}\int|r|\kappa_0(|r|)dr\cdot\int g(s)^2ds$ follows from $\int(s'r)\kappa_0(|r|)dr=0$.

Note that for $(X_1,X_2,X_3,X_4)$ mean-zero multivariate normal with covariances $\sigma_{ij}=\mathbb{E}[X_iX_j]$, $\mathbb{E}[(X_1X_2-\sigma_{12})(X_3X_4-\sigma_{34})]=\sigma_{14}\sigma_{23}+\sigma_{13}\sigma_{24}$. We have

$$\begin{aligned}\zeta_b^0(s,t) &= 2\mathbb{E}[L(s)L(t)]=|s|+|t|-|s+t|\\ \zeta_b^1(s,r,t) &= 2\mathbb{E}[(L(s+br)-L(s))L(t)]=|br+s|-|s|+|s-t|-|br+s-t|\\ \zeta_b^1(t,u,s) &= 2\mathbb{E}[(L(t+bu)-L(t))L(s)].\end{aligned}$$

Thus,

$$\begin{aligned}4\,\mathrm{Var}[Q_b]&=4\mathbb{E}\left[\left(Q_b-\mathbb{E}\left[Q_b\right]\right)^2\right]\\ &=\int_{\mathcal{I}_b}\int_{\mathcal{I}_b}\int\int b^{-2}\kappa_0(|r|)\kappa_0(|u|)[\zeta_b^0(s,t)\zeta_b(s,r,t,u)g(s+br)g(t+bu)\\ &\qquad +\zeta_b^1(s,r,t)\zeta_b^1(t,u,s)g(s+br)g(t+bu)]dr\cdot du\cdot dG(s)dG(t)\end{aligned}$$

Split the integral again into integrals over $\mathcal{R}_b^0$ and $\mathcal{R}_b^1$. For the integral over $\mathcal{R}_b^0$, note that for $|s-t|<2b$, $|\zeta_b^0(s,t)\zeta_b(s,r,t,u)|<Cb^2$ and $|\zeta_b^1(s,r,t)\zeta_b^1(t,u,s)|<Cb^2$ uniformly. At the same time, the area of integration for $\mathcal{R}_b^0$ is of order $b^d$, so the integral over $\mathcal{R}_b^0$ is of order $b^d \to 0$, and makes a vanishing contribution to $\mathrm{Var}[Q_b]$.

For the integral over $\mathcal{R}_b^1$, the term involving $\zeta_b^0(s,t)\zeta_b(s,r,t,u)$ is negligible as shown above, since $\sup_{s,t\in\mathcal{I}_b}\zeta_b^0(s,t)<\infty$. For the remaining term, apply a second order Taylor expansion to $\zeta_b^1(s,r,t)\zeta_b^1(t,u,s)g(s+br)g(t+bu)$

$$\zeta_b^1(s,r,t)\zeta_b^1(t,u,s)g(s+br)g(t+bu) \\ = \tfrac{1}{2}b^2 g(s)g(t)\left(\frac{s'r}{|s|}-\frac{(s-t)'r}{|s-t|}\right)\left(\frac{t'u}{|t|}-\frac{(t-s)'u}{|s-t|}\right)+\frac{b^3}{|s-t|^2}\Psi_b^1(s,r,t,u)$$

since $\zeta_0^1(s,r,t)=\zeta_0^1(t,u,s)=0$. By symmetry, for all $|s-t|>2b$,

$$\int \kappa_0(|r|)\left(\frac{s'r}{|s|}-\frac{(s-t)'r}{|s-t|}\right)dr=0$$

so using (S.2) we conclude $\mathrm{Var}[Q_b]\to 0$.

Finally, the result for $J_c$ follows, since the measure of $(J_c-J_c(0))$ is absolutely continuous with respect to the measure of $L$, and $J_c(0)$ has finite second moment. □

Lemma 7 is a special case of the following more general result applied with $p=1$ and $\psi(s)=1$. We will use the following notation: let $k:\mathcal{S}^0\times\mathcal{S}^0\mapsto\mathbb{R}$ be a continuous positive definite kernel (not necessarily equal to the covariance kernel of Lévy-Brownian Motion), and let $\mathbf{\Sigma}_n$ be the $n\times n$ matrix with $l,\ell$th element equal to $k(s_l^0,s_\ell^0)$. Let $\mathcal{L}_G^2$ be the Hilbert space of function $\mathcal{S}^0\mapsto\mathbb{R}$ with inner product $\langle f_1,f_2\rangle=\int f_1(s)f_2(s)dG(s)$. Define $L_k:\mathcal{L}_G^2\mapsto\mathcal{L}_G^2$ as the linear operator $L_k(f)(s)=\int f(r)k(r,s)dG(r)$, and $L_{k,n}=\int f(r)k(r,s)dG_n(r)$.

**Lemma S.1.** *Suppose the $p\times 1$ vector $x_l$ is such that $x_l=\psi(s_l^0)$ for $l=1,\ldots,n$ for some continuous function $\psi:\mathcal{S}^0\mapsto\mathbb{R}^p$, and $\int\psi(s)\psi(s)'dG_n(s)=H_n\to H$ for some positive definite matrix $H$. Let $M$ and $M_n$ be the projection operators $M_n(f)(s)=f(s)-\int\psi(r)'f(r)dG_n(r)H_n^{-1}\psi(s)$ and $M(f)(s)=f(s)-\int\psi(r)'f(r)dG(r)H^{-1}\psi(s)$. Let $\hat{k}_n$, and $\bar{k}$ be the kernels corresponding to the linear operators $M_nL_{k,n}M_n$ and $ML_kM$, respectively, so that the $(l,\ell)$ element of $\mathbf{M}_X\mathbf{\Sigma}_{n,L}\mathbf{M}_X$ is given by $\hat{k}_n(s_l^0,s_\ell^0)$. Let $\bar{k}(s,r)=\sum_{i=1}^\infty\bar{\nu}_i\bar{\varphi}_i(s)\bar{\varphi}_i(r)$ with $\int\bar{\varphi}_i(s)\bar{\varphi}_j(s)dG(s)=\mathbf{1}[i=j]$, $\bar{\nu}_i\geq\bar{\nu}_{i+1}\geq 0$ be the spectral decomposition of $\bar{k}$. Define $\hat{\varphi}_i(\cdot)=n^{-1}\hat{\nu}_i^{-1}\sum_{l=1}^n r_{i,l}\hat{k}_n(\cdot,s_l^0)$, where $(\hat{\nu}_i,(r_{i,1},\ldots,r_{i,n})')$ is the ith eigenvalue/eigenvector pair of $\mathbf{M}_X\mathbf{\Sigma}_n\mathbf{M}_X$. If $\bar{\nu}_1>\bar{\nu}_2>\ldots>\bar{\nu}_q>\bar{\nu}_{q+1}$ and Condition 1 holds, then for any $q\geq 1$, $\sup_{s\in\mathcal{S}^0,1\leq i\leq q}|\hat{\varphi}_i(s)-\bar{\varphi}_i(s)|\to 0$ and $\max_{1\leq i\leq q}|\hat{\nu}_i-\bar{\nu}_i|\to 0$.*

*Proof.* The proof follows from the same arguments as the proof of Lemma 6 in Müller and Watson (2022a). The two differences are (i) the generalization of the demeaning by the more general projection of $\psi$; and (ii) the replacement of the i.i.d. assumption for $s_l^0$ by $G_n \Rightarrow G$.

Set $k_0(s,r) = \bar{k}(s,r) + \psi(s)'H^{-1}\psi(r)$ and define the associated operators $L(f)(s) = \int f(r)k_0(r,s)dG(r)$, $L_n(f)(s) = \int f(r)k_0(r,s)dG_n(r)$, $\bar{L} = MLM$, $\bar{L}_n = ML_nM$ and $\hat{L}_n = M_nL_nM_n$. Note that $\bar{L} = ML_kM$ and $\hat{L}_n = M_nL_{k,n}M_n$. Let $\mathcal{H} \subset \mathcal{L}_G^2$ be the Reproducing Kernel Hilbert Space of functions $f : \mathcal{S}^0 \mapsto \mathbb{R}$ with kernel $k_0$ and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ satisfying $\langle f, k_0(\cdot, r)\rangle_{\mathcal{H}} = f(r)$ and associated norm $||f||_{\mathcal{H}}$. By Theorem 2.16 in Saitoh and Sawano (2016), $\mathcal{H}$ contains all functions of the form $a'\psi$ for $a \in \mathbb{R}^p$, so $\sup_{|a|=1} ||a'\psi||_{\mathcal{H}} < \infty$. Now proceed as in the proof of Lemma 6 of Müller and Watson (2022a) to argue that $\sup_{r\in\mathcal{S}^0} |f(r)| \leq \sqrt{\sup_{s\in\mathcal{S}^0} k_0(s,s)} \cdot ||f||_{\mathcal{H}}$, and

$$||Mf||_{\mathcal{H}} = ||f - \int \psi(r)'f(r)dG(r)H^{-1}\psi||_{\mathcal{H}} \leq ||f||_{\mathcal{H}} + \sup_{r\in\mathcal{S}^0} |f(r)| \cdot \sup_{r\in\mathcal{S}^0} |H^{-1}\psi(r)| \cdot \sup_{|a|=1} ||a'\psi||_{\mathcal{H}}$$

so $M : \mathcal{H} \mapsto \mathcal{H}$ is a bounded operator. By the same argument, so is $M_n$.

