Testing Coefficient Variability in Spatial Regression

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

This Draft: April 2024

Abstract

This paper develops a test for coefficient stability in spatial regressions. The test is designed to have good power for a wide range of persistent patterns of coefficient variation, be applicable in a wide range of spatial designs, and to accommodate both spatial correlation and spatial heteroskedasticity in regressors and regression errors. The test approximates the best local invariant test for coefficient stability in a Gaussian regression model with Lévy-Brown motion coefficient variation under the alternative, and is thus a spatial generalization of the Nyblom (1989) test of coefficient stability in time series regressions. An application to 1514 zip-code level bivariate regressions of U.S. socioeconomic variables reveals widespread coefficient instability.

Keywords: Spatial Correlation, Spatial Heterokesdasticity, Nyblom test
JEL: C12, C20

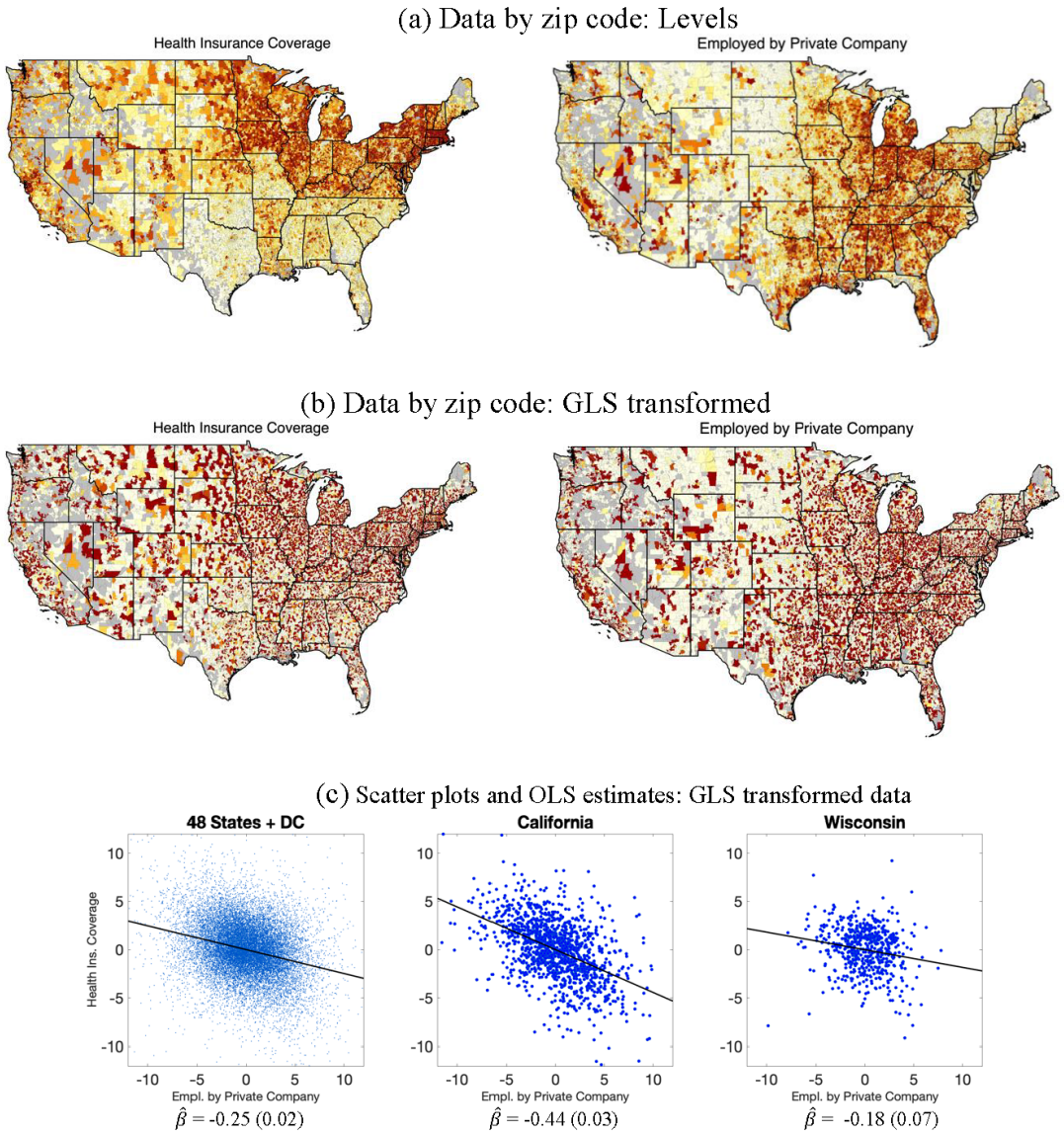
1 Introduction

This paper develops a test for coefficient stability in spatial regressions. The test is designed to have good power for a wide range of persistent patterns of coefficient variation, be applicable in a wide range of spatial designs, and allow for both spatial correlation and spatial heteroskedasticity in regressors and regression errors.

A specific empirical question helps motivate our analysis. Health insurance is not universal in the United States and is often provided by an employer. Are employees at private companies more or less likely to have health insurance than employees at other establishments (public, non-profit, self-employment, etc.)? And, if so, how large is the effect? Panel (a) of Figure 1 shows data on health insurance coverage (HIC) and private company employment (PCE) in roughly twenty thousand geographical regions ("zip codes") across the continental U.S. The data are discussed in detail in Section 4; for now we note that the data are ranks measured in percentiles, so for example, a value of 0.75 indicates that the zip code ranks in the 75th percentile for that variable. Consider the regression of HIC on PCE. Estimation of the slope coefficient β is complicated by the high degree of spatial correlation in the two variables. Time series regressions with highly serially correlated data can produce spurious regressions (Granger and Newbold (1974)) which yield unstable estimates, and spatial regressions with highly spatially correlated data suffer the same defect (Müller and Watson (2023)). Thus, before estimating β the data are spatially differenced using the GLS transformation advocated by Müller and Watson (2023), as shown in panel (b). Panel (c) shows a scatterplot of the transformed data with the resulting value of $\hat{\beta} = -0.25$, suggesting that PCE is associated with a reasonably large drop in health insurance coverage. Panel (c) also shows scatterplots using data only for California and Wisconsin. The estimated value of β in California is more than twice as large as in Wisconsin, and a Chow test of the null hypothesis of equal coefficients rejects with a t-statistic of 3.3. Should a researcher conclude that the HIC-PCE regression is unstable across the U.S.?

A moment's thought prompts caution reaching this conclusion. First, why consider only California and Wisconsin? Were these states chosen after "snooping" at the values of the OLS coefficients? If so, then the Chow test's critical value needs to be adjusted using a spatial analogue of the method developed by Andrews (1993) for time series regressions. More generally, how can one test for the wide variety of ways that regression coefficient might vary

Figure 1: Health Insurance Coverage and Employment by Private Companies



Notes: Panel (a) plots the health insurance coverage rate and the fraction of workers employed at private companies across 21,194 zip codes in the continental United States. Panel (b) shows the same data after applying a GLS transformation for Lévy Brownian motion spatial correlation. Panel (c) shows scatterplots using the GLS-transformed data from panel (b).

across space? In the HIC-PCE example, β might change discretely across state boundaries, or change more continuously with the type of employers in a region, or some combination of both. A useful test should have power against many different forms of spatial variation. A second concern is that socioeconomic data are not i.i.d. over space but typically exhibit spatial correlations and also spatial heteroskedasticity, such as a higher variance in some regions than others. Such variation might be mistaken for a spatially varying coefficient.

Coefficient stability tests that meet these challenges have been developed for time series applications, including the widely-used tests developed in Nyblom (1989) and Andrews (1993). While these tests were designed with particular types of time variation in mind (martingale variation in Nyblom (1989) and a single discrete break in Andrews (1993)) they are known to have good power for more general (persistent) patterns of coefficient variation (e.g., Elliott and Müller (2006)). And if one computes the critical value using Hansen's (2000) bootstrap scheme, then these tests are also robust against time-varying second moments, albeit at the cost of assuming serially uncorrelated regression errors.

The literature on testing for stability in spatial regressions is less well developed. Anselin (1990) derives a Chow test for spatial instability across two known regions, allowing for spatially correlated residuals. We are not aware of generalizations to testing for instability across two unknown regions. And, while there is an applied literature estimating "local" spatial regressions (see, for example, Fotheringham, Brunsdon, and Charlton (2002) and Fotheringham, Oshan, and Li (2024)), the stability tests using these methods that we are aware of are predicated on restrictive assumptions such as i.i.d. observations (e.g., the bootstrap methods discussed in Mei, Xu, and Wang (2016) or Fotheringham, Oshan, and Li (2024)). This paper's contribution is a spatial stability test that is locally best against martingale-like random coefficient variation in a canonical spatial regression model (in analogy to Nyblom (1989)), controls size under general distributional assumptions and spatial autocorrelation (in analogy to Andrews (1993)), and accommodates spatially varying second moments (in analogy to Hansen (2000)). The combination of robustness to correlated regression errors and second-moment instabilities is new also relative to the time series literature.

The paper is organized as follows. Section 2 lays out a canonical spatial regression model involving a dependent variable, y, a regressor of interest, x, with coefficient β and a vector of controls, z. In this model, the regressors are taken as fixed and the regression error is i.i.d. standard normal, making it straightforward to develop a test for spatial stability of β

with desirable optimality properties. Section 3 studies the large-sample properties of a version of this test under more realistic assumptions. Section 4 introduces a large spatial socioeconomic data set for the United States that is used to calibrate a set of simulation experiments to evaluate properties of the proposed test in realistic empirical settings. The section concludes by applying the proposed test to examine stability in over 1500 regressions involving variables from the dataset. It finds widespread spatial instability. The final section offers some concluding remarks. An online appendix includes a "user's guide" that summarizes the necessary calculations for computing the test statistic, its p-value and other statistics measuring the variability in the coefficients. The appendix also presents the required modifications of the various formulae for applications using instrumental variables.

2 A Canonical Spatial Regression Model with Varying Coefficients

We begin by studying a canonical spatial regression model with varying coefficients. The simple structure of the canonical model makes it easy to highlight several key features of the testing problem, and an optimal test in the canonical model follows from straightforward calculations.

Thus, consider a regression of y_l on a scalar regressor of interest x_l and $p-1$ controls z_l,

$$y_l = x_l\beta_l + z_l'\alpha + u_l, \;\; l = 1, \ldots, n, \tag{1}$$

where the observations (y_l, x_l, z_l) are associated with known spatial locations $s_l \in \mathcal{S} \subset \mathbb{R}^d$, for $d \geq 1$. We are interested in testing the null hypothesis that β_l is constant across space, that is

$$H_0 : \beta_l = \beta \text{ against } H_a : \beta_l \neq \beta_\ell \text{ for some } 1 \leq l, \ell \leq n. \tag{2}$$

In the canonical model, we assume that $u_l \sim iid\mathcal{N}(0,1)$ and $\{x_l, z_l\}$ are nonstochastic.

