Spatial Unit Roots

Ulrich K. Müller and Mark W. Watson
Department of Economics, Princeton University
Princeton, NJ, 08544
This Draft: November 2022

Abstract

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

Keywords: spatial correlation, spurious regression, Lévy-Brownian motion, functional central limit theorem

JEL: C12, C20


1 Introduction

Serial correlation complicates inference in time series regressions. When the serial correlation in the regressors and regression errors is weak, that is $I(0)$, inference can proceed as with i.i.d. sampling after using HAC/HAR standard errors that incorporate adjustments for serial correlation. However, when the serial correlation is strong, that is $I(1)$, HAC/HAR inference fails and OLS produces “spurious regressions” (Granger and Newbold (1974)) with estimators and test statistics behaving in non-standard ways (Phillips (1986)). Panel (a) of Figure 1 illustrates this well-known phenomenon: The realization of two independent random walks of length $n = 250$ are strongly correlated in sample, with a corresponding Newey and West (1987) $t$-statistic that is highly significant.

Variables measured over points in space exhibit correlation patterns that in many ways are analogous to serial correlation in time series, and this correlation also complicates inference in spatial regressions. There is a reasonably well-developed literature on the requisite spatial HAC/HAR corrections that are required in spatial regressions with weakly dependent stationary regressors and errors. However, much less is known about the implications of strong spatial correlation despite evidence suggesting its presence in many empirical applications in economics (Kelly (2019, 2020)). Panel (b) of Figure 1 illustrates the issue: The realization of two independent spatial “unit root” processes with values for each of the $n = 741$ U.S. commuter zones are strongly correlated in sample, and a $t$-statistic that is clustered by U.S. states is highly significant. This raises several natural questions. What is a natural spatial analogue of an $I(1)$ time series process, such as the process in Figure 1 (b)? Do such processes systematically induce spuriously significant regression coefficients? How can one test for $I(1)$ spatial persistence? And finally, is there a spatial analogue to the “first-differencing” transformation in time series that eliminates $I(1)$ persistence? This paper takes up these questions.

Throughout the paper we use spatial data and regressions from Chetty, Hendren, Kline, and Saez (2014) to illustrate the issues and methods. These authors construct an index of intergenerational mobility for the 741 commuting zones, and study its relationship to other socioeconomic factors by bivariate regressions with standard errors clustered by U.S. states. As an example, Figure 1 (c) plots their mobility index along with the teen labor force participation. The apparent similarity of this data with the simulated data of panel (b)

---

1 Conley (1999) is a leading example of spatial HAC inference. See Müller and Watson (2022a, 2022b) for a discussion of the post-Conley literature and new suggestions for inference in regression model with weak spatial dependence.
Figure 1: Strongly Dependent Data in Time and Space

(a) Independent Time Series Random Walks

(b) Independent Spatial Unit Root Processes

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

- Mobility Index
  - $R^2 = 0.40$
  - $t$-stat = 3.71

- Teen Labor Force Participation
  - $R^2 = 0.48$
  - $t$-stat = 8.31

- $R^2 = 0.14$
  - $t$-stat = 3.52
suggests that the issues considered in this paper are of empirical relevance. Much of our analysis parallels the analysis of persistent time series, but there is a notable difference worth highlighting at the outset. Time series analysis typically studies observations, say \( y_t \), observed at equidistant points in time, \( t = 1, 2, 3, \ldots \) where \( t \) indexes months, quarters, years, etc. Economic variables observed in space are not so neatly arranged. For example, geographical data may be collected at potentially arbitrary locations \( s_t \) within a given region \( S \) such as a U.S. state, and each state has its own unique shape. For the analysis to be useful in a wide range of spatial applications, we posit a model that assigns values to all locations that may potentially be observed. Thus, for the general problem with \( d \) spatial dimensions, we begin with a stochastic process \( Y_n(s) \) over \( s \in \mathbb{R}^d \), where \( d = 2 \) in the geography example. When \( d = 1 \), \( s \) could index time, so this is a time series model where \( Y_n(s) \) is a continuous time process and where the sample data correspond to realizations of \( y_t = Y_n(s_t) \) observed at potentially irregularly spaced points \( s_t \in \mathbb{R} \).

We thus follow the geostatistical tradition of positing a continuous parameter model of spatial variation, rather than modelling spatial dependence by spatial autoregressive (SAR) models of Cliff and Ord (1974) and Anselin (1988). There is a small literature on unit roots and spurious regression in SAR models, initiated by Fingleton (1999). SAR models require the definition of a spatial weights (or proximity) matrix, which is usually normalized so that its rows sum to unity (Ord (1975)). Under that normalization, the unit root SAR model is not well defined. Fingleton (1999) and Mur and Trívez (2003) modify the weight matrix so that one unit becomes unconnected. In contrast, Beenstock, Feldman, and Felsenstein (2012) abandon the row sum normalization, and Lauridsen and Kosfeld (2006) employ a generalized inverse to solve the model. The only authors that derive asymptotic distribution results are Lee and Yu (2009, 2013), who study the row normalized SAR model with a coefficient that converges to unity. They find that this model does not induce spurious regression effects of the type encountered in time series: OLS coefficients remain asymptotically normal, the regression \( R^2 \) converges in probability to zero, and t-statistics do not diverge. These results are markedly different from our findings based on a continuous parameter model \( Y_n(\cdot) \) of a spatial I(1) process.

With this background in place, the roadmap of the paper is as follows. Section 2 provides our definition of a spatial I(1) process. In time series models \((d = 1 \text{ in our notation})\), the

\[2\text{See Gelfand, Diggle, Guttorp, and Fuentes (2010) and Schabenberger and Gotway (2005) for useful overviews.} \]
canonical \( I(1) \) process is a Wiener process. Lévy-Brownian motion is a useful generalization of the Wiener process for \( d > 1 \), and Section 2 begins by reviewing its properties. In regular time series models, more general \( I(1) \) processes can be constructed by replacing the white noise increments of a random walk with a weakly correlated stationary series. For example, stationary ARMA\((p, q)\) noise yields a ARIMA\((p, 1, q)\) process. Section 2 defines the spatial \( I(1) \) process similarly 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 distribution of functions of \( I(1) \) processes can be approximated by the distributions of corresponding functions of Wiener processes. The FCLT is the core of such an approximation, and it provides the basis for large-sample inference using statistics constructed from realizations of \( I(1) \) processes. Section 2 provides a FCLT that is applicable to \( I(1) \) spatial processes. We also show how to appropriately generalize the \( I(1) \) model to a spatial “local-to-unity” process, and provide a corresponding FCLT result about its large sample behavior.

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

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

Section 5 takes up the problem of conducting inference about the degree of persistence in a scalar spatial variable. In particular, we construct spatial analogues of the time series “low-frequency” unit root and stationary tests of Müller and Watson (2008). In addition, we suggest a confidence interval for the mean reversion parameter in the spatial local-to-unity
model, analogous to the time series work by Stock (1991).

First-differencing an $I(1)$ time series yields an $I(0)$ process, so spurious time series regressions can be avoided by taking first differences of $I(1)$ variables. An analogous transformation for spatial $I(1)$ processes are the isotropic differences introduced in Section 4. Section 6 provides Monte Carlo evidence that regressions using isotropic differences do not suffer from spurious regression problems, and that valid inference can be conducted using the spatial-correlation robust methods developed in Müller and Watson (2022a, 2022b). But there are other intuitively plausible methods that may eliminate or mitigate problems associated with $I(1)$ variables in a spatial regression. These other methods include (i) low-pass and high-pass spectral regressions, (ii) regressions that incorporate small-area fixed effects, (iii) pooling estimates constructed from data in non-overlapping regions, and (iv) employing a GLS transformation based on Lévy-Brownian motion. Section 6 compares the coverage and efficiency of feasible confidence intervals from versions of these methods. We find the GLS transformation to be particularly efficient.

Section 7 offers some concluding remarks. The appendix contains all proofs.

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

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

2.1 Set-up and Notation

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

### 2.2 Lévy-Brownian Motion

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

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

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

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

with $|a| = \sqrt{a^2}$ for $a \in \mathbb{R}^d$, so in particular, $\text{Var}(L(s)) = |s|$ and $\text{Var}(L(s) - L(r)) = |s - r|$. When $d = 1$ and $s, r \geq 0$, the covariance function (1) simplifies to $E[L(s)L(r)] = \min(s,r)$, the covariance function of a Wiener process, so Lévy-Brownian motion reduces to a Wiener process. More generally, for any $d$, the process obtained along a line in $\mathbb{R}^d$, $W_{a,b}(s) = L(a + bs) - L(a)$, $a, b \in \mathbb{R}^d$, $|b| = 1$, $s \in \mathbb{R}$ is a Wiener process. Thus, $L$ is a natural embedding of the canonical time series model of strong persistence to the spatial case. Notice that Lévy-Brownian motion is **isotropic**, that is, $\text{Var}(L(s) - L(r))$ depends on $s, r$ only through $|s - r|$. Thus, Lévy-Brownian motion is invariant to rotations of the spatial axes, $L(Ors) \sim L(s)$, for any $d \times d$ rotation matrix $O$. Another generalization of the Wiener process to $d > 1$ is a **Brownian sheet** $\int_{\mathbb{R}^d} 1[0 \leq r \leq s]dW(r)$, $s \geq 0$, where the inequality $0 \leq r \leq s$ is to be understood element by element. The Brownian sheet is not isotropic, as is clearly visible in the sample realizations in Figure 2.2 for $d = 2$.\(^3\)

\(^3\)The generation of the left panel of Figure 2.2 as well as Figure 2.3 uses the computational approach of Stein (2002).
By Mercer’s Theorem, the covariance kernel (1) evaluated at $s, r \in S$ can be represented as

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

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

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

where the right hand side converges uniformly on $S$ with probability one (cf. Theorem 3.1.2 of Adler and Taylor (2007)). This result generalizes the corresponding observation in Phillips (1998) about representations of the Wiener process in terms of stochastically weighted averages of deterministic series. Figure 2.2 plots some of the eigenfunctions $\varphi_j$ for $S$ the continental U.S. and $G$ the uniform distribution.

A Wiener process can trivially be written as an integral over white noise, $W(s) = \int_0^s dW(r)$. Such a “moving average” representation exists for Lévy-Brownian motion for
all $d \geq 1$: from Stoll (1986) and Lindstrøm (1993)

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

(4)

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

### 2.3 Spatial I(1) Processes

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

$$Y_0(s) = \int h(r, s)B(r)dr.$$ 

(5)
Figure 4: Sample Realizations of $Y_0$ with Different Underlying $B$

$Y_0$ with $B \sim J_{100}$  

$Y_0$ with $B \sim J_{50,200}$

Note that if $B$ is isotropic, then so is $Y_0$.

Figure 2.3 plots two such $Y_0$ processes on the unit square with $B$ mean zero isotropic Gaussian processes of the type $J_{100}$ and $J_{50,200}$, which are characterized by a spectral density proportional to $1/(|\omega|^2 + 100^2)^{3/2}$ and $(|\omega|^2 + 50^2)^{3/2}/(|\omega|^2 + 200^2)^3$ for $\omega \in \mathbb{R}^d$, respectively (we discuss the $J_c$ process in greater detail in Section 2.5 below). As demonstrated in Figure 2.3, different characteristic of $B$ can induce quite different local behavior of $Y_0$.

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

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

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

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

Note that $\int_{\mathbb{R}^d} |h(r,s)| dr$ does not exist for $d > 1$, so $Y_0$ in (5) is not defined pathwise for every realization of $B$. However, $\int_{\mathbb{R}^d} h(r,s)^2 dr < \infty$, so under appropriate weak dependence
conditions on $B$, the integral that defines $Y_0$ can be shown to converge in a mean square sense. In particular, we have the following result.

**Lemma 1.** Under Condition 1, for all $d \geq 1$, $Y_0(\cdot)$ exists on any compact set $C \subset \mathbb{R}^d$ and has continuous sample paths with probability one.

Our definition of a spatial $I(1)$ process on $S$ is

$$Y_n(s) = Y_0(\lambda_n s), \quad s \in S$$

(6)

for some sequence of positive numbers $\lambda_n$, so that $\text{Var}[Y_n(s)] = O(\lambda_n)$. To make sense of this definition, consider first the $d = 1$ case with $s_t \in S = [0, 1]$. The continuous time analogue of the usual definition $y_t = \sum_{s=1}^{t} u_s$ for an $I(1)$ time series $y_t$ is given by $Y_n(s) = \int_0^{\lambda_n} B(r) dr = Y_0(\lambda_n s) = Y_n(s)$ with $\lambda_n = n$. In fact, with $u_t = \int_{t-1}^{t} B(r) dr$, the definitions coincide, since then $y_t = Y_n(t/n)$ for $t = 1, \ldots, n$. With potentially irregularly spaced $s_t \in S = [0, 1]$ with empirical distribution $G_n \Rightarrow G$ and $G$ absolutely continuous, the typical distance between two neighboring locations $s_t$ is still $O(n^{-1})$. A rate of $\lambda_n = n$ thus ensures that $Y_0$ is evaluated at substantially different locations, so this sampling scheme is “pure outfill”. A constant $\lambda_n = \lambda_0$, in contrast, corresponds to pure infill sampling, as $Y_0$ is then only ever evaluated over the fixed region $[0, \lambda_0]$. Finally, $\lambda_n \rightarrow \infty$ with $\lambda_n/n \rightarrow 0$ corresponds to a mixture of infill and outfill sampling, where neighboring locations evaluate $Y_0$ at ever closer values as $n \rightarrow \infty$, but the effective range of $\lambda_n s_t$ is the increasing interval $[0, \lambda_n]$.