From $\langle f, k_0(\cdot, r)\rangle_{\mathcal{H}} = f(r)$, we further obtain

$$\int \psi(r)f(r)(dG_n(r) - dG(r)) = \left\langle f, \int \psi(r)k_0(\cdot, r)(dG_n(r) - dG(r)) \right\rangle_{\mathcal{H}} \tag{S.3}$$

and for each component $\psi_i$ of $\psi$, $i = 1, \ldots, p$,

$$\begin{aligned} & \left\| \int \psi_i(r)k_0(\cdot, r)(dG_n(r) - dG(r)) \right\|_{\mathcal{H}}^2 \qquad \text{(S.4)} \\ = \;& \int\int \psi_i(s)k_0(s,r)\psi_i(r)(dG_n(s) - dG(s))(dG_n(r) - dG(r)) \\ = \;& \mathbb{E}[\psi_i(S_n)k_0(S_n, R_n)\psi_i(R_n) - \psi_i(S_n)k_0(S, R_n)\psi_i(R) \\ & \qquad -\psi_i(S)k_0(S_n, R)\psi_i(R_n) + \psi_i(S)k_0(S, R)\psi_i(R)] \\ \to \;& 0 \end{aligned}$$

where $(S_n, R_n)$ is a sequence of $\mathbb{R}^{2d}$ random variables with distribution $G_n \times G_n$ converging to $(S, R)$ with distribution $G \times G$. The convergence then follows since the $\mathbb{R}^{2d} \mapsto \mathbb{R}$ function $(s,r) \mapsto \psi_i(s)k_0(s,r)\psi_i(r)$ is continuous and bounded. Thus, by (S.3), (S.4) and Cauchy-Schwarz,

$$\sup_{||f||_{\mathcal{H}}\leq 1} \left| \int \psi(r)f(r)(dG_n(r) - dG(r)) \right| \to 0.$$

From $H_n^{-1} \to H^{-1}$ and $|\int \psi(r)f(r)dG_n(r)| \leq \sup_{r\in\mathcal{S}^0} |f(r)| \cdot \sup_{r\in\mathcal{S}^0} |\psi(r)| \leq \sup_{r\in\mathcal{S}^0} |\psi(r)| \sqrt{\sup_{s\in S} k_0(s,s)} \cdot ||f||_{\mathcal{H}}$, we conclude that with $\Delta_n(f) = H_n^{-1} \int \psi(r)f(r)dG_n(r) -$

$H^{-1}\int \psi(r)f(r)dG(r)$, $\sup_{||f||_{\mathcal{H}}\leq 1}|\Delta_n(f)| \to 0$. Thus

$$\sup_{||f||_{\mathcal{H}}\leq 1} ||(M_n - M)f||_{\mathcal{H}} = \left\|\Delta_n(f)'\psi\right\|_{\mathcal{H}} \leq \sup_{||f||_{\mathcal{H}}\leq 1} |\Delta_n(f)| \cdot \sup_{|a|=1} ||a'\psi||_{\mathcal{H}} \to 0.$$

The only remaining piece of the proof is to show that $||L_n - L||_{HS}^2 \to 0$ under the assumption of $G_n \Rightarrow G$, where for any Hilbert-Schmidt operator $A : \mathcal{H} \mapsto \mathcal{H}$, $||A||_{HS}^2 = \sum_{j\geq 1}\langle Ae_j, Ae_j\rangle_{\mathcal{H}}$ for an orthonormal base $e_j$. One choice for $e_j$ are the eigenfunctions scaled by the square root of the eigenvalues of the spectral decomposition of $k_0$, so that $k_0(r,s) = \sum_{j=1}^{\infty} e_j(r)e_j(s)$; see the discussion in the proof of Lemma 6 in Müller and Watson (2022a). We find

$$\begin{aligned}
||L_n - L||_{HS}^2 &= \sum_{j\geq 1}\left\langle \int e_j(s)k_0(s,\cdot)(dG_n(s) - dG(s)), \int e_j(s)k_0(s,\cdot)(dG_n(s) - dG(s))\right\rangle_{\mathcal{H}} \\
&= \int\int \left(\sum_{j\geq 1} e_j(s)e_j(r)\right) k_0(s,r)(dG_n(s) - dG(s))(dG_n(r) - dG(r)) \\
&= \int\int k_0(s,r)^2(dG_n(r) - dG(r))(dG_n(r) - dG(r)) \\
&= \mathbb{E}[k_0(S_n,R_n)^2 - k_0(S,R_n)^2 - k_0(S_n,R)^2 + k_0(S,R)^2] \to 0
\end{aligned}$$

where 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. □

**Lemma S.2.** *Assume the conditions of Lemma S.1 hold. Suppose $\tilde{x}_l = \psi_n(s_l^0)$, where the continuous functions $\psi_n : \mathcal{S}^0 \mapsto \mathbb{R}^p$ are such that $\sup_{s\in\mathcal{S}^0}|\psi_n(s) - \psi(s)| \to 0$, for some continuous function $\psi$. Define the the projection operator $\tilde{M}_n : \mathcal{L}_G^2 \mapsto \mathcal{L}_G^2$ as $\tilde{M}_n(f)(s) = f(s) - \int \psi_n(r)'f(r)dG_n(r)H_n^{-1}\psi_n(s)$, and let $\tilde{k}_n$ be the kernel corresponding to the linear operator $\tilde{M}_n L_{k,n}\tilde{M}_n$, so that the $(l,\ell)$ element of $\mathbf{M}_{\tilde{X}}\mathbf{\Sigma}_n\mathbf{M}_{\tilde{X}}$ is given by $\tilde{k}_n(s_l^0, s_\ell^0)$. Let $(\tilde{\nu}_i, (\tilde{r}_{i,1},\ldots,\tilde{r}_{i,n})')$ be the ith eigenvalue/eigenvector pair of $\mathbf{M}_{\tilde{X}}\mathbf{\Sigma}_n\mathbf{M}_{\tilde{X}}$, and define $\tilde{\varphi}_i(\cdot) = n^{-1}\tilde{\nu}_i^{-1}\sum_{l=1}^n \tilde{r}_{i,l}\tilde{k}_n(\cdot, s_l^0)$. Then $\sup_{s\in\mathcal{S}^0, 1\leq i\leq q}|\tilde{\varphi}_i(s) - \bar{\varphi}_i(s)| \to 0$ and $\max_{1\leq i\leq q}|\tilde{\nu}_i - \bar{\nu}_i| \to 0$.*

*Proof.* From standard arguments, we obtain $\int \psi_n(s)\psi_n(s)'dG_n(s) \to H$ and $\int \psi(s)\psi_n(s)'dG_n(s) \to H$. Thus, $||\mathbf{M}_{\tilde{X}} - \mathbf{M}_X|| \to 0$, and by a direct calculation, $\sup_{s,r\in\mathcal{S}^0}|\tilde{k}_n(r,s) - \hat{k}_n(r,s)| \to 0$, and $\sup_{s,r\in\mathcal{S}^0}|\hat{k}_n(r,s) - \bar{k}(r,s)| \to 0$ and thus $\sup_{s,r\in\mathcal{S}^0}|\tilde{k}_n(r,s) - \bar{k}(r,s)| \to 0$. Furthermore, proceeding as in the proof of Lemma S.1 shows that $||\mathbf{\Sigma}_n||$ converges to $\bar{\nu}_1$, the largest eigenvalue of the integral operator with kernel $\bar{k}$, so $||\mathbf{\Sigma}_n|| = O(1)$. Thus also $||\mathbf{M}_{\tilde{X}}\mathbf{\Sigma}_n\mathbf{M}_{\tilde{X}} - \mathbf{M}_X\mathbf{\Sigma}_n\mathbf{M}_X|| \to 0$, and from Weyl's inequality, $\max_{1\leq i\leq q}|\tilde{\nu}_i - \hat{\nu}_i| \to 0$. Since also $\max_{1\leq i\leq q}|\hat{\nu}_i - \bar{\nu}_i| \to 0$ from Lemma S.1, we can conclude that

$$\sup_{s\in\mathcal{S}^0}|(\tilde{\nu}_i^{-1} - \hat{\nu}_i^{-1})n^{-1}\sum_{l=1}^n r_{i,l}\hat{k}_n(s, s_l^0)| \leq |\tilde{\nu}_i^{-1} - \hat{\nu}_i^{-1}| \cdot \sup_{s\in\mathcal{S}^0}|\hat{\varphi}_i(s)| \cdot \sup_{s,r\in\mathcal{S}^0}|\hat{k}_n(r,s)| \to 0$$

where the inequality uses $r_{i,l} = \hat{\varphi}_i(s_l^0)$, and the convergence follows from the above results and $\sup_{s\in\mathcal{S}^0} |\hat{\varphi}_i(s)| \to \sup_{s\in\mathcal{S}^0} |\varphi_i(s)| < \infty$ from Lemma S.1. Also,

$$\sup_{s\in\mathcal{S}^0} |n^{-1}\sum_{l=1}^{n} r_{i,l}(\tilde{k}_n(s,s_l^0) - \hat{k}_n(s,s_l^0))| \leq \sup_{s\in\mathcal{S}^0} |\hat{\varphi}_i(s)| \cdot \sup_{r,s\in\mathcal{S}^0} |\tilde{k}_n(r,s) - \hat{k}(r,s)| \to 0.$$