Let $w_l = (x_l, z_l')'$ and $\delta = (\beta, \alpha')'$. Then (1) can be written as

$$y_l = w_l'\delta + e_l, \;\; e_l = u_l + (\beta_l - \beta)x_l, \;\; l = 1, \ldots, n. \tag{3}$$

For the purpose of learning about variation in β_l, the vector δ is a nuisance parameter, and

we focus on tests that are invariant to the transformations

$$y \to y + Wd \text{ for all } d \in \mathbb{R}^p$$

where $y = (y_1, \ldots, y_n)'$ and $W = (w_1, \ldots, w_n)'$, effectively eliminating δ from the analysis. One maximal invariant to these transformations is given by the residuals $\hat{e} = y - W\hat{\delta}$, where $\hat{\delta}$ is the OLS regression coefficient $\hat{\delta} = (W'W)^{-1}W'y$.

A standard calculation shows that the best invariant test of (2) against the alternative $\{\beta_l\}_{l=1}^n = \{\beta_l^1\}_{l=1}^n$ rejects the null hypothesis for large values of $\sum_{l=1}^n \beta_l^1 x_l \hat{e}_l$. Evidently, the optimal test depends on the particular configuration of the spatially varying $\beta_l's$, so there is no uniformly most powerful test, and any particular choice of $\{\beta_l^1\}_{l=1}^n$ leads to poor power properties under some alternative $\{\beta_l\}_{l=1}^n$.

This suggests using a weighted average power criterion over the various possible values of $\{\beta_l\}_{l=1}^n$. Maximizing weighted average power is equivalent to maximizing power against the alternative where β_l is random and drawn from the probability distribution that is proportional to the weighting function. We will use

$$H_a^* : \beta_l - \beta = \kappa L(s_l),\ l = 1, \ldots, n \tag{4}$$

where $\kappa > 0$ measures the size of the instability and $L(s)$ is a Lévy (1948)-Brownian motion on $\mathcal{S}$. Lévy-Brownian motion is a continuous parameter mean-zero Gaussian process with almost surely continuous sample paths and covariance kernel equal to $k_L(s,r) = \mathbb{E}[L(s)L(r)] = \frac{1}{2}(||r|| + ||s|| - ||s - r||)$. It provides an attractive weighting function because L induces a Wiener process along each line: For all $v_0, v_1 \in \mathbb{R}^d$ with $||v_1|| = 1$, $L(v_0 + tv_1) - L(v_0) \sim \mathbb{W}(t)$, where $\mathbb{W}(t)$ with $t \in \mathbb{R}$ is a standard scalar Wiener process. Thus, the parameter variation along each direction is a martingale, suitably generalizing Nyblom's (1989) assumption for parameter variation in time. Let Σ_L be the $n \times n$ covariance matrix of $(L(s_1), \ldots, L(s_n))'$.

As discussed in Chapter 5.5. of Ferguson (1967), the invariant test of H_0 that maximizes the slope of the power function against alternatives H_a^* at $\kappa = 0$ rejects for large values of $\partial \log(f(\hat{e}|\kappa))/\partial\kappa|_{\kappa=0}$, where $f(\hat{e}|\kappa)$ is the density of $\hat{e}$ under H_a^*. A calculation shows that this locally best invariant test equivalently rejects for large values of

$$\xi^* = n^{-1}\hat{e}' D_x \Sigma_L D_x \hat{e}$$

where $D_x = \text{diag}(x_1, \ldots, x_n)$, and the level α critical value is given by the $1-\alpha$ quantile of ξ^* that is induced by $e_l = u_l \sim iid\mathcal{N}(0,1)$.

Since $D_x\hat{e}$ is orthogonal to a constant, ξ^* can be rewritten as $\xi^* = n^{-1}\hat{e}'D_x\bar{\Sigma}_L D_x\hat{e}$, where $\bar{\Sigma}_L = M_1\Sigma_L M_1$ and M_1 is the $n \times n$ projection matrix that projects off the constant vector. Let $\bar{\Sigma}_L = R\Lambda R'$ be the spectral decomposition of $\bar{\Sigma}_L$, where the matrix of eigenvectors $R = (r_1, \ldots, r_n)$ is normalized to satisfy $n^{-1}R'R = I_n$, and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = 0$ are the eigenvalues of $\bar{\Sigma}_L$, scaled by n^{-1}. In this notation,

$$\xi^* = \sum_{j=1}^{n} \lambda_j \left(n^{-1/2} \sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l \right)^2 .$$

Figure 2 shows selected eigenvectors r_j for the locations of the empirical example of the introduction. The statistic ξ^* detects spatial variation in β_l by checking whether inner-products of the empirical scores $x_l\hat{e}_l$ and the eigenvectors r_j are more variable than expected given the randomness in u_l, with more weight given to inner-products with eigenvectors corresponding to the largest eigenvalues.

To simplify the asymptotic analysis in the general model introduced in the next section, we use an approximation to ξ^*,

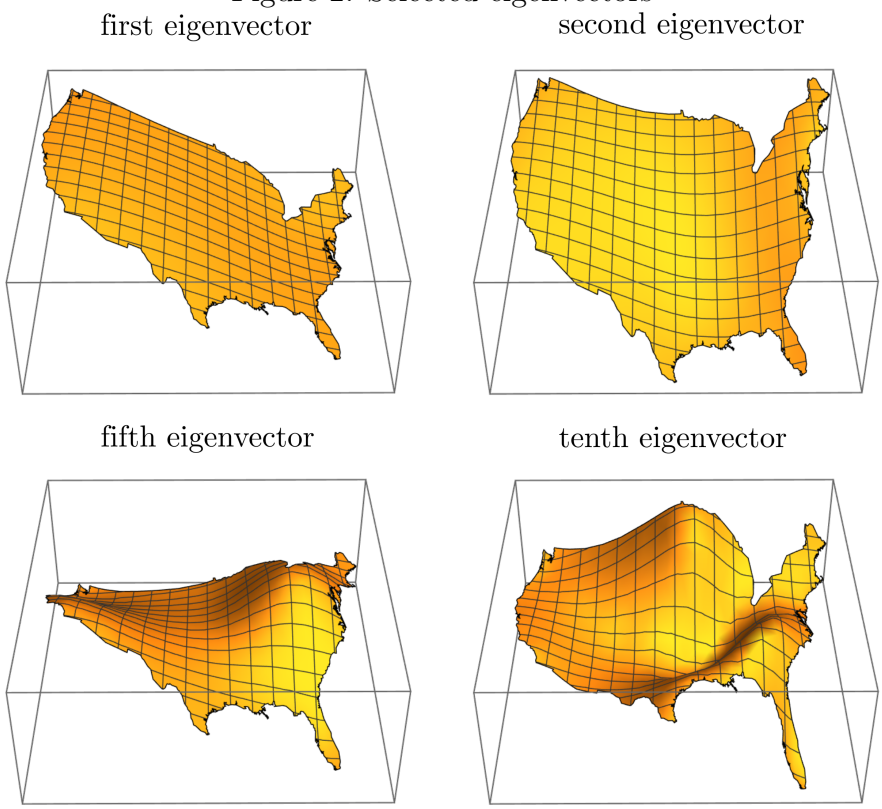
$$\xi_q = \sum_{j=1}^{q} \lambda_j \left(n^{-1/2} \sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l \right)^2 \tag{5}$$

for some fixed q, such as $q = 15$. The quality of the approximation of ξ^* by ξ_q depends on the distribution of eigenvalues λ_j of $\bar{\Sigma}_L$, which tend to decay rapidly. For instance, with the 21,194 zip-code locations from the empirical example of the introduction, $\lambda_1/\operatorname{tr}\bar{\Sigma}_L = 0.42$, $\lambda_2/\operatorname{tr}\bar{\Sigma}_L = 0.16$, $\lambda_{15}/\operatorname{tr}\bar{\Sigma}_L = 0.005$ and $\sum_{j=1}^{15}\lambda_j/\operatorname{tr}\bar{\Sigma}_L = 0.87$, so one would expect the loss of using ξ_q rather than ξ^* to be fairly minimal.

Before leaving the canonical model, we define two pieces of notation that will be useful in the following section. The first defines the $q \times 1$ vector Y_n and the diagonal matrix Λ_q as

$$Y_{n,j} = n^{-1/2} \sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l,\ Y_n = (Y_{n,1}, ..., Y_{n,q})', \text{ and } \Lambda_q = \text{diag}(\lambda_1, \ldots, \lambda_q), \tag{6}$$

Figure 2: Selected eigenvectors



Notes: The figure shows eigenvectors for the demeaned Lévy Brownian motion covariance matrix evaluated at the location of 21,194 zip codes in the continental United States.

so the test statistic ξ_q in (5) becomes the quadratic form

$$\xi_q = Y_n' \Lambda_q Y_n. \tag{7}$$

The second piece of notation writes $Y_{n,j}$ as a weighted average of e_l instead of the residuals $\hat{e}_l$. Using $\hat{e}_l = e_l - w_l'(\hat{\delta} - \delta)$, a calculation yields

$$Y_{n,j} = n^{-1/2} \sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l = n^{-1/2} \sum_{l=1}^{n} \bar{r}_{j,l}^{e\prime} w_l e_l \tag{8}$$

where $\bar{r}_{j,l}^{e}$ is the $p \times 1$ vector

$$\bar{r}_{j,l}^{e} = r_{j,l}^{e} - \left(n^{-1} \sum_{\ell=1}^{n} w_\ell w_\ell' \right)^{-1} n^{-1} \sum_{\ell=1}^{n} w_\ell w_\ell' r_{j,\ell}^{e} \quad \text{with } r_{j,l}^{e} = (r_{j,l}, 0, \ldots, 0)'. \tag{9}$$

The representation of the test statistic ξ_q in (7) with Y_n given in (8) provides a natural framework for the analysis of the distribution of ξ_q under the null and alternative hypothesis. In the canonical model, $e_l = u_l \sim iid\mathcal{N}(0,1)$ under the null hypothesis, so that Y_n is a zero-mean normal random variable with a covariance matrix that depends the weights $\bar{r}_j^e$, and ξ_q is a quadratic form that involves Y_n and the eigenvalues $\{\lambda_i\}_{i=1}^q$ that appear in Λ_q. Under the alternative, $e_l = u_l + x_l(\beta_l - \beta)$, so Y_n incorporates an additional term involving the spatially varying coefficients $\{\beta_l\}$, which could be considered as fixed constants (inducing a non-zero mean in Y_n) or realizations of Lévy-Brownian motion under (4) (augmenting the covariance matrix of Y_n). Thus, the distribution of the test statistic ξ_q under both the null and alternative hypothesis follows directly in the canonical model.

The large-sample distribution of ξ_q under more general assumptions can be deduced analogously using a central limit result for weighted averages of $w_l u_l$ and large sample limit results for the weights $\bar{r}_{j,l}^e$ and eigenvalues λ_i. As a practical matter, the most important modification of the analysis for the canonical model is the use of a spatial HAC estimator for the variance of Y_n The following section provides the analysis.