The exact same comments apply to the $d > 1$ case, except that now the typical distance between two neighboring locations is of order $O(n^{1/d})$. Thus, pure outfill sampling corresponds to $\lambda_n \propto n^{1/d}$, where $Y_0$ is effectively evaluated over the region $\lambda_n S$ of volume $O(n)$. A mixture of infill and outfill sampling arises with $\lambda_n \rightarrow \infty$, $\lambda_n/n^{1/d} \rightarrow 0$.

### 2.4 A Functional Central Limit Theorem

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

We assume the following central limit condition on $B$. 

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

We now present a FCLT that yields convergence of a suitably scaled general $I(1)$ process to Lévy-Brownian motion. The result requires some degree of outfill sampling, $\lambda_n \to \infty$, so intuitively, the theorem describes what happens if one “zooms far enough out” in Figure 2.3.

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

**Remark 2.1.** In practice, Theorem 2 can be used to argue that the distribution of a functional $\psi(Y_n)$ is well approximated by the distribution of $\psi(\lambda_n^{1/2} \omega L)$. Note, however, that this requires $\psi$ to be sufficiently continuous for the continuous mapping theorem to be applicable. For instance, recall that for $y_t$ a mean-zero zero $I(1)$ time series, a FCLT implies that $n^{-1/2}y_{[n]} \Rightarrow \omega W(\cdot)$, yet $y_t - y_{t-1} = u_t$ does not in general converge to a Gaussian variable. The same holds for our generalization to spatial $I(1)$ processes $Y_n$: As noted above, the distance between two typical neighboring locations is $O(n^{-1/d})$. For $d > 1$, the difference in the value of $Y_n(\cdot)$ evaluated at two such neighboring points $s$ and $s + n^{-1/d} a$, $a \in \mathbb{R}^d$ is given by

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

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

(7)

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

### 2.5 Spatial Local-to-Unity Processes

A large time series literature, initiated by Chan and Wei (1987) and Phillips (1987), concerns a generalization of the $I(1)$ model to the weakly mean reverting local-to-unity model. In this model, $y_t$ satisfies $n^{-1/2}(y_{[n]} - y_t) \Rightarrow \omega (J_c(\cdot) - J_c(0))$, with $J_c$ a stationary Ornstein-Uhlenbeck (OU) process with covariance kernel $\mathbb{E}[J_c(s)J_c(r)] = \exp[-c|s - r|/(2c)]$, $c > 0$. 

11
As demonstrated by Elliott (1999), \( J_c(\cdot) - J_c(0) \) converges to a Wiener process as \( c \to 0 \). We now generalize the spatial \( I(1) \) process defined above to an analogous local-to-unity spatial model.

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

Furthermore, from equation of (3.2.8) of Matérn (1986), \( J_c \) has the moving average representation

\[
J_c(s) = \int_{\mathbb{R}^d} h_c(r, s)dW(r)
\]

with \( h_c(r, s) = \kappa_{c,d}|s - r|^{(1-d)/4}K_{(1-d)/4}(c|s - r|) \) for a suitable choice of constant, where \( K_\nu \) is the modified Bessel function of the second kind, \( d \geq 1.4 \). Thus, replacing again the white noise term by the weakly dependent random field \( B \) as in (5), let

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

and define the spatial local-to-unity process on \( \mathcal{S} \) via

\[
Y_n(s) = Y_c/\lambda_n(\lambda_n s).
\]

As \( \lambda_n \) diverges and \( Y_c \) in (9) is evaluated at locations that are further and further apart, the degree of mean reversion shrinks at the rate \( \lambda_n^{-1} \). In that manner, the overall degree of mean reversion of \( Y_n \) over the fixed set \( \mathcal{S} \) converges as \( n \to \infty \). A calculation shows that \( \text{Var}[Y_n(s)] = O(\lambda_n) \), as for the spatial \( I(1) \) process.

Proceeding as for the \( I(1) \) model, the appendix shows that under Condition 1 \( Y_c(\cdot) \) and thus \( Y_n(\cdot) \) in (9) exist for all \( n \). Furthermore, under the conditions of Theorem 2, \( Y_n \) in (9) satisfies \( \lambda_n^{-1/2}Y_n(\cdot) \Rightarrow \omega J_c(\cdot) \).

**Remark 2.2.** These results allow for an extension of the central-limit results in Lahiri (2003) for weighted averages of weakly dependent covariance stationary spatial processes to weighted averages of spatial \( I(1) \) and local-to-unity processes.

---

\(^4\)For \( d = 1 \), the usual one-sided (causal) representation for a stationary OU process is \( J_c(s) = \int_s^\infty e^{-c(s-r)}dW(r) \). Equation (8) is an alternative two-sided (non-causal) representation when \( d = 1. \)
3 Spurious Regressions with Spatial $I(1)$ Variables

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

$$y_l = \alpha + x_l' \beta + u_l$$

for $l = 1, \ldots, n$, where $(y_l, x_l) = (Y_n(s_l), X_n(s_l)) \in \mathbb{R}^{p+1}$ 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 $\tilde{y}_l = y_l - n^{-1} \sum_{t=1}^n y_t$ denote the demeaned value of $y_l$ and similarly for $x_l$. Let $s_{\tilde{y}\tilde{y}} = n^{-1} \sum_{t=1}^n \tilde{y}_t^2$, $S_{\tilde{y}\tilde{x}} = n^{-1} \sum_{t=1}^n \tilde{x}_t \tilde{y}_t$ and $S_{\tilde{x}\tilde{y}} = n^{-1} \sum_{t=1}^n \tilde{x}_t \tilde{y}_t$. The OLS estimator is $\hat{\beta} = S_{\tilde{y}\tilde{x}} S_{\tilde{x}\tilde{y}}^{-1}$, the regression $R^2 = S_{\tilde{y}\tilde{y}} S_{\tilde{x}\tilde{y}}^{-1} S_{\tilde{x}\tilde{y}} / s_{\tilde{y}\tilde{y}}$, the OLS estimator for the variance of $u_l$ is $s_u^2 = n^{-p-1}(s_{\tilde{y}\tilde{y}} - S_{\tilde{y}\tilde{y}} S_{\tilde{x}\tilde{y}}^{-1} S_{\tilde{x}\tilde{y}})$, and the classical (non-spatial-correlation robust, homoskedastic) $F$-statistic for testing $H_0 : H \beta = 0$, where $H$ is a non-stochastic matrix with rank$(H) = m \leq p$, is $F_{\text{Hom}} = \frac{n-p-1}{m} \hat{\beta}' H' (H'S_{\tilde{x}\tilde{x}}H)^{-1} H \hat{\beta} / s_u^2$.

Suppose $(y_l, x_l) = (Y_n(s_l), X_n(s_l))$ follow spatial $I(1)$ processes with

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

(11)

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

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

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

(ii) $\tilde{\beta} \Rightarrow \Xi_{\tilde{y}\tilde{x}} / \Xi_{\tilde{x}\tilde{x}}$,

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

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

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

**Remark 3.1.** In the one-dimensional case with $d = 1$, $\mathcal{S}$ the unit interval and $G$ the uniform distribution, these results coincide with the spurious time-series regression limits derived in Phillips (1986). In the general spatial case, the limits are seen to depend on the spatial distribution of locations $G$ and its support $\mathcal{S}$. Section 6 provides numerical results for the behavior of $R^2$ for a large set of spatial designs with $d = 2$. 

13
Remark 3.2. An implication of part (v) of Theorem 3 is that the classical F-test statistic diverges to infinity so that \( P(F_{\text{Hom}} > cv) \to 1 \) for any \( cv \geq 0 \).

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

\[
F_{\text{HAC}} = \frac{n}{m} \tilde{\beta}' H'(HS_S^{-1} \hat{\Omega}_n S^{-1} H')^{-1} H \tilde{\beta}
\]

(12)

where \( \hat{\Omega}_n \) is a kernel-based estimator of \( \text{Var}(n^{-1/2} \sum \tilde{x}_t u_t) \) of the form

\[
\hat{\Omega}_n = n^{-1} \sum_{t, \ell=1}^{n} \kappa(b_n(s_t - s_{\ell})) e_t e_{\ell}'
\]

(13)

with \( e_t = \tilde{x}_t(y_t - \tilde{x}_t \tilde{\beta}) \), \( b_n \) a bandwidth (that may depend both on \( \{s_t\} \) and the data \( \{(y_t, x_t)\} \)) with \( b_n^{-1} = o_p(1) \) and \( \kappa : \mathbb{R}^d \to \mathbb{R} \) a kernel weighting function satisfying

\[
\sup_r |\kappa(r)| = \bar{\kappa} < \infty, \quad \lim_{\lambda \to \infty} \sup_{|a|=1} |\kappa(\lambda a)| = 0.
\]

(14)

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

The following result shows that inference using any such consistent spatial HAC estimator does not avoid spurious significance of the \( F_{\text{HAC}} \) test statistic.

Theorem 4. If \( G_n \Rightarrow G \) and (11) and (14) hold, then \( P(F_{\text{HAC}} > cv) \to 1 \) for any \( cv \geq 0 \).

Remark 3.3. Theorems 3 and 4 also hold for local-to-unity processes, that is, if \( [Y(\cdot), X(\cdot)] \) in (11) are \( p+1 \) independent processes of the type (8), with arbitrary and potentially different mean-reversion parameters \( c \). This is because the asymptotics that yield convergence to \( J_c(\cdot) \) are “pure infill” relative to the degree of mean reversion, and pure infill asymptotics are known to potentially lead to inconsistent parameter estimators (cf. Zhang and Zimmerman)
In contrast, no degree of “outfill” (or increasing domain) asymptotics can remedy the spurious regression effect in the $I(1)$ model, again just as in the time series case.

**Remark 3.4.** It follows from the Karhunen–Loève representation of $L$ in (3) and the FCLT result in Theorem 2 that the coefficients of a regressions of $\lambda_n^{-1/2}y_t$ on the eigenfunctions $[\varphi_1(s_t), \ldots, \varphi_p(s_t)]$ 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 $\lambda_n^{-1/2}y_t$ on smooth deterministic functions of $s_t$, say $\psi(s_t) \in \mathbb{R}^p$, converge to $\left(\int \psi(r)\psi(r)’dG(r)\right)^{-1}\omega\int \psi(r)L(r)dg(r)$ and are asymptotically significant as measured by a corresponding $F_{\text{Hom}}$ or $F_{\text{HAC}}$ statistic. Kelly (2019) observes such a phenomenon empirically in a number of applications with spatial data.

## 4 Isotropic Differences Regression

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

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

$$y^*_t = \frac{1}{n} \sum_{\ell \neq t} \kappa_b(|s_\ell - s_t|)(y_\ell - y_t)$$

(15)

for some weighting function $\kappa_b : \mathbb{R} \mapsto \mathbb{R}$ with $\kappa_b(x) = \kappa_0(x/b)$ for $b > 0$ and $\kappa_0(x) = 0$ for $|x| > 1$. We call this transformation “isotropic differencing”, as (15) averages over all

---

5At the same time, fixed-$b$ type spatial HAR inference (Bester, Conley, Hansen, and Vogelsang (2016), Sun and Kim (2012)) does not lead to diverging $F$-statistics, and the spatial correlation robust inference derived in Müller and Watson (2022a) explicitly accommodates some degree of “strong” persistence of the type exhibited by the spatial local-to-unity model for large enough $c$. See Section 6 below for corresponding numerical results.
differences that are within a ball of radius $b$ around $s_l$, so it is invariant to rotations of the data. The larger the bandwidth $b$, the more averaging is being employed.

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

**Theorem 5.** Let $\hat{\gamma}$ be the coefficient of an OLS regression of $y^*_l$ in (15) on $y_l$ for all $l$ such that $s_l \in I_b$. Suppose $y_t$ is a spatial local-to-unity process, $I_b$ has positive volume, and $\kappa_0$ has a finite number of discontinuity points and $G_n \Rightarrow G$. Then

$$
\hat{\gamma} \Rightarrow \int_{I_b} J_c(s) \int_S \kappa_b(|r-s|)(J_c(r)-J_c(s))dG(r)dG(s) \int_{I_b} J_c(s)^2 dG(s).
$$

(16)

**Remark 4.1.** Due to the long-range nature of the differences (15) with $b$ fixed, the long-run variance $\omega^2$ cancels in the limiting expression for $\hat{\gamma}$. One could hence use $\hat{\gamma}$ to learn about the degree of mean reversion $c$, analogous to what is suggested in Stock (1991). In order to simulate corresponding critical values, one could replace $G$ by the empirical distribution $G_n$ on the r.h.s. of (16), which is asymptotically justified (see the proof of Theorem 5).