Finally, since $\max_{1\leq i\leq q} |\tilde{\nu}_i - \bar{\nu}_i| \to 0$ and $\bar{\nu}_1 > \bar{\nu}_2 > \ldots > \bar{\nu}_q > \bar{\nu}_{q+1}$, we can apply Corollary 1 of Yu, Wang and Samworth (2015) and conclude that $n^{-1}\sum_{l=1}^{n}(\tilde{r}_{i,l} - r_{i,l})^2 \to 0$ for $i = 1,\ldots,q$. Applying Cauchy-Schwarz then yields

$$\sup_{s\in\mathcal{S}^0} |n^{-1}\sum_{l=1}^{n}(\tilde{r}_{i,l} - r_{i,l})\tilde{k}_n(s,s_l^0)|^2 \leq n^{-1}\sum_{l=1}^{n}(\tilde{r}_{i,l} - r_{i,l})^2 \cdot \sup_{s\in\mathcal{S}^0} n^{-1}\sum_{l=1}^{n}\tilde{k}_n(s,s_l^0)^2 \to 0$$

where the convergence follows from $n^{-1}\sum_{l=1}^{n}\tilde{k}_n(s,s_l^0)^2 \leq 2\sup_{r,s\in\mathcal{S}^0} |\bar{k}(r,s)|^2 + 2\sup_{s,r\in\mathcal{S}^0} |\tilde{k}_n(r,s) - \bar{k}(r,s)|^2 = O(1)$. □

**Theorem S.3.** *Suppose* $y_l = x_l'\beta + u_l$, $(x_l', u_l) = \lambda_n^{1/2}(X_n^0(s_l^0)', U_n^0(s_l^0)) \in \mathbb{R}^p \times \mathbb{R}$ *with* $(X_n^0(\cdot), U_n^0(\cdot))$ *satisfying (27), but* $X^0$ *is not necessarily independent of* $U^0$*. Let* $\mathbf{R}_n^X$ *be the* $n \times p$ *matrix of* $q$ *eigenvectors of* $\mathbf{M}_X\mathbf{\Sigma}_L\mathbf{M}_X$ *corresponding to the largest eigenvalues. Suppose for almost every realization of* $X^0$*, the largest* $q+1$ *eigenvalues of the kernel* $k_{X^0} : \mathcal{S}^0 \times \mathcal{S}^0 \mapsto \mathbb{R}$ *corresponding to the linear operator* $M_{X^0}L_kM_{X^0}$ *with* $M_{X^0}(f)(s) = f(s) - X^0(s)\left(\int X^0(r)X^0(r)'dG(r)\right)^{-1}\int X^0(r)'f(r)dG(r)$ *are distinct. If also Condition 1 holds, then*

$$\lambda_n^{-1/2}\mathbf{R}_n^{X\prime}\mathbf{Y}_n \Rightarrow \omega\int \boldsymbol{\varphi}_{X^0}(s)U^0(s)dG(s) \tag{S.5}$$

*where* $\boldsymbol{\varphi}_{X^0}(\cdot)$ *are the* $q$ *eigenfunctions of* $k_{X^0}$ *corresponding to the largest eigenvalues.*

*Furthermore, let* $\tilde{U}_n^0$ *be independent of* $(X_n^0, U_n^0)$*, and suppose* $\tilde{U}_n^0$ *satisfies* $\tilde{U}_n^0(\cdot) \Rightarrow \tilde{U}^0(\cdot)$ *with* $\tilde{U}^0 \sim U^0$*. Let* $\mathrm{cv}_n(X_n^0)$ *be the* $1-\alpha$ *quantile of the conditional distribution of* $\phi(\mathbf{R}_n^{X\prime}\tilde{\mathbf{U}}_n)$ *given* $\mathbf{R}_n^X$ *for some continuous function* $\phi : \mathbb{R}^q \mapsto \mathbb{R}$ *satisfying* $\phi(ax) = \phi(x)$ *for all* $a \neq 0$ *and* $x \in \mathbb{R}^q$*. Suppose that (i)* $X^0$ *is independent of* $U^0$*, (ii) for almost all realizations of* $X^0$ *the conditional distribution of* $\phi(\int \boldsymbol{\varphi}_{X^0}(s)U^0(s)dG(s))$ *is continuous. Then* $\mathbb{P}(\phi(\mathbf{R}_n^{X\prime}\mathbf{Y}_n) > \mathrm{cv}_n(X_n^0)) \to \alpha$*.*

*Proof.* We will show that $(\phi(\mathbf{R}_n^{X\prime}\mathbf{Y}_n), \mathrm{cv}_n(X_n^0)) \Rightarrow (\phi(\int \boldsymbol{\varphi}_X(s)U^0(s)dG(s)), q_{1-\alpha}^{\phi}(X^0))$ with $q_{1-\alpha}^{\phi}(X^0)$ the $1-\alpha$ quantile of $\phi(\int \boldsymbol{\varphi}_X(s)U^0(s)dG(s))$ conditional on $X^0$. The result then follows from the CMT applied to $\mathbf{1}[\phi(\mathbf{R}_n^{X\prime}\mathbf{Y}_n) > \mathrm{cv}_n(X_n^0)]$, and taking expectations.

Apply the almost sure representation theorem to argue that there exists a probability space $(\Omega_0, \mathfrak{F}_0, P_0)$ and associated random processes $X^*, U^*$ and $X_n^*, U_n^*$, $n \geq 1$ such that $(X_n^*, U_n^*) \sim (X_n^0, U_n^0)$, $(X^*, U^*) \sim (X^0, U^0)$ and $\sup_{s\in\mathcal{S}^0} |X_n^*(s) - X^*(s)| \overset{a.s.}{\to} 0$, $\sup_{s\in\mathcal{S}^0} |U_n^*(s) - U^*(s)| \overset{a.s.}{\to} 0$. Using the same arguments as in the proof of Theorem 3, and a realization by realization application of Lemma S.2, then yields

$$\lambda_n^{-1/2}\mathbf{R}_n^{X^*\prime}\mathbf{Y}_n^* \stackrel{a.s.}{\to} \omega \int \boldsymbol{\varphi}_{X^*}(s)U^*(s)dG(s) \sim \omega \int \boldsymbol{\varphi}_{X^*}(s)U^0(s)dG(s) \tag{S.6}$$

where $(\mathbf{R}_n^{X^*}, \mathbf{Y}_n^*)$ are defined analogously to $(\mathbf{R}_n^X, \mathbf{Y}_n)$ on $(\Omega_0, \mathfrak{F}_0, P_0)$, and $(\mathbf{R}_n^{X^*}, \mathbf{Y}_n^*) \sim (\mathbf{R}_n^X, \mathbf{Y}_n)$ by construction, so (S.5) holds.