3 Large-Sample Analysis

The analysis in this section proceeds in five steps. The first studies the large-sample behavior of the eigenvectors and eigenvalues (r_j, λ_j) that are used to form ξ_q. Recall that (r_j, λ_j) are constructed from $\bar{\Sigma}_L$, the demeaned version of the Lévy-Brownian motion covariance matrix, and therefore do not depend on the data $\{y_l, w_l\}$. However, they do depend on the locations $\{s_l\}$, and the analysis begins with assumptions about the distribution of these locations. The second step involves the large-sample behavior of weighted averages of terms such as $w_l u_l$ and $w_l w_l'$ that appear in Y_n. The third step combines the results from the previous steps to deduce the large-sample behavior of Y_n, and ξ_q. The fourth step in the analysis studies the properties of an estimator for parameters that characterize the limiting distribution of ξ_q. The fifth and final step of the analysis discusses estimates of parameter variation under the alternative including the parameter κ in (4) and the parameter path $\{\beta_l\}$.

Throughout we allow for a double-array structure where the values or distribution of $\{s_l\}_{l=1}^n$ and $\{(w_l, u_l)\}_{l=1}^n$ are allowed to vary with n, but this dependence on n is not made explicit in the notation.

3.1 Assumptions and the Large-Sample Distribution of ξ_q

3.1.1 Locations and Large-Sample Behavior of (r_j, λ_j)

We make the following assumption on the locations.

Condition 1. *The locations s_l are nonstochastic and elements of the compact set $\mathcal{S} \subset \mathbb{R}^d$, $d \geq 1$, and the empirical distribution G_n of $\{s_l\}_{l=1}^n$ converges weakly to G, an absolutely continuous distribution on $\mathcal{S}$ with bounded density g.*

Random locations can be accommodated by Condition 1 by conditioning, as long as the data generating process for the locations is independent of all other random elements of the regression model. The assumption that the empirical distribution converges does not seem overly limiting in this context; for instance, it would hold almost surely if the locations were sampled i.i.d. from density g.

Because Lévy-Brownian motion is self-similar, $L(\nu s) \sim \nu^{1/2} L(s)$ for all $\nu > 0$, the test statistic ξ_q and its critical value scale proportionally with the scale of $\{s_l\}_{l=1}^n$. The test based on ξ_q is thus invariant to the transformations $\{s_l\}_{l=1}^n \to \{\nu s_l\}_{l=1}^n$, $\nu > 0$. The assumption of a sample size independent sampling region $\mathcal{S}$ is thus merely for notational convenience; if the "original" locations s_l^* are obtained from increasing domain asymptotics with $s_l^* \in \mathcal{S}_n = \nu_n \mathcal{S} = \{s : \nu_n^{-1} s \in \mathcal{S}\}$ and ν_n is diverging, as in say Lahiri (2003), then the rescaled locations $s_l = \nu_n^{-1} s_l^*$ satisfy $\{s_l\}_{l=1}^n \subset \mathcal{S}$.

Condition 1 imposes sufficient structure to study the asymptotic behavior of the weights λ_j and the eigenvectors r_j that appear ξ_q: Define $\bar{L}(s) = L(s) - \int L(r) dG(r)$, a demeaned Lévy-Brownian Motion on $\mathcal{S}$. By Mercer's Theorem, the (continuous) covariance kernel $\bar{k}_L$ of $\bar{L}$ has a spectral decomposition

$$\bar{k}_L(r, s) = \mathbb{E}[\bar{L}(s)\bar{L}(r)] = \sum_{j=1}^{\infty} \lambda_j^0 \varphi_j(s) \varphi_j(r) \tag{10}$$

where $\lambda_1^0 \geq \lambda_2^0 \geq \ldots$, and the eigenfunctions $\varphi_j(s) = \lambda_j^{-1} \int \varphi_j(r) \bar{k}_L(r, s) dG(r)$ are continuous for $\lambda_j^0 > 0$, orthonormal $\int \varphi_i(s) \varphi_j(s) dG(s) = \mathbf{1}[i = j]$ and orthogonal to a constant, $\int \varphi_j(s) dG(s) = 0$, $j = 1, 2, \ldots$. Lemma S.1 in Müller and Watson (2023) establishes that under Condition 1 and for any finite q, the largest q eigenvalue-eigenvector pairs (λ_j, r_j) of $\bar{\Sigma}_L$

become well approximated by the corresponding eigenvalue-eigenfunction pairs of $\bar{k}_L$, that is

$$\sup_{j\leq q} |\lambda_j - \lambda_j^0| \quad \to \quad 0 \tag{11}$$

$$\sup_{1\leq l\leq n, j\leq q} |r_{j,l} - \varphi_j(s_l)| \quad \to \quad 0. \tag{12}$$

3.1.2 Large-Sample Behavior of Weighted Averages of $\{w_l u_l, w_l w_l'\}$

We now turn to an appropriate high-level assumption about the behavior of weighted averages of $w_l u_l$ and $w_l w_l'$.

Condition 2. *(a) There exists a function $\Omega_{wu} : \mathcal{S} \mapsto \mathbb{R}^{p\times p}$ and a positive sequence a_n such that for any uniformly convergent sequence of functions $h_n : \mathcal{S} \mapsto \mathbb{R}^p$ with (continuous) limit $h : \mathcal{S} \mapsto \mathbb{R}^p$,*

$$a_n^{-1/2} n^{-1/2} \sum_{l=1}^{n} h_n(s_l)' w_l u_l \Rightarrow \mathcal{N}\left(0, \int h(s)' \Omega_{wu}(s) h(s) dG(s)\right) \tag{13}$$

where $\Omega_{wu}(s)$ is positive semi-definite (p.s.d.) for all $s \in \mathcal{S}$, and $\sup_{s\in\mathcal{S}} ||\Omega_{wu}(s)|| < \infty$;

(b) there exists a p.s.d. matrix-valued function $\Omega_{ww} : \mathcal{S} \mapsto \mathbb{R}^{p\times p}$ with $\sup_{s\in\mathcal{S}} ||\Omega_{ww}(s)|| < \infty$ such that for any continuous $h : \mathcal{S} \mapsto \mathbb{R}^p$,

$$n^{-1} \sum_{l=1}^{n} w_l w_l' h(s_l) \xrightarrow{p} \int \Omega_{ww}(s) h(s) dG(s). \tag{14}$$

Condition 2(a) assumes a central limit theorem to hold for smoothly weighted averages of $w_l u_l$. It also implies joint convergence of any finite number of such weighted averages via the Cramér-Wold device. It is less strong than a functional central limit theorem since there is no (implicit) assumption on stochastic equicontinuity. The presence of a_n accommodates the possibility that the spatial dependence of $w_l u_l$ is strong enough to affect the rate of convergence.

Lahiri's (2003) Theorems 3.1 and 3.2, for example, imply Condition 2(a) for $h_n = h$, and $a_n^{-1/2} n^{-1/2} \sum_{l=1}^{n} (h_n(s_l) - h(s_l))' w_l u_l \xrightarrow{p} 0$ follows from arguments employed in his proof (see Lemma 12 of Müller and Watson (2022) for details). Note that even if $\{w_l u_l\}_{l=1}^n$ is strictly stationary, the asymptotic covariance matrix $\Omega_{wu}(s)$ is not necessarily constant: Lahiri's Theorem 3.2 yields $\Omega_{wu}(s)$ to be a weighted average of the variance and long run variance

of $w_l u_l$, with weights that are proportional to the density g. Thus, $\Omega_{wu}(s)$ is in general non-constant as soon as the (limiting) density of locations g is not uniform. Intuitively, the contribution to the variance from areas with many locations is amplified because the locations are necessarily close and hence more strongly spatially correlated.

Condition 2(b) assumes a law of large numbers for weighted averages of $w_l w_l'$. The presence of $\Omega_{ww}(s)$ in the expression for the limit allows for spatial nonstationarities, such as a different value for $\mathbb{E}[w_l w_l']$ in the east and the west of the sample. One could again apply Lahiri's (2003) results, or invoke the mixing conditions in Jenish and Prucha (2009) to obtain sufficient conditions for the convergence in part (b).

Finally, we impose one of two possible assumptions on the evolution of β_l under the alternative.

Condition 3. *(a)* $\beta_l - \beta = a_n^{1/2} n^{-1/2} b(s_l)$, $l = 1, \ldots, n$ *for some sample size independent continuous function* $b : \mathcal{S} \mapsto \mathbb{R}$*; or*

(b) $\beta_l - \beta = \kappa_n L(s_l)$ *with* $\kappa_n = a_n^{1/2} n^{-1/2} \gamma$ *for some sample size independent* $\gamma \geq 0$, $l = 1, \ldots, n$.

The local alternatives of the null hypothesis of stable coefficients are of order $O_p(a_n^{1/2} n^{-1/2})$, a rate which is related to the rate of convergence of the CLT in Condition 2(a). Intuitively, slower rates of convergence for the CLT degrade the signal-to-noise ratio in the weighted averages Y_n, making it harder to detect any instability, so the local alternative has to be relatively larger. Conditions 3(a) (and 2(b)) can be generalized to allow for a finite number of discontinuities in b to accommodate discrete breaks in β_l across regions in $\mathcal{S}$.

3.1.3 Limiting Distribution of Y_n and ξ_q

Combining these high level assumptions with the identities in (8) and (9) and additional arguments yields the following result.

Theorem 1. *(a) Under Conditions 1-2 and 3(a),* $a_n^{-1/2} Y_n \Rightarrow Y_0 + B$, *where* $Y_0 \sim \mathcal{N}(0, V_0)$, $B_i = \int \bar{\varphi}_i^e(s)' \Omega_{wx}(s) b(s) dG(s)$, $V_{0,i,j} = \int \bar{\varphi}_i^e(s)' \Omega_{wu}(s) \bar{\varphi}_j^e(s) dG(s)$, $\bar{\varphi}_j^e(s) = \varphi_j^e(s) - \left(\int \Omega_{ww}(r) dG(r)\right)^{-1} \int \Omega_{ww}(r) \varphi_j^e(s) dG(r)$, Ω_{wx} *is the first column of* Ω_{ww} *and* $\varphi_j^e(s) = (\varphi_j(s), 0, \ldots, 0)'$, $i, j = 1, \ldots, q$. *Furthermore,*

$$a_n^{-1} \xi_q \Rightarrow (Y_0 + B)' \Lambda_q^0 (Y_0 + B) \tag{15}$$

where $\Lambda_q^0 = \text{diag}(\lambda_1^0, \ldots, \lambda_q^0)$.