**Remark 4.2.** If a constant is included in the regression, then the same result holds with $J_c$ replaced by $\tilde{J}_c$ with $\tilde{J}_c(s) = J_c(s) - \int_{I_b} J_c(r)dG(r)$. Also, for a spatial $I(1)$ process $y_t$, (16) holds with $J_c$ replaced by $L$.

**Remark 4.3.** If the isotropic differences are computed with weights that are normalized to sum to one, $y^*_l = \sum_{t \neq l} \kappa_b(|s_l - s_t|)(y_t - y_t)/\sum_{t \neq l} \kappa_b(|s_l - s_t|)$, then (16) holds with $\kappa_b(|r-s|)$ replaced by $\kappa_b(|r-s|)/\int \kappa_b(|u-s|)dG(u)$.

By a change of variables, the numerator in (16) may be rewritten as

$$
b^d \int_{I_b} J_c(s) \int_{|r| \leq 1} \kappa_0(|r|)(J_c(s+br) - J_c(s))g(s+br)drdG(s)
$$

(17)

where $g$ is the density corresponding to the distribution $G$. As the bandwidth $b$ shrinks, the numerator in (16) thus becomes smaller for two reasons: mechanically, because the weight function $\kappa_b$ vanishes over most of $S$ (this is the term $b^d$ in (17)), and because the differences $J_c(r) - J_c(s)$ are computed with $r, s$ close to each other (cf. $J_c(s+br) - J_c(s)$, $|r| < 1$ in (17)). Such differences have small variance, suggesting that the numerator has little variability as $b \to 0$. The following result formalizes this intuition and establishes the limiting constant.
Theorem 6. Suppose the density $g$ of $G$ admits three bounded derivatives on $S$. Then under the assumptions of Theorem 5, as $b \to 0$,

$$b^{-d-1} \int_{I_b} J_c(s) \int_S \kappa_b(|r-s|)(J_c(r) - J_c(s))dG(r)dG(s) \xrightarrow{p} -\frac{1}{2} \int_{\mathbb{R}^d} |r|\kappa_0(|r|)dr. \quad (18)$$

Furthermore, $(18)$ continues to hold with $J_c(s)$ replaced by $J_c(s) - \bar{m}$ for any random variable $\bar{m}$ with $\mathbb{E}[\bar{m}^2] < \infty$, and also for $J_c$ replaced by $L$.

Remark 4.4. To get some intuition for the result in Theorem 6, consider a time series random walk $y_t = \sum_{s=1}^t \varepsilon_t$ with $\varepsilon_t \sim iid(0,1)$. As is well known, $n^{-1} \sum_{t=2}^{n-1} \Delta y_{t+1} y_t \Rightarrow \int_0^1 W(s)dW(s)$. Also, $n^{-1} \sum_{t=2}^{n-1} (y_{t-1} - y_t)y_t = n^{-1} \sum_{t=2}^{n-1} (-\Delta y_t)(y_{t-1} + \Delta y_t) \Rightarrow \int_0^1 W(s)dW(s) - 1$. Thus, $n^{-1} \sum_{t=2}^{n-1} (\Delta y_{t+1} + (y_{t-1} - y_t))y_t \xrightarrow{p} -1$. Treating the forward and backward difference as each receiving unit weight for a total of 2, this result accords with (18). Symmetric time series autoregressive estimators of this type are studied by Pantula, Gonzalez-Farias, and Fuller (1994) and Fuller (1996); see, for instance, Theorems 10.1.7 and 10.1.8 in the latter.

Remark 4.5. As $b \to 0$, the denominator in (16) converges to $\int_S J_c(s)^2 dG(s)$. Suppose we pick a weight function $\kappa_0$ such that $\int_{\mathbb{R}^d} |r|\kappa_0(|r|)dr = 1$. Taking $b \to 0$ limits after $n \to \infty$ limits yields a limiting distribution of $-b^{-d-1}\hat{\gamma}$ equal to $\frac{1}{2} \int_S J_c(s)^2 dG(s)$. Recalling that $E[J_c(s)^2] = (2c)^{-1}$, this suggests that $-b^{-d-1}\hat{\gamma}$ may be used to estimate the degree of mean reversion $c$ of the local-to-unity spatial process $y_t$. This is analogous to using $n(1 - \hat{\rho})$, with $\hat{\rho}$ the estimator of the largest autoregressive in a regular time series. The limiting distribution of $-b^{-d-1}\hat{\gamma}$ never takes on negative values, again mirroring the corresponding results for the symmetric autoregressive time series estimator $\hat{\rho}$ of Pantula, Gonzalez-Farias, and Fuller (1994) and Fuller (1996).

5 Inference for Spatial Persistence

This section develops methods to learn about the degree of spatial persistence.\textsuperscript{6} As discussed in the last section, it is possible to use regressions for that purpose. However, such an approach requires bandwidth and other choices, and it is difficult to reason about its efficiency. An alternative, non-regression based approach to learn about time series persistence is developed.

\textsuperscript{6}Previous approaches to testing for the presence of spatial correlation, such as Moran’s (1950) $I$ statistic or Geary’s (1954) $c$, require the specification of a spatial weight matrix and test the null hypothesis of zero spatial correlation.
in Müller and Watson (2008). This approach is based on the properties of \( q \) suitably chosen weighted averages, and it generalizes fairly directly to the spatial setting studied here.

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

5.1 Dimension Reduction by Weighted Averages

Let \( Y_n = (y_1, \ldots, y_n)' \) and let \( \Sigma_{n,L} \) be the \( n \times n \) covariance matrix induced by Lévy-Brownian motion \( y_t = L(s_t) \), conditional on \( \{s_t\}_{t=1}^{n} \). We are interested in tests that are invariant to translation shifts \( Y_n \rightarrow Y_n + a1 \), where \( 1 \) is a vector of ones. We therefore seek weighted averages of \( Y_n \) that sum to zero. Let \( M = I_n - 1(1')^{-1}1' \), and let \( R_n \) be the \( n \times q \) matrix of eigenvectors of \( M\Sigma_{n,L}M \) corresponding to the \( q \) largest eigenvalues, where \( R_n \) satisfies \( n^{-1}R_n'\Sigma_{n,L}R_n = I_q \). If \( Y_n \sim (0, \Sigma_{n,L}) \), the columns of \( R_n \) extract the \( q \) linear combinations of \( Y_n \) with the largest variance. Let \( Z_n = R_n'MY_n = R_n'Y_n \), a \( q \times 1 \) random vector, denote the associated weighted averages of the data, where the final equality holds because \( R_n'1 = 0 \). As in Müller and Watson (2008), we treat \( Z_n \) as the effective observation, that is, we seek to conduct inference about the persistence properties of \( Y_n \) with a test that is a function of \( Z_n \) only.

Different models for persistence in \( Y_n \) imply different values for \( \text{Var}(Z_n) = \Omega_n \). Consider first the generic problem of testing \( H_0 : \Omega = \Omega_0 \) versus \( H_a : \Omega = \Omega_a \) when \( Z_n \sim \mathcal{N}(0, \Omega) \). A standard calculation shows that the most powerful level \( \alpha \) scale invariant test rejects for large values of

\[
\frac{Z_n'(\Omega_0)^{-1}Z_n}{Z_n'(\Omega_a)^{-1}Z_n} \leq (19)
\]

with a critical value that equal to the \( 1 - \alpha \) quantile of (19) under the null distribution.
$Z_n \sim \mathcal{N}(0, \Omega_0)$.

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

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

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

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

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

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

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

(20)

Lemma 6 of Müller and Watson (2022a), building of the work of Rosasco, Belkin, and Vito (2010), shows that under the assumption that $s_i$ is i.i.d. with distribution $G$ and $\hat{\nu}_1 > \hat{\nu}_2 > \ldots > \hat{\nu}_q$, $(\hat{\nu}_i, \hat{\phi}_i)$ converge to $(\bar{\nu}_i, \phi_i)$, $i = 1, \ldots, q$, and the lemma also provides corresponding convergence rates. The following result does away with the i.i.d. assumption on the generation of the locations $s_i$, but rather assumes that the non-stochastic sequence of locations $\{s_l\}_{l=1}^{n}$ is
such that \( G_n \Rightarrow G \). This condition holds for almost all realizations of \( \{ s_t \}_{t=1}^n \) if \( s_t \sim G \) is i.i.d. by the Glivenko-Cantelli Theorem, so in this sense, the following result is stronger, albeit at the cost of not providing convergence rates.

**Lemma 7.** Suppose \( \bar{\nu}_1 > \bar{\nu}_2 > \ldots > \bar{\nu}_q > \bar{\nu}_{q+1} \) and \( G_n \Rightarrow G \). Then for any \( q \geq 1 \),
\[
\sup_{s \in S, 1 \leq i \leq q} |\tilde{\varphi}_i(s) - \bar{\varphi}(s)| \to 0 \quad \text{and} \quad \max_{1 \leq i \leq q} |\tilde{\nu}_i - \bar{\nu}_i| \to 0.
\]

### 5.2 Tests of the \( I(1) \) Null Hypothesis

With this background, consider the problem of testing the \( I(1) \) null hypothesis against the local-to-unity alternative. The canonical form of these models are \( y_t = L(s_t) \) and \( y_t = J(c)_t \). This yields \( Y_n \sim \mathcal{N}(0, \Sigma_{0,L}) \) and \( Y_n \sim \mathcal{N}(0, \Sigma_n(c)) \), respectively, with the \( l, \ell \) element of \( \Sigma_n(c) \) equal to \( \exp[-c(s_l - s_\ell)]/(2c) \). Thus, optimal tests in this problem are of the form (19) with \( \Omega_0 = \Omega_{n,L} = R_n^{\prime} \Sigma_{n,L} R_n \) and \( \Omega_a = \Omega_n(c_a) = R_n^{\prime} \Sigma_n(c_a) R_n \) for some \( c_a > 0 \). This yields the test statistic
\[
LFUR_n = \frac{Z_n^{\prime} \Omega_{n,L}^{-1} Z_n}{Z_n^{\prime} \Omega_n^{-1}(c_a) Z_n},
\]
where the notation emphasizes that this is the spatial analogue of the time series low-frequency unit root test (LFUR) from Müller and Watson (2008). In the Gaussian AR(1) time series model with parameter \( \rho \), the null is \( \rho = 1 \) and the alternative is \( \rho = 1 - c_a/n \). To determine a value of \( c_a \) that ensures good power for a wide range of values of \( c \), we follow King (1987) and choose \( c_a \) such that a 5% level test has 50% power.

By construction, this test is valid under the canonical \( H_0 \) model \( Y_n \sim \mathcal{N}(0, \Sigma_{L,n}) \). But by the FCLT in Theorem 2, Lemma 7 and the CMT, \( \lambda_n^{-1/2} n^{-1} Z_n \Rightarrow \mathcal{N}(0, \omega^2 \text{diag}(\bar{\nu}_1, \ldots, \bar{\nu}_q)) \) for the entire class of \( I(1) \) processes that satisfy the conditions of Theorem 2 (which includes the canonical model). Since the test (19) is scale invariant, the scale parameters \( \lambda_n^{-1/2} n^{-1} \) and \( \omega^2 \) cancel. Thus, the critical value computed from the canonical model converges to the asymptotically correct critical value for generic \( I(1) \) processes, and the test is asymptotically valid.

### 5.3 Tests of the \( I(0) \) Null Hypothesis

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

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

of no more than 0.03. Denote the corresponding cut-off value of \( c \) by \( c_{0.03} \). The canonical version of the testing problem then becomes \( H_0 : \Omega = \Omega_n(c), \ c \geq c_{0.03} \) against \( H_a : \Omega = \Omega_n(c) + g_a^2 \Omega_{n,L}, \ c \geq c_{0.03}, \ g_a \in \mathbb{R} \). This form of alternative, a sum of a stationary and \( I(1) \) process, also motivates the time series stationary tests in Nyblom (1989), Kwiatkowski, Phillips, Schmidt, and Shin (1992), etc. The larger the scale \( g_a \) of the Lévy-Brownian motion under the alternative, the easier it is to discriminate the two hypotheses, so \( g_a \) can again be chosen using the 50% power rule. The stationarity testing problem is complicated by the presence of the additional nuisance parameter \( c \) that indexes the covariance matrix \( \Sigma_n(c) \) in both the null and alternative. Here numerical experimentation revealed that in many configurations of locations, picking \( c = c_{0.001} \) under both \( H_0 \) and \( H_a \) works well in the sense of generating a test statistic (19) that has a 95% quantile that is fairly constant as a function of \( c \geq c_{0.03} \). Thus, the stationary test rejects if

\[
\text{LFST}_n = \frac{Z_n' \Omega_n(c_{0.001})^{-1} Z_n}{Z_n' \Omega_n(c_{0.001}) + g_a^2 \Omega_{n,L}} Z_n = \frac{\chi^2_{n}}{\text{inv} \chi^2_{n}}
\]

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

By the same arguments applied to the LFUR\(_n\) test, this stationarity test remains valid in large samples under the general local-to-unity model (9) for \( c \geq c_{0.03} \). A more subtle question is whether it also remains valid under generic weak dependence, defined as \( y_t = B(\lambda_n s_t) \), with \( \lambda_n \rightarrow \infty \) 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 \( s_t \sim G \) is i.i.d.), a suitably scaled version of \( Z_n \) becomes Gaussian, but not necessarily with covariance matrix proportional to \( I_q \). In the spatial case, the effect of weak dependence on the covariance of smoothly weighted averages is generically more subtle than a multiplication by the scalar
long-run variance. The LFST\textsubscript{n} test still remains valid, since for every \( n \), its critical value is chosen to be valid for all \( c \geq c_{0.03} \), so it is also valid under all sequences of \( c_{n} \to \infty \), including those that induce the different possible limits identified by Lahiri’s (2003) CLT.

