Econometrica, Vol. 92, No. 5 (September, 2024), 1661–1695

SPATIAL UNIT ROOTS AND SPURIOUS REGRESSION

ULRICH K. MÜLLER
Department of Economics, Princeton University

MARK W. WATSON
Department of Economics, Princeton University

This paper proposes a model for, and investigates the consequences of, strong spatial dependence in economic variables. Our findings echo those of the corresponding "unit root" time series literature: Spatial unit root processes induce spuriously significant regression results, even with clustered standard errors or spatial HAC corrections. We develop large-sample valid unit root and stationarity tests that can detect such strong spatial dependence. Finally, we use simulations to study strategies for valid inference in regressions with persistent spatial data, such as spatial analogues of first-differencing transformations. Regressions from Chetty, Hendren, Kline, and Saez (2014) are used to illustrate the issues and methods.

KEYWORDS: Spatial correlation, Lévy–Brownian motion, functional central limit theorem.

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 realizations 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 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 realizations 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.

Ulrich K. Müller: umueller@princeton.edu
Mark W. Watson: mwatson@princeton.edu

We thank two referees and participants at several seminars for helpful comments and suggestions. Müller acknowledges support from the National Science Foundation grant SES-2242455.

[1]Conley (1999) is a leading example of spatial HAC inference. See Müller and Watson (2022, 2023) for a discussion of the post-Conley literature and new suggestions for inference in regression models with weak spatial dependence.

 https://doi.org/10.3982/ECTA21654

This paper provides a systematic analysis of such strongly correlated spatial processes, and their impact on regression analysis. To this end, we define a natural spatial analogue of time series $I(1)$ processes, such as the process in Figure 1(b), and study their asymptotic behavior. We find that spatial $I(1)$ processes systematically induce spuriously significant regression coefficients, even if standard errors are clustered, one includes cluster fixed effects, or the standard errors are corrected by using a spatial HAC estimator. To avoid this potential pitfall, it is important to know whether a given variable exhibits such strong $I(1)$ spatial dependence, and we develop corresponding unit root and stationarity tests. Finally, spurious regression in time series can be avoided by using first differences of $I(1)$ processes. We investigate a number of spatial analogues of such "first-differencing" transformations in a simulation study, and find a particular (parameter-free) GLS transformation to be remarkably effective at yielding short and approximately correctly sized confidence intervals.

The practical implications are threefold. First, researchers should be mindful that strong spatial dependence in variables invalidates standard regression inference, even with clustered standard errors and clustered fixed effects. Second, one can detect such strong dependence using the spatial unit root and stationarity tests developed here. And third, a potential solution to restoring valid regression inference is to spatially difference strongly dependent series using the aforementioned GLS transformation.

Throughout the paper, we use spatial data and regressions from Chetty et al. (2014) to illustrate the issues and methods. These authors constructed an index of intergenerational mobility for commuting zones in the United States, and studied 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 of panel (b) highlights the empirical relevance of the issues and methods presented here.

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, $t = 1, 2, 3, \dots$, where t indexes months, quarters, years, etc. Economic variables observed in space are typically not so neatly arranged. For example, geographical data may be collected at potentially arbitrary locations s_l within a given region 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 locations measures the dissimilarity of the characteristic. We thus follow the geostatistical tradition of positing a continuous parameter model of spatial variation, rather than, say, model dependence by spatial autoregressive (SAR) models of Cliff and Ord (1974) and Anselin (1988).[2]

The roadmap of the paper is as follows. Section 2 provides our definition of a spatial $I(1)$ process, the basic building block of our analysis. In time series ($d = 1$ in our notation), the canonical $I(1)$ process is a Wiener process. Lévy–Brownian motion is a useful

[2]Gelfand et al. (2010) provided a useful overview. There is a small literature on unit roots and spurious regression in SAR models that we discuss in Section 5.

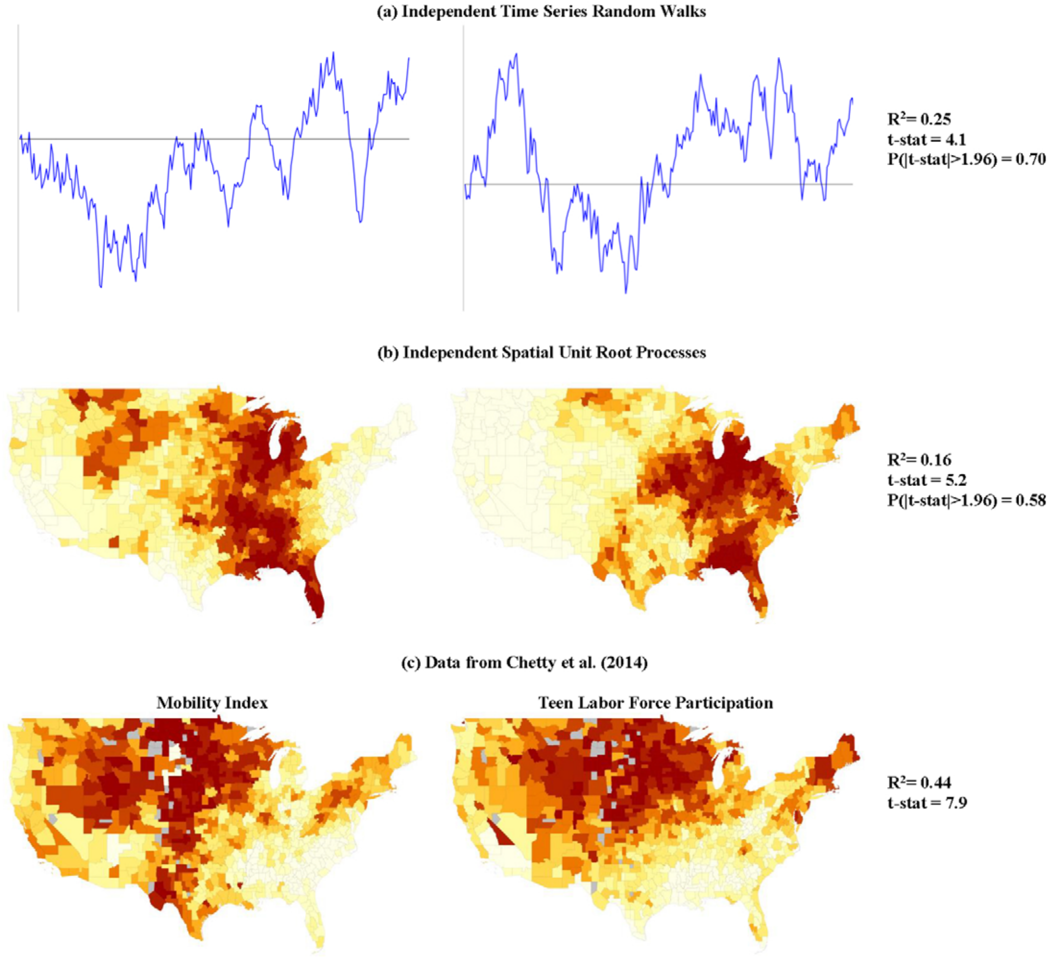

FIGURE 1.—Strongly dependent data in time and space. Notes: Panel (a) shows realizations from two independent Gaussian random walks with $T = 250$ together with the sample R^2 and HAC t-statistic. Also shown is the probability that the t-statistic from such a spurious regression exceeds 1.96 in absolute value. Panel (b) shows corresponding results from independent spatial unit root processes over $n = 722$ U.S. commuting zones. Panel (c) shows two variables from Chetty et al. (2014).

generalization of the Wiener process for $d > 1$, and Section 2.2 begins by reviewing its properties. More general time series $I(1)$ 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 an ARIMA$(p, 1, q)$ process. Section 2.3 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 2.4 provides a FCLT that is applicable to spatial $I(1)$ processes.

The time series literature has also considered so-called "local-to-unity" processes that are characterized by weak mean reversion as measured by a parameter $c > 0$. For small values of c, these processes behave very nearly like $I(1)$ processes, and for large c, they share many characteristics of weakly dependent $I(0)$ series. The local-to-unity model thus usefully turns the $I(0)/I(1)$ dichotomy into a continuous spectrum of dependence, which is also helpful for the systematic study of the robustness of inference procedures. Section 2.5 shows how to analogously generalize the spatial $I(1)$ model to a spatial "local-to-unity" process, and also provides 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 vector-valued Lévy–Brownian motion. Regression F-statistics (including HAC and clustered versions) diverge to infinity, even for local-to-unity processes. The bottom line is that researchers should be wary of spurious regressions using spatial data, just as they are using time series data.

Section 4 takes up the problem of conducting inference about the degree of spatial persistence in a scalar variable. The degree of spatial persistence can be interesting in its own right, or it can be motivated by a concern about spurious inference that arises with strongly persistent spatial variables. Several diagnostic tests are developed. 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 derive 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 a regression, yielding spatial analogues of the residual-based cointegration tests of 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. Section 5 uses simulations to study several spatial differencing methods including nearest-neighbor differences, local demeaning, local fixed effects, and a GLS transformation. When combined with spatial HAR standard errors, several of these differencing methods mitigate or eliminate spurious regression problems. We find the GLS transformation to be particularly effective. In contrast, inference methods that rely on clustered standard errors perform very poorly, even when cluster fixed effects are included.

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

2. SPATIAL $I(1)$ PROCESSES AND THEIR LIMITS

This section is divided into five subsections. The first defines some notation for the spatial environment. The second reviews Lévy–Brownian motion, a spatial generalization of the Wiener process. The third 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 which serves as a benchmark mean-reverting, but highly persistent spatial process.

2.1. *Setup and Notation*

Our analysis requires three ingredients: (i) the spatial sampling region under consideration, denoted by $\mathcal{S}$; (ii) the observed locations, $s_l \in \mathcal{S}$; and (iii) the stochastic process Y, that is defined on $\mathcal{S}$. Taken together, these ingredients describe the observations

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

We discuss the sampling region and observed locations in this subsection. The stochastic process Y is discussed in the following two subsections.

We utilize a large-sample framework and assume that the locations s_l, $l = 1, \dots, n$, are non-stochastic (or, equivalently, are independent of all other random elements). The locations are allowed to depend on n in a double-array fashion, but we abstract from this dependence 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 λ_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 $\mathcal{S}^0$.*

A familiar example helps clarify the sampling framework: consider a regularly spaced time series process observed at time periods $l = 1, \dots, n$, 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 example to a general spatial setting with a general prototypical sampling region $\mathcal{S}^0 \subset \mathbb{R}^d$ that grows at an arbitrary rate λ_n.

2.2. *Lévy–Brownian Motion*

Consider the usual time series $I(1)$ process $y_t = \sum_{s=1}^t u_s$, $t = 1, \dots, n$, where u_t is mean zero, covariance stationary, and weakly dependent (i.e., 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 next 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 plays 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

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

[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.

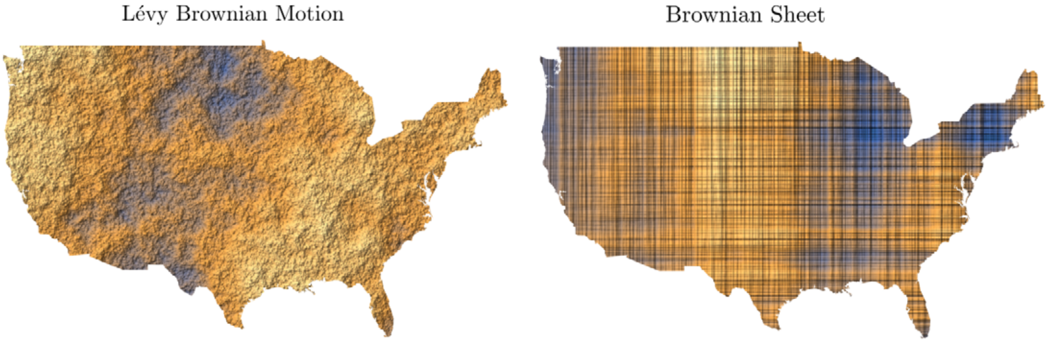

FIGURE 2.—Sample realizations of stochastic processes for $d = 2$.

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. 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 $|s - r|$, so Lévy–Brownian motion is invariant to rotations of the spatial axes, that is, $L(Os) \sim L(s)$ for any $d \times d$ rotation matrix O.

The left panel of Figure 2 plots a realization of L on the sampling region $\mathcal{S}_n$ representing the 48 contiguous U.S. states. The right panel shows a realization of a more familiar 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 isotropic, as is apparent from the sample realization. We therefore find Lévy–Brownian motion a more appealing generalization of the Wiener process for most applications, and thus define $y_l = L(s_l)$ as the *canonical unit root process* for $d > 1$.