(b) Under Conditions 1-2 and 3(b), $a_n^{-1/2} Y_n \Rightarrow Y_0 + \gamma Y_1$ *where* $Y_{1,j} = \int \bar{\varphi}_j^e(s)' \Omega_{wx}(s) L(s) dG(s)$ *and* $Y_1 \sim \mathcal{N}(0, V_1)$ *independent of* Y_0 *with* $V_{1,i,j} = \int\int \bar{\varphi}_i^e(s) \Omega_{wx}(s) \bar{k}_L(s,r) \bar{\varphi}_j^e(r)' \Omega_{wx}(r) dG(s) dG(r)$, $i, j = 1, \ldots, q$. *Furthermore,*

$$a_n^{-1} \xi_q \Rightarrow (Y_0 + \gamma Y_1)' \Lambda_q^0 (Y_0 + \gamma Y_1). \tag{16}$$

The limiting distribution of ξ_q in part (b) is a quadratic form in q jointly normal variables, with a covariance matrix $V_0 + \gamma^2 V_1$ that is a somewhat complicated function of the eigenfunctions φ_j of the covariance kernel of Lévy-Brownian motion, and the asymptotic behavior of weighted averages of $w_l u_l$ and $w_l w_l'$. It is instructive to consider a baseline case where $\Omega_{ww}(s) = \Omega_{wu}(s) = I_p$. From the orthogonality of the eigenfunctions φ_i, we obtain $V_0 = I_q$ and $V_1 = \Lambda_q^0$, so under Condition 3(b)

$$a_n^{-1} \xi_q \Rightarrow Z'(\Lambda_q^0 + \gamma^2 (\Lambda_q^0)^2) Z = \sum_{j=1}^{q} (\lambda_j^0 + (\gamma \lambda_j^0)^2) Z_j^2 \quad \text{with } Z_j \sim iid \mathcal{N}(0,1),$$

a weighted average of independent chi-squared random variables (cf. equation (3.3) of Nyblom (1989)). The presence of Lévy-Brownian motion type-variability of β_l is seen to affect the asymptotic distribution of ξ_q via the squared eigenvalues in this baseline case. Thus, if the eigenvalues decay quickly, very little power will be lost by the truncation at the first q eigenvalues for all moderately large q.

3.2 Estimation of the Parameters Characterizing the Limiting Distribution of Y_n and ξ_q

To test the null hypothesis of coefficient stability based on ξ_q, one needs to estimate Λ_q^0 and V_0, and if we want to conduct inference about κ_n under Condition 3(b), then we also need an estimator for V_1.

Consistent estimation of Λ_q^0 by Λ_q is immediate from (11). Consistent estimation of V_1 is achieved by the straightforward plug-in estimator with elements

$$\hat{V}_{1,i,j} = n^{-2} \sum_{l,\ell=1}^{n} \bar{r}_{i,l}^{e\prime} w_l x_l \bar{k}_n(s_l, s_\ell) \bar{r}_{j,\ell}^{e\prime} w_\ell x_\ell \xrightarrow{p} V_{1,i,j}$$

as shown in Theorem 2 below.

Estimation of V_0 is harder, as it needs to account for potential spatial correlation in $w_l u_l$. We suggest the kernel estimator

$$\hat{V}_{0,i,j} = n^{-1} \sum_{l,\ell=1}^{n} \hat{v}_{l,i} k_c(s_l, s_\ell) \hat{v}_{\ell,j}, \quad \hat{v}_{l,j} = \bar{r}_{j,l}^{e\prime} w_l \hat{e}_l \tag{17}$$

with kernel

$$k_c(s,r) = \exp(-c||s-r||) = \sum_{j=1}^{\infty} \psi_j(s)\psi_j(r). \tag{18}$$

Here ψ_j are the eigenfunctions of k_c on $\mathcal{S}$ satisfying $\int \psi_j(s)\psi_i(s)dG(s) = 0$ for $i \neq j$, scaled by the square root of the eigenvalues. For any fixed c, this kernel estimator puts non-zero weight on all cross products $\hat{v}_{l,j}\hat{v}_{\ell,j}$, even those whose distance $||s_l - s_\ell||$ is a positive fraction of the diameter of the sampling region $\mathcal{S}$. As such, $\hat{V}_0$ is a "fixed-b" HAR estimator in the spirit of Kiefer and Vogelsang (2005) and considered in the spatial context in Bester, Conley, Hansen, and Vogelsang (2016).

Let $\zeta_{k,j,l} = \psi_k(s_l)\bar{r}_{j,l}^e$. Then

$$\sum_{l=1}^{n} \psi_k(s_l)\hat{v}_{l,j} = \sum_{l=1}^{n} \zeta_{k,j,l}' w_l \hat{e}_l = \sum_{l=1}^{n} \bar{\zeta}_{k,j,l}' w_l e_l$$

where in analogy to (9)

$$\bar{\zeta}_{k,j,l} = \zeta_{k,j,l}' - \left(n^{-1} \sum_{\ell=1}^{n} w_\ell w_\ell' \right)^{-1} n^{-1} \sum_{\ell=1}^{n} w_\ell w_\ell' \zeta_{k,j,\ell}.$$

Theorem 2. *(a) Under Conditions 1-2 and 3(b),* $\hat{V}_1 \xrightarrow{p} V_1$.

(b) Under Conditions 1-3, $a_n^{-1}\hat{V}_0 \Rightarrow \tilde{V}_0^c$ *where*

$$\tilde{V}_{0,i,j}^c = \sum_{k=1}^{\infty} \left(\int \bar{\zeta}_{k,i}(s)' dJ(s) \right) \left(\int \bar{\zeta}_{k,j}(s)' dJ(s) \right) \tag{19}$$

with $\bar{\zeta}_{k,j}(s) = \zeta_{k,j}(s) - \left(\int \Omega_{ww}(r)dG(r) \right)^{-1} \int \Omega_{ww}(r)\zeta_{k,j}(r)dG(r)$, $\zeta_{k,j}(s) = \psi_k(s)\bar{\varphi}_j^e(s)$, *and under Condition 3(a),* $dJ(s) = g(s)^{1/2}\Omega_{wu}(s)^{1/2}d\mathbb{W}_p(s) + \Omega_{wx}(s)b(s)g(s)ds$ *whereas under Condition 3(b),* $dJ(s) = g(s)^{1/2}\Omega_{wu}(s)^{1/2}d\mathbb{W}_p(s) + \gamma\Omega_{wx}(s)L(s)g(s)ds$, *with* $\mathbb{W}_p$ *a* $p \times 1$ *stan-*

dard Wiener process on $\mathcal{S}$, $i, j = 1, \ldots, q$.

(c)

$$\tilde{V}_0^c \xrightarrow{p} V_0 \text{ as } c \to \infty. \tag{20}$$

Part (b) of Theorem 2 derives the limiting distribution of $\hat{V}_0$, suitably scaled. For any fixed c, this distribution is non-degenerate, as one would expect from a fixed-b kernel variance estimator. At the same time, part (c) shows that for a large enough c, the limiting distribution becomes arbitrarily close to being a point mass at the target V_0. In other words, for large enough but fixed c, Theorem 2 shows that the estimator $a_n^{-1}\hat{V}_0$ has negligible asymptotic bias and sampling variability.

One could presumably obtain a formally consistent estimator $a_n^{-1}\hat{V}_0$ by letting c slowly increase with n. We do not do so for three reasons: First, the introduction of such an arbitrary slowly diverging sequence does not change the small sample behavior of the resulting $\hat{V}_0$, so acknowledging its variability is arguably a more honest approach to inference. Second, the derivation of consistency under such a slowly diverging sequence $c = c_n$ requires additional regularity conditions beyond Condition 2, while the provably excellent properties of $\hat{V}_0$ under large but fixed c do not. Lastly, choosing c unnecessarily large reduces the robustness of inference against spatial correlation in $\{w_l u_l\}$, as it increases the potential bias in $\hat{V}_0$ that arises from downweighing cross products $\hat{v}_{l,j}\hat{v}_{\ell,j}$ whose distance $||s_l - s_\ell||$ is small but non-negligible.

Remark 3.1. In practice, the estimator $\hat{V}_0$ requires a choice of c for the kernel k_c in (18). We find it useful to parameterize c in terms of the implied average value of $k_c(s_l, s_\ell)$ over distinct locations. Thus, let $c_{\bar{\rho}}$ solve $\bar{\rho} = \frac{1}{n(n-1)} \sum_l \sum_{\ell \neq l} k_c(s_l, s_\ell)$, so that larger values of $\bar{\rho}$ lead to more diffuse kernels and $\bar{\rho} = 0$ yields the kernel appropriate for spatially uncorrelated observations. In the empirical analysis in the next section we consider values of $\bar{\rho}$ between 0 and 0.03. These values produce estimates of $\hat{V}_0$ with a reasonably small amount of sampling variability in the canonical model: the coefficient of variation of trace$(\hat{V}_0)$ ranges from approximately 0.02 to 0.15 for $0 \leq \bar{\rho} \leq 0.03$. For comparision, recall that $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n Z_i^2$, the variance estimator in the classical Gaussian model with $Z_i \sim iid\mathcal{N}(0, 1)$, has a coefficient of variation of $(2/n)^{1/2}$, which takes on values less than 0.15 for $n \geq 85$.

Remark 3.2. The estimator $\hat{V}_0$ can also be obtained from a bootstrap scheme: Conditional on the data, let η be a mean zero stochastic process on $\mathcal{S}$ with covariance kernel

$\mathbb{E}[\eta(r)\eta(s)] = k_c(s,r)$. Let $u_l^* = \hat{e}_l\eta(s_l)$ be the fixed regressor multiplier bootstrap draw of the OLS innovations u_l. The natural bootstrap approximation to the distribution of $Y_{n,j} = n^{-1/2}\sum_{l=1}^n r_{j,l}x_l\hat{e}_l$ under the null hypothesis is given by $Y_j^* = n^{-1/2}\sum_{l=1}^n r_{j,l}x_l\hat{u}_l^*$, where $\hat{u}_l^*$ are the OLS residuals of a linear regression of u_l^* on w_l. A straightforward calculation shows that the resulting bootstrap covariance matrix estimator of $Y^* = (Y_1^*, \ldots, Y_q^*)$ is numerically identical to $\hat{V}_0$. If in addition, the bootstrap multiplier process $\eta(\cdot)$ is Gaussian, then $Y^* \sim \mathcal{N}(0, \hat{V}_0)$ conditional on the data.

Hansen (2000) suggests a closely related fixed regressor multiplier bootstrap approximation for the distribution of tests statistics for the null hypothesis of coefficient stability in time series regressions, albeit with i.i.d. multipliers, as his model has u_l serially independent. Also see Conley, Gonçalves, Kim, and Perron (2023) for an application of the dependent multiplier bootstrap in spatial econometrics in another context.

3.3 Asymptotic Inference About Spatial Variation

3.3.1 Tests of Coefficient Stability

Theorem 1 yields that under the null hypothesis (2),

$$a_n^{-1}\xi_q \Rightarrow Y_0'\Lambda_q^0 Y_0,\ Y_0 \sim \mathcal{N}(0, V_0)$$

and Theorem 2 shows that $a_n^{-1}\hat{V}_0$ is an accurate estimator of V_0 for large enough c. Noting that the (generally unknown) rate a_n^{-1} cancels, we can hence simply compute ξ_q as defined in (5) and compare it to the $1-\alpha$ quantile of $Y_0^{*\prime}\Lambda_q Y_0^*$, where $Y_0^* \sim \mathcal{N}(0, \hat{V}_0)$ conditional on the data. For large enough c, this test will will have an asymptotic null rejection rate arbitrarily close to α under Conditions 1 and 2.