**Theorem 8.** If \( y_{t} = B(\lambda_{n} s_{1}) \) and \( \lambda_{n} \to \infty \) with \( \lambda_{n}^{d} / n \to a \in [0, \infty) \), then under the assumptions of Lahiri’s CLT in Theorem 3.2, \( \limsup_{n \to \infty} P(\text{LFST}_{n} \geq \text{cv}_{n}^{\text{LFST}}) \leq \alpha \).

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

### 5.4 Confidence Sets for \( \bar{p}(c) \)

A closely related problem is the construction of a confidence set for \( c \), the parameter in the spatial local-to-unity model. As usual, one may obtain a \( 100(1 - \alpha)\% \) confidence set by collecting the values of \( c_{0} \) for which a family of \( \alpha \)-level tests of the form \( H_{0} : c = c_{0} \) does not reject. What is more, if this family of tests is optimal against the alternative that \( c \) is drawn from some probability distribution \( \Pi \), then the classic result in Pratt (1961) implies that the resulting confidence interval has the smallest \( \Pi \)-weighted expected length.

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

\[
\int_{0}^{1} \frac{\det(\Omega_{n}(c_{r}))^{-1/2}(Z_{n}' \Omega_{n}(c_{r})^{-1}Z_{n})^{-q/2} d\bar{p}}{(Z_{n}' \Omega_{n}(c_{\bar{p}_{0}})^{-1}Z_{n})^{-q/2}} \tag{24}
\]

do not exceed the \( 1 - \alpha \) quantile of (24) under \( Z_{n} \sim N(0, \Omega_{n}(c_{\bar{p}_{0}})) \). The large sample validity of this confidence set for \( \bar{p} > 0 \) in the general local-to-unity model (9) follows from the same arguments as the large sample validity of the LFUR\textsubscript{n} test discussed above.
5.5 Residual Based Tests

Now suppose we want to conduct inference about the persistence properties of the disturbance \( u_t \) in a linear regression \( y_t = x_t' \beta + u_t \) (where here a potential constant is part of \( x_t \)). The above results are then not directly applicable, since with \( \beta \) unknown, \( u_t \) is unobserved.

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

With \( \mathbf{u}_n \) independent of \( \mathbf{X}_n \), we can simply condition on the realization of \( \mathbf{X}_n \), and apply the above tests with \( \mathbf{Z}_n^X \) in place of \( \mathbf{Z}_n \).

In order to invoke the asymptotic arguments above in such an approach, one needs to generalize the eigenfunction convergence of Lemma 7; see Lemma 11 in the appendix.

More substantively, the assumption of \( \mathbf{u}_n \) to be independent of the entire set of regressors \( \mathbf{X}_n \) is quite strong. Without that assumption, there is statistical dependence between the eigenvectors \( \mathbf{R}_n^X \) and \( \mathbf{u}_n \), invalidating an analysis that conditions on \( \mathbf{R}_n^X \). It turns out that for large sample validity, it suffices to assume asymptotic independence between \( \mathbf{X}_n \) and \( \mathbf{u}_n \) in the sense that for some appropriate \( p \times p \) scaling matrices \( \Lambda_n \) and scalar sequence \( \lambda_n \), we have the convergence

\[
\begin{bmatrix}
\Lambda_n^{-1/2} \mathbf{X}_n(\cdot) \\
\lambda_n^{-1/2} \mathbf{U}_n(\cdot)
\end{bmatrix} \Rightarrow \begin{bmatrix}
\mathbf{X}(\cdot) \\
\omega \mathbf{U}(\cdot)
\end{bmatrix}
\]  

(25)

on \( \mathcal{S} \), with \( \omega > 0 \) and \( \mathbf{X} \) independent of \( \mathbf{U} \). See Theorem 12 in the appendix.

This result has a particularly noteworthy implication for the test of the null hypothesis of no cointegration among the \( p + 1 \) variables \( (x_l, y_l) \), that is, for the spatial analogue of Engle and Granger’s (1987) residual based test of cointegration. Here the assumption is that under the null hypothesis, \( (x_l', y_l') = (\mathbf{X}_n(s_l'), \mathbf{Y}_n(s_l')) \) with \( \lambda_n^{-1/2}(\mathbf{X}_n(\cdot)', \mathbf{Y}_n(\cdot)) \Rightarrow \Phi L_{X,Y}(\cdot) \), where \( L_{X,Y} \) is a vector of \( p + 1 \) independent Lévy-Brownian Motions, and \( \Phi \) is an arbitrary full rank \( (p + 1) \times (p + 1) \) matrix. Noting that \( O \mathbf{L}_{X,Y} \sim L_{X,Y} \) for any \( (p + 1) \times (p + 1) \) rotation matrix \( \mathbf{O} \), it is without loss of generality to assume that \( \Phi \) is lower triangular. Letting \( \beta \) be equal to the first \( p \) elements in the last row of \( \Phi \) then yields that \( \mathbf{U}_n(\cdot) = \mathbf{Y}_n(\cdot) - \mathbf{X}_n(\cdot)' \beta \) satisfies (25) with \( \mathbf{U} \) a scalar Lévy-Brownian motion independent of the \( p \) dimensional Lévy-Brownian motion \( \mathbf{X} \).

To implement such a level \( \alpha \) test of the null hypothesis of no spatial cointegration in
practice, one computes the LFUR\(_n\) statistic (21) with \(Z_n^X\) in place of \(Z_n\), \(\Omega_0 = R_n^X\Sigma_{i,n} R_n^X\) and \(\Omega_1 = R_n^X\Sigma_n(c_a) R_n^X\), and compares it to \(1 - \alpha\) quantile of the statistic under \(Y_n \sim \mathcal{N}(0, \Sigma_{n,L})\).

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

Chetty, Hendren, Kline, and Saez (2014) use administrative records on the incomes of more than 40 million children and parents to study intergenerational income mobility in the United States. They construct an index of mobility for each of the 741 commuter zones in the United States and investigate the relationship between mobility and other factors by regressing their mobility index on variables such as racial segregation, school quality and so forth. They find large and statistically significant correlations between their absolute mobility index (AMI) and many socioeconomic indicators. One might suspect that the variables used in their regressions are spatially correlated, and this raises the question of the robustness of their results to sampling error associated with this spatial persistence. This question is answered in Table 1.\(^7\)

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

### 6 Regressions with Transformed Spatial Variables

To avoid spurious regression effects using \(I(1)\) time series data, researchers routinely estimate regressions using first differences of the original variables and rely on HAC/HAR inference methods to account for any remaining \(I(0)\) autocorrelation. The best approach for regressions involving spatial \(I(1)\) variables is not so obvious, and we explore a number of possibilities.\(^7\)

\(^7\)The variables are chosen from Figure VIII in Chetty, Hendren, Kline, and Saez (2014). The data are taken from their comprehensive replication materials.
Table 1: Spatial Persistence of Variables in Chetty et al. (2014)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Spatial Persistence Statistics</th>
<th>Regression of the AMI onto Variable</th>
<th>( \hat{\beta} ) [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p )-Value for Test</td>
<td>95% CI for ( \hat{\rho} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( I(1) ) Null</td>
<td>( I(0) ) Null</td>
<td>[0.14, 1.00]</td>
</tr>
<tr>
<td>Absolute Mobility Index</td>
<td>0.08</td>
<td>&lt;0.01</td>
<td>[0.61, 0.85]</td>
</tr>
<tr>
<td>Frac. Black Residents</td>
<td>0.02</td>
<td>0.01</td>
<td>[0.62, 0.71]</td>
</tr>
<tr>
<td>Racial Segregation</td>
<td>0.07</td>
<td>0.02</td>
<td>[0.65, 1.00]</td>
</tr>
<tr>
<td>Segregation of Poverty</td>
<td>0.13</td>
<td>0.04</td>
<td>[0.66, 1.00]</td>
</tr>
<tr>
<td>Frac &lt; 15 Mins to Work</td>
<td>0.69</td>
<td>&lt;0.01</td>
<td>[0.41, 0.61]</td>
</tr>
<tr>
<td>Mean Household Income</td>
<td>0.02</td>
<td>0.18</td>
<td>[0.60, 1.00]</td>
</tr>
<tr>
<td>Gini</td>
<td>0.56</td>
<td>&lt;0.01</td>
<td>[0.40, 1.00]</td>
</tr>
<tr>
<td>Top 1 Perc. Inc. Share</td>
<td>0.60</td>
<td>0.03</td>
<td>[0.63, 1.00]</td>
</tr>
<tr>
<td>Student-Teacher Ratio</td>
<td>0.03</td>
<td>0.16</td>
<td>[0.04, 0.87]</td>
</tr>
<tr>
<td>Test Scores (Inc. adjusted)</td>
<td>0.40</td>
<td>0.07</td>
<td>[0.27, 1.00]</td>
</tr>
<tr>
<td>High School Dropout</td>
<td>0.63</td>
<td>0.02</td>
<td>[0.40, 1.00]</td>
</tr>
<tr>
<td>Social Capital Index</td>
<td>0.73</td>
<td>&lt;0.01</td>
<td>[0.58, 1.00]</td>
</tr>
<tr>
<td>Frac Religious</td>
<td>0.11</td>
<td>0.03</td>
<td>[0.15, 1.00]</td>
</tr>
<tr>
<td>Violent Crime Rate</td>
<td>0.52</td>
<td>0.04</td>
<td>[0.58, 1.00]</td>
</tr>
<tr>
<td>Frac. Single Mothers</td>
<td>0.03</td>
<td>&lt;0.01</td>
<td>[0.65, 0.88]</td>
</tr>
<tr>
<td>Divorce Rate</td>
<td>&lt;0.01</td>
<td>0.21</td>
<td>[0.62, 0.53]</td>
</tr>
<tr>
<td>Frac. Married</td>
<td>0.09</td>
<td>0.07</td>
<td>[0.12, 1.00]</td>
</tr>
<tr>
<td>Local Tax Rate</td>
<td>0.01</td>
<td>0.25</td>
<td>[0.01, 0.59]</td>
</tr>
<tr>
<td>Colleges per Capita</td>
<td>0.57</td>
<td>0.10</td>
<td>[0.00, 1.00]</td>
</tr>
<tr>
<td>College Tuition</td>
<td>0.21</td>
<td>&lt;0.01</td>
<td>[0.15, 1.00]</td>
</tr>
<tr>
<td>Coll. Grad. Rate (Inc. Adjusted)</td>
<td>0.46</td>
<td>0.01</td>
<td>[0.34, 1.00]</td>
</tr>
<tr>
<td>Manufacturing Share</td>
<td>0.04</td>
<td>&lt;0.01</td>
<td>[0.10, 1.00]</td>
</tr>
<tr>
<td>Chinese Import Growth</td>
<td>0.02</td>
<td>0.07</td>
<td>[0.01, 0.58]</td>
</tr>
<tr>
<td>Teenage LFP Rate</td>
<td>0.28</td>
<td>&lt;0.01</td>
<td>[0.20, 1.00]</td>
</tr>
<tr>
<td>Migration Inflow</td>
<td>0.06</td>
<td>0.11</td>
<td>[0.00, 1.00]</td>
</tr>
<tr>
<td>Migration Outflow</td>
<td>0.05</td>
<td>0.02</td>
<td>[0.07, 1.00]</td>
</tr>
<tr>
<td>Frac. Foreign Born</td>
<td>0.44</td>
<td>0.02</td>
<td>[0.35, 1.00]</td>
</tr>
</tbody>
</table>

Notes: The first two columns show p-values for tests of the \( I(1) \) and \( I(0) \) null hypotheses using the statistics (21) and (23). The third column shows the 95% confidence for \( \hat{\rho} \) constructed by inverting the tests in (24). The final two columns show estimated regression coefficients (\( \hat{\beta} \)) and nominal 95% confidence intervals from regression of the Absolute Mobility Index (AMI) onto each of the variables in the table, where the first column of results uses the levels of the variables with nominal 95% confidence intervals constructed using standard errors clustered by state, and the second column shows the LBM-GLS estimates and CSPC 95% confidence intervals. The variables are standardized (in levels) to have mean zero and unit standard deviation.
Using simulated data, we assess whether regressions with such transformed data and spatial HAR corrections yields (approximately) valid inference.