The further result now follows if we can show that also $\text{cv}_n(X_n^*) \stackrel{a.s.}{\to} q_{1-\alpha}^{\phi}(X^*)$, since almost sure convergence implies convergence in distribution. To that end, note there exists a separate probability space $(\Omega_1, \mathfrak{F}_1, P_1)$ with associated sequences of random process $\tilde{U}^*$ and $\tilde{U}_n^*$ and such that $\tilde{U}_n^* \sim \tilde{U}_n^0$, $\tilde{U}^* \sim \tilde{U}^0 \sim U^0$ and $\sup_{s\in S^0} |\tilde{U}_n^*(s) - \tilde{U}^*(s)| \stackrel{a.s.}{\to} 0$. Form the product space $(\Omega_0 \times \Omega_1, \mathfrak{F}_0 \otimes \mathfrak{F}_1, P_0 \times P_1)$, so that on this new space, $(X^*, \{X_n^*\}_{n=1}^{\infty})$ is independent of $(\tilde{U}^*, \{\tilde{U}_n^*\}_{n=1}^{\infty})$ by construction. Use the same arguments as for (S.6) obtain that for $P_0$-almost all $\omega_0 \in \Omega_0$ and $P_1$-almost all $\omega_1 \in \Omega_1$, in obvious notation,

$$\lambda_n^{-1/2}\mathbf{R}_n^{X^*\prime}\tilde{\mathbf{U}}_n^* \to \int \boldsymbol{\varphi}_{X^*}(s)\tilde{U}^*(s)dG(s)$$

jointly with (S.6). But almost sure convergence implies convergence in distribution, and $\tilde{U}^* \sim U^0$, so for $P_0$-almost all $\omega_0 \in \Omega_0$, the distribution of $\lambda_n^{-1/2}\mathbf{R}_n^{X^*\prime}\tilde{\mathbf{U}}_n^*$ induced by $P_1$ converges to the conditional distribution of $\int \boldsymbol{\varphi}_{X^*}(s)U^0(s)dG(s)$ given $X^*$. Since $\phi$ is continuous and the conditional distribution is assumed continuous, this implies that for all such $\omega_0$, $\text{cv}_n(X_n^0) \stackrel{a.s.}{\to} q_{1-\alpha}^{\phi}(X^*)$. Thus $(\phi(\mathbf{R}_n^{X\prime}\mathbf{Y}_n), \text{cv}_n(X_n^0)) \sim (\phi(\mathbf{R}_n^{X^*\prime}\mathbf{Y}_n^*), \text{cv}_n(X_n^*)) \stackrel{a.s.}{\to} (\phi(\int \boldsymbol{\varphi}_{X^*}(s)U^*(s)dG(s)), q_{1-\alpha}^{\phi}(X^*)) \sim (\phi(\int \boldsymbol{\varphi}_{X^0}(s)U^0(s)dG(s)), q_{1-\alpha}^{\phi}(X^0))$, and the result follows, because almost sure convergence implies convergence in distribution. □

In applications, the theorem justifies use of a critical value for the test statistic $\phi(\mathbf{R}_n^{X\prime}\mathbf{Y}_n)$ that is equal to the $1-\alpha$ quantile of $\phi(\mathbf{R}_n^{X\prime}\tilde{\mathbf{U}}_n)$ conditional on $\mathbf{R}_n^X$, for some (pseudo-) random variable draws of $\tilde{u}_l = \tilde{\mathbf{U}}_n(s_l^0)$ that induce the same limiting process as the actual regression errors $u_l$. Since $\phi$ is assumed scale invariant, the scaling of $\tilde{u}_l$ is immaterial in this construction.

**Proof of Theorem 8:**

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

$$\lambda_n^{d/2} n^{-1}\mathbf{Z}_n \Rightarrow \mathcal{N}\left(\mathbf{0}, a\sigma_B(0)\int \bar{\boldsymbol{\varphi}}(s)\bar{\boldsymbol{\varphi}}(s)'dG(s) + \omega^2 \int \bar{\boldsymbol{\varphi}}(s)\bar{\boldsymbol{\varphi}}(s)'g(s)dG(s)\right) \tag{S.7}$$

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 $\text{LFST}_n$ statistic is scale invariant, its limiting distribution under (S.7) 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 $\liminf_{n\to\infty} \text{cv}_n^{\text{LFST}}$ is at least as large as the $1-\alpha$ quantile, say $\text{cv}_\chi^{\text{LFST}}$, of the (continuous) asymptotic distribution of $\text{LFST}_n$ for this value of $\chi$.

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 $\lambda_n c_* \geq c_{0.03}$, $\text{cv}_n^{\text{LFST}}$ is such that the $\text{LFST}_n$ test controls size under $B = J_{c^*}$. But since $B = J_{c_*}$ satisfies the assumptions of Lahiri (2003), this model induces the same limit (S.7), so its $1 - a$ quantile converges to $\text{cv}_\chi^{\text{LFST}}$, and the result follows. □

## S.3 Detailed Monte Carlo Results

The following tables summarize the distributions of the null rejection probability and average length of confidence intervals for each method and DGP across the 96 spatial designs described in Section 6.

Entries show the median across spatial locations and the values in parentheses are $5^{\text{th}}$ and $95^{\text{th}}$ percentiles.

# Method: OLS (C-SCPC)

Null Rejection Probability: k = 1

| DGP | |
|---|---|
| Levy-BM | 0.227 (0.202,0.267) |
| I(1) c=0.01 | 0.243 (0.217,0.276) |
| I(1) c=0.03 | 0.271 (0.243,0.312) |
| I(1) Matern | 0.249 (0.227,0.284) |
| J c=0.03 | 0.035 (0.032,0.040) |
| J c = 0.50 | 0.145 (0.131,0.168) |
| Br. Sheet | 0.254 (0.218,0.302) |

Null Rejection Probability: k = 5

| DGP | |
|---|---|
| Levy-BM | 0.196 (0.183,0.213) |
| I(1) c=0.01 | 0.198 (0.185,0.211) |
| I(1) c=0.03 | 0.225 (0.210,0.243) |
| I(1) Matern | 0.202 (0.189,0.218) |
| J c=0.03 | 0.038 (0.033,0.042) |
| J c = 0.50 | 0.145 (0.132,0.156) |
| Br. Sheet | 0.233 (0.205,0.259) |

Average Length: k = 1

| DGP | |
|---|---|
| Levy-BM | 1.133 (1.071,1.204) |
| I(1) c=0.01 | 1.338 (1.262,1.437) |
| I(1) c=0.03 | 1.419 (1.346,1.495) |
| I(1) Matern | 1.385 (1.325,1.453) |
| J c=0.03 | 0.497 (0.488,0.507) |
| J c = 0.50 | 1.030 (0.995,1.095) |
| Br. Sheet | 1.071 (1.003,1.146) |

Average Length: k = 5

| DGP | |
|---|---|
| Levy-BM | 0.854 (0.828,0.884) |
| I(1) c=0.01 | 1.101 (1.060,1.141) |
| I(1) c=0.03 | 1.181 (1.130,1.244) |
| I(1) Matern | 1.168 (1.101,1.219) |
| J c=0.03 | 0.484 (0.478,0.489) |
| J c = 0.50 | 0.833 (0.807,0.869) |
| Br. Sheet | 0.801 (0.750,0.858) |

## Method: Isotropic difference (C-SCPC)

Null Rejection Probability: k = 1

| DGP | b =0.030 | b =0.060 | b =0.090 | b =0.120 | b =0.150 |
|---|---|---|---|---|---|
| Levy-BM | 0.020 (0.016,0.024) | 0.022 (0.017,0.027) | 0.028 (0.023,0.035) | 0.034 (0.027,0.044) | 0.040 (0.033,0.055) |
| I(1) c=0.01 | 0.056 (0.046,0.065) | 0.045 (0.041,0.051) | 0.045 (0.039,0.056) | 0.049 (0.042,0.065) | 0.056 (0.045,0.074) |
| I(1) c=0.03 | 0.097 (0.080,0.112) | 0.079 (0.069,0.089) | 0.071 (0.062,0.083) | 0.072 (0.060,0.093) | 0.076 (0.063,0.105) |
| I(1) Matern | 0.079 (0.067,0.089) | 0.065 (0.057,0.073) | 0.059 (0.054,0.065) | 0.060 (0.053,0.073) | 0.065 (0.056,0.086) |
| J c=0.03 | 0.019 (0.015,0.024) | 0.021 (0.017,0.026) | 0.026 (0.021,0.031) | 0.029 (0.024,0.035) | 0.033 (0.027,0.038) |
| J c = 0.50 | 0.020 (0.016,0.024) | 0.022 (0.018,0.028) | 0.027 (0.022,0.034) | 0.033 (0.026,0.045) | 0.038 (0.031,0.055) |
| Br. Sheet | 0.042 (0.033,0.066) | 0.067 (0.050,0.117) | 0.092 (0.071,0.153) | 0.109 (0.086,0.175) | 0.120 (0.096,0.185) |