3.3.2 Confidence Interval and Median Unbiased Estimator of the Magnitude of Instability

Now consider the problem of constructing a confidence interval for the scale of the Lévy-Brownian motion-type instability $\kappa_n = n^{-1/2}a_n^{1/2}\gamma$ under Conditions 1-2 and 3(b). For a given value of $\kappa_n = \kappa_n^0$, Theorems 1 and 2 show that the quantiles of ξ_q are well approximated by the quantiles of

$$(Y_0^* + n^{1/2}\kappa_n^0 Y_1^*)'\Lambda_q(Y_0^* + n^{1/2}\kappa_n^0 Y_1^*) \tag{21}$$

with $Y_1^* \sim \mathcal{N}(0, \hat{V}_1)$ independent of $Y_0^* \sim \mathcal{N}(0, \hat{V}_0)$ conditional on the data, at least as long as κ_n^0 is not too large and c is large enough. Denote the quantile function of the distribution of (21) by $\hat{q}(\cdot|\kappa_n^0)$. A level $1-\alpha$ interval estimator for κ_n, $\alpha < 1/2$, is given by the interval $[\hat{\kappa}_n^L, \hat{\kappa}_n^U]$ where κ_n^L solves $\hat{q}(1-\alpha/2|\hat{\kappa}_n^L) = \xi_q$ and $\hat{q}(\alpha/2|\hat{\kappa}_n^U) = \xi_q$. The asymptotic coverage of this interval will become arbitrarily close to $1-\alpha$ if Conditions 1-2 and 3(b) hold, and c is sufficiently large.

One might alternatively seek a point estimator of κ_n that is approximately median unbiased. Such an estimator $\hat{\kappa}_n^{MU}$ is simply obtained by solving $\hat{q}(0.5|\hat{\kappa}_n^{MU}) = \xi_q$. Again, if Conditions 1-2 and 3 (b) hold then for large enough c, $\mathbb{P}(\hat{\kappa}_n^{MU} > \kappa_n)$ becomes arbitrarily close to 0.5 as $n \to \infty$. See Stock and Watson (1998) for analogous calculations in unstable time series regressions.

3.3.3 Estimates of the Parameter Path

Finally, one might be interested in estimating the value of $\beta_l = \beta + \kappa_n L(s_l)$ for $l = 1, \ldots, n$ directly. From Theorem 1(b), we have $a_n^{-1/2} Y_n \Rightarrow Y = Y_0 + \gamma Y_1$ where $Y_0 \sim \mathcal{N}(0, V_0)$ is independent of Y_1 and $Y_{1,i} = \int \bar{\varphi}_i^e(s)' \Omega_{wx}(s) L(s) dG(s)$. Inspection of the proof of Theorem 1(b) shows that this convergence holds jointly with the trivial convergence $L \Rightarrow L$, that is, the statistical dependence between Y and L approximates the small sample relationship between Y_n and the spatial variation in β_l.

Note that $\mathbb{E}[L(s)Y] = \gamma \sigma_{LY}(s)$ for $s \in \mathcal{S}$, where $\sigma_{LY}(s)$ has ith element $\int \bar{\varphi}_i^e(r)' \Omega_{wx}(r) k_L(s,r) dG(r)$. By the usual formula for the conditional normal distribution, we hence obtain $\mathbb{E}[L(s)|Y] = \gamma \sigma_{LY}(s)'(V_0 + \gamma^2 V_1)^{-1} Y$. The small sample version is readily obtained by replacing all quantities by sample analogues

$$\hat{\beta}_l = \hat{\beta} + n^{1/2} \kappa_n^2 \hat{\sigma}_{LY}(s_l)'(\hat{V}_0 + n\kappa_n^2 \hat{V}_1)^{-1} Y_n$$

where the ith element of $\hat{\sigma}_{LY}(s)$ is given by $n^{-1} \sum_{l=1}^n \bar{r}_{i,l}^{e\prime} w_l x_l k_L(s_l, s)$.

Implementing this estimator requires choosing a value for κ_n, and a simple approach relies on $\hat{\kappa}_n^{MU}$. Alternatively, one could average the estimators $\hat{\beta}_l$ over the posterior distribution for κ_n using some prior and the likelihood based on the asymptotic normality of Y_n, which is proportional to $\det(\hat{V}_0 + n\kappa_n^2 \hat{V}_1)^{-1/2} \exp[-\frac{1}{2} Y_n'(\hat{V}_0 + n\kappa_n^2 \hat{V}_1)^{-1} Y_n]$.

4 Monte Carlo Results and Empirical Analysis

Section 1 introduced data on health insurance coverage (HIC) and private company employment (PCE) across zip codes in the continental United States. In this section we use these variables together with sixty other socioeconomic variables to investigate two questions: How well do the methods developed in Sections 2 and 3 perform in realistic environments calibrated to these data? And how stable or unstable are the bivariate relationships between these socioeconomic variables in the U.S.?

We begin with a description of the data.

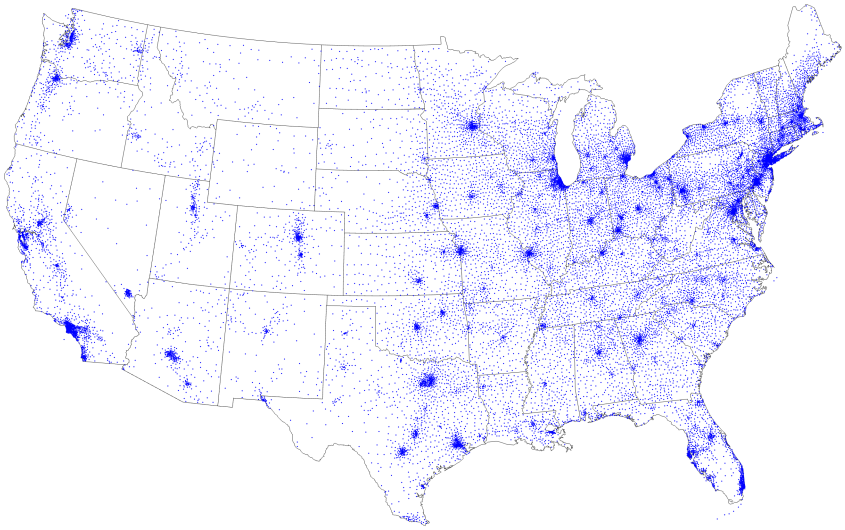
4.1 Data Description

All data are from the American Community Survey, 5-year estimates from 2018-2022, for the zip codes regions ("zcta") making up the contiguous 48 states and the District of Columbia. The dataset contains sixty-two variables measuring population, educational attainment, income, employment, race, citizenship, health, marital status, mobility, and a handful of other indicators. The online appendix provides a detailed description of the variables. The underlying dataset is a balanced panel of roughly thirty thousand zip codes. Zip codes containing a small number of observations (generally 250 or fewer) were merged with adjacent zip codes, resulting in a balanced panel of $n = 21,194$ regions. The (approximate) center of each region was used as its location, s_l, and distances between regions are measured by the great circle formula. For simplicity, we continue to refer to these regions as zip codes. Figure 3 plots their locations.

The raw data were transformed in three ways. First, in most cases, the variables were scaled by the relevant population in the region. For example, HIC was measured as the fraction of the zip code's population with health insurance, and PCE was measured as the fraction of workers over the age of 16 employed at private companies. Second, each variable was then converted to a percentile over the universe of zip codes. For example, a value of 0.75 for a variable in a given zip code indicates that this zip code ranks in the 75th percentile for this variable. This means that a regression coefficient of, say $\beta = 0.10$, implies that the percentile rank of the y-variable is predicted to increase by 1 percentage point when the rank of the x-variable increases by 10 percentage points.

The resulting variables typically exhibited strong spatial persistence as indicated by the

Figure 3: Zip code locations in the dataset



Notes: Each dot shows the approximated center of one of the 21,194 zip codes in the dataset.

spatial unit roots test from Müller and Watson (2023). For example, for the HIC and PCE variables plotted in panel (a) of Figure 1 the spatial unit root tests had p-values of 0.50 and 0.25, respectively; the unit root null hypothesis was not rejected at the 5% level for 45 of the remaining 60 variables. As noted in the introduction, strong spatial persistence can lead to spurious regressions, and to mitigate this concern all of the variables were transformed using the same spatial differencing GLS transformation used for HIC and PCE in panel (b) of Figure 1.

By selecting pairs of variables from the dataset we could obtain many bivariate regressions: with 62 variables, it is possible to construct 3,782 bivariate regressions. But many of these are uninteresting — half merely interchange the y and x variables, and several others involve variables that are closely related by construction (e.g., the fraction of married adults and the fraction of divorced adults). As described in the online appendix, we used simple rules to eliminate pairs of closely-related variables, and for the remaining pairs random assignment was used to determine which variable was y and which was x. This process resulted in 1,514 bivariate regressions that are used in the exercises reported below.

4.2 Size, Power and Choice of Variance Estimator

The first set of exercises investigate the size and power properties of the ξ_q test. It focuses on three questions. First, how does spatial correlation and spatial heteroskedasticity affect size and power? Second, how does the choice of kernel used to construct the variance estimator $\hat{V}_0$ from (17) affect size and power? Third, does the test have power to detect "discrete" spatial-breaks in the coefficients in addition to the Lévy-Brownian motion variation underlying the design of the test? All experiments involve bivariate regressions of the form $y_l = \alpha + x_l\beta_l + u_l$ where the locations s_l indexing the observations are given by the zip-code locations shown in Figure 3.

4.2.1 Monte Carlo Design

The experiments involve three ingredients: (i) the kernel used to compute $\hat{V}_0$, (ii) the process generating the regressors and error $\{x_l, u_l\}$, and (iii) the process generating the coefficients $\{\beta_l\}$. We discuss these in turn.

As described in Remark 3.1, $\hat{V}_0$ is computed using the exponential kernel $k_c(r,s) = \exp(-c||r-s||)$, with $c = c_{\overline{\rho}}$ parameterized in terms of $\overline{\rho}$, the implied average value of $k_c(r,s)$ between locations. Our experiments use kernels corresponding to the three values $\bar{\rho} \in \{0, 0.015, 0.03\}$. The kernels are denoted k_0, $k_{0.015}$ and $k_{0.03}$.

The data generating processes (DGP) for $\{x_l, u_l\}$ also use the covariance kernel k_c, but in a different way: they involve the zero-mean Gaussian processes η with covariance kernel $\mathbb{E}[\eta(r)\eta(s)] = k_c(r,s)$, denoted $\eta \sim \mathcal{G}_c$. For this process, smaller values of c correspond to larger spatial correlation. In our dataset, locations are points in the continental U.S., and it is natural to measure spatial correlation in terms of its half-life, that is the distance that yields a spatial correlation of 1/2. Our DGPs for $\{x_l, u_l\}$ will utilize $\mathcal{G}_c$ processes with c that induces half-lives of 0, 10, 25, 50, and 100 kilometers.