6.1 Simulation Design

We are interested in inference about $\beta_1$, the first element of $\beta$, in the linear regression (10), maintaining throughout that $Y_n$ is independent of $X_n = (x_1, \ldots, x_n)'$. The simulated data sets have $n = 400$ observations and differ both in their distribution of locations $\{s_l\}$ and the distribution of $(Y_n, X_n)$. Spatial locations are drawn from the 48-U.S. States design used in Müller and Watson (2022a, 2022b). Specifically, for each of the 48 contiguous U.S. states, we draw 2 sets of 400 locations at random uniformly within the boundaries of the state. Conditional on each of these 96 location set draws, we consider seven distributions for $(Y_n, X_n)$, for $p = 1$ and $p = 5$. In each of those, the $p + 1$ columns of $(Y_n, X_n)$ are independent and identically distributed. The seven distributions for $y_l$ (which are also used to generate each element of $x_l$) are:

- **DGP1**: $y_l = L(s_l)$, Lévy-Brownian motion;
- **DGP2**: $y_l \sim I(1)$ as in (6) with $B = J_c$ and $c = c_{0.01}$, so the average pairwise correlation of $\{B(s_l)\}_{l=1}^n$ is 0.01;
- **DGP3**: $y_l \sim I(1)$ with $B = J_c$ and $c = c_{0.03}$;
- **DGP4**: $y_l \sim I(1)$ with $B$ a Gaussian process with Matérn covariance function equal to $\mathbb{E}[B(s)B(r)] = (1 + c\Delta + (c\Delta)^2/3)\exp(-c\Delta)$ for $\Delta = |s - r|$ and $c$ such that the average pairwise correlation of $\{B(s_l)\}_{l=1}^n$ is equal to $\bar{\rho} = 0.03$,
- **DGP5**: $y_l = J_c(s_l)$ with $c = c_{0.03}$;
- **DGP6**: $y_l = J_c(s_l)$ with $c = c_{0.50}$;
- **DGP7**: $y_l = \int_{\mathbb{R}^2} 1[0 \leq r \leq s_l]dW(r)$, so dependence is induced by a Brownian sheet.

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

6.2 Data Transformations

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

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

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

*Isotropic Differences:* This is the transformation (15) which we apply for bandwidths $b = 0.03, 0.06, \ldots, 0.15$, where the locations $\{s_l\}$ are scale normalized so that $\max_{l,t} |s_l - s_t| = 1$.

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

$$ y^*_l = y_l - \frac{\sum_{i=1}^m 1[s_l \in R_i] \sum_{\ell \neq l} 1[s_\ell \in R_i] y_\ell}{\sum_{i=1}^m 1[s_l \in R_i] \sum_{\ell \neq l} 1[s_\ell \in R_i]}.$$

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

*LBM-GLS:* First differences in a regular time series are a GLS transformation under the canonical random walk model of $I(1)$ persistence. Recall from the last section that $\Sigma_{n,L}$ is the $n \times n$ covariance matrix of $Y_n$ induced by a Lévy-Brownian motion, the canonical $I(1)$ model of spatial persistence. With $Y_n \sim \mathcal{N}(\mu 1, \Sigma_{n,L})$,

$$ Y^*_n = (M \Sigma_{n,L} M)^{-1/2} Y_n \sim \mathcal{N}(0, M) $$

(26)

where $(M \Sigma_{n,L} M)^{-1/2}$ is the Moore-Penrose generalized inverse of $(M \Sigma_{n,L} M)^{1/2}$. This GLS transformation converts $Y_n$ into a set of demeaned i.i.d. random variables $Y^*_n$. In a more general $I(1)$ model, this is no longer true, but given the FCLT in Theorem 2, it is plausible that this LBM-GLS transformation induces enough stationarity for spatial HAR inference to be reliable. Figure 6.2 illustrates the GLS transformations for the data in Figure 1.
Figure 5: Transformed Strongly Dependent Data

(a) First Difference of Independent Time Series Random Walks

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

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

Mobility Index

Teen Labor Force Participation
Remark 6.1. The LBM-GLS estimator of $\beta$ is closely related to the OLS estimator obtained after controlling for a smooth spline function, say $\eta(s)$, in the regression. To see this, write the stacked regression as $Y_n = X_n \beta + \eta + e$, where $Y_n$ is $n \times 1$, $X_n$ is $n \times p$, $\beta$ is $p \times 1$, $\eta$ is $n \times 1$ with $\eta_l = \eta(s_l)$, and $e$ is a vector of errors. Estimation of $\beta$ subject to a smoothness constraint on $\eta(\cdot)$ can be accomplished by solving the penalized least squares problem $\min_{\beta, \eta} \left( (Y_n - X_n \beta - \eta)^T (Y_n - X_n \beta - \eta) + \lambda^{-1} \eta^T \Omega^{-1} \eta \right)$ for an appropriately chosen matrix $\Omega$ that captures the smoothness properties of $\eta(\cdot)$. The solution to the problem yields the estimator $\hat{\beta} = [X_n^T (I + \lambda^{-1} \Omega)^{-1} X_n]^{-1} [X_n^T (I + \lambda^{-1} \Omega)^{-1} Y_n]$, which is recognized as the GLS estimator using the error covariance matrix $I + \lambda^{-1} \Omega$. The location-invariant GLS estimator uses $(MY_n, MX_n)$ in place of $(Y_n, X_n)$. The use of $\Sigma_{n,L}$ for $\Omega$ imposes a Lévy-Brownian motion smoothness prior of $\eta$; this is a spatial generalization of the Wiener process smoothing prior that yields a quadratic smoothing spline when $d = 1$. For our designs, we found that inference using this estimator and spatial HAR standard errors performed best using $\lambda = 0$, which coincides with the LBM-GLS method.

Low-pass Eigenvector Transformation: Recall that the $q \times 1$ vector $Z_n$ in the previous section was defined as $Z_n = R_n^T Y_n$, where $R_n$ collects the eigenvectors of $M \Sigma_{n,L} M$ corresponding to the $q < n$ largest eigenvalues $\hat{\nu}_n$. By construction, if $Y_n \sim N(\mu_1, \Sigma_{n,L})$, then

$$Y_n^* = \text{diag}(\hat{\nu}_n)^{-1/2} Z_n \sim N(0, n^{-1} I_q). \quad (27)$$

Here $Y_n^*$ is $q \times 1$, rather than $n \times 1$, as in the previous transformations. Note that setting $q = n - 1$ amounts to the LBM-GLS transformation (26). The potential advantage of using a smaller, fixed $q$ is that the CMT and the FCLT imply that the (suitably scaled) $Y_n^*$ in (27) has an asymptotic $N(0, I_q)$ distribution in the general $I(1)$ model. Thus, classical Gaussian small sample inference in the regression of $Y_n^*$ on $X_n^*$ is asymptotically justified with this transformation in the general $I(1)$ model. This approach is analogous to what is suggested by Müller and Watson (2017) for persistent time series. We implement this with $q = 10, 20, 50$.

High-pass Eigenvector Transformation: An alternative to using the first $q$ principal components in $Z_n$ is to use the remaining $n - 1 - q$ principal components, say $Y_n^* = \hat{R}_n^T Y_n$, where

Kelly (2022) proposes a spline-augmented OLS estimator for inference in spatial regressions and presents simulation evidence showing that it improves size control compared to using spatial HAC standard errors and OLS with untransformed regressors. This remark reiterates well-known results on the relationship between smoothing splines, Bayes smoothness priors, and maximum likelihood estimators (cf. Engle and Watson (1988)).
$\tilde{R}_n$ collects the eigenvectors of $M\Sigma_{n,I}M$ corresponding to the $n-1-q$ smallest eigenvalues. The rationale for this approach is that eliminating the first $q$ principal components purges the data of the large-variance components associated with spatial $I(1)$ persistence. Alternatively, in the context of Remark 6.1, the resulting regression controls for smooth spatial function spanned by the columns of $R_n$. We implement this with $q = 5, 10, 20, 50$.

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

These estimators of $\beta_1$ are used in conjunction with three types of standard errors: We present results using heteroskedasticity robust standard errors for the LBM-GLS and cluster fixed effects estimators. For the latter estimator, we also consider clustered standard errors. Finally, for all but the last two methods, we use spatial HAR standard errors and critical values suggested by Müller and Watson (2022b). This so-called C-SCPC-method is calibrated to control size under spatial dependence with an average correlation of no more than 0.03 (and, by “conditioning” on the regressor, it is by construction more conservative than the method developed in Müller and Watson (2022a)).

### 6.3 Simulation Results

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

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

Table 3 summarizes the expected length of the resulting (non-size corrected) 95% confi-
Table 2: Rejection Frequency for Nominal 5% Tests (median over 96 spatial designs)

<table>
<thead>
<tr>
<th>Method</th>
<th>Lévy-BM</th>
<th>(I(1)_{\max})</th>
<th>(I(1)_{\max})</th>
<th>(I(1)_{\text{Record}})</th>
<th>(J_{s,\alpha})</th>
<th>(J_{s,\alpha})</th>
<th>Br. Sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS (C-SCPC)</td>
<td>0.29</td>
<td>0.32</td>
<td>0.35</td>
<td>0.33</td>
<td>0.04</td>
<td>0.20</td>
<td>0.33</td>
</tr>
<tr>
<td>Isotropic difference (C-SCPC)</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Cluster fixed-effects (cluster)</td>
<td>0.08</td>
<td>0.24</td>
<td>0.35</td>
<td>0.30</td>
<td>0.07</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Cluster fixed-effects (C-SCPC)</td>
<td>0.05</td>
<td>0.08</td>
<td>0.12</td>
<td>0.10</td>
<td>0.04</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>LBM-GLS</td>
<td>0.05</td>
<td>0.26</td>
<td>0.59</td>
<td>0.38</td>
<td>0.06</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>LBM-GLS (C-SCPC)</td>
<td>0.03</td>
<td>0.05</td>
<td>0.07</td>
<td>0.06</td>
<td>0.03</td>
<td>0.03</td>
<td>0.09</td>
</tr>
<tr>
<td>Low-pass Eigenvector</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.08</td>
<td>0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>High-pass Eigenvector (C-SCPC)</td>
<td>0.05</td>
<td>0.10</td>
<td>0.13</td>
<td>0.14</td>
<td>0.05</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>Irbragimov-Müller</td>
<td>0.08</td>
<td>0.13</td>
<td>0.15</td>
<td>0.13</td>
<td>0.05</td>
<td>0.07</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Addendum: Avg. \(R^2\)

<table>
<thead>
<tr>
<th></th>
<th>(b = 1)</th>
<th>(b = 5)</th>
<th>(b = 10)</th>
<th>(b = 25)</th>
<th>(b = 50)</th>
<th>(b = 100)</th>
<th>(b = 250)</th>
<th>(b = 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS (C-SCPC)</td>
<td>0.29</td>
<td>0.32</td>
<td>0.36</td>
<td>0.33</td>
<td>0.05</td>
<td>0.22</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Isotropic difference (C-SCPC)</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.06</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Cluster fixed-effects (cluster)</td>
<td>0.08</td>
<td>0.24</td>
<td>0.35</td>
<td>0.31</td>
<td>0.07</td>
<td>0.08</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Cluster fixed-effects (C-SCPC)</td>
<td>0.05</td>
<td>0.08</td>
<td>0.11</td>
<td>0.09</td>
<td>0.05</td>
<td>0.05</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>LBM-GLS</td>
<td>0.05</td>
<td>0.26</td>
<td>0.59</td>
<td>0.38</td>
<td>0.06</td>
<td>0.05</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>LBM-GLS (C-SCPC)</td>
<td>0.03</td>
<td>0.05</td>
<td>0.07</td>
<td>0.06</td>
<td>0.03</td>
<td>0.03</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Low-pass Eigenvector</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.08</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>High-pass Eigenvector (C-SCPC)</td>
<td>0.06</td>
<td>0.11</td>
<td>0.14</td>
<td>0.15</td>
<td>0.05</td>
<td>0.06</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Irbragimov-Müller</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Addendum: Avg. \(R^2\)