2.2.1. *Two Representations of Lévy–Brownian Motion*

We take advantage of 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

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

where (ν_j, φ_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

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

where the right-hand side converges uniformly on $\mathcal{S}^0$ with probability 1 (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.

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),

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

where $\kappa_d > 0$ is a scalar chosen so that $\mathrm{Var}(L(s)) = 1$ when $|s| = 1$. The moving average weights $h(r, s)$ have poles of opposing signs at $|s - r|$ and $|r|$ for $d > 1$, but we do not have good intuition for their exact form.

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 et al. (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 socioeconomic 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 utilize the moving average representation (5) of Lévy–Brownian motion, replace the white noise innovations $dW(r)$ by a weakly dependent random field B, and 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.

The integral $\int_{\mathbb{R}^d} |h(r, s)|\, dr$ does not exist for $d > 1$, so Y in (6) is not defined path-wise 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. We make the following assumption.

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

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

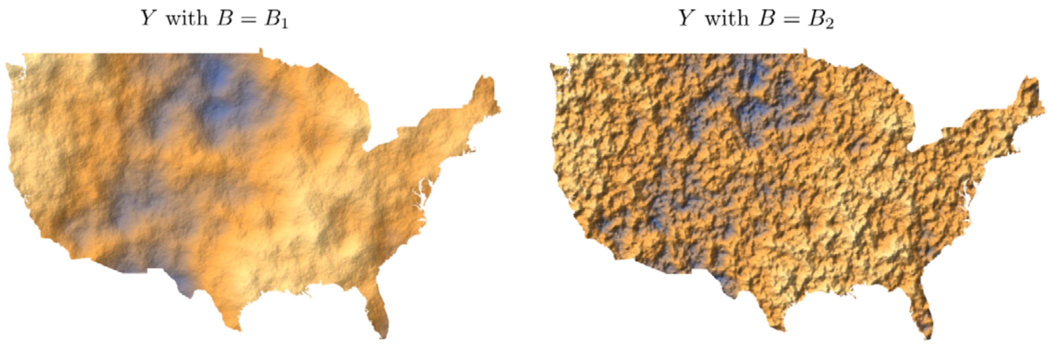

FIGURE 3.—Sample realizations of Y with different underlying B. Notes: B_1 and B_2 are zero-mean Gaussian processes with spectral densities $f_1(\omega) \propto 1/(|\omega|^2 + 100^2)^{3/2}$ and $f_2(\omega) \propto (|\omega|^2 + 50^2)^{3/2}/(|\omega|^2 + 200^2)^3$ for $\omega \in \mathbb{R}^2$, respectively, where the width of the contiguous United States is normalized to unity.

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.

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 continuous sample paths with probability* 1.

Figure 3 plots realizations from two spatial $I(1)$ 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 (see Appendix C for details). Evidently, 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

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

We make the following assumption about the process B.

CONDITION 3: *For some positive sequence $\zeta_n \to \infty$, let $\mathcal{R}_n = [-\zeta_n, \zeta_n]^d \subset \mathbb{R}^d$. If $f_n : \mathbb{R}^d \mapsto \mathbb{R}$ is a sequence of functions such that* $\limsup_{n\to\infty} \sup_{r\in\mathcal{R}_n} \zeta_n^{d/2}|f_n(r)| < \infty$ *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.

With this background, we can state a FCLT for spatial data.

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 λ_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 proportional to λ_n^d, so the average number of observations per unit of volume is proportional to n/λ_n^d, and $\lambda_n \asymp n^{1/d}$ corresponds to pure outfill. The theorem holds for any mixture of infill and outfill sampling as long as the overall sampling region diverges, $\lambda_n \to \infty$.

REMARK 2.2: 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 that imply 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 (LTU) model. LTU models exhibit strong persistence, but are stationary, and serve as a convenient bridge between $I(0)$ and $I(1)$ processes. In time series, the LTU model is employed to derive confidence intervals for autoregressive roots near unity (Stock (1991)), for local power analysis of unit root tests (Elliott, Rothenberg, and Stock (1996)), to understand the implications of local deviations from exact unit root specifications for inference about cointegration (Elliott (1998)), and for long-horizon economic forecasting (Müller and Watson (2016)), to mention just a handful of applications. Related spatial applications are easy to imagine and we consider some of these below. Here, we provide the requisite spatial generalization of the LTU model.

In the time series LTU model, y_t satisfies $n^{-1/2}(y_{\lfloor \cdot n \rfloor} - y_1) \Rightarrow \omega(J_c(\cdot) - J_c(0))$, with J_c a stationary Ornstein–Uhlenbeck (OU) process with covariance kernel $\mathbb{E}[J_c(s)J_c(r)] = \exp[-c|s-r|]/(2c)$, $c > 0$. 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)).

To generalize the LTU model 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 recognized as a member 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 (3.2.8) of Matérn (1986), J_c has the moving average representation

$$J_c(s) = \int_{\mathbb{R}^d} h_c(r, s)\,dW(r), \tag{8}$$

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 $\kappa_{c,d}$, where K_ν is the modified Bessel function of the second kind, $d \geq 1$.[4] We proceed as in the spatial $I(1)$ model (6) and replace the white noise term by the weakly dependent random field B,

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

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

and define the *spatial local-to-unity process* on $\mathcal{S}_n$ as the sequence of processes Y_{c/λ_n}. In this definition, the parameter c/λ_n is a drifting sequence, generalizing the corresponding local-to-unity time series device in which the largest autoregressive root is parameterized as $\rho_n = 1 - c/n$. The rate of this drift is such that the overall degree of mean reversion of

$$Y^0_{n,c}(s) = \lambda_n^{-1/2} Y_{c/\lambda_n}(\lambda_n s), \quad s \in \mathcal{S}^0 \tag{9}$$

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

In the Appendix, we show that under Condition 2, Y_{c/λ_n} exists on $\mathcal{S}_n$ for all n. Furthermore, under the conditions of Theorem 2, $Y^0_{n,c}$ in (9) satisfies $Y^0_{n,c}(\cdot) \Rightarrow \omega J_c(\cdot)$.

3. SPURIOUS REGRESSIONS WITH SPATIAL $I(1)$ VARIABLES

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

$$y_l = \alpha + x_l'\beta + u_l \tag{10}$$

for $l = 1, \dots, n$, 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 $R^2 = S_{\tilde{x}\tilde{y}}'S_{\tilde{x}\tilde{x}}^{-1}S_{\tilde{x}\tilde{y}}/s_{\tilde{y}\tilde{y}}$, and the classical (non-spatial-correlation robust, homoscedastic) F-statistic for testing $H_0: H\beta = 0$, where H is a non-stochastic matrix with $\operatorname{rank}(H) = m \le p$, is $F^{\text{Hom}} = \frac{n}{m}\hat{\beta}'H'(H'S_{\tilde{x}\tilde{x}}^{-1}H)^{-1}H\hat{\beta}/s_u^2$ with $s_u^2 = \frac{n}{n-p-1}(s_{\tilde{y}\tilde{y}} - S_{\tilde{x}\tilde{y}}'S_{\tilde{x}\tilde{x}}^{-1}S_{\tilde{x}\tilde{y}})$.

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

$$\begin{bmatrix} Y_n^0(\cdot) \\ X_n^0(\cdot) \end{bmatrix} = \begin{bmatrix} \lambda_n^{-1/2}Y(\lambda_n\cdot) \\ \lambda_n^{-1/2}X(\lambda_n\cdot) \end{bmatrix} \Rightarrow \begin{bmatrix} Y^0(\cdot) \\ X^0(\cdot) \end{bmatrix} \tag{11}$$

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 (11),*

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

An implication of part (iv) 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 $\text{cv} \ge 0$. We now show that this extends to statistics computed with heteroscedasticity and HAC-corrected standard errors (cf. Phillips (1998) for a corresponding time series result).

Consider the class of correlation-robust-HAC F-statistics

$$F^{\text{HAC}} = \frac{n}{m}\hat{\beta}'H'\left(HS_{\tilde{x}\tilde{x}}^{-1}\hat{\Omega}_n S_{\tilde{x}\tilde{x}}^{-1}H'\right)^{-1}H\hat{\beta}, \tag{12}$$

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

$$\hat{\Omega}_n = n^{-1}\sum_{l,\ell=1}^{n}\kappa\big(b_n(s_l - s_\ell)\big)e_l e_\ell', \tag{13}$$

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

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

The assumption of $\lambda_n^{-1}b_n^{-1} = o_p(1)$ ensures that in large samples, $\hat{\Omega}_n$ in (13) puts negligible weight on pairs of locations with $\lambda_n^{-1}|s_l - s_\ell| > \varepsilon$, for all positive ε. 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, heteroscedasticity robust standard errors correspond to $\kappa(r) = \mathbf{1}[r = 0]$, which also satisfies (14).

Alternatively, researchers sometimes employ clustered standard errors over larger regions to account for spatial dependence. The corresponding F^{clust} statistic has the same form as F^{HAC} in (12) with $\hat{\Omega}_n$ replaced by $\hat{\Omega}_n^{\mathrm{clust}} = n^{-1}\sum_{j=1}^{n_C}(\sum_{l\in C_j} e_l)(\sum_{l\in C_j} e_l)'$, where $C_j \subset \{1, 2, \ldots, n\}$ collects the observations 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 $\hat{\Omega}_n^{\mathrm{clust}}$ under weak dependence.

THEOREM 4: *Under Condition* 1, (11), *and* (14), $\mathbb{P}(F^{\mathrm{HAC}} > \mathrm{cv}) \to 1$ *and* $\mathbb{P}(F^{\mathrm{clust}} > \mathrm{cv}) \to 1$ *for any* $\mathrm{cv} \ge 0$.

REMARK 3.1: In contrast to spatial HAC inference, fixed-b type spatial HAR inference (Sun and Kim (2012), Bester, Conley, Hansen, and Vogelsang (2016)) does not lead to diverging F-statistics, and the spatial correlation robust inference derived in Müller and Watson (2022) explicitly accommodates some degree of "strong" persistence of the type exhibited by the spatial local-to-unity model for large enough c.

REMARK 3.2: Theorems 3 and 4 also hold for local-to-unity processes, that is, if $[Y^0(\cdot), X^0(\cdot)]$ in (11) are $p+1$ independent processes of the type (8), with arbitrary and potentially different mean-reversion parameters c. It is thus not only non-stationary unit root-type spatial dependence that leads to spurious regression, but standard cluster and HAC corrections are inadequate also for stationary forms of strong spatial dependence.

REMARK 3.3: It follows from the Karhunen–Loève representation of L in (4) and the FCLT result in Theorem 2 that the coefficients of a regression 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 $(\int\psi(r)\psi(r)'\,dG(r))^{-1}\omega\int\psi(r)L(r)\,dG(r)$ and are asymptotically significant as measured by a corresponding F^{Hom}, F^{HAC}, or F^{clust} statistic. Kelly (2019) observed such a phenomenon empirically in a number of applications with spatial data.

REMARK 3.4: In some spatial applications, it might be natural to consider the model

$$y_l = a_n^{-1} Y(s_l) + \varepsilon_l, \quad \varepsilon_l \sim \text{iid}(0, \sigma_\varepsilon^2), \tag{15}$$

where $Y(\cdot)$ is a spatial $I(1)$ process as in (11) independent of $\{\varepsilon_l\}$, and a_n is a nondecreasing positive sequence. In this model, the observations are noisy measurements of the $I(1)$ "local level" $a_n^{-1} Y(s_l)$. For instance, y_l might measure the income of a household at location s_l that is the sum of an idiosyncratic component (ε_l), and a highly persistent spatial component ($a_n^{-1} Y(s_l)$).

For $a_n = o(\lambda_n^{1/2})$, $Y_n^*(s_l) = y_l$ in (15) satisfies $\lambda_n^{-1/2} Y_n^*(\lambda_n \cdot) \Rightarrow Y^0(\cdot)$, just as in the model without the idiosyncratic noise. This is analogous to the classic Beveridge and Nelson (1981) observation: an $I(1)$ time series remains $I(1)$ even when noise is added. Our baseline spatial $I(1)$ model thus accommodates idiosyncratic noise, as long as the $I(1)$ component is shrinking more slowly than $\lambda_n^{-1/2}$. This rate corresponds to the local level component being of greater overall variability across all of $\mathcal{S}_n$ compared to the variation induced by ε_l.