With this background the three data generating processes for $\{x_l, u_l\}$ are:

- DGP1: x_l and u_l are generated by independent $\mathcal{G}_c$ processes.
- DGP2: x_l is randomly selected from the 62 standardized variables in our dataset and u_l follows a $\mathcal{G}_c$ process.
- DGP3: $\{y_l^o, x_l^o\}$ are a pair of series from the list of 1,514 bivariate regressions in our

dataset. Then $x_l = x_l^o \eta_{x,l}$ and $u_l = y_l^o \eta_{u,l}$ where $\eta_{x,l}$ and $\eta_{u,l}$ follow independent $\mathcal{G}_c$ processes.

In DGP1, the regressors and errors are spatially correlated (as indexed by the half-life used to determine c), but spatially homoskedastic. In DGP2, x_l are actual data (with real-world spatial correlation and heteroskedasticity), while u_l is spatially correlated but homoskedastic. In DGP3, both x_l and u_l inherit the actual data's spatial heteroskedasticity, but the data's spatial correlation is muted by multiplication with the η variables.

The final ingredient is the data generating process for β_l. We consider three. The first is the null model with $\beta_l = \beta$ (and where invariance makes the value of β irrelevant). The second generates β_l from the scaled Lévy-Brownian motion process with $\beta_l \sim \kappa L(s_l)$, where the parameter κ indexes the size of spatial variation in β_l. In our application it is convenient to measure this size in terms of the standard deviation of changes in β_l over distances of 1000 kilometers, that is $\sigma_{\Delta 1000\text{km}} = \text{std}(\beta(s) - \beta(r))$ for $||s - r|| = 1000$km. We chose four values of κ corresponding to $\sigma_{\Delta 1000km} \in \{0.01, 0.025, 0.05, 0.10\}$. The third model allows β_l to take on different values in different regions of the country. We define regions in three ways: (i) by the 48 states plus the District of Columbia, (ii) by the nine Census regions, and (iii) by the eastern and western regions of the U.S.[1] In these experiments the value of β_l is $iid\mathcal{N}(0, \sigma_\beta^2)$ across regions. In the experiments with spatially varying coefficients, the $\{x_l, u_l\}$ are generated by DGP-3 with $\{\eta_{x,l}, \eta_{u,l}\}$ independent $iid\mathcal{N}(0, 1)$ random variables.

4.2.2 Results

Table 1 summarizes the results for size and for power against Lévy-Brownian motion variation in the coefficients. The results show the standard trade-off between size control and power: In the models with no spatial correlation (half-life equal to zero), the k_0 kernel yields a size of 5% for each DGP and better power than the $k_{0.015}$ and $k_{0.03}$ kernels, but size control with the k_0 kernel deteriorates precipitously as spatial correlation increases (see the results DGP-1 in panel (a)). In contrast, $k_{0.015}$ controls size well for moderate spatial correlation (half-life less than 25 kilometers) and $k_{0.03}$ for somewhat more severe spatial correlation (but at the cost of further reductions in power).

[1]For state-wide regions, the number of zip-code observations ranges from 18 (in the District of Columbia) to 1,381 (in Texas); for Census regions, the number of observations range from 1,273 to 3,649; fo the east-west regions there are 8,192 zip codes in the west and 13,002 in the east.

Table 1: Size and power of nominal 5% tests

(a) Size of nominal 5% tests

Half-life	DGP1			DGP2			DGP3		
(in KM)	k_0	$k_{.015}$	$k_{.03}$	k_0	$k_{.015}$	$k_{.03}$	k_0	$k_{.015}$	$k_{.03}$
0	0.05	0.04	0.04	0.05	0.04	0.04	0.05	0.04	0.03
10	0.75	0.04	0.03	0.00	0.03	0.03	0.05	0.04	0.03
25	0.99	0.07	0.04	0.00	0.02	0.02	0.05	0.04	0.03
50	1.0	0.12	0.07	0.00	0.01	0.01	0.05	0.03	0.03
100	1.0	0.26	0.15	0.00	0.00	0.00	0.05	0.03	0.03

(b) Power of nominal 5% tests

$\sigma_{\Delta 1000km}$	DGP1			DGP2			DGP3		
	k_0	$k_{.015}$	$k_{.03}$	k_0	$k_{.015}$	$k_{.03}$	k_0	$k_{.015}$	$k_{.03}$
0.01	0.13	0.11	0.10	0.13	0.10	0.09	0.11	0.09	0.07
0.025	0.45	0.40	0.37	0.46	0.39	0.35	0.34	0.29	0.26
0.05	0.83	0.76	0.72	0.80	0.76	0.65	0.65	0.59	0.54
0.10	0.97	0.93	0.89	0.97	0.93	0.85	0.90	0.84	0.79

Notes: See the text for description of the DGPs. k_0, $k_{0.015}$ and $k_{0.03}$ denote kernels with c chosen to yield $\bar{\rho} \in \{0, 0.015, 0.03\}$. The rejection frequency ("power") shown in panel (b) are note size-adjusted and correspond to sizes given in panel (a) for DGP3 and half-life equal to zero.

Perhaps most intriguingly, Table 1 shows that the tests are under-sized for DGP-2, most notably for the k_0 kernel. This reflects negative spatial correlation in the GLS-transformed variables that are used as the regressors in this experiment. (The negative correlation is inherited by the cross-products $x_l u_l$ when u_l is positively spatially correlated.) One potential explanation for the negative spatial correlation in x is sampling error in the zip-code level American Community Survey data. Classical sampling error adds spatially uncorrelated noise to the *levels* of the variables, which induces negative spatial correlations in the transformed data, analogous to over-differencing in time series. One takeaway from these results is that, for the GLS-transformed data that we will use in the regressions reported below, substantial positive spatial correlation does not seem to be present. With this in mind, we use the $k_{0.015}$ kernel in the empirical analysis, both when estimating V_0 and in the construction of t-tests for regression coefficients.

We do not present detailed results for the model with discrete shifts in β, but the experiments indicate that the test has power to detect these discrete shifts, albeit more so for the East-West and Census regions than for individual States. For example, with σ_β denoting the

Table 2: Summary of Results from Bivariate Spatial Regressions

		Quantile				
		0.05	0.25	0.50	0.75	0.95
		(a) OLS estimates				
$\lvert t_{\hat{\beta}} \rvert$		0.63	3.75	8.28	14.60	29.36
$\lvert \hat{\beta} \rvert$		0.01	0.05	0.11	0.22	0.45
		(b) Spatial variation in β				
ξ_{15} p-value		<0.00	0.02	0.07	0.20	0.52
$\sigma_{\Delta^{1000km}}$ $(\hat{\kappa}^{MU})$		0.00	0.03	0.05	0.09	0.18

Notes: The table shows selected quantiles for results from 1,514 regressions.

standard deviation of β_l across regions, the test has 50% power when $\sigma_\beta \in \{0.05, 0.06, 0.22\}$ for the {East-West, Census, State} designs. As expected, everything else being equal, the more spatially persistent the evolution of β_l under the alternative, the more powerful the test.

4.3 Spatial Instability in Bivariate Socioeconomic Regressions in the U.S.

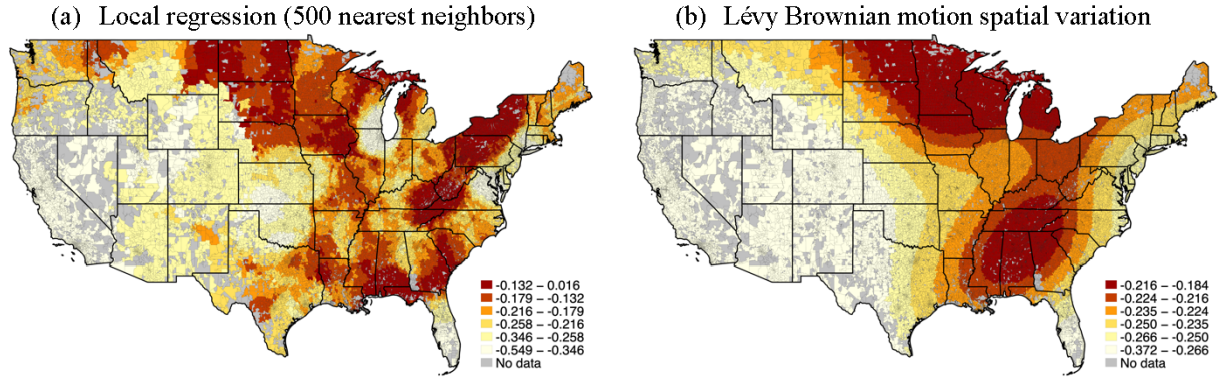
We now turn to the empirical analysis of the 1,514 bivariate spatial regressions involving the 62 socioeconomic variables in the dataset. For each of the regressions we computed four statistics: (i) $\hat{\beta}$, the OLS estimate of β, (ii) $t_{\hat{\beta}} = \hat{\beta}/se(\hat{\beta})$, the associated t-statistic, (iii) ξ_{15}, the SVP test statistic using $q = 15$, and (iv) $\hat{\kappa}^{MU}$, the median unbiased estimate of κ. The results are summarized in Table 2.

The first two rows of the table show the unsurprising fact that most socioeconomic variables are closely related: the absolute value of the t-statistic exceeds 2 in nearly 90 percent of the regressions, the median absolute t-statistic is exceeds 8, and many values of $\hat{\beta}$ are large in a real-world sense with a median value of 0.11.

The results also suggest substantial instability in the regression coefficients: the null hypothesis of spatial stability is rejected at the 5% level for over 40 percent of the regressions and at the 10% level for nearly 60 percent. The median unbiased estimates of κ suggest that the instability can be large: converted to values of the standard deviation of changes in β_l over 1000km (that is $\sigma_{\Delta 1000\text{km}}$), more than half the regressions have estimates of $\sigma_{\Delta 1000\text{km}}$ that exceed 0.05.

We end this section with one final set of calculations: estimates of β_l over the U.S. for

Figure 4: Estimates of β in the HIC-PCE regression



Notes: Panel (a) shows results using local regressions with 500 nearest neighbors. Panel (b) shows the sample paths estimated using Gaussion regression formulae in Section 3.3.3.

the HIC-PEC regression introduced in Figure 1. Figure 4 presents two sets of estimates. The first, shown in panel (a), uses local regressions centered at each of 21,194 zip codes and its 500 nearest neighbors. This is an example of a "geographically weighted regression" often used in applied work (Fotheringham, Brunsdon, and Charlton (2002)), and suggests an interesting pattern of spatial variation, with HIC and PEC strongly negatively correlated in the west, but much less so in the upper Midwest, the Ohio valley and portions of the south. The Lévy-Brownian motion model imposes smoothness on the evolution of β_l and panel (b) shows estimates that impose this Lévy-Brownian motion prior and estimates β_l using the weighted averages Y_n, as described in Section 3.3.3, setting κ equal to the median unbiased estimator. It indicates a similar geographic pattern for the coefficients, although with less variation reflecting, in part, the smaller sampling error inherent in optimal signal extraction methods.

5 Concluding Remarks

This paper has proposed a test for coefficient stability in spatial regressions. The test is straightforward to construct and allows for spatial correlation in both regressors and regression errors, as well as for second-moment nonstationarities. Simulation experiments suggest that the test performs well in realistic empirical settings. In an application to bivariate socioeconomic regression across U.S. zip codes, we find evidence of substantial spatial variation

in the coefficients.

