Testing Coefficient Variability in Spatial Regression *

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

This Draft: March 2026

Abstract

This paper develops a test for coefficient variability 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 accommodate spatial correlation in regressors and regression errors. The test approximates the best local invariant test for coefficient stability in a Gaussian regression model with martingale-like coefficient variation under the alternative, and is thus a spatial generalization of the Nyblom (1989) test of coefficient stability in time series regressions.

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

*We thank two referees for constructive comments. Müller acknowledges financial support from the National Science Foundation grant SES-2242455.

1 Introduction

Variation in regression coefficients across time or space complicates the interpretation of empirical findings. At best, the OLS coefficient estimator measures an average population regression coefficient, but this average no longer corresponds to the best linear predictor conditional on time or location. If the regression aims to measure the causal effect of a potential policy intervention, then the heterogeneity of the effect typically influences the desirability of the policy. Furthermore, observing instability of coefficients within a given context renders the extrapolation of the "average" results to other contexts even more problematic.

For these reasons researchers routinely test for temporal stability of the coefficients in time series regressions. A wide array of tests, originating with Chow (1960) and Quandt (1960) and including widely used tests by Nyblom (1989) and Andrews (1993), have been developed for this purpose.

Methods for testing for stability in spatial regressions are far 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)), to our knowledge, stability tests using these methods rely 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)).

In this paper we derive a test for coefficient stability in spatial regressions that is designed to have good power for a wide range of persistent patterns of coefficient variation, is applicable in a wide range of spatial designs, and accommodates spatial correlation in regressors and regression errors. The proposed test statistic approximates the locally best test for the null hypothesis of coefficient stability against an alternative of martingale-like random coefficient variation in a canonical spatial regression model (in analogy to Nyblom (1989)) and controls size under general distributional assumptions and spatial autocorrelation (in analogy to Andrews (1993)).

Our analysis proceeds as follows. Section 2 lays out a canonical spatial regression model involving a dependent variable, y, a scalar 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, simplifying the development of an optimal test for spatial stability of β. The form of the optimal test and its power depend on how β varies under the alternative hypothesis. For instance, a Chow test is appropriate if β undergoes a discrete break across a known spatial boundary. However, if the boundary is unknown, the problem becomes more complex. In time series, a sequential Chow test, like the Quandt or "Sup-Wald" test, can be used, with large-sample critical values derived by Andrews (1993).

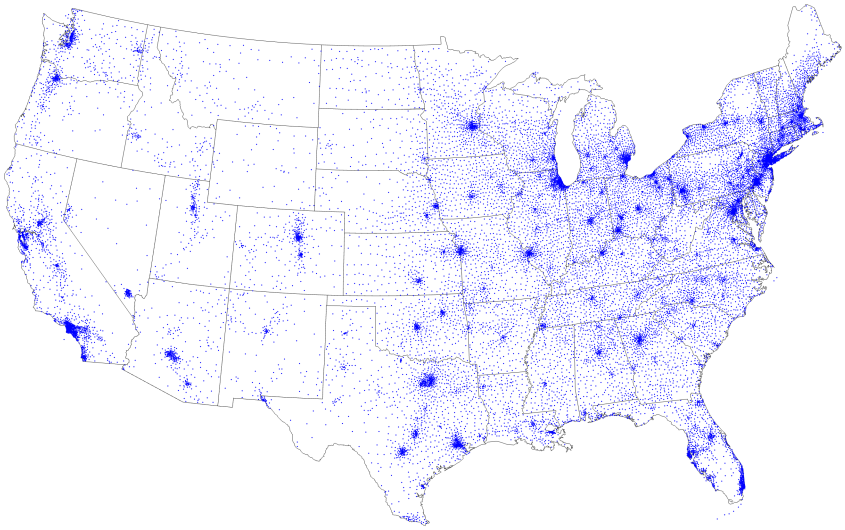
In spatial regressions, however, the myriad of ways that observations can be divided makes such an approach less attractive, and extending these tests to allow for multiple discrete breaks becomes even more challenging. An alternative approach, common in time series analyses, postulates that β evolves as a random walk under the alternative. Nyblom (1989) developed an optimal test for local time variation of this sort. Importantly, in time series regressions, tests for random-walk variation or tests for discrete breaks at unknown break points have similar power properties across a wide range of persistent patterns of coefficient variation. (See Elliott and Müller (2006) for the associated theory and Stock and Watson (1996) for empirics). With these considerations in mind, Section 2 derives a locally optimal test for (spatial) random walk variation in the scalar coefficient β.