Much smaller variation of the local level can still be problematic due to averaging: Consider a rate $a_n = a_0 \lambda_n^{1/2} n^{1/2}$ in (15), where the idiosyncratic variation strongly dominates the variation in the local level component (in fact, this rate is the local asymptotic alternative to test whether y_l is a pure i.i.d. process). In this case, the smooth averages $\sum_{l=1}^n \psi(\lambda_n^{-1} s_l) y_l$ will be affected by both the local level and the noise, $n^{-1/2} \sum_{l=1}^n \psi(\lambda_n^{-1} s_l) y_l \Rightarrow a_0^{-1} \int \psi(s) Y^0(s)\, dG(s) + Z$, with $Z \sim \mathcal{N}(0, \sigma_Z^2)$ and $\sigma_Z^2 = \sigma_\varepsilon^2 \int \psi(s)^2 dG(s)$, and where Z and Y^0 are independent. Thus, a regression of y_l on smoothly varying spatial characteristics, including spatial $I(1)$ processes, still suffers from spurious regression effects unless a_0 is very large.

Finally, if the regressors are not smooth in this sense because they are also generated by the local level model (15) and $a_n = a_0 \lambda_n^{1/2} n^{1/4}$ for both the regressors and the regressand, then a calculation shows that we again obtain spurious regression effects that dissipate only as $a_0 \to \infty$.

In summary, spurious regression remains a concern also for data that is characterized by "small" spatial $I(1)$ components.

4. INFERENCE FOR SPATIAL PERSISTENCE

In time series applications, researchers routinely rely on a set of procedures to diagnose strong persistence. These include unit root tests, "stationarity" tests, and confidence intervals for the largest roots in autoregressive polynomials. In this section, we develop analogous diagnostic tools for spatial variables.

The autoregressive representation for a discrete-time $I(1)$ time series process has a unit root in its autoregressive polynomial, making it straightforward to test for $I(1)$ persistence using Dickey–Fuller or related unit root tests. Spatial $I(1)$ processes as defined here do not have an analogous autoregressive representation, so these tests do not directly generalize. Similarly, popular time series stationarity tests (e.g., Nyblom (1989), Kwiatkowski, Phillips, Schmidt, and Shin (1992), and Elliott and Müller (2006)) are not directly applicable to spatial data. An alternative, non-regression-based approach to learn about time series persistence was 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.[5]

The intuition underlying this approach is straightforward: 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. Focusing 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.

The FCLT result in Theorem 2 suggests that little is lost by developing the various tests for the canonical models from Section 2 with $y_l = L(s_l)$ or $y_l = J_c(s_l)$, respectively, and for clarity we take this approach in Sections 4.1–4.5. Section 4.6 discusses the asymptotic validity of the tests under the more general assumptions for $I(1)$ and LTU processes given in Section 2, as well as for more general $I(0)$ processes.

4.1. *Dimension Reduction by Weighted Averages*

The suggested tests depend on the data only through q weighted averages. This section establishes corresponding notation and briefly reviews standard results for hypothesis tests under Gaussianity.

Let $\mathbf{Y} = (y_1, \dots, y_n)'$ and let $\mathbf{\Sigma}_L$ be the $n \times n$ covariance matrix of $\mathbf{Y}$ induced by Lévy–Brownian motion $y_l = L(s_l)$. We are interested in tests that are invariant to translation shifts $\mathbf{Y} \to \mathbf{Y} + a\mathbf{1}$, where $\mathbf{1}$ is a vector of ones. Such tests can be constructed from weighted averages of $\mathbf{Y}$ that sum to zero. Let $\mathbf{M} = \mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$, and $\mathbf{MY}$ denote the demeaned values of y. With $\mathrm{Var}(\mathbf{Y}) = \mathbf{\Sigma}_L$, we have $\mathrm{Var}(\mathbf{MY}) = \mathbf{M\Sigma}_L\mathbf{M}$. Let $\mathbf{R}$ be the $n \times q$ matrix of eigenvectors of $\mathbf{M\Sigma}_L\mathbf{M}$ corresponding to the q largest eigenvalues, where $\mathbf{R}$ satisfies $n^{-1}\mathbf{R}'\mathbf{R} = \mathbf{I}_q$. The columns of $\mathbf{R}$ extract the q linear combinations of $\mathbf{MY}$ with the largest variance; that is,

$$\mathbf{Z} = \mathbf{R}'\mathbf{MY} = \mathbf{R}'\mathbf{Y}$$

are the q largest principal components of $\mathbf{MY}$ under $\mathbf{MY} \sim (\mathbf{0}, \mathbf{M\Sigma}_L\mathbf{M})$. As in Müller and Watson (2008), we treat $\mathbf{Z}$ as the effective observation, that is, we seek to conduct inference about the persistence properties of $\mathbf{Y}$ with a test that is a function of $\mathbf{Z}$ only.

Different models for persistence in $\mathbf{Y}$ imply different values for $\mathrm{Var}(\mathbf{Z}) = \mathbf{\Omega}$, and this means that we can discriminate between the models by testing hypotheses concerning the value of $\mathbf{\Omega}$. Thus, consider the generic problem of testing $H_0 : \mathbf{\Omega} = \mathbf{\Omega}_0$ versus $H_a : \mathbf{\Omega} = \mathbf{\Omega}_a$ when $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$. A standard calculation shows that the most powerful level α scale invariant test rejects for large values of

$$\frac{\mathbf{Z}'\mathbf{\Omega}_0^{-1}\mathbf{Z}}{\mathbf{Z}'\mathbf{\Omega}_a^{-1}\mathbf{Z}} \tag{16}$$

with a critical value that equals the $1 - \alpha$ quantile of (16) under the null distribution $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_0)$.

[5]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.

Inference of this type depends on q, the number of weighted averages included in $\mathbf{Z}$. The choice of q faces a classic efficiency versus 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 set $q = 15$ in our numerical analysis.

4.2. *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, where the canonical models are $y_l = L(s_l)$ and $y_l = J_c(s_l)$. This yields $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_L)$ and $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(c))$, respectively, with the l, ℓ element of $\mathbf{\Sigma}(c)$ equal to $\exp[-c|s_l - s_\ell|]/(2c)$. Optimal tests in this problem are of the form (16) with $\mathbf{\Omega}_0 = \mathbf{\Omega}_L = \mathbf{R}'\mathbf{\Sigma}_L\mathbf{R}$ and $\mathbf{\Omega}_a = \mathbf{\Omega}(c_a) = \mathbf{R}'\mathbf{\Sigma}(c_a)\mathbf{R}$ for some $c_a > 0$. This yields the test statistic

$$\mathrm{LFUR} = \frac{\mathbf{Z}'\mathbf{\Omega}_L^{-1}\mathbf{Z}}{\mathbf{Z}'\mathbf{\Omega}^{-1}(c_a)\mathbf{Z}}. \tag{17}$$

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.[6] (We label the statistic “LFUR” because it is the spatial generalization of the low-frequency unit root test proposed in Müller and Watson (2008).)

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

Now consider a corresponding spatial stationarity test based on $\mathbf{Z}$. Here, we seek a test of the null hypothesis that y_l exhibits weak spatial correlation, and this requires a definition of “weak” correlation. One useful gauge for the strength of correlation is whether HAR inference about the mean remains valid. The inference derived in Müller and Watson (2022) remains valid by construction in the $\mathbf{\Sigma}(c)$ model for (all large enough) values of c that induce an average pairwise correlation

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

of no more than $\bar{\rho} = 0.03$. 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 $H_0 : \mathbf{\Omega} = \mathbf{\Omega}(c)$, $c \geq c_{0.03}$ against $H_a : \mathbf{\Omega} = \mathbf{\Omega}(c) + g_a^2\mathbf{\Omega}_L$, $g_a > 0$, where the form of the alternative, a sum of a stationary and $I(1)$ process, is standard in time series stationarity tests (e.g., see Nyblom (1989) and Kwiatkowski et al. (1992)). 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 $\mathbf{\Omega}(c)$ 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

[6]The well-known DF-GLS test of Elliott, Rothenberg, and Stock (1996) also requires a choice of c_a, which is chosen in the same fashion.

works well in the sense of generating a test statistic (16) that has a 95% quantile that is fairly constant as a function of $c \geq c_{0.03}$. Thus, the stationary test rejects if

$$\text{LFST} = \frac{\mathbf{Z}'\boldsymbol{\Omega}(c_{0.001})^{-1}\mathbf{Z}}{\mathbf{Z}'\big[\boldsymbol{\Omega}(c_{0.001}) + g_a^2\boldsymbol{\Omega}_L\big]^{-1}\mathbf{Z}} \tag{19}$$

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

REMARK 4.1: Suppose the $p \times 1$ vector x_l is spatially cointegrated of order 1 with cointegrating vector β_0, that is, $\beta_0' x_l \sim I(0)$, but $\beta' x_l \sim I(1)$ for all β that are not proportional to β_0. An asymptotic level $1-\alpha$ confidence set for β_0 can then be formed by collecting those values of b for which the level α LFST test does not reject the $I(0)$ null hypothesis when applied to the variable $b' x_l$. 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).

4.4. *Confidence Sets for the Local-to-Unity Parameter and Half-Life of Persistence*

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 α-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 Π, the classic result in Pratt (1961) implies that the resulting confidence interval has the smallest Π-weighted expected length.

An easily interpretable transformation of the parameter c is given by the half-life, that is, the distance Δ at which the correlation $\exp[-c\Delta]$ is equal to 1/2, which is $h(c) = \ln 2/c$. With Π such that the implied weighting of h is uniform on $[0, \Delta_{\max}]$ with $\Delta_{\max} = \max_{l,\ell} |s_l - s_\ell^0|$, the average length minimizing scale-invariant confidence interval collects the values of h_0 for which the test based on

$$\frac{\int_0^{\Delta_{\max}} \det\big(\boldsymbol{\Omega}(\ln 2/h)\big)^{-1/2}\big(\mathbf{Z}'\boldsymbol{\Omega}(\ln 2/h)^{-1}\mathbf{Z}\big)^{-q/2}\,dh}{\big(\mathbf{Z}'\boldsymbol{\Omega}(\ln 2/h_0)^{-1}\mathbf{Z}\big)^{-q/2}} \tag{20}$$

does not exceed the $1-\alpha$ quantile of (20) under $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}(\ln 2/h_0))$.

4.5. *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 β unknown, u_l is unobserved.

There is an easy solution to this problem if $\mathbf{u} = (u_1, \dots, u_n)'$ is independent of $\mathbf{X} = (x_1, \dots, x_n)'$. Namely, one can simply base inference on weighted averages of $\mathbf{Y}$ with weights that are orthogonal to $\mathbf{X}$. Let $\mathbf{R}_X$ be the $n \times q$ matrix of the eigenvectors of

$\mathbf{M}_X\mathbf{\Sigma}_L\mathbf{M}_X$ corresponding to the largest q eigenvalues, where $\mathbf{M}_X = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $n^{-1}\mathbf{R}_X'\mathbf{R}_X = \mathbf{I}_q$. Then, by construction, $\mathbf{R}_X'\mathbf{X} = \mathbf{0}$, so that $\mathbf{Z}_X = \mathbf{R}_X'\mathbf{Y} = \mathbf{R}_X'\mathbf{u}$. With $\mathbf{u}$ independent of $\mathbf{X}$, one can simply condition on the realization of $\mathbf{X}$, and apply the above tests with $\mathbf{Z}_X$ in place of $\mathbf{Z}$.

REMARK 4.2: A noteworthy application is 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).[7] To implement such a level α test of the null hypothesis of no spatial cointegration in practice, one computes the LFUR statistic (17) using $\mathbf{Z}_X$ in place of $\mathbf{Z}$, $\mathbf{\Omega}_0 = \mathbf{R}_X'\mathbf{\Sigma}_L\mathbf{R}_X$ and $\mathbf{\Omega}_1 = \mathbf{R}_X'\mathbf{\Sigma}(c_a)\mathbf{R}_X$, and compares it to $1-\alpha$ quantile of the statistic under $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_L)$, where c_a can again be chosen by the 50% power rule.

4.6. *Large-Sample Analysis*

In Sections 4.1–4.5, the data were assumed to follow the canonical $I(1)$ or LTU processes with $y_l = L(s_l)$ or $y_l = J_c(s_l)$. The large-sample validity of these procedures under the general $I(1)$ and LTU processes defined in Section 2 follow from the FCLT result in Theorem 2. Furthermore, the validity of the LFST test under a more general $I(0)$ null hypothesis can be established by invoking the spatial CLT in Lahiri (2003). Details are provided in the Appendix. Here we highlight three features of that analysis.