Null Rejection Probability: k = 5

| DGP | b =0.030 | b =0.060 | b =0.090 | b =0.120 | b =0.150 |
|---|---|---|---|---|---|
| Levy-BM | 0.023 (0.019,0.028) | 0.024 (0.020,0.030) | 0.029 (0.025,0.038) | 0.035 (0.029,0.049) | 0.042 (0.032,0.058) |
| I(1) c=0.01 | 0.059 (0.050,0.069) | 0.047 (0.042,0.052) | 0.045 (0.039,0.053) | 0.048 (0.042,0.064) | 0.053 (0.045,0.076) |
| I(1) c=0.03 | 0.096 (0.082,0.105) | 0.077 (0.069,0.088) | 0.068 (0.062,0.075) | 0.067 (0.060,0.081) | 0.071 (0.062,0.092) |
| I(1) Matern | 0.080 (0.069,0.089) | 0.064 (0.057,0.072) | 0.058 (0.051,0.065) | 0.058 (0.050,0.071) | 0.063 (0.053,0.081) |
| J c=0.03 | 0.022 (0.017,0.025) | 0.023 (0.019,0.028) | 0.026 (0.022,0.032) | 0.030 (0.025,0.037) | 0.032 (0.028,0.040) |
| J c = 0.50 | 0.022 (0.019,0.026) | 0.024 (0.019,0.028) | 0.028 (0.023,0.036) | 0.033 (0.028,0.045) | 0.039 (0.032,0.056) |
| Br. Sheet | 0.047 (0.037,0.079) | 0.072 (0.055,0.131) | 0.090 (0.072,0.162) | 0.108 (0.086,0.175) | 0.120 (0.097,0.182) |

Average Length: k = 1

| DGP | b =0.030 | b =0.060 | b =0.090 | b =0.120 | b =0.150 |
|---|---|---|---|---|---|
| Levy-BM | 0.465 (0.410,0.533) | 0.415 (0.384,0.454) | 0.428 (0.400,0.483) | 0.473 (0.433,0.563) | 0.531 (0.475,0.625) |
| I(1) c=0.01 | 0.705 (0.640,0.783) | 0.636 (0.588,0.686) | 0.644 (0.599,0.720) | 0.701 (0.634,0.809) | 0.762 (0.690,0.893) |
| I(1) c=0.03 | 0.824 (0.772,0.932) | 0.736 (0.683,0.822) | 0.729 (0.680,0.791) | 0.770 (0.712,0.858) | 0.838 (0.770,0.947) |
| I(1) Matern | 0.843 (0.779,0.928) | 0.746 (0.699,0.837) | 0.733 (0.689,0.803) | 0.764 (0.723,0.868) | 0.819 (0.766,0.958) |
| J c=0.03 | 0.465 (0.405,0.517) | 0.403 (0.377,0.435) | 0.404 (0.374,0.436) | 0.418 (0.389,0.463) | 0.436 (0.405,0.483) |
| J c = 0.50 | 0.462 (0.417,0.541) | 0.411 (0.382,0.441) | 0.426 (0.399,0.472) | 0.467 (0.431,0.552) | 0.518 (0.473,0.620) |
| Br. Sheet | 0.536 (0.478,0.595) | 0.498 (0.468,0.543) | 0.517 (0.486,0.569) | 0.542 (0.510,0.610) | 0.575 (0.543,0.661) |

Average Length: k = 5

| DGP | b =0.030 | b =0.060 | b =0.090 | b =0.120 | b =0.150 |
|---|---|---|---|---|---|
| Levy-BM | 0.449 (0.405,0.502) | 0.402 (0.376,0.435) | 0.425 (0.394,0.472) | 0.468 (0.427,0.534) | 0.514 (0.471,0.600) |
| I(1) c=0.01 | 0.661 (0.607,0.711) | 0.606 (0.570,0.656) | 0.633 (0.591,0.706) | 0.691 (0.641,0.797) | 0.756 (0.703,0.868) |
| I(1) c=0.03 | 0.779 (0.716,0.817) | 0.705 (0.658,0.745) | 0.717 (0.676,0.785) | 0.774 (0.723,0.885) | 0.844 (0.775,0.965) |
| I(1) Matern | 0.786 (0.738,0.859) | 0.721 (0.684,0.772) | 0.728 (0.690,0.785) | 0.777 (0.728,0.868) | 0.839 (0.778,0.942) |
| J c=0.03 | 0.456 (0.408,0.506) | 0.393 (0.372,0.425) | 0.397 (0.374,0.429) | 0.415 (0.387,0.450) | 0.433 (0.403,0.471) |
| J c = 0.50 | 0.449 (0.408,0.495) | 0.403 (0.377,0.430) | 0.422 (0.391,0.476) | 0.464 (0.430,0.542) | 0.512 (0.471,0.605) |
| Br. Sheet | 0.506 (0.464,0.562) | 0.480 (0.452,0.517) | 0.498 (0.464,0.535) | 0.527 (0.489,0.576) | 0.557 (0.519,0.624) |

## Method: Cluster fixed-effects (clustered standard error)

Null Rejection Probability: k = 1

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.168 (0.155,0.178) | 0.139 (0.130,0.148) | 0.105 (0.098,0.111) | 0.076 (0.072,0.082) |
| I(1) c=0.01 | 0.263 (0.239,0.277) | 0.281 (0.263,0.296) | 0.285 (0.261,0.310) | 0.238 (0.217,0.257) |
| I(1) c=0.03 | 0.350 (0.331,0.367) | 0.390 (0.363,0.412) | 0.412 (0.391,0.435) | 0.369 (0.336,0.391) |
| I(1) Matern | 0.305 (0.284,0.322) | 0.339 (0.318,0.360) | 0.364 (0.337,0.390) | 0.326 (0.296,0.347) |
| J c=0.03 | 0.092 (0.086,0.097) | 0.080 (0.076,0.085) | 0.070 (0.066,0.075) | 0.066 (0.061,0.070) |
| J c = 0.50 | 0.140 (0.132,0.149) | 0.117 (0.109,0.124) | 0.093 (0.087,0.100) | 0.075 (0.070,0.081) |
| Br. Sheet | 0.282 (0.243,0.339) | 0.258 (0.219,0.310) | 0.213 (0.185,0.262) | 0.133 (0.116,0.162) |

Null Rejection Probability: k = 5

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.175 (0.164,0.185) | 0.142 (0.130,0.151) | 0.109 (0.101,0.116) | 0.083 (0.078,0.088) |
| I(1) c=0.01 | 0.271 (0.255,0.287) | 0.283 (0.265,0.297) | 0.284 (0.268,0.298) | 0.243 (0.230,0.265) |
| I(1) c=0.03 | 0.348 (0.326,0.366) | 0.378 (0.356,0.399) | 0.398 (0.374,0.414) | 0.356 (0.338,0.375) |
| I(1) Matern | 0.311 (0.288,0.328) | 0.338 (0.315,0.355) | 0.363 (0.339,0.378) | 0.327 (0.306,0.342) |
| J c=0.03 | 0.097 (0.092,0.104) | 0.084 (0.079,0.090) | 0.074 (0.071,0.079) | 0.072 (0.068,0.076) |
| J c = 0.50 | 0.149 (0.142,0.160) | 0.123 (0.115,0.133) | 0.098 (0.092,0.105) | 0.079 (0.074,0.084) |
| Br. Sheet | 0.295 (0.256,0.340) | 0.266 (0.231,0.312) | 0.221 (0.188,0.285) | 0.141 (0.124,0.178) |

Average Length: k = 1

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.353 (0.342,0.364) | 0.307 (0.299,0.317) | 0.294 (0.288,0.301) | 0.355 (0.347,0.364) |
| I(1) c=0.01 | 0.474 (0.458,0.495) | 0.412 (0.396,0.431) | 0.382 (0.365,0.408) | 0.442 (0.420,0.474) |
| I(1) c=0.03 | 0.501 (0.480,0.520) | 0.424 (0.406,0.448) | 0.389 (0.363,0.410) | 0.441 (0.415,0.469) |
| I(1) Matern | 0.497 (0.481,0.515) | 0.430 (0.407,0.453) | 0.392 (0.369,0.414) | 0.450 (0.422,0.478) |
| J c=0.03 | 0.275 (0.271,0.280) | 0.264 (0.260,0.269) | 0.272 (0.267,0.276) | 0.342 (0.337,0.348) |
| J c = 0.50 | 0.343 (0.335,0.352) | 0.302 (0.295,0.312) | 0.291 (0.287,0.297) | 0.354 (0.347,0.361) |
| Br. Sheet | 0.376 (0.358,0.402) | 0.329 (0.312,0.351) | 0.314 (0.303,0.328) | 0.380 (0.361,0.398) |

Average Length: k = 5

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.334 (0.327,0.344) | 0.297 (0.289,0.307) | 0.288 (0.283,0.295) | 0.349 (0.343,0.356) |
| I(1) c=0.01 | 0.448 (0.438,0.463) | 0.395 (0.386,0.411) | 0.371 (0.358,0.383) | 0.424 (0.407,0.444) |
| I(1) c=0.03 | 0.476 (0.463,0.492) | 0.411 (0.399,0.426) | 0.382 (0.369,0.395) | 0.431 (0.412,0.450) |
| I(1) Matern | 0.475 (0.463,0.491) | 0.414 (0.400,0.429) | 0.384 (0.366,0.398) | 0.433 (0.413,0.454) |
| J c=0.03 | 0.269 (0.264,0.274) | 0.260 (0.255,0.265) | 0.269 (0.264,0.273) | 0.338 (0.333,0.344) |
| J c = 0.50 | 0.327 (0.320,0.336) | 0.293 (0.287,0.300) | 0.286 (0.282,0.292) | 0.347 (0.342,0.352) |
| Br. Sheet | 0.347 (0.337,0.362) | 0.313 (0.303,0.326) | 0.305 (0.294,0.316) | 0.370 (0.355,0.381) |