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

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Lévy-BM</th>
<th>(I(1)_{\max})</th>
<th>(I(1)_{\max})</th>
<th>(I(1)_{\text{Record}})</th>
<th>(J_{s,\alpha})</th>
<th>(J_{s,\alpha})</th>
<th>Br. Sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isotropic difference (C-SCPC)</td>
<td>0.53</td>
<td>0.70</td>
<td>0.73</td>
<td>0.73</td>
<td>0.44</td>
<td>0.52</td>
<td>0.54</td>
</tr>
<tr>
<td>Cluster-FE (CSCPC)</td>
<td>0.55</td>
<td>0.81</td>
<td>0.91</td>
<td>0.88</td>
<td>0.43</td>
<td>0.39</td>
<td>0.50</td>
</tr>
<tr>
<td>LBM-GLS (C-SCPC)</td>
<td>0.25</td>
<td>0.42</td>
<td>0.54</td>
<td>0.55</td>
<td>0.26</td>
<td>0.26</td>
<td>0.33</td>
</tr>
<tr>
<td>Low-pass Eigenvector</td>
<td>1.51</td>
<td>1.51</td>
<td>1.51</td>
<td>1.51</td>
<td>0.57</td>
<td>0.57</td>
<td>1.51</td>
</tr>
<tr>
<td>Irbragimov-Müller</td>
<td>0.42</td>
<td>0.87</td>
<td>1.00</td>
<td>0.96</td>
<td>0.37</td>
<td>0.41</td>
<td>0.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(b = 1)</th>
<th>(b = 5)</th>
<th>(b = 10)</th>
<th>(b = 25)</th>
<th>(b = 50)</th>
<th>(b = 100)</th>
<th>(b = 250)</th>
<th>(b = 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isotropic difference (C-SCPC)</td>
<td>0.51</td>
<td>0.69</td>
<td>0.77</td>
<td>0.78</td>
<td>0.43</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>Cluster fixed-effects (C-SCPC)</td>
<td>0.35</td>
<td>0.42</td>
<td>0.48</td>
<td>0.47</td>
<td>0.34</td>
<td>0.35</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>LBM-GLS (C-SCPC)</td>
<td>0.26</td>
<td>0.42</td>
<td>0.54</td>
<td>0.55</td>
<td>0.27</td>
<td>0.26</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Low-pass Eigenvector</td>
<td>2.30</td>
<td>2.30</td>
<td>2.30</td>
<td>2.30</td>
<td>0.60</td>
<td>0.60</td>
<td>2.30</td>
<td>2.30</td>
</tr>
<tr>
<td>Irbragimov-Müller</td>
<td>0.46</td>
<td>0.65</td>
<td>0.67</td>
<td>0.78</td>
<td>0.36</td>
<td>0.48</td>
<td>0.49</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Notes: Entries are the median average length across the 96 spatial designs described in the text. See Table 2 for additional comment notes.
dence intervals. It leaves out the levels-OLS estimator, clustered-standard error fixed effects and high-pass eigenvector transform methods because of their significant size distortions.

There are two key takeaways from the tables. First, isotropic differences and LBM-GLS implemented with HAR standard errors have reasonably good size properties in all designs, as does the eigenvalue transformation. Second, LBM-GLS (with HAR standard errors) produces confidence intervals with the smallest average length. These results, along with the observation that LBM-GLS does not require the choice of a bandwidth or other parameter, suggests that it dominates the other methods considered here.

Remark 6.2. The final rows in Table 2 show the average value of the $R^2$ in the levels-regression (10). (This is the median value across the 96 spatial designs.) These $R^2$ values are large for the $I(1)$ models, consistent with the implications of Theorem 3, and for the local-to-unity model with $c = c_{0.50}$, consistent with the discussion in Remark 3.3.

6.4 Regressions in Chetty et al. (2014)

We now return to the results in Table 1. As noted in Section 5.6 the first three columns of the table suggest substantial spatial correlation in many of the variables. The final two columns summarize results from the regression of the Absolute Mobility Index (the first variable in the table) onto each of the other variables. These regressions were reported in Figure VII of Chetty, Hendren, Kline, and Saez (2014). The penultimate column in the Table 1 (labeled “Level(Cluster)” ) reports the levels-OLS estimate of the regression coefficient with a nominal 95% confidence interval computed using standard errors clustered at the state level. These results are reported in Chetty, Hendren, Kline, and Saez (2014). The final column in the table shows results using LBM-GLS with a C-SCPC 95% confidence interval.

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

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

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

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

Proof of Lemma 1: By the Corollary on page 48 of Adler (2010), the result holds if for some $m > 2d$, \( \mathbb{E} \left[ (Y_0(s) - Y_0(r))^{2m} \right] \leq C|s-r|^m \) for some $C$. Let $m > 2d$ and apply Condition 1 to obtain

\[
\mathbb{E}[(Y_0(s) - Y_0(r))^{2m}] \leq C_m \left( \int_{\mathbb{R}^d} (h(u, s) - h(u, r))^2 du \right)^m
\]

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

The latter follows by Theorem 23.7 of Kallenberg (2021) from (28) and

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

from a first-order Taylor expansion of the function $x \mapsto \exp(-cx)$. \(\square\)

Proof of Theorem 2: Consider first the claim for the convergence for the LTU process (9). From

\[
\int h_c(r, 0)^2 dr = (2c)^{-1} = \lambda^d \int h_c(\lambda r, 0)^2 dr = \lambda^{(1+d)/2} \frac{\kappa_{c,d}^2}{\kappa_{\lambda c,d}^2} \int h_{\lambda c}(r, 0)^2 dr = \lambda^{(1+d)/2} \frac{\kappa_{c,d}^2}{\kappa_{\lambda c,d}^2} (2c\lambda)^{-1}
\]

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

\[
Y_n(s) = \lambda^{(1-d)/2} \int_{\mathbb{R}^d} h_c(\lambda_n^{-1} r, s) B(r) dr.
\]

We show convergence of finite dimensional distributions and tightness of the process $\lambda_n^{-1/2} Y_n$. The latter follows by Theorem 23.7 of Kallenberg (2021) from (28) and $J_c(0) \sim \mathcal{N}(0, (2c)^{-1})$. For the former, consider for $t_1, \ldots, t_k \in S$, the $k \times 1$ vector $\lambda_n^{-1/2}(Y_n(t_1), \ldots, Y_n(t_k))$. By the Cramér-Wold device, it suffices to establish the convergence $X_n = \lambda_n^{-1/2} \sum_{j=1}^k v_j Y_n(t_j) \Rightarrow \sum_{j=1}^k v_j \omega c J_c(t_j)$ for $(v_1, \ldots, v_k) \in \mathbb{R}^k$. Let $f_v(r) = \sum_{j=1}^k v_j h_c(r, t_j)$, so that from (8), \( \sum_{j=1}^k v_j J_c(t_j) \sim \mathcal{N}(0, \int_{\mathbb{R}^d} f_v(r)^2 du) \)

and from (29)

\[
X_n = \lambda_n^{-d/2} \int_{\mathbb{R}^d} f_v(\lambda_n^{-1} r) B(r) du.
\]
For $\varepsilon > 0$, define $f^\varepsilon_v(r) = f_v(r)1[|r| < 1/\varepsilon] \prod_{j=1}^k 1[|t_j - r| > \varepsilon]$ and let

$$X^\varepsilon_n = \lambda_n^{-d/2} \int_{\mathbb{R}^d} f^\varepsilon_v(\lambda_n^{-1}r)B(r)dr.$$ 

From Condition 1 we find

$$\mathbb{E}[(X^\varepsilon_n - X_n)^2] = \lambda_n^{-d} \mathbb{E} \left[ \left( \int_{\mathbb{R}^d} (f_v(\lambda_n^{-1}r) - f^\varepsilon_v(\lambda_n^{-1}r))B(r)dr \right)^2 \right]$$

$$\leq C_2 \lambda_n^{-d} \int_{\mathbb{R}^d} (f_v(\lambda_n^{-1}r) - f^\varepsilon_v(\lambda_n^{-1}r))^2 dr$$

$$= C_2 \int_{\mathbb{R}^d} (f_v(r) - f^\varepsilon_v(r))^2 dr.$$ 

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

Furthermore

$$\mathbb{E}[(X^\varepsilon_n)^2] = \lambda_n^{-d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f^\varepsilon_v(\lambda_n^{-1}r)f^\varepsilon_v(\lambda_n^{-1}s)\sigma_B(s-r)dr ds$$

$$= \int_{\mathbb{R}^d} \sigma_B(s) \int_{\mathbb{R}^d} f^\varepsilon_v(r)f^\varepsilon_v(r + \lambda_n^{-1}s)dr ds$$

$$\to \int_{\mathbb{R}^d} \sigma_B(s) ds \int_{\mathbb{R}^d} f^\varepsilon_v(r)^2 dr$$

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

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

$$X^\varepsilon_n \Rightarrow \mathcal{N} \left( 0, \int_{\mathbb{R}^d} \sigma_B(s) ds \int_{\mathbb{R}^d} f^\varepsilon_v(r)^2 dr \right).$$

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

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

Consider the convergence $\lambda_n^{-1} n^{-1} \sum_{l=1}^n x_t y_l \Rightarrow \int X(s)Y(s)dG(s)$. By the Skorohod almost sure representation theorem (see, for instance, Theorem 11.7.2 of Dudley (2002)), there exist random
elements $(Y_n^*(\cdot), X_n^*(\cdot))$ such that $\sup_{s \in \mathcal{S}} |(Y_n^*(s) - Y^*(s), X_n^*(s) - X^*(s))| \xrightarrow{a.s.} 0$, $(Y^*(\cdot), X^*(\cdot)) \sim (Y(\cdot), X(\cdot))$ and $\lambda_n^{-1/2}(Y_n^*(\cdot), X_n^*(\cdot)) \sim (Y_n^*(\cdot), X_n^*(\cdot))$ for $n = 1, 2, \ldots$. Thus it suffices to show the claim for $\int X_n^*(s)Y_n^*(s)dG_n(s) = n^{-1} \sum_{i=1}^n X_n^*(s_i)Y_n^*(s_i) \sim \lambda_n^{-1} n^{-1} \sum_{i=1}^n X_n(s_i)Y_n(s_i)$. We have
\[
\left| \int (X_n^*(s)Y_n^*(s) - X^*(s)Y^*(s))dG_n(s) \right| \leq \sup_{s \in \mathcal{S}} |(Y_n^*(s) - Y^*(s), X_n^*(s) - X^*(s))| \xrightarrow{a.s.} 0
\]
so it suffices to show the claim for $\int X^*(s)Y^*(s)dG_n(s)$. Now almost all realizations of the $\mathbb{R}^k \mapsto \mathbb{R}$ function $s \mapsto X^*(s)Y^*(s)$ on $\mathcal{S}$ 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 $\int X^*(s)Y^*(s)dG_n(s) \xrightarrow{a.s.} \int X^*(s)Y^*(s)dG(s)$. But almost sure convergence implies convergence in distribution, so the desired result follows. The other terms are dealt with in the same manner. □

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

**Lemma 9.** For any $\delta > 0$, $\lim_{n \to \infty} \sup_{r \in \mathcal{S}} G_n\{s : |s - r| \leq \delta\} \leq \sup_{r \in \mathcal{S}} G\{r : |s - r| \leq \delta\}$, where $G_n(A)$ and $G(A)$ is the measure that is assigned to the Borel set $A \subset \mathbb{R}^d$ by the distributions $G_n$ and $G$, respectively.

**Proof.** Let $a = \sup_{r \in \mathcal{S}} G\{r : |s - r| \leq \delta\}$. Suppose otherwise. Then there exists $\varepsilon > 0$ and a sequence $r_n$ such that
\[
\limsup_{n \to \infty} \sup_{r \in \mathcal{S}} G_n\{s : |s - r| \leq \delta\} = \lim_{n \to \infty} G_n\{s : |s - r_n| \leq \delta\} \geq a + \varepsilon.
\]
Let $\delta' > \delta$ be such that $\sup_{r \in \mathcal{S}} G\{r : |s - r| \leq \delta'\} \leq a + \varepsilon/2$. Since $\mathcal{S}$ is compact, $r_n \to r_0$ along some subsequence. Along that subsequence, for all $n$ large enough so that $|r_n - r_0| < \delta' - \delta$, we have
\[
G_n\{s : |s - r_n| \leq \delta\} \leq G_n\{s : |s - r_0| \leq \delta'\} \to G\{s : |s - r_0| \leq \delta'\} \leq a + \varepsilon/2
\]
yielding the desired contradiction. □

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

Now since $\lambda_n^{-1}S_{\hat{\beta}x} \
onumber \Rightarrow \Xi_{\hat{\beta}x}$ and $\Xi_{\hat{\beta}x}$ is full rank with probability one, it suffices to show that $n^{-1}\lambda_n^{-2}||\Omega_n|| \xrightarrow{P} 0$.