REMARK 4.3: The asymptotic analysis in Section 2 included the long-run standard deviation ω and the spatial scale factor λ_n, but they have been ignored in this section. This is due to two features of the proposed tests. First, the tests are scale invariant, so they are unaffected by any scaling of the data. Second, our choices for c_a, $c_{0.001}$, and g_a in the tests LFUR and LFST induce invariance to the scale of the locations $\{s_l\}$, and thus λ_n.

REMARK 4.4: A complication in directly applying the FCLT, CLT, and continuous mapping theorem to show the large-sample normality of $\mathbf{Z} = \mathbf{R}'\mathbf{Y}$ is that the eigenvector weights $\mathbf{R}$ depend on the n sample locations s_l. Lemma 5 in the Appendix studies the large-sample behavior of these eigenvectors and shows that they converge to the eigenfunctions of the covariance kernel of demeaned Lévy–Brownian motion in a well-defined sense, building on closely related results of Rosasco, Belkin, and De Vito (2010) and Müller and Watson (2022).

REMARK 4.5: The asymptotic validity of the residual-based tests only requires a form of asymptotic independence of $\mathbf{u}$ and $\mathbf{X}$, and this independence always holds under the null hypothesis of no cointegration for the Engle and Granger (1987)-type test.

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

Chetty et al. (2014) used administrative records on the incomes of more than 40 million children and parents to study intergenerational income mobility in the United States.

[7]In the canonical model, the assumption is that under the null hypothesis, $(x_l', y_l) = (X(s_l)', Y(s_l)) = \Phi L_{X,Y}(s_l)$, where $L_{X,Y}$ is a vector of $p+1$ independent Lévy–Brownian motions, and Φ is an arbitrary full rank $(p+1)\times(p+1)$ matrix. In this model, the assumption that u_l follows a Lévy–Brownian motion and is independent of x_l follows directly by defining β as $\Sigma_{XX}^{-1}\Sigma_{XY}$, where Σ_{XX} and Σ_{XY} are the appropriate blocks of $\Sigma = \Phi\Phi'$.

They constructed an index of mobility for each of the commuter zones in the United States and investigated the relationship between mobility and other factors by regressing their mobility index on variables such as racial segregation, school quality, and so forth. They found 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 I.[8]

The first three columns in the table apply the tests outlined in this section to gauge the spatial correlation in the socioeconomic variables for the contiguous 48 states, which contain 722 of the 741 commuting zones used by Chetty et al. The results indicate that these variables exhibit substantial spatial correlation: 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, while wide, suggest a high degree of spatial persistence. The remaining columns of the table investigate the robustness of the regression results reported in Chetty et al. to this spatial correlation. We discuss these columns after introducing additional analysis in the next section.

5. REGRESSIONS WITH TRANSFORMED SPATIAL VARIABLES

To avoid spurious regression effects using $I(1)$ time series data, researchers routinely estimate regressions using first differences of the original variables and rely on HAC/HAR inference methods to account for any remaining $I(0)$ autocorrelation. The best approach for regressions involving spatial $I(1)$ variables is not so obvious. There are several ways to spatially difference the data, and confidence intervals can be constructed using various spatial HAC/HAR or clustered standard errors. In this section, we use simulations to study the properties of regressions estimated using a variety of spatial differencing methods and confidence intervals constructed using spatial HAR and clustered standard errors. We stress that the analysis in this section relies on simulations rather than the kind of formal analysis that was the basis of earlier sections. In that sense, this section's conclusions are tentative; that said, we think they provide useful guidance for applied work and direction for future analytic work.

The simulations used here mimic the Chetty et al. regressions shown in Table I, and therefore involve univariate regressions $y_l = x_l\beta + u_l$, with locations fixed at the (centers of the) $n = 720$ commuting zones plotted in Figure 1.[9] In all experiments, the sample data (in levels) are standardized to have a zero sample mean and unit standard deviation so that intercepts are excluded from the regression and β is measured in standard deviation units. The experiments differ in the spatial persistence of the data, the method for spatial differencing the data, and the standard errors used to construct test statistics and confidence intervals.

[8]The variables are chosen from Figure VIII in Chetty et al. (2014). The data are taken from their comprehensive replication materials.

[9]To avoid arguably artificial effects in our spatial AR data generating process, the experiments exclude two of the 722 commuting zones because they correspond to islands (Nantucket, MA and Friday Harbor, WA) that are non-adjacent to other commuting zones.

TABLE I

VARIABLES AND REGRESSIONS FROM CHETTY ET AL. (2014).

	Spatial Persistence Statistics			Regression of AMI onto Variable				
	p-value for Test of		Half-life	Levels			LBM-GLS	
Variable	$I(1)$ Null	$I(0)$ Null	95% CI	R^2	$\hat{\beta}$ [95% CI] Cluster	p-value Resid. $I(1)$ Test	R^2	$\hat{\beta}$ [95% CI] C-SCPC
Absolute Mobility Index (AMI)	0.38	<0.01	[0.09, ∞]	NA	NA	NA	NA	NA
Frac. Black Residents	0.11	<0.01	[0.04, ∞]	0.36	−0.60 [−0.73, −0.47]	0.22	0.10	−0.42 [−0.70, −0.15]
Racial Segregation	0.01	0.13	[0.00, 0.28]	0.14	−0.38 [−0.46, −0.29]	0.29	0.18	−0.24 [−0.35, −0.12]
Segregation of Poverty	0.28	0.03	[0.05, ∞]	0.18	−0.43 [−0.55, −0.30]	0.27	0.15	−0.21 [−0.37, −0.05]
Frac. < 15 Mins to Work	0.57	<0.01	[0.14, ∞]	0.48	0.69 [0.55, 0.84]	0.14	0.15	0.37 [0.08, 0.65]
Mean Household Income	0.13	0.14	[0.02, ∞]	0.00	0.05 [−0.10, 0.19]	0.38	0.00	−0.01 [−0.25, 0.24]
Gini Coefficient	0.78	<0.01	[0.26, ∞]	0.37	−0.60 [−0.78, −0.43]	0.23	0.10	−0.22 [−0.38, −0.05]
Top 1 Perc. Inc. Share	0.31	0.02	[0.07, ∞]	0.04	−0.21 [−0.36, −0.06]	0.36	0.02	−0.07 [−0.13, −0.00]
Student-Teacher Ratio	0.23	0.13	[0.05, ∞]	0.12	−0.35 [−0.55, −0.15]	0.45	0.03	−0.17 [−0.44, 0.11]
Test Scores	0.30	0.06	[0.07, ∞]	0.34	0.58 [0.40, 0.76]	0.41	0.30	0.42 [0.16, 0.68]
High School Dropout	0.09	0.02	[0.03, ∞]	0.34	−0.58 [−0.74, −0.41]	0.49	0.21	−0.29 [−0.55, −0.03]
Social Capital Index	0.72	<0.01	[0.22, ∞]	0.41	0.64 [0.46, 0.82]	0.29	0.08	0.28 [−0.02, 0.58]
Frac. Religious	0.27	0.04	[0.07, ∞]	0.28	0.53 [0.36, 0.70]	0.26	0.14	0.32 [0.14, 0.50]
Violent Crime Rate	0.55	0.02	[0.15, ∞]	0.21	−0.45 [−0.67, −0.24]	0.34	0.04	−0.14 [−0.26, −0.03]
Frac. Single Mothers	0.18	<0.01	[0.05, ∞]	0.59	−0.77 [−0.92, −0.63]	0.11	0.52	−0.60 [−0.94, −0.27]
Divorce Rate	0.05	0.17	[0.02, ∞]	0.27	−0.52 [−0.70, −0.33]	0.49	0.26	−0.37 [−0.63, −0.11]
Frac. Married	0.05	0.08	[0.01, ∞]	0.31	0.56 [0.43, 0.68]	0.22	0.31	0.35 [0.11, 0.59]
Local Tax Rate	0.02	0.24	[0.01, 0.49]	0.12	0.35 [0.21, 0.48]	0.40	0.01	0.07 [−0.09, 0.23]
Colleges per Capita	0.24	0.07	[0.06, ∞]	0.06	0.24 [−0.01, 0.48]	0.27	0.00	0.01 [−0.23, 0.25]
College Tuition	0.38	<0.01	[0.09, ∞]	0.00	−0.02 [−0.15, 0.11]	0.28	0.00	0.01 [−0.05, 0.08]
Coll. Grad. Rate	0.04	0.03	[0.00, ∞]	0.02	0.15 [0.03, 0.28]	0.35	0.03	0.08 [0.01, 0.15]
Manufacturing Share	0.20	<0.01	[0.06, ∞]	0.09	−0.30 [−0.47, −0.13]	0.37	0.01	0.07 [−0.09, 0.23]
Chinese Import Growth	0.02	0.07	[0.02, 0.46]	0.03	−0.17 [−0.33, −0.02]	0.37	0.00	0.03 [−0.01, 0.06]
Teenage LFP Rate	0.51	<0.01	[0.12, ∞]	0.44	0.66 [0.50, 0.82]	0.28	0.04	0.26 [−0.06, 0.58]

(*Continues*)

TABLE I

Continued.

	Spatial Persistence Statistics			Regression of AMI onto Variable				
	p-value for Test of		Half-life	Levels			LBM-GLS	
Variable	$I(1)$ Null	$I(0)$ Null	95% CI	R^2	$\hat{\beta}$ [95% CI] Cluster	p-value Resid. $I(1)$ Test	R^2	$\hat{\beta}$ [95% CI] C-SCPC
Migration Inflow	0.30	0.08	$[0.00, \infty]$	0.07	−0.27 [−0.42, −0.13]	0.32	0.02	−0.12 [−0.27, 0.04]
Migration Outlflow	0.34	0.01	$[0.08, \infty]$	0.03	−0.16 [−0.30, −0.02]	0.37	0.01	−0.08 [−0.16, 0.01]
Frac. Foreign Born	0.56	0.04	$[0.17, \infty]$	0.00	−0.03 [−0.16, 0.09]	0.39	0.02	−0.12 [−0.29, 0.06]

Note: The first two columns show p-values for tests of the $I(1)$ and $I(0)$ null hypotheses using the statistics (17) and (19). The third column shows the 95% confidence interval for the half-life, h, constructed by inverting the tests in (20). The half-life values shown are relative to the largest distance between commuter zones in the data set, 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 $I(1)$ test 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.

5.1. *Data Transformations and Spatial Standard Errors*

The experiments involve regressions estimated by OLS after transforming $\{y_l, x_l\}$ using the same spatial difference transformation. Denote the transformed data by $\{y_l^*, x_l^*\}$. We consider five methods, described here for y_l^*.

Levels: $y_l^* = y_l$. The resulting estimator is the OLS levels regression studied in Section 3.

Nearest Neighbor (NN) Differences: $y_l^* = y_l - y_{\ell(l)}$, where $s_{\ell(l)}$ is the location closest to s_l.

Isotropic Differences: y_l^* is the deviation of y_l from the average value of y in a circular (isotropic) neighborhood of s_l with radius equal to b. Specifically, let $m_l = \sum_{j\neq l} \mathbf{1}[|s_l - s_j| < b]$ denote the number of locations within a distance b of s_l and $\overline{y}_l(b) = m_l^{-1} \sum_{j\neq l} \mathbf{1}[|s_l - s_j| < b] y_j$ denote the average value of y at these locations. If $m_l > 0$, then $y_l^* = y_l - \overline{y}_l(b)$ and $y_l^* = 0$ otherwise (equivalently, the observation is dropped from the regression). In the experiments, we scale the locations so that $\max_{l,\ell} |s_l - s_\ell| = 1$ and use two values, $b = 0.04$ and $b = 0.08$. When $b = 0.04$, $m_l = 0$ for 1% of the locations, and no locations are dropped when $b = 0.08$.

Cluster Fixed Effects: Here, the data are partitioned into m clusters and cluster fixed effects are included in the regression. Equivalently, the regressions are estimated by OLS using data that are demeaned within each cluster. Our experiments use three choices of clusters. First, because the locations are commuting zones, we follow a common practice in applied work and cluster by U.S. state. Alternatively, we use a more agnostic approach that partitions the sampling region into clusters by applying the k-means algorithm to the locations $\{s_l\}$, and this is implemented for $m = 60$ and $m = 240$ clusters.