We conclude by considering the question: Why should an empirical researcher test for coefficient variation in their spatial regression? One answer is that spatial variation might be a direct research question of interest. But beyond that obvious answer, the researcher should be concerned about the interpretation of the OLS estimand in a constant-coefficient regression. With spatial variation, the constant-coefficient estimand measures a spatial average population regression coefficient. This average does not correspond to the best linear predictor conditional on a location. And, if the regression coefficent is meant to measure the causal effect of a policy intervention, spatial heterogeneity might well affect the policy's desirability. Finally, instability of the coefficients within a sampled region makes extrapolation of the "average" effect to other regions problematic, undermining external validity. To our mind, these considerations support the routine use of coefficient stability tests in spatial regressions, and we hope that the test proposed here is useful in that regard.

A Proofs

Proof of Theorem 1:
For future reference, note that for any sequence of functions $h_n : \mathcal{S} \mapsto \mathbb{R}$ that converges uniformly to a continuous function h, $\sup_{s\in\mathcal{S}} |h_n(s) - h(s)| \to 0$, we have $||n^{-1}\sum_{l=1}^n w_l w_l'(h_n(s_l) - h(s_l))|| \leq \sup_{s\in\mathcal{S}} ||h_n(s) - h(s)|| \cdot n^{-1}\sum_{l=1}^n w_l w_l' \xrightarrow{p} 0$ via (14), so that

$$n^{-1}\sum_{l=1}^n w_l w_l' h_n(s_l) \xrightarrow{p} \int_{\mathcal{S}} \Omega_{ww}(s)h(s)dG(s). \tag{22}$$

(a) For arbitrary $v_1, \ldots, v_q \in \mathbb{R}$, we have $a_n^{-1/2}\sum_{j=1}^q v_j Y_{n,j} = a_n^{-1/2}n^{-1/2}\sum_{l=1}^n r_{v,l}' w_l e_l$ with $r_{v,l} = \sum_{j=1}^q v_j \bar{r}_{j,l}^e$. Given (12) and Condition 2(b), it follows that $\sup_l |r_{v,l} - \varphi_v(s_l)| \xrightarrow{p} 0$, where $\varphi_v(s) = \sum_{j=1}^q v_j \bar{\varphi}_j^e(s)$. We thus have

$$\begin{aligned} a_n^{-1/2}\sum_{j=1}^q v_j Y_{n,j} &= a_n^{-1/2}n^{-1/2}\sum_{l=1}^n r_{v,l}' w_l u_l + n^{-1}\sum_{l=1}^n r_{v,l}' w_l x_l b(s_l) \\ &\Rightarrow \mathcal{N}\left(\int \varphi_v(s)'\Omega_{wx}(s)b(s)dG(s), \int \varphi_v(s)'\Omega_{wu}(s)\varphi_v(s)dG(s)\right) \\ &\sim \sum_{j=1}^q v_j(B_j + Y_{0,j}) \end{aligned}$$

where the convergence follows from applying Condition 2(a) and (22). The convergence $a_n^{-1/2}Y_n \Rightarrow Y_0 + B$ now follows from the Cramér-Wold device, and (15) is a consequence of the continuous mapping theorem.

(b) Let $Q_n(L) = n^{-1}\sum_{l=1} r_{v,l}' w_l x_l L(s_l)$, where $r_{v,l}$ is defined in the proof of part (a). We now show that $Q_n(L) \xrightarrow{p} Q_0(L) = \int \varphi_v(s)'\Omega_{wx}(s)L(s)dG(s)$. The argument is as follows: because Lévy-Brownian motion has almost surely continuous sample paths, L can be viewed without loss of generality as a random element that takes values in the space of continuous functions on $\mathcal{S}$, equipped with the sup norm. Thus, for every $\varepsilon > 0$, there exists a compact set $\mathcal{C}_\varepsilon$ of continuous functions $\mathcal{S} \mapsto \mathbb{R}$ such that $\mathbb{P}(L \in \mathcal{C}_\varepsilon) > 1 - \varepsilon$. On this set $\mathcal{C}_\varepsilon$, the convergence in probability of $Q_n(h) = n^{-1}\sum_{l=1}^n r_{v,l}' w_l x_l h(s_l)$ to $Q_0(h) = \int \varphi_v(s)'\Omega_{wx}(s)h(s)dG(s)$ is uniform, that is, for all $\varepsilon_0 > 0$, $\sup_{h\in\mathcal{C}_\varepsilon} \mathbb{P}(|Q_n(h) - Q_0(h)| > \varepsilon_0) \to 0$. (Suppose otherwise. Then there exists a $\varepsilon_1 > 0$ and a subsequence of functions $h_{n'} \in \mathcal{C}_\varepsilon$ such that along that subsequence, $\mathbb{P}(|Q_{n'}(h_{n'}) - Q_0(h_{n'})| > \varepsilon_0) > \varepsilon_1$. Since $\mathcal{C}_\varepsilon$ is compact, this subsequence $h_{n'}$ has a further subsequence $h_{n''}$ that converges, and along that subsequence, $\mathbb{P}(|Q_{n''}(h_{n''}) - Q_0(h_{n''})| > \varepsilon_0) \to 0$ by virtue of (22), a contradiction.) We conclude that $\mathbf{1}[L \in \mathcal{C}_m]Q_n(L) \xrightarrow{p} \mathbf{1}[L \in \mathcal{C}_m]Q_0(L)$, and since $\mathbb{P}(L \in \mathcal{C}_\varepsilon) > 1 - \varepsilon$ for arbitrary $\varepsilon > 0$, also $Q_n(L) \xrightarrow{p} Q_0(L)$. □

Proof of Theorem 2:

(a) In the notation of the proof of Theorem 1, let $\omega_{v,n}^2 = n^{-2}\sum_{l,\ell=1}^n r'_{v,l} w_l x_l \bar{k}_n(s_l, s_\ell) r'_{v,\ell} w_\ell x_\ell$, where $\bar{k}_n(s_l, s_\ell) = \sum_{j=1}^n \lambda_j r_{j,l} r_{j,\ell}$ is the l,ℓth element of $\bar{\Sigma}_L$. It suffices to show that $\omega_{v,n}^2 \xrightarrow{p} \omega_v^2 = \int\int \varphi_v(s)'\Omega_{wx}(s)\bar{k}_L(s,r)\varphi_v(r)\Omega_{wx}(r)dG(s)dG(r)$.

Let $m > p$ not depend on n. Recall that for an arbitrary p.s.d. matrix Σ with largest eigenvalue $\bar{\lambda}_\Sigma$ and arbitrary conformable vector v, $v'\Sigma v \leq \bar{\lambda}_\Sigma ||v||^2$. Thus $|\omega_{v,n}^2 - \sum_{j=1}^m \lambda_j \left(n^{-1}\sum_{l=1}^n r'_{v,l} w_l x_l r_{j,l}\right)^2| \leq \lambda_{m+1}\left(n^{-1}\sum_{l=1}^n r'_{v,l} w_l x_l\right)^2$. Using (12) and (22), $n^{-1}\sum_{l=1}^n r'_{v,l} w_l x_l \xrightarrow{p} \int \varphi_v(s)\Omega_{wx}(s)dG(s)$, and from (11), $\lambda_{m+1} \to \lambda_{m+1}^0$. Thus, for every $\varepsilon > 0$, we can choose m large enough so that $\mathbb{P}(\lambda_{m+1}\left(n^{-1}\sum_{l=1}^n r'_{v,l} w_l x_l\right)^2 > \varepsilon) \to 0$. Furthermore, applying (11), (12) and (22), we have $n^{-1}\sum_{l=1}^n r'_{v,l} w_l x_l r_{j,l} \xrightarrow{p} \int \varphi_v(s)\Omega_{wx}(s)\varphi_j(s)dG(s)$, so that

$$\sum_{j=1}^m \lambda_j \left(n^{-1}\sum_{l=1}^n r'_{v,l} w_l x_l r_{j,l}\right)^2 \xrightarrow{p} \omega_{v,m}^2$$
$$= \int\int \varphi_v(s)'\Omega_{wx}(s)\left(\sum_{j=1}^m \lambda_j^0 \varphi_j(s)\varphi_j(r)\right)\varphi_v(r)\Omega_{wx}(r)dG(s)dG(r).$$

From (10), also $\lim_{m\to\infty}|\omega_v^2 - \omega_{v,m}^2| = 0$. Since m was arbitrary, $\omega_{v,n}^2 \xrightarrow{p} \omega_v^2$ follows.

(b) We consider first the limit under Condition 3(a). Let $\eta(\cdot)$ be a mean-zero Gaussian process on $\mathcal{S}$ with covariance kernel $k_c(s,r)$, which is continuous almost surely. By Theorem 3.1.2 of Adler and Taylor (2007), we can set $\eta(s) = \sum_{k=1}^\infty \psi_k(s)Z_k$, $Z_k \sim iid\mathcal{N}(0,1)$ where the infinite sum converges uniformly almost surely. Thus, for any $\varepsilon > 0$, there exists a finite m such that with $\eta_m(s) = \sum_{k=m+1}^\infty \psi_k(s)Z_k$, there is a compact set $\mathcal{C}_m$ of continuous functions with respect to the sup norm such that $\mathbb{P}(\eta_m \in \mathcal{C}_m) > 1-\varepsilon$ and on $\mathcal{C}_m$, $\sup_{s\in\mathcal{S}}|\eta_m(s)| \leq \varepsilon$. We have $\mathbb{E}[\eta_m(s)\eta_m(r)] = \sum_{k=m+1}^\infty \psi_k(s)\psi_k(r)$ so that $k_c(s,r) = \sum_{k=1}^m \psi_k(s)\psi_k(r) + \mathbb{E}[\eta_m(s)\eta_m(r)]$, and

$$a_n^{-1}n^{-1}\sum_{l,\ell}^n \hat{v}_{l,i}\left(\sum_{k=1}^m \psi_k(s_l)\psi_k(s_\ell)\right)\hat{v}_{\ell,j} \Rightarrow \sum_{k=1}^m \left(\int \bar{\zeta}_{k,i}(s)'dJ(s)\right)\left(\int \bar{\zeta}_{k,j}(s)'dJ(s)\right) \qquad (23)$$

by the same arguments employed in the proof of Theorem 1. Since $\mathcal{C}_m$ is compact, the convergence $a_n^{-1/2}n^{-1/2}\sum_l h(s_l)\hat{v}_{l,j} \Rightarrow \mathcal{N}(0, \int h(s)'\Omega_{wu}(s)h(s)dG(s))$ holds uniformly over $h \in \mathcal{C}_m$ by the same reasoning as employed in the proof of Theorem 1(b). Furthermore, by construction of $\mathcal{C}_m$, for all $h \in \mathcal{C}_m$, $\int h(s)'\Omega_{wu}(s)h(s)ds \leq \varepsilon^2 \int \Omega_{wu}(s)dG(s)$, so that $\mathbb{P}((\mathbf{1}[\eta_m \in \mathcal{C}_m]a_n^{-1/2}n^{-1/2}\sum_l \eta_m(s_l)\hat{v}_{l,j})^2 >$