By the FCLT and CMT, $\lambda_n^{-1}e_n(\cdot) \Rightarrow e_0(\cdot) = (\hat{Y}(\cdot) - \bar{X}(\cdot)\Xi_{\hat{\beta}x}^{-1}\Xi_{\hat{\beta}y})\bar{X}(\cdot)$, so that $\lambda_n^{-1}\sup_l |e_n(s_l)| \Rightarrow \sup_{s \in \mathcal{S}} |e_0(s)|$, and $\lambda_n^{-1}\sup_l |e_n(s_l)| = O_p(1)$. We have
\[
\lambda_n^{-2}n^{-2} \left| \sum_{l,\ell=1}^n \kappa(b_n(s_l - s_\ell))e_n(s_l)e_n(s_\ell)' \right| \leq \lambda_n^{-2} \sup_l |e_n(s_l)|^2 \cdot n^{-2} \sum_{l,\ell=1}^n |\kappa(b_n(s_l - s_\ell))|
\]

36
and
\[ \sum_{l, \ell=1}^{n} |\kappa(b_n(s_l - s_\ell))| \leq \kappa \sum_{l, \ell=1}^{n} 1[|s_l - s_\ell| \leq b_n^{-1/2}] + \sum_{l, \ell=1}^{n} 1[|s_l - s_\ell| > b_n^{-1/2}]|\kappa(b_n(s_l - s_\ell))|.
\]

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

by (14) and \( b_n^{-1} = o_p(1) \). Furthermore, since \( G \) is continuous, for every \( \varepsilon > 0 \), there exists a \( \delta > 0 \) such that \( \sup_{r \in S} G(\{s : |s - r| \leq \delta\}) \leq \varepsilon \). Note that
\[ n^{-2} \sum_{l, \ell=1}^{n} 1[|s_l - s_\ell| \leq b_n^{-1/2}] \leq \sup_{r \in S} G_n(\{s : |s - r| \leq b_n^{-1/2}\}) \leq \sup_{r \in S} G_n(\{s : |s - r| \leq \delta\}) + \mathbb{P}(b_n^{-1/2} > \delta).
\]

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

Proof of Theorem 5: Clearly,
\[  \hat{\gamma} = \frac{\int_{T_b} \int Y_n(s) \kappa_b(|s - r|)(Y_n(r) - Y_n(s))dG_n(r)dG_n(s)}{\int_{T_b} Y_n(s)^2dG_n(s)} \quad (30) \]

and proceeding as in the proof of Theorem 3 shows that it suffices to show the claim with \( Y_n(s) \) replaced by \( Y^*(s) = \omega J_c(s) \) in (30). Denote the resulting expression by \( \hat{\gamma}^* \), we have
\[
\hat{\gamma}^* = \frac{\mathbb{E}[1[S_n \in T_b]Y^*(S_n)\kappa_b(|S_n - R_n|)(Y^*(R_n) - Y^*(S_n))|Y^*]}{\mathbb{E}[1[S_n \in T_b]Y^*(S_n)^2|Y^*]} \quad a.s.
\]
\[ = \frac{\mathbb{E}[1[S \in T_b]Y^*(S)\kappa_b(|S - R|)(Y^*(R) - Y^*(S))|Y^*]}{\mathbb{E}[1[S \in T_b]Y^*(S)^2|Y^*]} \quad a.s.
\]
\[ = \frac{\int_{T_b} J_c(s) \kappa_b(|s - r|)(J_c(r) - J_c(s))dG(r)dG(s)}{\int_{T_b} J_c(s)^2dG(s)} \]

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

Proof of Theorem 6: We first show the result for \( L \) in place of \( J_c \).
As a first step, we show that replacing $L(s)$ by $L(s) - \tilde{m}$ induces a $o_p(1)$ difference, where the convergences throughout the proof are with respect to $b \to 0$. By Cauchy-Schwarz, the second moment of the difference is bounded above by

$$\mathbb{E} \left[ \left( b^{-d-1} \tilde{m} \int_{I_b} \int \kappa_b(|s-r|)(L(r) - L(s))dG(r)dG(s) \right)^2 \right]$$

$$\leq \mathbb{E}[\tilde{m}^2] \mathbb{E} \left[ \left( b^{-d-1} \int_{I_b} \int \kappa_b(|s-r|)(L(r) - L(s))dG(r)dG(s) \right)^2 \right].$$

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

$$\int_l^{u-b} \int_{l+1}^u \kappa_b(|s-r|)(L(r) - L(s))dG(r)dG(s) = \int_l^u h_b(r)L(r)dG(r)$$

with $h_b(r) = \int_l^u (1)[l+b \leq s \leq u-b]\kappa_b(|s-r|) - 1[l+r \leq u-b]\kappa_b(|s-r|))dG(s)$. By inspection, for all small enough $b$, $h_b(r) = 0$ for $r \in [l+2b, u-2b]$, $\int_l^{l+2b} h_b(r)dG(r) = \int_u^{u-2b} h_b(r)dG(r) = 0$ and $\sup_r |h_b(r)| \leq Cb$. Thus

$$\mathbb{E} \left[ \left( \int_l^u h_b(r)L(r)dG(r) \right)^2 \right] = \int_l^u \int_l^u h_b(r)h_b(s)\min(r,s)dG(r)dG(s)$$

$$= \int_l^{l+2b} \int_l^{l+2b} h_b(r)h_b(s)(\min(r,s) - l)dG(r)dG(s)$$

$$+ \int_u^{u-2b} \int_u^{u-2b} h_b(r)h_b(s)(\min(r,s) - u)dG(r)dG(s)$$

$$= O(b^5)$$

so the desired result follows.

For $d > 1$,

$$c_b^2 = \mathbb{E} \left[ \left( b^{-d-1} \int_{I_b} \int \kappa_b(|s-r|)(L(r) - L(s))dG(r)dG(s) \right)^2 \right]$$

$$= \mathbb{E} \left[ \left( b^{-1} \int_{I_b} \int \kappa_0(|r|)(L(s + br) - L(s))g(s + br)drdG(s) \right)^2 \right]$$

$$= \int_{I_b} \int_{I_b} \int \int b^{-2} \kappa_0(|r|) \kappa_0(|u|) \zeta_b(s, r, t, u)g(s + br)g(t + bu)dr \cdot du \cdot dG(s)dG(t)$$

with

$$4\zeta_b(s, r, t, u) = 4\mathbb{E}[(L(s + br) - L(s))(L(t + bu) - L(t))]$$

38
Now split the integral over \( dG(s) \) and \( dG(t) \) into a piece \( R_b^0 = \{ s, t : s, t \in I_b, |s - t| < 2b \} \) and \( R_b^1 = (I_b \times I_b) \setminus R_b^0 \). For the integral over \( R_b^0 \), note that for \(|s - t| < 2b\), \(|b(s, r, t, u)| < Cb\) for some \( C > 0 \). At the same time, the area of integration for \( R_b^0 \) is of order \( b^d \). So with \( g \) and \( \kappa_0 \) bounded, the integral over \( R_b^0 \) is of order \( b^{d-1} \rightarrow 0 \), and makes a vanishing contribution to \( c_b^2 \).

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

\[
\frac{\partial}{\partial x} |\omega + xv| = \frac{(\omega + xv)'v}{|\omega + xv|} \quad \frac{\partial^2}{\partial x^2} |\omega + xv| = -\frac{((\omega + xv)'v)^2}{|\omega + xv|^3} + \frac{v'v}{|\omega + xv|} \quad \frac{\partial^3}{\partial x^3} |\omega + xv| = 3 \frac{(\omega + xv)'v)^3}{|\omega + xv|^5} - 3 \frac{(\omega + xv)'v'v}{|\omega + xv|^3}.
\]

For the integral over \( R_b^1 \) where \(|s - t| \geq 2b\), employ a second order Taylor expansion with respect to \( r \) and \( t \), around \( b = 0 \). Since \( \frac{\partial}{\partial b} \frac{\partial^3}{\partial x^3} |\omega + xv| |\omega + xv|^5 \) is of order \( O(b^3) \), we have

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

Furthermore,

\[
\int_{I_b} \int_{I_b} \min \left( \frac{b^3}{|s-t|^2}, \frac{b}{2} \right) dG(s) dG(t) \leq C \int_{|s| < C} \min \left( \frac{b^3}{|s|^2}, b \right) ds = C \int_0^C x^{d-1} \min \left( \frac{b^3}{x^2}, b \right) dx = O(b^3 \log(b))
\]

so the result follows.

Given this first result, it is without loss of generality to assume that \( S \) does not contain the origin. Let \( Q_b \) be the left hand side of (18). We will show that \( Q_b \) converges to zero in mean square. We have

\[
\mathbb{E}[Q_b] = \frac{b^{-1}}{2} \int_{I_b} \int_{I_b} \kappa_0(|r|)(|s + br| - |s| - b|r|)g(s + br)drdG(s).
\]

By a first order Taylor expansion, for \(|s| \geq 2b\),

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

39
and $\mathbb{E}[Q_b] \rightarrow -\frac{1}{2} \int |r| \kappa_0(|r|) dr$ follows from $\int (s' r) \kappa_0(|r|) dr = 0$.

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

\[
\begin{align*}
\zeta^0_b(s, t) &= 2 \mathbb{E}[L(s)L(t)] = |s| + |t| - |s + t| \\
\zeta^1_b(s, r, t) &= 2 \mathbb{E}[(L(s + br) - L(s))L(t)] = |br + s| - |s| + |s - t| - |br + s - t| \\
\zeta^1_b(t, u, s) &= 2 \mathbb{E}[(L(t + bu) - L(r))L(s)]
\end{align*}
\]

Thus,

\[
4 \text{Var}[Q_b] = \int_{I_0} \int_{I_0} \int \int b^{-2} \kappa_0(|r|) \kappa_0(|u|) (\zeta^0_b(s, t) \zeta_b(s, r, t, u))g(s + br)g(t + bu) \\
+ \zeta^1_b(s, r, t) \zeta^1_b(t, u, s) (s + br)g(t + bu) \text{d}r \cdot \text{d}u \cdot \text{d}G(s) \text{d}G(t)
\]

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

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

\[
\zeta^1_b(s, r, t) \zeta^1_b(t, u, s) (s + br)g(t + bu) = b^2 g(s)g(t) \left( \frac{s' r}{|s|} - \frac{(s - t)' r}{|s - t|} \right) \left( \frac{t' u}{|t|} - \frac{(t - s)' u}{|t - s|} \right) + b^3 \frac{1}{|s - t|^2} \Psi_b(s, r, u)
\]

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

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

so $\text{Var}[Q_b] \rightarrow 0$.

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

Lemma 7 is a special case of the following more general result with $p = 1$ and $\psi(s) = 1$.

**Lemma 10.** Let $L^2_0$ be the Hilbert space of function $S \mapsto \mathbb{R}$ with inner product $\langle f_1, f_2 \rangle = \int f_1(s) f_2(s) dG(s)$. Let $L_k : L^2_0 \mapsto L^2_0$ be the linear operator $L_k(f)(s) = \int f(r) k(r, s) dG(r)$, and $L_{k,n} = \int f(r) k(r, s) dG_n(r)$. Suppose the $p \times 1$ vector $x_l$ is such that $x_l = \psi(s_l)$ for $l = 1, \ldots, n$ for
some continuous function $\psi : S \mapsto \mathbb{R}^p$, and $\int \psi(s)\psi(s)\prime dG_n(s) = H_n \to H$ for some positive definite matrix $H$. Let $M$ and $M_n$ be the projection operators $M_n(f) = f - \int f(s) dG_n(s)$ and $M(f) = f - \int f(s) dG(r)H^{-1}\psi(s)$. Let $k_n$ and $k$ be the kernels corresponding to the linear operators $M_nL_{k,n}M_n$ and $ML_kM$, respectively, so that the $(i,\ell)$ element of $M_X\Sigma_nLM_X$ is given by $\hat{k}_n(s_i, s_\ell)$. Let $\hat{k}(s, r) = \sum_{i=1}^{\infty} \hat{\psi}_i(s)\hat{\psi}_i(r)$ with $\hat{\psi}_i(s)\hat{\psi}_i(s)dG(s) = 1[i = j]$, $\hat{\psi}_i \geq \hat{\psi}_{i+1} \geq 0$ be the spectral decomposition of $k$. Define $\hat{\psi}_i(\cdot) = n^{-1}\sum_{i=1}^{n} r_i, \hat{k}(\cdot,s_i)$, where $(\hat{\psi}_i, (r_i,\ldots,r_{i,n})')$ is the $i$th eigenvalue/eigenvector pair of $M_X\Sigma_nLM_X$. If $\hat{\psi}_1 > \hat{\psi}_2 > \ldots > \hat{\psi}_q > \hat{\psi}_{q+1}$ and $G_n \Rightarrow G$, then for any $q \geq 1$, $\sup_{s \in S, 1 \leq i \leq q} |\hat{\psi}_i(s) - \hat{\psi}(s)| \to 0$ and $\max_{1 \leq i \leq q} |\hat{\psi}_i - \hat{\psi}_i| \to 0$.