LBM-GLS: In a time series regression, first differences correspond to a GLS transformation under the canonical random walk model for $I(1)$ persistence. This motivates a GLS transformation based on Lévy–Brownian motion, the canonical spatial $I(1)$ model. We label this LBM-GLS. Recall that the data used in the regression are demeaned, so that under the canonical model $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{M}\mathbf{\Sigma}_L\mathbf{M})$, where $\mathbf{M} = \mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$ and $\mathbf{\Sigma}_L$ is the covariance matrix induced by Lévy–Brownian motion. The corresponding GLS transform sets

$$\mathbf{Y}^* = (\mathbf{M}\mathbf{\Sigma}_L\mathbf{M})^{-1/2}\mathbf{Y}, \tag{21}$$

where $(\mathbf{M}\mathbf{\Sigma}_L\mathbf{M})^{-1/2}$ is the Moore–Penrose generalized inverse of $(\mathbf{M}\mathbf{\Sigma}_L\mathbf{M})^{1/2}$.[10] In the canonical model, this GLS transformation converts $\mathbf{Y}$ into a set of demeaned i.i.d. random variables $\mathbf{Y}^*$. This is no longer true in the more general $I(1)$ model, but given the FCLT in Theorem 2, it is plausible that the LBM-GLS transformation induces enough stationarity for spatial HAR inference to be reliable. As an example, Figure 4 illustrates the GLS transformations for the levels data plotted in panel (c) of Figure 1 and shows that the LBM-GLS transformed data exhibit far less spatial persistence.

These estimators of β are used in conjunction with two types of standard errors. The first are spatial HAR standard errors and critical values suggested by Müller and Watson (2023). 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 (2022)). For the fixed effects estimator, we employ the cluster version of the C-SCPC method, and also report results using traditional clustered standard errors.

[10]We choose the symmetric square root $(\mathbf{M}\mathbf{\Sigma}_L\mathbf{M})^{1/2}$ to ensure that $\mathbf{Y}^*$ shares the same spatial coordinate basis as $\mathbf{Y}$. This is useful when constructing spatial correlation standard errors for the transformed data.

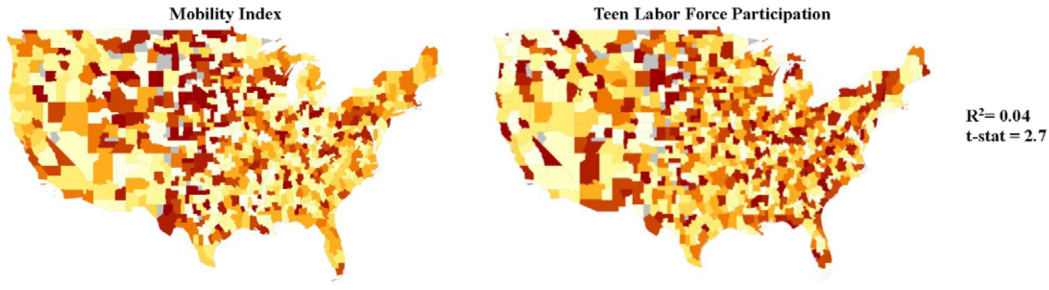

FIGURE 4.—LBM-GLS transformed data.

5.2. *Data Generating Processes*

The experiments differ in their distribution of $(\mathbf{Y}, \mathbf{X})$, where $\mathbf{Y}$ and $\mathbf{X}$ are independent and identically distributed and are generated by one of nine models.

The first four models (DGP1–DGP4) use $y_l = Y(s_l)$, where Y are $I(1)$ processes: DGP1 is Lévy–Brownian motion and DGP2–DGP4 use $Y \sim I(1)$ as in (6) with different models for B. DGP2 uses $B = J_c$ and $c = c_{0.01}$, so the average pairwise correlation (18) of $\{B(s_l)\}_{l=1}^n$ is $\bar{\rho} = 0.01$, and DGP3 is the same with $c = c_{0.03}$. For DGP4, B is 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 $\bar{\rho} = 0.03$.

The next two models (DGP5 and DGP6) use $y_l = J_c(s_l)$ with $c = c_{0.03}$ in DGP5 and $c = c_{0.50}$ in DGP6. These models exhibit less than $I(1)$ persistence, much less so in DGP5, and are included to examine the potential effects of "over-differencing."

The final three models (DGP7–DGP9) generate highly persistent data, but are outside the class of $I(1)$ models introduced in Section 2. DGP7 is the Brownian sheet with $y_l = \int_{\mathbb{R}^2} \mathbf{1}[0 \le r \le s_l]\, dW(r)$. DGP8 and DGP9 generate data from the spatial autoregression (SAR) model $\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \mathbf{U}$, where $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{W}$ is the adjacency matrix for the commuting zones with row sums normalized to unity. DGP8 uses $\rho = 0.99$ and DGP9 uses $\rho = 0.999$.

REMARK 5.1: The models generate data with different degrees of spatial persistence. One way to gauge their relative persistence is by comparing their rejection frequencies for the spatial unit root ($I(1)$) and stationarity ($I(0)$) tests introduced in Section 4. These are shown in panel (a) of Table II. By design, the rejection frequency for Lévy–Brownian motion for the $I(1)$ test is 5% and similarly for the $J_{c_{0.03}}$ process for the $I(0)$ test. Using this gauge, the other $I(1)$ process are more persistent than Lévy–Brownian motion (with smaller $I(1)$ and larger $I(0)$ rejection frequencies) as is the SAR model with $\rho = 0.999$.

REMARK 5.2: 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). With the row sums of $\mathbf{W}$ normalized to unity, the unit root SAR model is not well-defined (hence our use of $\rho = 0.99$ and 0.999 in DGP8 and DGP9). Lee and Yu (2009, 2013) studied asymptotic properties of the row normalized SAR model with a SAR coefficient that converges to unity. For particular forms of the matrix $\mathbf{W}$, they found 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

TABLE II

SUMMARY OF SIMULATION EXPERIMENTS.

	Data Generating Process									
	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	$\text{SAR}_{0.99}$	$\text{SAR}_{0.999}$	
Test	(a) Rejection Frequency for $I(1)$ and $I(0)$ Tests									
LFUR ($I(1)$ null)	0.05	0.03	0.02	0.03	0.81	0.13	0.13	0.14	0.02	
LFST ($I(0)$ null)	0.82	0.86	0.89	0.87	0.05	0.67	0.49	0.61	0.89	
Method	(b) C-SCPC SEs: Rejection Frequency and *Average Length of Confidence Interval*									Avg
OLS	0.28	0.3	0.33	0.31	0.04	0.18	0.39	0.17	0.39	0.26
	0.88	*1.01*	*1.05*	*1.03*	*0.44*	*0.85*	*0.76*	*0.98*	*1.08*	*0.90*
NN-Difference	0.07	0.14	0.19	0.16	0.05	0.06	0.21	0.09	0.21	0.13
	0.25	*0.46*	*0.57*	*0.56*	*0.23*	*0.24*	*0.30*	*0.41*	*0.60*	*0.40*
Iso-Difference ($b = 0.04$)	0.05	0.11	0.17	0.13	0.04	0.05	0.17	0.12	0.28	0.13
	0.22	*0.36*	*0.46*	*0.47*	*0.20*	*0.21*	*0.32*	*0.42*	*0.63*	*0.37*
Iso-Difference: ($b = 0.08$)	0.04	0.07	0.11	0.08	0.03	0.04	0.16	0.06	0.14	0.08
	0.34	*0.53*	*0.61*	*0.60*	*0.28*	*0.32*	*0.41*	*0.60*	*0.83*	*0.50*
Fixed Effects (State)	0.16	0.20	0.25	0.22	0.08	0.13	0.26	0.15	0.25	0.19
	0.44	*0.65*	*0.72*	*0.69*	*0.28*	*0.40*	*0.54*	*0.61*	*0.80*	*0.57*
Fixed Effects ($m = 60$)	0.05	0.08	0.13	0.10	0.03	0.04	0.14	0.06	0.17	0.09
	0.44	*0.71*	*0.83*	*0.77*	*0.31*	*0.40*	*0.52*	*0.61*	*0.99*	*0.62*
Fixed Effects ($m = 240$)	0.05	0.10	0.15	0.12	0.04	0.05	0.14	0.05	0.12	0.09
	0.33	*0.61*	*0.72*	*0.68*	*0.27*	*0.30*	*0.41*	*0.50*	*0.69*	*0.50*
LBM-GLS	0.03	0.07	0.10	0.07	0.03	0.03	0.14	0.06	0.18	0.08
	0.20	*0.33*	*0.44*	*0.46*	*0.20*	*0.20*	*0.30*	*0.40*	*0.63*	*0.35*
Method	(c) Clustered SEs: Rejection Frequency									Avg
OLS (State)	0.58	0.61	0.64	0.62	0.18	0.49	0.62	0.46	0.69	0.54
Fixed Effects: State	0.23	0.31	0.38	0.34	0.12	0.19	0.36	0.20	0.36	0.28
Fixed Effects ($m = 60$)	0.21	0.32	0.45	0.38	0.10	0.15	0.33	0.20	0.41	0.28
Fixed Effects ($m = 240$)	0.14	0.44	0.55	0.47	0.08	0.11	0.28	0.23	0.41	0.30

Note: Panel (a) shows the rejection frequency for nominal 5% $I(1)$ and $I(0)$ tests developed in Section 4. Panel (b) shows rejection frequencies for nominal 5% tests of the null that $\beta = 0$ using the transformation shown in the first column and C-SCPC standard errors and critical values. Also shown (in italics) is the average length of the (non-size-adjusted) nominal 95% confidence interval for β. Panel (c) shows the rejection frequency for nominal 5% tests of $\beta = 0$ using clustered standard errors and normal critical values. The data generating processes are defined in the text.

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) generated data from a version of the SAR model that is well-defined with a unit SAR coefficient and presented Monte Carlo results suggesting spurious regression phenomena. Our DGP8-DGP9 designs are similar to his. Using a related model, Rossi and Lieberman (2023) derived non-standard large-sample distributions for the estimated SAR coefficient for particular specifications of the SAR weight matrix, and suggested that model-specific non-standard results will hold more generally.

5.3. *Simulation Results*

Panels (b) and (c) of Table II summarize the null rejection frequency of nominal 5% level tests for each method and each DGP. Panel (b) also shows the expected length of confidence intervals. There are three clear takeaways from the table. First, spatial differencing used in conjunction with C-SCPC inference significantly reduces the size distortions from strong spatial persistence, particularly using a relatively large isotropic bandwidth ($b = 0.08$) or a small number of clusters ($m = 60$). Second, inference using traditional clustered standard errors results in much larger size distortions than C-SCPC inference. Even the inclusion of fixed effects in conjunction with clustered standard errors still induces highly misleading inference, no matter the size of the clusters. This latter effect is specific to the spatial case: A time series random walk that is demeaned over each of m non-overlapping segments yields independent processes across segments, but a Lévy–Brownian motion that is demeaned over m non-overlapping regions in $\mathbb{R}^d$, $d > 1$ does not in general. Third, LBM-GLS (with a second step C-SCPC correction) controls size as well or better than the other methods and 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 cluster size, suggests that it dominates the other methods considered here.

5.4. *Regressions in Chetty et al. (2014)*

We now return to Table I. As noted in Section 4.7, 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).[11] The first set of statistics are for regressions using the levels of the variables, and the second set uses the LBM-GLS transformed variables.

We highlight three results. First, the residuals from the levels regressions 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 β and the regression R^2 tend to be smaller in magnitude than in the levels regression. Third, while the OLS and LBM-GLS results differ, 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.

6. CONCLUDING REMARKS

Applied researchers are well aware of the pitfalls of conducting inference with persistent time series data. Variables are routinely tested for extreme time series persistence using unit root or related tests, and are often differenced to avoid spurious regression effects.

This paper demonstrates that inference with highly persistent spatial data is equally fraught: HAC or clustered standard errors fail in the presence of strong spatial correlation, 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 regression inference by suitably transforming the spatial variables, combined with spatial HAR corrections to accommodate any residual weak correlations. The theory here is much less complete, however. It would

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

be especially desirable for future research to obtain a good theoretical understanding of the most promising of these transformations, namely the GLS approach using the canonical spatial unit root model as a baseline.