## Method: Cluster fixed-effects (C-SCPC)

Null Rejection Probability: k = 1

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.053 (0.045,0.062) | 0.056 (0.044,0.073) | 0.056 (0.046,0.065) | 0.047 (0.040,0.061) |
| I(1) c=0.01 | 0.084 (0.076,0.094) | 0.097 (0.080,0.129) | 0.112 (0.091,0.132) | 0.102 (0.081,0.133) |
| I(1) c=0.03 | 0.122 (0.112,0.134) | 0.132 (0.116,0.175) | 0.157 (0.134,0.183) | 0.150 (0.126,0.173) |
| I(1) Matern | 0.098 (0.089,0.112) | 0.114 (0.097,0.149) | 0.134 (0.118,0.166) | 0.127 (0.107,0.150) |
| J c=0.03 | 0.030 (0.026,0.035) | 0.034 (0.029,0.044) | 0.041 (0.034,0.049) | 0.042 (0.035,0.049) |
| J c = 0.50 | 0.043 (0.039,0.051) | 0.048 (0.039,0.064) | 0.050 (0.041,0.062) | 0.046 (0.040,0.059) |
| Br. Sheet | 0.104 (0.082,0.145) | 0.106 (0.080,0.150) | 0.105 (0.081,0.150) | 0.076 (0.058,0.103) |

Null Rejection Probability: k = 5

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.053 (0.048,0.061) | 0.056 (0.046,0.073) | 0.060 (0.048,0.076) | 0.051 (0.041,0.063) |
| I(1) c=0.01 | 0.080 (0.073,0.090) | 0.098 (0.078,0.139) | 0.122 (0.104,0.147) | 0.116 (0.097,0.136) |
| I(1) c=0.03 | 0.107 (0.096,0.118) | 0.129 (0.107,0.178) | 0.177 (0.153,0.202) | 0.165 (0.146,0.188) |
| I(1) Matern | 0.090 (0.081,0.103) | 0.102 (0.088,0.157) | 0.149 (0.129,0.179) | 0.144 (0.130,0.163) |
| J c=0.03 | 0.030 (0.026,0.035) | 0.036 (0.030,0.047) | 0.043 (0.034,0.054) | 0.046 (0.040,0.058) |
| J c = 0.50 | 0.044 (0.038,0.051) | 0.050 (0.039,0.067) | 0.055 (0.047,0.067) | 0.050 (0.042,0.059) |
| Br. Sheet | 0.106 (0.090,0.145) | 0.115 (0.085,0.163) | 0.109 (0.086,0.152) | 0.083 (0.064,0.114) |

Average Length: k = 1

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.550 (0.530,0.572) | 0.447 (0.420,0.468) | 0.393 (0.373,0.411) | 0.453 (0.431,0.471) |
| I(1) c=0.01 | 0.809 (0.774,0.841) | 0.697 (0.648,0.744) | 0.627 (0.588,0.664) | 0.683 (0.632,0.723) |
| I(1) c=0.03 | 0.907 (0.876,0.942) | 0.813 (0.745,0.854) | 0.737 (0.683,0.775) | 0.773 (0.721,0.820) |
| I(1) Matern | 0.878 (0.848,0.916) | 0.773 (0.712,0.822) | 0.708 (0.647,0.747) | 0.763 (0.716,0.806) |
| J c=0.03 | 0.405 (0.392,0.418) | 0.370 (0.348,0.382) | 0.352 (0.337,0.365) | 0.433 (0.409,0.458) |
| J c = 0.50 | 0.527 (0.510,0.546) | 0.433 (0.411,0.451) | 0.386 (0.368,0.401) | 0.449 (0.424,0.466) |
| Br. Sheet | 0.620 (0.576,0.675) | 0.514 (0.476,0.559) | 0.455 (0.419,0.485) | 0.502 (0.454,0.532) |

Average Length: k = 5

| DGP | m = 30 | m = 60 | m = 120 | m = 240 |
|---|---|---|---|---|
| Levy-BM | 0.530 (0.513,0.548) | 0.433 (0.410,0.455) | 0.379 (0.363,0.397) | 0.444 (0.424,0.467) |
| I(1) c=0.01 | 0.795 (0.765,0.820) | 0.680 (0.595,0.722) | 0.586 (0.552,0.615) | 0.626 (0.596,0.672) |
| I(1) c=0.03 | 0.901 (0.872,0.922) | 0.778 (0.687,0.824) | 0.659 (0.622,0.692) | 0.702 (0.651,0.741) |
| I(1) Matern | 0.878 (0.842,0.906) | 0.780 (0.661,0.814) | 0.655 (0.618,0.685) | 0.699 (0.665,0.731) |
| J c=0.03 | 0.401 (0.390,0.412) | 0.363 (0.344,0.378) | 0.346 (0.330,0.362) | 0.427 (0.404,0.444) |
| J c = 0.50 | 0.513 (0.500,0.527) | 0.420 (0.394,0.445) | 0.375 (0.356,0.390) | 0.439 (0.421,0.459) |
| Br. Sheet | 0.583 (0.556,0.619) | 0.493 (0.454,0.532) | 0.435 (0.408,0.461) | 0.483 (0.450,0.516) |

# Method: LBM-GLS

Null Rejection Probability: k = 1

| DGP | |
|---|---|
| Levy-BM | 0.053 (0.049,0.057) |
| I(1) c=0.01 | 0.256 (0.244,0.267) |
| I(1) c=0.03 | 0.392 (0.374,0.412) |
| I(1) Matern | 0.379 (0.359,0.396) |
| J c=0.03 | 0.058 (0.055,0.062) |
| J c = 0.50 | 0.053 (0.050,0.056) |
| Br. Sheet | 0.234 (0.204,0.298) |

Null Rejection Probability: k = 5

| DGP | |
|---|---|
| Levy-BM | 0.054 (0.051,0.058) |
| I(1) c=0.01 | 0.257 (0.243,0.268) |
| I(1) c=0.03 | 0.392 (0.377,0.408) |
| I(1) Matern | 0.380 (0.363,0.400) |
| J c=0.03 | 0.060 (0.056,0.063) |
| J c = 0.50 | 0.054 (0.051,0.057) |
| Br. Sheet | 0.234 (0.206,0.300) |

Average Length: k = 1

| DGP | |
|---|---|
| Levy-BM | 0.195 (0.195,0.195) |
| I(1) c=0.01 | 0.212 (0.209,0.215) |
| I(1) c=0.03 | 0.224 (0.219,0.231) |
| I(1) Matern | 0.222 (0.215,0.229) |
| J c=0.03 | 0.196 (0.195,0.196) |
| J c = 0.50 | 0.195 (0.195,0.196) |
| Br. Sheet | 0.208 (0.199,0.213) |

Average Length: k = 5

| DGP | |
|---|---|
| Levy-BM | 0.195 (0.195,0.195) |
| I(1) c=0.01 | 0.212 (0.208,0.214) |
| I(1) c=0.03 | 0.224 (0.218,0.229) |
| I(1) Matern | 0.223 (0.218,0.228) |
| J c=0.03 | 0.196 (0.195,0.196) |
| J c = 0.50 | 0.195 (0.195,0.195) |
| Br. Sheet | 0.208 (0.199,0.212) |

# Method: LBM-GLS (C-SCPC)

Null Rejection Probability: k = 1

| DGP | |
|---|---|
| Levy-BM | 0.030 (0.027,0.035) |
| I(1) c=0.01 | 0.049 (0.043,0.055) |
| I(1) c=0.03 | 0.069 (0.060,0.076) |
| I(1) Matern | 0.059 (0.051,0.066) |
| J c=0.03 | 0.029 (0.025,0.033) |
| J c = 0.50 | 0.030 (0.027,0.035) |
| Br. Sheet | 0.088 (0.072,0.125) |

Null Rejection Probability: k = 5

| DGP | |
|---|---|
| Levy-BM | 0.031 (0.027,0.035) |
| I(1) c=0.01 | 0.050 (0.043,0.056) |
| I(1) c=0.03 | 0.069 (0.061,0.078) |
| I(1) Matern | 0.059 (0.052,0.067) |
| J c=0.03 | 0.029 (0.025,0.033) |
| J c = 0.50 | 0.030 (0.027,0.034) |
| Br. Sheet | 0.085 (0.072,0.132) |