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

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

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

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

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

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

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

$$\left\| \int \psi_i(r)k_0(\cdot, r)(dG_n(r) - dG(r)) \right\|^2_{\mathcal{H}}$$

$$= \int \int \psi_i(s)k_0(s, r)\psi_i(r)'(dG_n(s) - dG(s))(dG_n(r) - dG(r))$$

$$= \mathbb{E}[\psi_i(S_n)k_0(S_n, R_n)\psi_i(R_n) - \psi_i(S_n)k_0(S, R_n)\psi_i(R) - \psi_i(S)k_0(S, R)\psi_i(R_n) + \psi_i(S)k_0(S, R)\psi_i(R)]$$

$$\to 0$$

where for each $n$, $S_n, R_n$ are independent random variables both with distribution $G_n$, and $S, R$ are two independent random variables with distribution $G$. The convergence then follows since the
$\mathbb{R}^{2d} \mapsto \mathbb{R}$ function $(s, r) \mapsto \psi_i(r)k_0(s, r)\psi_j(s)$ is continuous and bounded. Thus, by (31) and (32) and the Cauchy-Schwarz inequality,

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

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

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

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

$$||L_n - L||_{HS}^2 = \sum_{j \geq 1} \left( \int e_j(s)k_0(s, \cdot)(dG_n(s) - dG(s)), \int e_j(s)k_0(s, \cdot)(dG_n(s) - dG(s)) \right)_{\mathcal{H}}$$

$$= \int \int \left( \sum_{j \geq 1} e_j(s) e_j(r) \right) k_0(s, r)(dG_n(s) - dG(s))(dG_n(r) - dG(r))$$

$$= \int \int k_0(s, r)^2(dG_n(r) - dG(r))(dG_n(r) - dG(r))$$

$$= \mathbb{E}[k_0(S_n, R_n)^2 - k_0(S, R_n)^2 - k_0(S_n, R)^2 + k_0(S, R)^2] \to 0$$

where the change of the order of integration and summation is justified by Fubini’s Theorem, and the convergence follows, since the $R^{2d} \mapsto R$ function $(s, r) \mapsto k_0(s, r)^2$ is bounded and continuous. \hfill $\Box$

**Lemma 11.** Suppose $\tilde{x}_i = \psi_n(s_i)$, where the continuous functions $\psi_n : S \mapsto \mathbb{R}^p$ are such that $\sup_{s \in S}|\psi_n(s) - \psi(s)| \to 0$, for some continuous function $\psi$. Define the the projection operator $\tilde{M}_n : \mathcal{L}_G^2 \mapsto \mathcal{L}_G^2$ as $\tilde{M}_n(f)(s) = f(s) - \int \psi_n(r)'f(r)dG_n(r)H_n^{-1}\psi_n(s)$, and let $\tilde{k}_n$ be the kernel corresponding to the linear operator $\tilde{M}_nL_k\tilde{M}_n$, so that the $(l, \ell)$ element of $\tilde{M}_n\Sigma_nL\tilde{M}_n$ is given by $\tilde{k}_n(s_l, s_\ell)$. Let $(\tilde{\nu}_1, (\tilde{\nu}_l, \ldots, \tilde{\nu}_n))'$ be the $i$th eigenvalue/eigenvector pair of $\tilde{M}_n\Sigma_nL\tilde{M}_n$, and define $\tilde{\varphi}_i(\cdot) = n^{-1}\tilde{\nu}_i^{-1}\sum_{l=1}^n \tilde{\nu}_l\tilde{k}_n(\cdot, s_l)$. Under the conditions of Lemma 10, $\sup_{s \in S, 1 \leq i \leq q} |\tilde{\varphi}_i(s) - \tilde{\varphi}(s)| \to 0$ and $\max_{1 \leq i \leq q} |\tilde{\nu}_i - \tilde{\nu}_i| \to 0$.

**Proof.** From standard arguments, we obtain $\int \psi_n(s)\psi_n(s)'dG_n(s) \to H$ and $\int \psi(s)\psi_n(s)'dG_n(s) \to H$. Thus, $||M_n - M|| \to 0$, and by a direct calculation, $\sup_{s, r \in S} |\tilde{k}_n(r, s) - \tilde{k}_n(r, s)| \to 0$, and $\sup_{s, r \in S} |\tilde{k}_n(r, s) - \tilde{k}(r, s)| \to 0$ and thus $\sup_{s, r \in S} |\tilde{k}_n(r, s) - \tilde{k}_n(s, r)| \to 0$. Furthermore, proceeding as
in the proof of Lemma 10 shows that \(|\Sigma_{n,L}|| converges to \(\tilde{\nu}_1\), the largest eigenvalue of the integral operator with kernel \(\tilde{k}\), so \(|\Sigma_{n,L}|| = O(1)\). Thus also \(|M_X \Sigma_{n,L} M_X - M_X \Sigma_{n,L} M_X|| \to 0\), and from Weyl’s inequality, \(\max_{1 \leq i \leq q} |\tilde{\nu}_i - \tilde{\nu}_i| \to 0\). Since also \(\max_{1 \leq i \leq q} |\tilde{\nu}_i - \tilde{\nu}_i| \to 0\) from Lemma 10, we can conclude that

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

where the inequality uses \(r_{i,l} = \tilde{\varphi}_i(s_l)\), and the convergence follows from the above results and \(\sup_{s \in \mathcal{S}} |\tilde{\varphi}_i(s)| \to \sup_{s \in \mathcal{S}} |\tilde{\varphi}_i(s)| < \infty\) from Lemma 10. Also,

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

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

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

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

**Theorem 12.** Suppose \(y_l = x_l^T \beta + u_l\), \((x_l, u_l) = (X_n(s_l), U_n(s_l)) \in \mathbb{R}^p \times \mathbb{R}\) with \((X_n(\cdot), U_n(\cdot))\) satisfying (25), but \(X\) is not necessarily independent of \(U\). Let \(R_n^X\) be the \(n \times p\) matrix of \(q\) eigenvectors of \(M_X \Sigma_{n,L} M_X\) corresponding to the largest eigenvalues. Suppose for almost every realization of \(X\), the largest \(q + 1\) eigenvalues of the kernel \(k_X : \mathcal{S} \times \mathcal{S} \to \mathbb{R}\) corresponding to the linear operator \(M_X L_k M_X\) with \(M_X(f)(s) = f(s) - X(s) (\int X(r) X(r') dG(r))^{-1} \int X(r') f(r) dG(r)\) are distinct. If further \(G_n \Rightarrow G\), then

\[
\lambda_n^{-1/2} R_n^X Y_n \Rightarrow \omega \int \varphi_X(s) U(s) dG(s)
\]

where \(\varphi_X(\cdot)\) are the \(q\) eigenfunctions of \(k_X\) corresponding to the largest eigenvalues.

Furthermore, let \(U_n\) be independent of \((X_n, U_n)\), and suppose \(U_n\) satisfies \(\lambda_n^{-1/2} U_n(\cdot) \Rightarrow \tilde{U}(\cdot)\) with \(\tilde{U} \sim U\). Let \(cv_n(X_n)\) be the \(1 - \alpha\) quantile of the conditional distribution of \(\phi(R_n^X U_n)\) given \(X_n\) for some continuous function \(\phi : \mathbb{R}^q \to \mathbb{R}\) satisfying \(\phi(cx) = \phi(x)\) for all \(c \neq 0\) and \(x \in \mathbb{R}^q\). Suppose that (i) \(X\) is independent of \(U\), (ii) for almost all realizations of \(X\) the conditional distribution of \(\phi(\int \varphi_X(s) U(s) dG(s))\) is continuous. Then \(P(\phi(R_n^X Y_n) > cv_n(X_n)) \to \alpha\).

**Proof.** We will show that \((\phi(R_n^X Y_n), cv_n(X_n)) \Rightarrow (\phi(\int \varphi_X(s) U(s) dG(s)), q_{1-\alpha}(X))\) with \(q_{1-\alpha}(X)\) the \(1 - \alpha\) quantile of \(\phi(\int \varphi_X(s) U(s) dG(s))\) conditional on \(X\). The result then follows from the continuous mapping theorem applied to \(1[\phi(R_n^X Y_n) > cv_n(X_n)]\), and taking expectations.
Apply the almost sure representation theorem to argue that there exists a probability space \((\Omega_0, \mathcal{F}_0, P_0)\) and associated random processes \(X^*, U^*\) and \(X_n^*, U_n^*, n \geq 1\) such that \((X_n^*, U_n^*) \sim (\Lambda_n^{-1/2} X_n, \lambda_n^{-1/2} U_n)\), \((X^*, U^*) \sim (X, U)\) and \(\sup_{s \in S} |X_n^*(s) - X^*(s)| \overset{a.s.}{\rightarrow} 0\), \(\sup_{s \in S} |U_n^*(s) - U^*(s)| \overset{a.s.}{\rightarrow} 0\). Using the same arguments as in the proof of Theorem 3, and a realization by realization application of Lemma 11, then yields

\[
\lambda_n^{-1/2} R_n^{X^*} Y_n^{*} \overset{a.s.}{\rightarrow} \omega \int \varphi_{X^*}(s) U^*(s) dG(s) = \omega \int \varphi_{X^*}(s) U(s) dG(s) \tag{34}
\]

where \((R_n^{X^*}, Y_n^*)\) are defined analogously to \((R_n^X, Y_n)\) on \((\Omega_0, \mathcal{F}_0, P_0)\), and \((R_n^{X^*}, Y_n^*) \sim (R_n^X, Y_n)\) by construction. The result now follows if we can show that also \(cv_n(X_n^*) \overset{a.s.}{\rightarrow} q^\phi_{1-\alpha}(X^*)\), since almost sure convergence implies convergence in distribution.

To that end, note there exists a separate probability space \((\Omega_1, \mathcal{F}_1, P_1)\) with associated sequences of random processes \(\tilde{U}^*\) and \(\tilde{U}_n^*\) such that \(\tilde{U}_n^* \sim \lambda_n^{-1/2} \tilde{U}_n, \tilde{U}^* \sim \tilde{U} \sim U\) and \(\sup_{s \in S} |\tilde{U}_n^*(s) - \tilde{U}^*(s)| \overset{a.s.}{\rightarrow} 0\). Form the product space \((\Omega_0 \times \Omega_1, \mathcal{F}_0 \otimes \mathcal{F}_1, P_0 \times P_1)\), so that on this new space, \((X^*, X_n^*)\) is independent of \((\tilde{U}^*, \tilde{U}_n^*)\) by construction. Use the same arguments as for (34) 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,

\[
\lambda_n^{-1/2} R_n^{X^*} \tilde{U}_n^* \overset{a.s.}{\rightarrow} \int \varphi_{X^*}(s) \tilde{U}^*(s) dG(s)
\]

jointly with (34). But almost sure convergence implies convergence in distribution, and \(\tilde{U}^* \sim U\), so for \(P_0\)-almost all \(\omega_0 \in \Omega_0\), the distribution of \(\lambda_n^{-1/2} R_n^{X^*} \tilde{U}_n^*\) induced by \(P_1\) converges to the conditional distribution of \(\int \varphi_{X^*}(s) U(s) dG(s)\) given \(X^*\). Since \(\phi\) is continuous and the conditional distribution is assumed continuous, this implies that for all such \(\omega_0, cv_n(X_n) \overset{a.s.}{\rightarrow} q^\phi_{1-\alpha}(X^*)\). Thus \((\phi(R_n^X Y_n), cv_n(X_n)) \sim (\phi(R_n^{X^*} Y_n^*), cv_n(X_n^*)) \overset{a.s.}{\rightarrow} (\phi(\int \varphi_{X^*}(s) U^*(s) dG(s)), q^\phi_{1-\alpha}(X^*)) \sim (\phi(\int \varphi_{X^*}(s) U(s) dG(s)), q^\phi_{1-\alpha}(X))\), and the result follows, because almost sure convergence implies convergence in distribution.

**Proof of Theorem 8:**

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

\[
\lambda_n^{d/2} n^{-1} Z_n \Rightarrow N\left(0, a \sigma_B(0) \int \tilde{\varphi}(s) \varphi(s) dG(s) + \omega^2 \int \tilde{\varphi}(s) \varphi(s) g(s) dG(s)\right) \tag{35}
\]

where \(\varphi = (\varphi_1, \ldots, \varphi_q)\), \(\omega^2 = \int_{\mathbb{R}^d} \sigma_B(s) ds\) and \(g\) is the density of the distribution \(G\). Since the LFST statistic is scale invariant, its limiting distribution under (35) 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 \(\liminf_{n \to \infty} cv_n^{\text{LFST}}\) is at least as large as the \(1 - \alpha\) quantile, say \(cv^{\text{LFST}}_\chi\), of the (continuous) asymptotic distribution of LFST for this value of \(\chi\).

44
Note that for $B = J_c$, $\sigma_B(0)/\omega^2 = K_d c^{1+d}$ for some $K_d > 0$. For $a > 0$, let $c_*$ be such $K_d c_*^{1+d} = \chi/a$, and let $c_* = 1$ otherwise. For all $n$ sufficiently large so that $\lambda_n c_* \geq 0.03$, $\text{cv}_{n}^{\text{LFST}}$ 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 (35), so its $1 - a$ quantile converges to $\text{cv}_{\chi}^{\text{LFST}}$, and the result follows. □

References