APPENDIX A: PROOFS

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

$$\mathbb{E}\big[\big(Y(s)-Y(r)\big)^{2m}\big] \leq C_m\bigg(\int_{\mathbb{R}^d}\big(h(u,s)-h(u,r)\big)^2\,du\bigg)^m = C_m|s-r|^m,$$

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

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

$$\begin{aligned}\mathbb{E}\big[\big(Y_c(s)-Y_c(r)\big)^{2m}\big] &\leq C_m\bigg(\int_{\mathbb{R}^d}\big(h_c(u,s)-h_c(u,r)\big)^2\,du\bigg)^m \\ &= C_m\mathbb{E}\big[\big(J_c(s)-J_c(r)\big)^2\big]^m, \qquad (22)\end{aligned}$$

where the last equality follows from the representation (8) of J_c, and $\mathbb{E}[(J_c(s)-J_c(r))^2] = (1-\exp(-c|s-r|))/c \leq |s-r|$. *Q.E.D.*

PROOF OF THEOREM 2: Consider first the claim for the convergence for the LTU process (9). From

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

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

$$Y_n^0(s) = \lambda_n^{-d/2}\int_{\mathbb{R}^d} h_c\big(\lambda_n^{-1}r,s\big)B(r)\,dr, \quad s \in \mathcal{S}^0. \qquad (23)$$

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 (22), and

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

where the inequality invokes Condition 2. For the former, consider the $p\times 1$ vector $(Y_n^0(t_1),\dots,Y_n^0(t_p))$ for arbitrary $t_1,\dots,t_p \in \mathcal{S}^0$. By the Cramér–Wold device, it suffices to establish the convergence $X_n = \sum_{j=1}^p v_jY_n^0(t_j) \Rightarrow \sum_{j=1}^p v_j\omega J_c(t_j)$ for $(v_1,\dots,v_p) \in \mathbb{R}^p$. Let $f_v(r) = \sum_{j=1}^p v_jh_c(r,t_j)$, so that from (8), $\sum_{j=1}^p v_jJ_c(t_j) \sim \mathcal{N}(0,\int_{\mathbb{R}^d} f_v(r)^2\,dr)$, and from (23), $X_n = \lambda_n^{-d/2}\int_{\mathbb{R}^d} f_v(\lambda_n^{-1}r)B(r)\,dr$.

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

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

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

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

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

Finally, note that f_v^ε is bounded and $f_v^\varepsilon(\lambda_n^{-1}r) = 0$ for $|r| > \lambda_n/\varepsilon$. Thus, using Condition 3 with $\zeta_n = \lambda_n$, $X_n^\varepsilon \Rightarrow \mathcal{N}(0, \int_{\mathbb{R}^d}\sigma_B(s)\,ds\int_{\mathbb{R}^d} f_v^\varepsilon(r)^2\,dr)$. The result for the LTU process (9) 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), $Y_n^0(s) = \lambda_n^{-d/2}\int_{\mathbb{R}^d} h(\lambda_n^{-1}r, s) \times B(r)\,dr$, and $Y_n^0(0) = 0$, so the result follows from the same steps. *Q.E.D.*

PROOF OF THEOREM 3: The results follow straightforwardly from the continuous mapping theorem (CMT) if we can show that $\lambda_n^{-1/2}n^{-1}\sum_{l=1}^n y_l \Rightarrow \int Y^0(s)\,dG(s)$, $\lambda_n^{-1/2}n^{-1} \times \sum_{l=1}^n x_l \Rightarrow \int X^0(s)\,dG(s)$, $\lambda_n^{-1}n^{-1}\sum_{l=1}^n x_l y_l \Rightarrow \int X^0(s)Y^0(s)\,dG(s)$, and $\lambda_n^{-1}n^{-1} \times \sum_{l=1}^n x_l x_l' \Rightarrow \int X^0(s)X^0(s)'\,dG(s)$.

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 $\int X_n^*(s)Y_n^*(s)\,dG_n(s) = n^{-1}\sum_{l=1}^n X_n^*(s_l) \times Y_n^*(s_l) \sim n^{-1}\sum_{l=1}^n X_n^0(s_l)Y_n^0(s_l) = \lambda_n^{-1}n^{-1}\sum_{l=1}^n x_l y_l$. We have

$$\begin{aligned}&\bigg|\int\big(X_n^*(s)Y_n^*(s) - X^*(s)Y^*(s)\big)\,dG_n(s)\bigg| \\ &\quad\le \sup_{s\in\mathcal{S}^0}\big(|X_n^*(s)|\cdot|Y_n^*(s) - Y^*(s)| + |Y^*(s)|\cdot|X_n^*(s) - X^*(s)|\big) \overset{a.s.}{\to} 0,\end{aligned}$$

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, $\int X^*(s)Y^*(s)\,dG_n(s) \stackrel{a.s.}{\to} \int X^*(s)Y^*(s)\,dG(s)$. But almost sure convergence implies convergence in distribution, so the desired result follows. The argument for the other terms is analogous. *Q.E.D.*

PROOF OF THEOREM 4: From Theorem 3 and the CMT, $H\hat{\beta} \Rightarrow H\Xi_{\tilde{x}\tilde{x}}^{-1}\Xi_{\tilde{x}\tilde{y}}$ with the r.h.s. non-zero with probability 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 1, it suffices to show $n^{-1}\lambda_n^{-2}\|\hat{\Omega}_n\| \stackrel{p}{\to} 0$.

Let $\tilde{Y}_n^0(\cdot) = Y_n^0(\cdot) - \int Y_n^0(s)\,dG_n(s)$, $\tilde{X}_n^0(\cdot) = X_n^0(\cdot) - \int X_n^0(s)\,dG_n(s)$, and $e_n^0(\cdot) = (\tilde{Y}_n^0(\cdot) - \tilde{X}_n^0(\cdot)'\hat{\beta})\tilde{X}_n^0(\cdot)$, so that $e_l = \lambda_n e_n^0(\lambda_n^{-1}s_l)$. By (11), 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 $\lambda_n^{-1}\sup_{l\le n}|e_l| = O_p(1)$. Consider first the HAC estimator. We have

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

and with $b_n^0 = \lambda_n b_n$ and $s_l^0 = \lambda_n^{-1}s_l$, $\sum_{l,\ell=1}^n|\kappa(b_n(s_l-s_\ell))| = \sum_{l,\ell=1}^n|\kappa(b_n^0(s_l^0-s_\ell^0))|$ and

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

Now

$$n^{-2}\sum_{l,\ell=1}^{n}\mathbf{1}\big[\big|s_l^0-s_\ell^0\big| > \big(b_n^0\big)^{-1/2}\big]\big|\kappa\big(b_n^0\big(s_l^0-s_\ell^0\big)\big)\big| \le \sup_{|a|\ge\sqrt{b_n^0}}\big|\kappa(a)\big| \stackrel{p}{\to} 0$$

by (14) and $1/b_n^0 = o_p(1)$. Let $\varepsilon > 0$. Then, for all large enough n,

$$\begin{aligned} n^{-2}\sum_{l,\ell=1}^{n}\mathbf{1}\big[\big|s_l^0-s_\ell^0\big| \le \big(b_n^0\big)^{-1/2}\big] &\le \int\int\mathbf{1}\big[|s-r|\le\varepsilon\big]\,dG_n(s)\,dG_n(r) \\ &\to \int\int\mathbf{1}\big[|s-r|\le\varepsilon\big]\,dG(s)\,dG(r), \end{aligned}$$

where the convergence follows because Condition 1 implies that the product measure $G_n\times G_n$ on $\mathcal{S}_0\times\mathcal{S}_0$ converges weakly to $G\times G$, and G is absolutely continuous. Again invoking the absolute continuity of G, we obtain $\lim_{\varepsilon\searrow 0}\int\int\mathbf{1}[|s-r|\le\varepsilon]\,dG(s)\,dG(r) = 0$, so $n^{-2}\sum_{l,\ell=1}^n\mathbf{1}[|s_l^0-s_\ell^0|\le(b_n^0)^{-1/2}]\to 0$, and the result follows.

For the cluster estimator, we similarly have $\lambda_n^{-2}n^{-2}\|\sum_{j=1}^{n_C}(\sum_{l\in C_j}^n e_l)(\sum_{l\in C_j}^n e_l)'\| \le \lambda_n^{-2}(\sup_{l\le n}|e_l|)^2\cdot n^{-2}\sum_{j=1}^{n_C}|C_j|^2$ 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). *Q.E.D.*

APPENDIX B: ASYMPTOTIC VALIDITY OF TESTS OF DEGREE OF PERSISTENCE OF SECTION 4

We first establish a result about the convergence of eigenvectors. The following lemma is similar to Lemma 6 of Müller and Watson (2022) (also see Rosasco, Belkin, and De Vito (2010)). The two differences are the replacement of the i.i.d. assumption for s_l^0 by $G_n \Rightarrow G$, and that it allows for a larger class of projections. The latter generalization is necessary for the analysis of residual-based tests discussed in greater detail below.

We will use the following notation: let $k : \mathcal{S}^0 \times \mathcal{S}^0 \mapsto \mathbb{R}$ be the covariance kernel of Lévy–Brownian motion (although the following derivations will go through for any continuous positive definite kernel), and let $\mathbf{\Sigma}_n$ be the $n \times n$ matrix with l, ℓth 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 5: *Suppose the $p \times 1$ vector x_l is such that $x_l = \psi(s_l^0)$ for $l = 1, \dots, n$ 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*

$$M_n(f)(s) = f(s) - \int \psi(r)' f(r)\, dG_n(r) H_n^{-1}\psi(s),$$

$$M(f)(s) = f(s) - \int \psi(r)' f(r)\, dG(r) H^{-1}\psi(s).$$

Let $\hat{k}_n$ and $\bar{k}$ be the kernels corresponding to the linear operators $M_n L_{k,n} M_n$ and $M L_k M$, respectively, so that the (l, ℓ) element of $\mathbf{M}_X \mathbf{\Sigma}_n \mathbf{M}_X$ is given by $\hat{k}_n(s_l^0, s_\ell^0)$. Let

$$\bar{k}(s,r) = \sum_{i=1}^{\infty} \bar{\nu}_i \bar{\varphi}_i(s)\bar{\varphi}_i(r) \tag{24}$$

with $\int \bar{\varphi}_i(s)\bar{\varphi}_j(s)\, dG(s) = \mathbf{1}[i=j]$, $\bar{\nu}_i \ge \bar{\nu}_{i+1} \ge 0$ be the spectral decomposition of $\bar{k}$. Further, let $\sum_{i=1}^n \hat{\nu}_i \mathbf{r}_i \mathbf{r}_i'$, $\mathbf{r}_i = (r_{i,1}, \dots, r_{i,n})'$ be the spectral decomposition of $\mathbf{M}_X \mathbf{\Sigma}_n \mathbf{M}_X$, where $\mathbf{r}_i'\mathbf{r}_i = n$ and $\hat{\nu}_i \ge \hat{\nu}_{i+1}$, $i = 1, \dots, n-1$ are the eigenvalues, scaled by n^{-1}. Define

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

If $\bar{\nu}_1 > \bar{\nu}_2 > \cdots > \bar{\nu}_q > \bar{\nu}_{q+1}$ and Condition 1 holds, then, for any $q \ge 1$, $\sup_{s\in\mathcal{S}^0, 1\le i\le q} |\hat{\varphi}_i(s) - \bar{\varphi}_i(s)| \to 0$ and $\max_{1\le i\le q} |\hat{\nu}_i - \bar{\nu}_i| \to 0$.