Average Length: k = 1

| DGP | |
|---|---|
| Levy-BM | 0.254 (0.251,0.257) |
| I(1) c=0.01 | 0.419 (0.408,0.430) |
| I(1) c=0.03 | 0.541 (0.524,0.559) |
| I(1) Matern | 0.545 (0.523,0.562) |
| J c=0.03 | 0.264 (0.260,0.266) |
| J c = 0.50 | 0.255 (0.252,0.258) |
| Br. Sheet | 0.333 (0.319,0.349) |

Average Length: k = 5

| DGP | |
|---|---|
| Levy-BM | 0.256 (0.253,0.258) |
| I(1) c=0.01 | 0.419 (0.408,0.430) |
| I(1) c=0.03 | 0.536 (0.517,0.553) |
| I(1) Matern | 0.547 (0.528,0.565) |
| J c=0.03 | 0.266 (0.262,0.268) |
| J c = 0.50 | 0.257 (0.253,0.259) |
| Br. Sheet | 0.335 (0.320,0.347) |

## Method: Low-pass eigenvector

Null Rejection Probability: k = 1

| DGP | q = 10 | q = 20 | q = 50 |
|---|---|---|---|
| Levy-BM | 0.050 (0.046,0.054) | 0.050 (0.047,0.054) | 0.050 (0.046,0.053) |
| I(1) c=0.01 | 0.051 (0.047,0.054) | 0.052 (0.049,0.056) | 0.064 (0.060,0.068) |
| I(1) c=0.03 | 0.053 (0.050,0.057) | 0.063 (0.058,0.067) | 0.105 (0.099,0.110) |
| I(1) Matern | 0.051 (0.047,0.055) | 0.055 (0.052,0.059) | 0.082 (0.077,0.087) |
| J c=0.03 | 0.100 (0.093,0.107) | 0.094 (0.088,0.099) | 0.078 (0.074,0.083) |
| J c = 0.50 | 0.056 (0.052,0.060) | 0.054 (0.050,0.059) | 0.052 (0.048,0.055) |
| Br. Sheet | 0.128 (0.095,0.171) | 0.160 (0.120,0.209) | 0.210 (0.170,0.272) |

Null Rejection Probability: k = 5

| DGP | q = 10 | q = 20 | q = 50 |
|---|---|---|---|
| Levy-BM | 0.050 (0.046,0.054) | 0.050 (0.046,0.054) | 0.050 (0.048,0.054) |
| I(1) c=0.01 | 0.050 (0.047,0.054) | 0.051 (0.048,0.056) | 0.062 (0.059,0.066) |
| I(1) c=0.03 | 0.052 (0.048,0.055) | 0.060 (0.057,0.063) | 0.101 (0.096,0.107) |
| I(1) Matern | 0.050 (0.046,0.053) | 0.054 (0.050,0.058) | 0.080 (0.074,0.085) |
| J c=0.03 | 0.095 (0.089,0.100) | 0.095 (0.088,0.099) | 0.079 (0.075,0.083) |
| J c = 0.50 | 0.054 (0.050,0.057) | 0.054 (0.050,0.057) | 0.052 (0.048,0.055) |
| Br. Sheet | 0.104 (0.080,0.135) | 0.147 (0.119,0.180) | 0.201 (0.168,0.243) |

Average Length: k = 1

| DGP | q = 10 | q = 20 | q = 50 |
|---|---|---|---|
| Levy-BM | 1.507 (1.499,1.515) | 0.960 (0.957,0.963) | 0.574 (0.573,0.575) |
| I(1) c=0.01 | 1.508 (1.500,1.515) | 0.960 (0.956,0.964) | 0.574 (0.573,0.575) |
| I(1) c=0.03 | 1.507 (1.500,1.516) | 0.960 (0.957,0.964) | 0.574 (0.573,0.576) |
| I(1) Matern | 1.508 (1.499,1.518) | 0.961 (0.956,0.964) | 0.574 (0.572,0.576) |
| J c=0.03 | 1.509 (1.496,1.517) | 0.960 (0.956,0.964) | 0.574 (0.573,0.576) |
| J c = 0.50 | 1.508 (1.499,1.513) | 0.960 (0.957,0.963) | 0.574 (0.573,0.575) |
| Br. Sheet | 1.507 (1.498,1.518) | 0.959 (0.956,0.967) | 0.574 (0.572,0.576) |

Average Length: k = 5

| DGP | q = 10 | q = 20 | q = 50 |
|---|---|---|---|
| Levy-BM | 2.299 (2.279,2.317) | 1.101 (1.095,1.106) | 0.601 (0.599,0.602) |
| I(1) c=0.01 | 2.297 (2.280,2.313) | 1.100 (1.096,1.105) | 0.600 (0.599,0.602) |
| I(1) c=0.03 | 2.297 (2.276,2.317) | 1.100 (1.095,1.105) | 0.600 (0.599,0.602) |
| I(1) Matern | 2.298 (2.283,2.316) | 1.101 (1.095,1.105) | 0.600 (0.599,0.602) |
| J c=0.03 | 2.297 (2.274,2.318) | 1.100 (1.095,1.104) | 0.600 (0.599,0.602) |
| J c = 0.50 | 2.300 (2.282,2.320) | 1.101 (1.096,1.106) | 0.600 (0.599,0.602) |
| Br. Sheet | 2.300 (2.283,2.322) | 1.101 (1.095,1.106) | 0.600 (0.598,0.602) |

## Method: High-pass eigenvector (C-SCPC)

Null Rejection Probability: k = 1

| DGP | q = 5 | q = 10 | q = 20 | q = 50 | q = 100 |
|---|---|---|---|---|---|
| Levy-BM | 0.129 (0.117,0.139) | 0.095 (0.087,0.103) | 0.070 (0.063,0.078) | 0.050 (0.045,0.056) | 0.042 (0.037,0.046) |
| I(1) c=0.01 | 0.174 (0.160,0.184) | 0.141 (0.132,0.152) | 0.118 (0.106,0.128) | 0.090 (0.081,0.099) | 0.069 (0.061,0.078) |
| I(1) c=0.03 | 0.215 (0.205,0.234) | 0.183 (0.168,0.197) | 0.150 (0.137,0.167) | 0.111 (0.096,0.128) | 0.081 (0.071,0.097) |
| I(1) Matern | 0.193 (0.180,0.206) | 0.165 (0.152,0.180) | 0.146 (0.131,0.159) | 0.118 (0.106,0.136) | 0.097 (0.077,0.121) |
| J c=0.03 | 0.050 (0.045,0.054) | 0.051 (0.046,0.056) | 0.050 (0.045,0.055) | 0.045 (0.040,0.049) | 0.040 (0.035,0.044) |
| J c = 0.50 | 0.120 (0.112,0.133) | 0.093 (0.086,0.099) | 0.070 (0.064,0.076) | 0.050 (0.045,0.055) | 0.041 (0.037,0.047) |
| Br. Sheet | 0.213 (0.186,0.270) | 0.192 (0.163,0.246) | 0.167 (0.141,0.221) | 0.132 (0.113,0.174) | 0.099 (0.084,0.136) |

Null Rejection Probability: k = 5

| DGP | q = 5 | q = 10 | q = 20 | q = 50 | q = 100 |
|---|---|---|---|---|---|
| Levy-BM | 0.125 (0.116,0.134) | 0.093 (0.087,0.101) | 0.070 (0.065,0.078) | 0.051 (0.045,0.057) | 0.041 (0.037,0.046) |
| I(1) c=0.01 | 0.161 (0.151,0.170) | 0.135 (0.125,0.147) | 0.114 (0.106,0.126) | 0.089 (0.078,0.102) | 0.068 (0.061,0.078) |
| I(1) c=0.03 | 0.200 (0.187,0.212) | 0.173 (0.161,0.184) | 0.144 (0.133,0.158) | 0.108 (0.094,0.126) | 0.082 (0.070,0.097) |
| I(1) Matern | 0.179 (0.167,0.188) | 0.157 (0.147,0.168) | 0.139 (0.129,0.153) | 0.118 (0.104,0.134) | 0.095 (0.080,0.113) |
| J c=0.03 | 0.051 (0.046,0.054) | 0.051 (0.048,0.056) | 0.051 (0.046,0.054) | 0.045 (0.042,0.051) | 0.040 (0.036,0.044) |
| J c = 0.50 | 0.117 (0.108,0.128) | 0.090 (0.085,0.096) | 0.069 (0.063,0.074) | 0.050 (0.045,0.057) | 0.041 (0.037,0.045) |
| Br. Sheet | 0.203 (0.182,0.249) | 0.183 (0.161,0.232) | 0.160 (0.140,0.214) | 0.129 (0.108,0.174) | 0.100 (0.083,0.138) |