$2\varepsilon^2 \int \Omega_{wu}(s)dG(s)) \to 0$. We conclude that

$$\begin{aligned} & \limsup_{n\to\infty} \mathbb{P}\left(a_n^{-1} n^{-1} \sum_{l,\ell}^{n} \hat{v}_{l,j} \left(\sum_{k=m+1}^{\infty} \psi_k(s_l)\psi_k(s_\ell) \right) \hat{v}_{\ell,j} > 2\varepsilon^2 \int \Omega_{wu}(s)dG(s) \right) \\ = & \limsup_{n\to\infty} \mathbb{P}\left(\mathbb{E}\left[a_n^{-1} n^{-1} \sum_{l,\ell}^{n} \hat{v}_{l,j}\eta_m(s_l)\eta_m(s_\ell)\hat{v}_{\ell,j} | \{\hat{v}_{l,j}\}_{l=1}^{n} \right] > 2\varepsilon^2 \int \Omega_{wu}(s)dG(s) \right) \leq 2\varepsilon \end{aligned} \quad (24)$$

Furthermore, note that

$$\int \bar{\zeta}_{k,i}(s)' dJ(s) = \int \psi_k(s) d\tilde{J}_i(s) \quad (25)$$

where

$$\begin{aligned} d\tilde{J}_i(s) &= \bar{\varphi}_i^e(s)' dJ(s) - [\bar{\varphi}_i^e(s)'\Omega_{ww}(s)g(s) \left(\int \Omega_{ww}(r)dG(r) \right)^{-1} \int J(r)dr]ds \\ &= g(s)^{1/2}\bar{\varphi}_i^e(s)'\Omega_{wu}(s)^{1/2} d\mathbb{W}_p(s) + \tilde{B}_i(s)dG(s) \end{aligned}$$

with

$$\begin{aligned} \tilde{B}_i(s) &= \bar{\varphi}_i^e(s)'\Omega_{wx}(s)b(s) \\ &\quad -\bar{\varphi}_i^e(s)'\Omega_{ww}(s) \left(\int \Omega_{ww}(r)dG(r) \right)^{-1} \left(\int g(r)^{1/2}\Omega_{wu}(r)^{1/2} d\mathbb{W}_p(r) + \int \Omega_{wx}(r)b(r)dG(r) \right) \end{aligned}$$

a scalar Gaussian random process on $\mathcal{S}$ with almost surely bounded sample paths.

By the Cauchy-Schwarz inequality, $(\tilde{V}_{0,i,j}^c)^2 \leq \tilde{V}_{0,i,i}^c \tilde{V}_{0,j,j}^c$. From $(a+b) \leq 2a^2 + 2b^2$ and (25) , we have

$$\begin{aligned} \sum_{k=m+1}^{\infty} \left(\int \bar{\zeta}_{k,i}(s)' dJ(s) \right)^2 \leq & 2 \sum_{k=m+1}^{\infty} \left(\int \psi_k(s) g(s)^{1/2} \bar{\varphi}_i^e(s)' \Omega_{wu}(s)^{1/2} d\mathbb{W}(s) \right)^2 \\ & + 2 \sum_{k=m+1}^{\infty} \left(\int \psi_k(s) \tilde{B}_i(s) dG(s) \right)^2 \end{aligned}$$

and with $k_c^m(r,s) = \sum_{k=m+1}^{\infty} \psi_k(s)\psi_k(r)$

$$\begin{aligned} \mathbb{E}\left[\sum_{k=m+1}^{\infty} \left(\int \psi_k(s) g(s)^{1/2} \bar{\varphi}_i^e(s)' \Omega_{wu}(s)^{1/2} d\mathbb{W}(s) \right)^2 \right] &= \int k_c^m(s,s) \bar{\varphi}_i^e(s)' \Omega_{wu}(s) \bar{\varphi}_i^e(s) dG(s) \\ \sum_{k=m+1}^{\infty} \left(\int \psi_k(s) \tilde{B}_i(s) dG(s) \right)^2 &= \int\int \tilde{B}_i(r) k_c^m(r,s) B_i(s) dG(r) dG(s). \end{aligned}$$

As $m \to \infty$, $\sup_{s,r\in\mathcal{S}} |k_c^m(r,s)| \to 0$ by Mercer's Theorem, so $\sum_{k=m+1}^{\infty} \left(\int \bar{\zeta}_{k,i}(s)' dJ(s)\right)\left(\int \bar{\zeta}_{k,j}(s)' dJ(s)\right) \xrightarrow{p} 0$ as $m \to \infty$. Combining this result with (23) and (24) yields (19).

The result under Condition 3(b) follows analogously, using the same reasoning as employed in the proof of Theorem 1(b).

(c) Again, consider the limit under Condition 3(a). Let

$$\begin{aligned}
V_{i,j}^* &= \sum_{k=1}^{\infty} \left(\int \psi_k(s) g(s)^{1/2} \bar{\varphi}_i^e(s)' \Omega_{wu}(s)^{1/2} d\mathbb{W}(s)\right)\left(\int \psi_k(s) g(s)^{1/2} \bar{\varphi}_j^e(s)' \Omega_{wu}(s)^{1/2} d\mathbb{W}(s)\right), \\
R_{i,j}^{BB} &= \sum_{k=1}^{\infty} \left(\int \psi_k(s) \tilde{B}_i(s) dG(s)\right)\left(\int \psi_k(s) \tilde{B}_j(s) dG(s)\right), \\
R_{i,j}^{BW} &= \sum_{k=1}^{\infty} \left(\int \psi_k(s) \tilde{B}_i(s) ds\right)\left(\int \psi_k(s) g(s)^{1/2} \bar{\varphi}_j^e(s)' \Omega_{wu}(s)^{1/2} d\mathbb{W}(s)\right).
\end{aligned}$$

Then

$$\tilde{V}_{0,i,j}^c = V_{i,j}^* + R_{i,j}^{BB} + 2R_{i,j}^{BW}.$$

By the Cauchy-Schwarz inequality, $(R_{i,j}^{BB})^2 \leq R_{i,i}^{BB} R_{j,j}^{BB}$ and

$$\begin{aligned}
R_{i,i}^{BB} &= \int\int \tilde{B}_i(r) k_c(r,s) \tilde{B}_i(s) dG(r) dG(s) \\
&\leq \sup_{s\in\mathcal{S}} \tilde{B}_i(s)^2 \cdot \int\int |k_c(s,r)| dG(r) dG(s).
\end{aligned}$$

Similarly, the Cauchy-Schwarz inequality, $(R_{i,j}^{BW})^2 \leq R_{i,i}^{BB} V_{j,j}^*$. We further have

$$\mathbb{E}[V_{i,j}^*] = \int k_c(s,s) \bar{\varphi}_i^e(s)' \Omega_{wu}(s) \bar{\varphi}_i^e(s) dG(s) = \int \bar{\varphi}_i^e(s)' \Omega_{wu}(s) \bar{\varphi}_i^e(s) dG(s)$$

since $k_c(s,s) = 1$. Recall that for X_i, $i = 1, \ldots, 4$ jointly mean-zero normal random variables with covariances σ_{ij}

$$\mathbb{E}[(X_1 X_2 - \sigma_{12})(X_3 X_4 - \sigma_{34})] = \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23}.$$

Thus

$$\begin{aligned}
\mathrm{Var}[V_{i,j}^*] &= \sum_{k,l=1}^{\infty} \left(\int \psi_k(s)\psi_l(s) \bar{\varphi}_i^e(s)' \Omega_{wu}(s) \bar{\varphi}_i^e(s) dG(s)\right)\left(\int \psi_k(s)\psi_l(s) \bar{\varphi}_j^e(s)' \Omega_{wu}(s) \bar{\varphi}_j^e(s) dG(s)\right) \\
&\quad + \sum_{k,l=1}^{\infty} \left(\int \psi_k(s)\psi_l(s) \bar{\varphi}_i^e(s)' \Omega_{wu}(s) \bar{\varphi}_j^e(s) dG(s)\right)^2
\end{aligned}$$

$$
\begin{aligned}
&= \int\int \bar{\varphi}_i^e(s)'\Omega_{wu}(s)\bar{\varphi}_i^e(s)k_c(s,r)^2\bar{\varphi}_j^e(r)'\Omega_{wu}(r)\bar{\varphi}_j^e(r)dG(r)dG(s) \\
&\quad + \int\int \bar{\varphi}_i^e(s)'\Omega_{wu}(s)\bar{\varphi}_j^e(s)k_c(s,r)^2\bar{\varphi}_i^e(r)'\Omega_{wu}(r)\bar{\varphi}_j^e(r)dG(r)dG(s) \\
&\leq 2\sup_{s\in\mathcal{S}}||\Omega_{wu}(s)|| \cdot \sup_{1\leq i\leq q, s\in\mathcal{S}}||\bar{\varphi}_i^e(s)||^4 \cdot \int\int k_c(s,r)^2 dG(r)dG(s).
\end{aligned}
$$

Since

$$
\int\int |k_c(s,r)|dG(r)dG(s) \to 0 \text{ and } \int k_c(s,r)^2 dG(r)dG(s) \to 0
$$

as $c \to \infty$, we conclude $\tilde{V}_{0,i,j}^c \xrightarrow{p} \mathbb{E}[V_{i,j}^*] = \int \bar{\varphi}_i^e(s)'\Omega_{wu}(s)\bar{\varphi}_i^e(s)dG(s)$. The result under Condition 3(b) follows analogously. □

References

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

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Anselin, L. (1990): "Spatial Dependence and Structural Instability in Applied Regression Analysis," *Journal of Regional Science*, 30(2), 185–207.

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

Conley, T. G., S. Gonçalves, M. S. Kim, and B. Perron (2023): "Bootstrap inference under cross-sectional dependence," *Quantitative Economics*, 14, 511–569.

Elliott, G., and U. K. Müller (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73, 907–940.

Ferguson, T. S. (1967): *Mathematical Statistics — A Decision Theoretic Approach*. Academic Press, New York and London.

Fotheringham, A. S., C. Brunsdon, and M. Charlton (2002): *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley and Sons.

Fotheringham, A. S., T. M. Oshan, and Z. Li (2024): *Multiscale Geographically Weighted Regression, Theory and Practice*. CRC Press.

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

Hansen, B. E. (2000): "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115.

Jenish, N., and I. R. Prucha (2009): "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays and Random Fields," *Journal of Econometrics*.

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

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

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

Mei, C.-L., M. Xu, and N. Wang (2016): "A Bootstrap Test for Constant Coefficients in Geographically Weighted Regression Models," *International Journal of Geographical Information Science*.

Müller, U. K., and M. W. Watson (2022): "Spatial Correlation Robust Inference," *Econometrica*, 90, 2901–2935.

——— (2023): "Spatial Unit Roots," *Manuscript, Princeton University*.

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

Stock, J. H., and M. W. Watson (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.