PROOF: 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 (2022) 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

$$\begin{aligned}\|Mf\|_{\mathcal{H}} &= \left\| f - \int \psi(r)' f(r)\, dG(r) H^{-1}\psi \right\|_{\mathcal{H}} \\ &\leq \|f\|_{\mathcal{H}} + \sup_{r\in\mathcal{S}^0}\big|f(r)\big| \cdot \sup_{r\in\mathcal{S}^0}\big|H^{-1}\psi(r)\big| \cdot \sup_{|a|=1}\big\|a'\psi\big\|_{\mathcal{H}},\end{aligned}$$

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

$$\int \psi(r) f(r)\big(dG_n(r) - dG(r)\big) = \left\langle f, \int \psi(r) k_0(\cdot, r)\big(dG_n(r) - dG(r)\big)\right\rangle_{\mathcal{H}}, \tag{25}$$

and for each component ψ_i of ψ, $i = 1, \ldots, p$,

$$\begin{aligned}&\left\| \int \psi_i(r) k_0(\cdot, r)\big(dG_n(r) - dG(r)\big)\right\|_{\mathcal{H}}^2 \\ &= \int\int \psi_i(s) k_0(s, r)\psi_i(r)\big(dG_n(s) - dG(s)\big)\big(dG_n(r) - dG(r)\big) \\ &= \mathbb{E}\big[\psi_i(S_n)k_0(S_n, R_n)\psi_i(R_n) - \psi_i(S_n)k_0(S_n, R)\psi_i(R) \\ &\quad - \psi_i(S)k_0(S, R_n)\psi_i(R_n) + \psi_i(S)k_0(S, R)\psi_i(R)\big] \\ &\to 0,\end{aligned} \tag{26}$$

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 (25), (26), and Cauchy–Schwarz,

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

From $H_n^{-1} \to H^{-1}$ and

$$\left|\int \psi(r) f(r)\, dG_n(r)\right| \leq \sup_{r\in\mathcal{S}^0}\big|f(r)\big| \cdot \sup_{r\in\mathcal{S}^0}\big|\psi(r)\big| \leq \sup_{r\in\mathcal{S}^0}\big|\psi(r)\big|\sqrt{\sup_{s\in\mathcal{S}^0} k_0(s,s)} \cdot \|f\|_{\mathcal{H}},$$

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,

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

The only remaining piece of the proof is to show that $\|L_n - L\|_{\text{HS}}^2 \to 0$ under the assumption of $G_n \Rightarrow G$, where for any Hilbert–Schmidt operator $A : \mathcal{H} \mapsto \mathcal{H}$, $\|A\|_{\text{HS}}^2 =$

$\sum_{j\geq 1}\langle Ae_j, Ae_j\rangle_{\mathcal{H}}$ 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 (2022). We find

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

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. *Q.E.D.*

Consider now the asymptotic validity of the LFUR statistic. Specializing Lemma 5 to $p=1$ and $x_l=\psi(s)=1$, note that the l,ith element of the matrix of eigenvectors $\mathbf{R}$ is given by $\hat{\varphi}_i(s_l)$. Lemma 5 shows that $\sup_{s\in\mathcal{S}^0,1\leq i\leq q}|\hat{\varphi}_i(s)-\bar{\varphi}_i(s)|\to 0$ with $\bar{\varphi}_i$ the (continuous) eigenfunctions of $\bar{k}$, which is recognized as the covariance kernel of $\bar{L}(s)=L(s)-\int L(r)\,dG(r)$. The same arguments as in the proof of Theorem 3 then yield that under the null hypothesis of $y_l\sim I(1)$,

$$\lambda_n^{-1/2}n^{-1}\mathbf{Z}=\lambda_n^{-1/2}n^{-1}\mathbf{R}'\mathbf{Y} \quad\Rightarrow\quad \omega\int\bar{\boldsymbol{\varphi}}(s)L(s)\,dG(s)\sim\mathcal{N}\big(0,\omega^2\,\mathrm{diag}(\bar{\nu}_1,\dots,\bar{\nu}_q)\big),$$

where $\bar{\boldsymbol{\varphi}}(s)=(\bar{\varphi}_1(s),\dots,\bar{\varphi}_q(s))'$ and the equality in distribution follows from $\int\bar{\boldsymbol{\varphi}}(s)L(s)\,dG(s)=\int\bar{\boldsymbol{\varphi}}(s)\bar{L}(s)\,dG(s)$ and (24). Since the LFUR statistic is scale invariant, the scale parameters $\lambda_n^{-1/2}n^{-1}$ and ω^2 vanish, and the critical value computed from the canonical model converges to the asymptotically correct critical value for generic spatial $I(1)$ processes.

By the same arguments, the LFST test is large-sample valid under the general local-to-unity model (9) 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 $\mathbf{Z}$ 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 LFST 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 6: *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, $\limsup_{n\to\infty}\mathbb{P}(\mathrm{LFST}\geq\mathrm{cv}^{\mathrm{LFST}})\leq\alpha$.

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

$$\lambda_n^{d/2} n^{-1} \mathbf{Z} \Rightarrow \mathcal{N}\left(\mathbf{0}, a\sigma_B(0) \int \bar{\boldsymbol{\varphi}}(s)\bar{\boldsymbol{\varphi}}(s)'\, dG(s) + \omega^2 \int \bar{\boldsymbol{\varphi}}(s)\bar{\boldsymbol{\varphi}}(s)' g(s)\, dG(s)\right), \tag{27}$$

where $\bar{\boldsymbol{\varphi}} = (\bar{\varphi}_1, \dots, \bar{\varphi}_q)$ are the eigenfunctions of the covariance kernel of demeaned Lévy–Brownian motion associated with the largest q eigenvalues, $\omega^2 = \int_{\mathbb{R}^d} \sigma_B(s)\, ds$, and g is the density of the distribution G. Since the LFST statistic is scale invariant, its limiting distribution under (27) 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 $\liminf_{n\to\infty} \text{cv}^{\text{LFST}}$ is at least as large as the $1-\alpha$ quantile, say $\text{cv}_\chi^{\text{LFST}}$, of the (continuous) limit distribution of LFST for this value of χ.

Note that for $B = J_c$ with J_c as defined in Section 2.5, $\sigma_B(0)/\omega^2 = K_d c^d$ for some $K_d > 0$. Let c_* be such $K_d c_*^d = \chi/a$ if $a > 0$, and let $c_* = 1$ otherwise. For n sufficiently large, $\lambda_n c_* \ge c_{0.03}$ and cv^{LFST} is such that the LFST test controls size under $B = J_{c^*}$. But since $B = J_{c_*}$ satisfies the assumptions of Lahiri (2003), this model induces the same limit (27), so its $1-a$ quantile converges to $\text{cv}_\chi^{\text{LFST}}$, and the result follows. *Q.E.D.*

We now turn to the validity of residual-based tests. We first need a further generalization of the convergence of eigenvectors that accommodates projections off converging sequences of functions.

LEMMA 7: *Assume the conditions of Lemma 5 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 ψ. 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, ℓ) element of $\mathbf{M}_{\tilde{X}} \boldsymbol{\Sigma}_n \mathbf{M}_{\tilde{X}}$ is given by $\tilde{k}_n(s_l^0, s_\ell^0)$. Let $\sum_{i=1}^n \tilde{\nu}_i \tilde{\mathbf{r}}_i \tilde{\mathbf{r}}_i'$, $\tilde{\mathbf{r}}_i = (\tilde{r}_{i,1}, \dots, \tilde{r}_{i,n})'$ be the spectral decomposition of $\mathbf{M}_{\tilde{X}} \boldsymbol{\Sigma}_n \mathbf{M}_{\tilde{X}}$, where $\tilde{\mathbf{r}}_i' \tilde{\mathbf{r}}_i = n$ and $\tilde{\nu}_i \ge \tilde{\nu}_{i+1}$, and define $\tilde{\varphi}_i(\cdot) = n^{-1} \tilde{\nu}_i^{-1} \sum_{l=1}^n \tilde{r}_{i,l} \tilde{k}_n(\cdot, s_l^0)$, $i = 1, \dots, n-1$. Then $\sup_{s\in\mathcal{S}^0, 1\le i\le q} |\tilde{\varphi}_i(s) - \bar{\varphi}_i(s)| \to 0$ and $\max_{1\le i\le q} |\tilde{\nu}_i - \bar{\nu}_i| \to 0$.

PROOF: From standard arguments, we obtain

$$\int \psi_n(s)\psi_n(s)'\, dG_n(s) \to H \quad \text{and} \quad \int \psi(s)\psi_n(s)'\, dG_n(s) \to H.$$

Thus, $\|\mathbf{M}_{\tilde{X}} - \mathbf{M}_X\| \to 0$, and by a direct calculation, $\sup_{r,s\in\mathcal{S}^0} |\tilde{k}_n(r, s) - \hat{k}_n(r, s)| \to 0$, and $\sup_{r,s\in\mathcal{S}^0} |\hat{k}_n(r, s) - \bar{k}(r, s)| \to 0$ and thus $\sup_{r,s\in\mathcal{S}^0} |\tilde{k}_n(r, s) - \bar{k}(r, s)| \to 0$. Furthermore, proceeding as in the proof of Lemma 5 shows that $n^{-1}\|\boldsymbol{\Sigma}_n\|$ converges to ν_1, the largest eigenvalue of the integral operator with kernel k, so $n^{-1}\|\boldsymbol{\Sigma}_n\| = O(1)$, where $\|\cdot\|$ is the spectral norm. Thus also $n^{-1}\|\mathbf{M}_{\tilde{X}} \boldsymbol{\Sigma}_n \mathbf{M}_{\tilde{X}} - \mathbf{M}_X \boldsymbol{\Sigma}_n \mathbf{M}_X\| \to 0$, and from Weyl's inequality, $\max_{1\le i\le q} |\tilde{\nu}_i - \hat{\nu}_i| \to 0$. Since also $\max_{1\le i\le q} |\hat{\nu}_i - \bar{\nu}_i| \to 0$ from Lemma 5, we can conclude that $\max_{1\le i\le q} |\tilde{\nu}_i - \bar{\nu}_i| \to 0$. For the claim about the eigenfunctions, first note that

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

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 5. Also,

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

Finally, since $\max_{1\le i\le q} |\tilde{\nu}_i - \bar{\nu}_i| \to 0$ and $\bar{\nu}_1 > \bar{\nu}_2 > \cdots > \bar{\nu}_q > \bar{\nu}_{q+1}$, 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 $i = 1, \dots, q$. Applying Cauchy–Schwarz then yields

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

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

THEOREM 8: *Suppose $y_l = x_l'\beta + u_l$, $(x_l', u_l) = (X_n(s_l)', U_n(s_l)) \in \mathbb{R}^p \times \mathbb{R}$ with $(X_n(\cdot), U_n(\cdot))$ satisfying*

$$\begin{bmatrix} X_n^0(\cdot) \\ U_n^0(\cdot) \end{bmatrix} = \begin{bmatrix} \lambda_n^{-1/2} X_n(\lambda_n \cdot) \\ \lambda_n^{-1/2} U_n(\lambda_n \cdot) \end{bmatrix} \Rightarrow \begin{bmatrix} X^0(\cdot) \\ U^0(\cdot) \end{bmatrix} \tag{28}$$

on $\mathcal{S}^0$. Let $\mathbf{R}_X$ be the $n \times q$ matrix of q eigenvectors of $\mathbf{M}_X\mathbf{\Sigma}_n\mathbf{M}_X$ corresponding to the largest eigenvalues, normalized to satisfy $n^{-1}\mathbf{R}_X'\mathbf{R}_X = \mathbf{I}_q$. Suppose for almost all realizations 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

$$M_{X^0}(f)(s) = f(s) - \int X^0(r)'f(r)\,dG(r)\left(\int X^0(r)X^0(r)'\,dG(r)\right)^{-1} X^0(s)$$

are distinct. If also Condition 1 holds, then

$$\lambda_n^{-1/2} n^{-1}\mathbf{R}_X'\mathbf{Y} \Rightarrow \int \boldsymbol{\varphi}_{X^0}(s)U^0(s)\,dG(s), \tag{29}$$

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

Furthermore, let $\tilde{U}_n$ be independent of (X_n, U_n), and suppose $\tilde{U}_n$ satisfies $\tilde{U}_n^0(\cdot) = \lambda_n^{-1/2}\tilde{U}_n(\lambda_n\cdot) \Rightarrow \tilde{U}^0(\cdot)$ with $\tilde{U}^0 \sim U^0$. Let $\mathrm{cv}_n(X_n)$ be the $1 - \alpha$ quantile of the conditional distribution of $\phi(\mathbf{R}_X'\tilde{\mathbf{U}})$ given $\mathbf{R}_X$ for some continuous function $\phi : \mathbb{R}^q \mapsto \mathbb{R}$ satisfying $\phi(ax) = \phi(x)$ for all $a \neq 0$ and $x \in \mathbb{R}^q$, and $\tilde{\mathbf{U}} = (\tilde{U}_n(s_1), \dots, \tilde{U}_n(s_n))'$. Suppose that X^0 is independent of U^0, and 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 $\mathbb{P}(\phi(\mathbf{R}_X'\mathbf{Y}) > \mathrm{cv}_n(X_n)) \to \alpha$.

PROOF: 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^*, $n \ge 1$, such that $(X_n^*, U_n^*) \sim (X_n^0, U_n^0)$, $(X^*, U^*) \sim (X^0, U^0)$ and $\sup_{s\in\mathcal{S}^0} |X_n^*(s) - X^*(s)| \stackrel{a.s.}{\to} 0$, $\sup_{s\in\mathcal{S}^0} |U_n^*(s) - U^*(s)| \stackrel{a.s.}{\to} 0$. For clarity, we note the dependence of $\mathbf{R}_X$ and $\mathbf{Y}$ on n in the following. Using the same arguments as in the proof of Theorem 3, for P_0-almost all realizations $\omega_0 \in \Omega_0$, application of Lemma 7 then yields

$$\lambda_n^{-1/2} n^{-1} \mathbf{R}_{X,n}^{*\prime} \mathbf{Y}_n^* \to \int \boldsymbol{\varphi}_{X^*}(s) U^*(s)\, dG(s), \tag{30}$$

where $(\mathbf{R}_{X,n}^*, \mathbf{Y}_n^*)$ are defined analogously to $(\mathbf{R}_{X,n}, \mathbf{Y}_n)$ on $(\Omega_0, \mathfrak{F}_0, P_0)$, and $(\mathbf{R}_{X,n}^*, \mathbf{Y}_n^*) \sim (\mathbf{R}_{X,n}, \mathbf{Y}_n)$ for all $n \geq 0$ by construction. Since $(X^*, U^*) \sim (X^0, U^0)$, (29) follows.