Average Length: k = 1

| DGP | q = 5 | q = 10 | q = 20 | q = 50 | q = 100 |
|---|---|---|---|---|---|
| Levy-BM | 0.565 (0.552,0.578) | 0.467 (0.459,0.476) | 0.391 (0.382,0.399) | 0.328 (0.322,0.335) | 0.320 (0.314,0.325) |
| I(1) c=0.01 | 0.744 (0.720,0.770) | 0.647 (0.625,0.671) | 0.558 (0.541,0.576) | 0.464 (0.450,0.479) | 0.428 (0.414,0.441) |
| I(1) c=0.03 | 0.789 (0.755,0.825) | 0.690 (0.656,0.715) | 0.587 (0.567,0.617) | 0.489 (0.468,0.510) | 0.449 (0.427,0.466) |
| I(1) Matern | 0.788 (0.759,0.820) | 0.690 (0.666,0.720) | 0.607 (0.581,0.628) | 0.521 (0.492,0.542) | 0.501 (0.466,0.522) |
| J c=0.03 | 0.419 (0.412,0.425) | 0.388 (0.381,0.394) | 0.353 (0.349,0.359) | 0.318 (0.314,0.324) | 0.317 (0.313,0.322) |
| J c = 0.50 | 0.558 (0.542,0.575) | 0.465 (0.455,0.475) | 0.389 (0.383,0.399) | 0.329 (0.322,0.334) | 0.320 (0.314,0.325) |
| Br. Sheet | 0.592 (0.571,0.614) | 0.523 (0.498,0.551) | 0.472 (0.449,0.498) | 0.423 (0.401,0.447) | 0.409 (0.394,0.427) |

Average Length: k = 5

| DGP | q = 5 | q = 10 | q = 20 | q = 50 | q = 100 |
|---|---|---|---|---|---|
| Levy-BM | 0.535 (0.524,0.549) | 0.456 (0.447,0.467) | 0.386 (0.380,0.394) | 0.329 (0.322,0.334) | 0.322 (0.317,0.327) |
| I(1) c=0.01 | 0.735 (0.711,0.755) | 0.647 (0.623,0.665) | 0.557 (0.542,0.576) | 0.465 (0.448,0.480) | 0.428 (0.413,0.443) |
| I(1) c=0.03 | 0.792 (0.764,0.820) | 0.693 (0.667,0.719) | 0.594 (0.573,0.618) | 0.491 (0.471,0.511) | 0.448 (0.430,0.467) |
| I(1) Matern | 0.786 (0.765,0.812) | 0.697 (0.677,0.725) | 0.613 (0.594,0.634) | 0.526 (0.498,0.548) | 0.501 (0.477,0.525) |
| J c=0.03 | 0.412 (0.406,0.418) | 0.383 (0.378,0.389) | 0.352 (0.346,0.357) | 0.318 (0.314,0.324) | 0.319 (0.314,0.325) |
| J c = 0.50 | 0.533 (0.520,0.545) | 0.455 (0.446,0.462) | 0.387 (0.380,0.393) | 0.329 (0.323,0.335) | 0.322 (0.316,0.327) |
| Br. Sheet | 0.551 (0.529,0.575) | 0.498 (0.476,0.514) | 0.454 (0.436,0.471) | 0.413 (0.398,0.430) | 0.404 (0.390,0.417) |

# Method: Ibragimov-Müller

Null Rejection Probability: k = 1

| DGP | m = 10 | m = 20 | m = 50 |
|---|---|---|---|
| Levy-BM | 0.105 (0.090,0.117) | 0.105 (0.096,0.114) | 0.080 (0.072,0.087) |
| I(1) c=0.01 | 0.125 (0.110,0.137) | 0.144 (0.131,0.157) | 0.154 (0.137,0.168) |
| I(1) c=0.03 | 0.152 (0.130,0.166) | 0.193 (0.174,0.207) | 0.235 (0.211,0.254) |
| I(1) Matern | 0.134 (0.115,0.147) | 0.163 (0.149,0.175) | 0.193 (0.179,0.206) |
| J c=0.03 | 0.062 (0.058,0.067) | 0.062 (0.056,0.067) | 0.053 (0.047,0.058) |
| J c = 0.50 | 0.088 (0.081,0.095) | 0.087 (0.082,0.094) | 0.070 (0.063,0.077) |
| Br. Sheet | 0.182 (0.132,0.223) | 0.198 (0.156,0.234) | 0.158 (0.131,0.193) |

Null Rejection Probability: k = 5

| DGP | m = 10 | m = 20 | m = 50 |
|---|---|---|---|
| Levy-BM | 0.084 (0.077,0.091) | 0.076 (0.068,0.082) | 0.048 (0.043,0.052) |
| I(1) c=0.01 | 0.091 (0.082,0.098) | 0.091 (0.081,0.098) | 0.062 (0.057,0.069) |
| I(1) c=0.03 | 0.104 (0.092,0.114) | 0.114 (0.102,0.121) | 0.080 (0.072,0.087) |
| I(1) Matern | 0.092 (0.082,0.101) | 0.098 (0.088,0.105) | 0.072 (0.064,0.079) |
| J c=0.03 | 0.060 (0.055,0.064) | 0.055 (0.049,0.060) | 0.043 (0.039,0.047) |
| J c = 0.50 | 0.075 (0.070,0.081) | 0.068 (0.060,0.072) | 0.045 (0.042,0.051) |
| Br. Sheet | 0.142 (0.115,0.169) | 0.135 (0.106,0.159) | 0.070 (0.063,0.085) |

Average Length: k = 1

| DGP | m = 10 | m = 20 | m = 50 |
|---|---|---|---|
| Levy-BM | 0.587 (0.575,0.599) | 0.442 (0.432,0.470) | 0.418 (0.379,0.479) |
| I(1) c=0.01 | 0.871 (0.851,0.899) | 0.696 (0.680,0.722) | 0.627 (0.583,0.707) |
| I(1) c=0.03 | 1.004 (0.976,1.046) | 0.798 (0.780,0.824) | 0.701 (0.656,0.789) |
| I(1) Matern | 0.964 (0.942,0.988) | 0.782 (0.762,0.807) | 0.709 (0.664,0.768) |
| J c=0.03 | 0.365 (0.357,0.376) | 0.330 (0.320,0.345) | 0.375 (0.329,0.438) |
| J c = 0.50 | 0.550 (0.537,0.564) | 0.428 (0.417,0.442) | 0.407 (0.376,0.455) |
| Br. Sheet | 0.609 (0.590,0.643) | 0.468 (0.454,0.488) | 0.435 (0.397,0.473) |

Average Length: k = 5

| DGP | m = 10 | m = 20 | m = 50 |
|---|---|---|---|
| Levy-BM | 0.480 (0.472,0.491) | 0.411 (0.393,0.481) | 0.461 (0.385,0.539) |
| I(1) c=0.01 | 0.755 (0.740,0.770) | 0.647 (0.614,0.727) | 0.652 (0.552,0.767) |
| I(1) c=0.03 | 0.867 (0.851,0.886) | 0.730 (0.708,0.805) | 0.668 (0.577,0.821) |
| I(1) Matern | 0.883 (0.863,0.910) | 0.783 (0.761,0.873) | 0.780 (0.617,0.935) |
| J c=0.03 | 0.359 (0.349,0.372) | 0.357 (0.338,0.448) | 0.454 (0.359,0.527) |
| J c = 0.50 | 0.467 (0.457,0.477) | 0.404 (0.383,0.466) | 0.478 (0.396,0.571) |
| Br. Sheet | 0.503 (0.495,0.522) | 0.435 (0.413,0.531) | 0.487 (0.390,0.571) |

## $R^2$ values in OLS regression

k = 1

| DGP | |
|---|---|
| Levy-BM | 0.137 (0.125,0.162) |
| I(1) c=0.01 | 0.179 (0.166,0.204) |
| I(1) c=0.03 | 0.208 (0.197,0.238) |
| I(1) Matern | 0.192 (0.179,0.215) |
| J c=0.03 | 0.010 (0.010,0.011) |
| J c = 0.50 | 0.085 (0.079,0.099) |
| Br. Sheet | 0.139 (0.117,0.161) |

k = 5

| DGP | |
|---|---|
| Levy-BM | 0.434 (0.419,0.471) |
| I(1) c=0.01 | 0.561 (0.548,0.592) |
| I(1) c=0.03 | 0.638 (0.626,0.664) |
| I(1) Matern | 0.595 (0.584,0.625) |
| J c=0.03 | 0.049 (0.047,0.050) |
| J c = 0.50 | 0.314 (0.298,0.354) |
| Br. Sheet | 0.443 (0.404,0.471) |

## Additional References

SAITOH, S., AND Y. SAWANO (2016): *Theory of Reproducing Kernels and Applications.* Springer, New York.

STEIN, M. L. (2002): "Fast and Exact Simulation of Fractional Brownian Surfaces," *Journal of Computational and Graphical Statistics,* 11, 587-599.

YU, Y., T. WANG AND R. J. SAMWORTH (2015): "A useful variant of the Davis-Kahan theorem for statisticians," *Biometrika,* 102, 315-323.