For the further result, we will show that

$$\left(\phi\left(\mathbf{R}_{X,n}' \mathbf{Y}_n\right), \mathrm{cv}_n(X_n)\right) \Rightarrow \left(\phi\left(\int \boldsymbol{\varphi}_{X^0}(s) U^0(s)\, dG(s)\right), q_{1-\alpha}^{\phi}\left(X^0\right)\right) \tag{31}$$

with $q_{1-\alpha}^{\phi}(X^0)$ the $1-\alpha$ quantile of $\phi(\int \boldsymbol{\varphi}_{X^0}(s) U^0(s)\, dG(s))$ conditional on X^0. Given (31), $\mathbb{P}(\phi(\mathbf{R}_{X,n}' \mathbf{Y}_n) > \mathrm{cv}_n(X_n)) \to \alpha$ follows from the CMT applied to $\mathbf{1}[\phi(\mathbf{R}_{X,n}' \mathbf{Y}_n) > \mathrm{cv}_n(X_n)]$, and taking expectations.

Construct a probability space $(\Omega_1, \mathfrak{F}_1, P_1)$ with associated sequences of random processes $\tilde{U}^*$ 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 $\sup_{s \in \mathcal{S}^0} |\tilde{U}_n^*(s) - \tilde{U}^*(s)| \overset{a.s.}{\to} 0$. 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 (30) to 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, $\lambda_n^{-1/2} n^{-1} \mathbf{R}_{X,n}^{*\prime} \tilde{\mathbf{U}}_n^* \to \int \boldsymbol{\varphi}_{X^*}(s) \tilde{U}^*(s)\, dG(s)$ jointly with (30). But almost sure convergence implies convergence in distribution, so for P_0-almost all $\omega_0 \in \Omega_0$, the distribution of $\lambda_n^{-1/2} n^{-1} \mathbf{R}_{X,n}^{*\prime} \tilde{\mathbf{U}}_n^*$ induced by P_1 converges to the conditional distribution of $\int \boldsymbol{\varphi}_{X^*}(s) \tilde{U}^*(s)\, dG(s) \sim \int \boldsymbol{\varphi}_{X^*}(s) U^*(s)\, dG(s)$ given X^*. Since ϕ is continuous and the conditional distribution is assumed continuous, this implies that for all such ω_0, $\mathrm{cv}_n(X_n^*) \to q_{1-\alpha}^{\phi}(X^*)$. Thus, $(\phi(\mathbf{R}_{X,n}' \mathbf{Y}_n), \mathrm{cv}_n(X_n)) \sim (\phi(\mathbf{R}_{X,n}^{*\prime} \mathbf{Y}_n^*), \mathrm{cv}_n(X_n^*)) \overset{a.s.}{\to} (\phi(\int \boldsymbol{\varphi}_{X^*}(s) U^*(s)\, dG(s)), q_{1-\alpha}^{\phi}(X^*)) \sim (\phi(\int \boldsymbol{\varphi}_{X^0}(s) U^0(s)\, dG(s)), q_{1-\alpha}^{\phi}(X^0))$, and (31) follows. *Q.E.D.*

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

For the application to a residual-based cointegration test, suppose $(x_l', y_l) = (X(s_l)', Y(s_l))$ and $\lambda_n^{-1/2}(X(\lambda_n \cdot)', Y(\lambda_n \cdot))' \Rightarrow \Phi L_{X,Y}(s_l)$, where $L_{X,Y}$ is a vector of $p+1$ independent Lévy–Brownian motions. Then, as discussed in Footnote 7, it is without loss of generality to assume that there exists β such that (28) holds with U^0 independent of X^0, even though there is no assumption that $u_l = y_l - x_l' \beta$ is independent of x_l, as claimed in Remark 4.5. The critical value construction of Remark 4.2 is thus large-sample valid for general spatial $I(1)$ processes. The same reasoning applies if the regression contains additional deterministic terms, such as a constant or smooth functions of the locations.

APPENDIX C: GENERATION OF FIGURES 2 AND 3

For the left panel of Figure 2 and Figure 3, 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 3. These spectral densities are isotropic, so the covariance functions satisfy $\mathbb{E}[Y_i(r)Y_i(s)] = \sigma_i(|r - s|)$ with

$$\sigma_i(x) = \int_0^\infty J_0^B(\omega x)\tilde{f}_i(\omega)\, d\omega,$$

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

REFERENCES

ADLER, ROBERT J. (2010): *The Geometry of Random Fields*. Classics in Applied Mathematics. Philadelphia: SIAM. [1684]

ADLER, ROBERT J., AND JONATHAN E. TAYLOR (2007): *Random Fields and Geometry*. Springer Monographs in Mathematics. New York: Springer. [1666]

ANSELIN, LUC (1988): *Spatial Econometrics: Methods and Models*. Kluwer. [1662]

BESTER, C. ALLAN, TIMOTHY G. CONLEY, CHRISTIAN B. HANSEN, AND TIMOTHY J. VOGELSANG (2016): "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators," *Econometric Theory*, 32, 154–186. [1671]

BEVERIDGE, STEPHEN, AND CHARLES R. NELSON (1981): "A New Approach to Decomposition of Economics Time Series Into Permanent and Transitory Components With Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 7, 151–174. [1672]

CHAN, NGAI H., AND CHING-ZHONG WEI (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063. [1669]

CHETTY, RAJ, NATHANIEL HENDREN, PATRICK KLINE, AND EMMANUEL SAEZ (2014): "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 129, 1553–1623. [1661-1663,1667,1676-1678,1683]

CLIFF, ANDREW D., AND JOHN K. ORD (1974): *Spatial Autocorrelation*. London: Pion. [1662]

CONLEY, TIMOTHY G. (1999): "GMM Estimation With Cross Sectional Dependence," *Journal of Econometrics*, 92, 1–45. [1661,1662,1671]

DEO, CHANDRAKANT M. (1975): "A Functional Central Limit Theorem for Stationary Random Fields," *The Annals of Probability*, 3, 708–715. [1669]

DUDLEY, RICHARD M. (2002): *Real Analysis and Probability*. Camebridge, UK: Cambridge University Press. [1685]

ELLIOTT, GRAHAM (1998): "The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots," *Econometrica*, 66, 149–158. [1669]

——— (1999): "Efficient Tests for a Unit Root When the Initial Observation Is Drawn From Its Unconditional Distribution," *International Economic Review*, 40, 767–783. [1669]

ELLIOTT, GRAHAM, AND ULRICH K. MÜLLER (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73, 907–940. [1672]

ELLIOTT, GRAHAM, THOMAS J. ROTHENBERG, AND JAMES H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836. [1669,1674]

ENGLE, ROBERT F., AND CLIVE W. J. GRANGER (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276. [1664,1676]

FINGLETON, BERNARD (1999): "Spurious Spatial Regression: Some Monte Carlo Results With a Spatial Unit Root and Spatial Cointegration," *Journal of Regional Science*, 39, 1–19. [1681,1682]

GEARY, ROY C. (1954): "The Contiguity Ratio and Statistical Mapping," *The Incorporated Statistician*, 5, 115–145. [1673]

GELFAND, ALAN E., PETER DIGGLE, PETER GUTTORP, AND MONTSERRAT FUENTES (2010): *Handbook of Spatial Statistics*. CRC Press. [1662]

GRANGER, CLIVE W. J., AND PAUL NEWBOLD (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120. [1661]

HANSEN, BRUCE, AND SEOJEONG LEE (2019): "Asymptotic Theory for Clustered Samples," *Journal of Econometrics*, 210, 268–290. [1671,1686]

IVANOV, ALEXANDER V., AND NIKOLAI N. LEONENKO (1989): *Statistical Analysis of Random Fields*. Dordrecht: Kluwer Academic Publishers. [1668,1693]

KALLENBERG, OLAV (2021): *Foundations of Modern Probability*. Springer. [1684]

KELLY, MORGAN (2019): "The Standard Errors of Persistence," University College Dublin WP19/13. [1661, 1671]

——— (2020): "Understanding Persistence," CPER Discussion Paper DP15246. [1661]

KING, MAXWELL L. (1987): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218. [1674]

KWIATKOWSKI, DENIS, PETER C. B. PHILLIPS, PETER SCHMIDT, AND YONGCHEOL SHIN (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159–178. [1672,1674]

LAHIRI, SOUMENDRA N. (2003): "Central Limit Theorems for Weighted Sums of a Spatial Process Under a Class of Stochastic and Fixed Designs," *Sankhya*, 65 (2), 356–388. [1665,1676,1689,1690]

LEE, LUNG-FEI, AND JIHAI YU (2009): "Spatial Nonstationarity and Spurious Regression: The Case With a Row-Normalized Spatial Weights Matrix," *Spatial Economic Analysis*, 4, 301–327. [1681]

——— (2013): "Near Unit Root in the Spatial Autoregressive Model," *Spatial Economic Analysis*, 8, 314–351. [1681]

LÉVY, PAUL (1948): *Processus stochastiques et mouvement brownien*. Gauthier-Vilars. [1665]

LINDSTRØM, TOM (1993): "Fractional Brownian Fields as Integrals of White Noise," *Bulletin of the London Mathematical Society*, 25, 83–88. [1667]

MATÉRN, BERTIL (1986): *Spatial Variation*. Lecture Notes in Statistics, Vol. 36. Berlin: Springer. [1669]

MORAN, PATRICK A. P. (1950): "Notes on Continuous Stochastic Phenomena," *Biometrika*, 37, 17–23. [1673]

MÜLLER, ULRICH K., AND MARK W. WATSON (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016. [1664,1672-1675]

——— (2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81. [1675]

——— (2016): "Measuring Uncertainty About Long-Run Predictions," *Review of Economic Studies*, 83, 1711–1740. [1669]

——— (2022): "Spatial Correlation Robust Inference," *Econometrica*, 90, 2901–2935. [1661,1671,1674,1676, 1680,1687-1690]

——— (2023): "Spatial Correlation Robust Inference in Linear Regression and Panel Models," *Journal of Business and Economic Statistics*, 41 (4), 1050–1064. [1661,1680]

NEWEY, WHITNEY K., AND KENNETH D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [1661]

NYBLOM, JUKKA (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230. [1672,1674]

PHILLIPS, PETER C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340. [1661,1664,1670]

——— (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547. [1669]

——— (1998): "New Tools for Understanding Spurious Regression," *Econometrica*, 66, 1299–1325. [1667, 1670,1671]

PHILLIPS, PETER C. B., AND SAM OULIARIS (1990): "Asymptotic Properties of Residual Based Test for Cointegration," *Econometrica*, 58, 165–193. [1676]

PRATT, JOHN W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 549–567. [1675]

ROSASCO, LORENZO, MIKHAIL BELKIN, AND ERNESTO DE VITO (2010): "On Learning With Integral Operators," *Journal of Machine Learning Research*, 11 (30), 905–934. [1676,1687]

ROSSI, FRANCESCA, AND OFFER LIEBERMAN (2023): "Spatial Autoregressions With an Extended Parameter Space and Similarity-Based Weights," *Journal of Econometrics*, 235, 1770–1798. [1681,1682]

SAITOH, SABUROU, AND YOSHIHIRO SAWANO (2016): *Theory of Reproducing Kernels and Applications*. New York: Springer. [1687]

STEIN, MICHAEL L. (2002): "Fast and Exact Simulation of Fractional Brownian Surfaces," *Journal of Computational and Graphical Statistics*, 11, 587–599. [1693]

STOCK, JAMES H. (1991): "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459. [1664,1669]

SUN, YIXIAO, AND MIN SEONG KIM (2012): "Asymptotic F-Test in a GMM Framework With Cross-Sectional Dependence," *Review of Economics and Statistics*, 91 (1), 210–233. [1671]

WRIGHT, JONATHAN H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business and Economic Statistics*, 18, 211–222. [1675]

YU, YI, TENGYAO WANG, AND RICHARD J. SAMWORTH (2015): "A Useful Variant of the Davis–Kahan Theorem for Statisticians," *Biometrika*, 102, 315–323. [1691]

Co-editor Guido W. Imbens handled this manuscript.

Manuscript received 1 March, 2023; final version accepted 28 June, 2024; available online 10 July, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.11199509. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.