Section 3 introduces a large-sample framework that allows us to study the properties of this test. The framework accommodates non-Gaussian spatially correlated processes observed at locations sampled from a wide range of spatial designs. We characterize the large sample properties of the test both under the null hypothesis and under local alternatives that include deterministic coefficient spatial variation (including discrete breaks) and random walk spatial variation. A crucial precursor to these results is the analysis in Lahiri (2003), who provides a central limit theorem for weighted averages of spatially correlated data and an associated limiting covariance matrix.

Building on these results, Section 4 derives a feasible test that is large-sample valid under general forms of weak spatial correlation. We also derive an estimator for the degree of spatial variation in β. Specifically, under the assumption that β evolves as a spatial random walk, we compute a (nearly-) median unbiased estimator for the standard deviation of changes in β over the spatial domain.

Sections 3 and 4 imply that the size and power of the proposed test depend on the spatial correlation of the data and the distribution of locations where the variables are measured. Section 5 examines these effects through a series of experiments with designs that are cali-

Figure 1: 21,194 Zip Code Locations



brated to match regressions involving socioeconomic variables observed at the zip code level across the United States. The experiments use 61 different socioeconomic variables from the American Community Survey (ACS), measured across over 21,000 zip codes in the contiguous 48 U.S. states. This allows us to design experiments with data exhibiting diverse distributions and patterns of spatial dependence. Figure 1 shows the locations of these zip codes, a spatial design we utilize at several points throughout the paper to illustrate key concepts.

Section 5 also uses these data to investigate instability in bivariate relationships between socioeconomic variables across the United States. We explore this by examining roughly 3,500 regression involving the 61 socioeconomic variables used in the simulation experiments. We find substantial evidence of instability when considering the continental U.S. as a whole, but little evidence of instability within individual states. Evidently (and perhaps not surprisingly), reduced-form socioeconomic relationships vary across states and regions in the U.S.

The final section offers some concluding remarks, focusing on modifications of the test in instrumental variable regressions and extensions of the analysis to test for stability in multiple coefficients (i.e., for vector-valued β). An appendix provides proofs and details underlying some of the numerical calculations. An additional online appendix describes the ACS data used in Section 5 and includes a user's guide summarizing the steps for carrying out the

proposed test. A Matlab script that implements the test is available in the paper's replication files.

2 A Canonical Spatial Regression Model with Varying Coefficients

We begin by studying a canonical spatial regression model with a regression coefficient that is potentially varying across space. The simple structure of this 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 a $(p-1)$-dimensional vector of 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 $d \geq 1$. (In time series applications, $d = 1$, while in geographic applications like those shown in Figure 1, $d = 2$.) 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 this canonical model, we assume that $u_l \sim iid\mathcal{N}(0, 1)$ and $\{x_l, z_l\}$ are nonstochastic. These assumptions are relaxed in the following sections.

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

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

For the purpose of learning about spatial variation in β_l, the vector δ is a nuisance parameter. We focus on tests that are invariant to the transformations

$$y \to y + Wa \text{ for all } a \in \mathbb{R}^p$$

where $y = (y_1, \ldots, y_n)'$ and $W = (w_1, \ldots, w_n)'$; this eliminates the parameter 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 $\sum_{l=1}^n \beta_l^1 x_l \hat{e}_l$. Because the optimal test depends on the particular configuration of the spatially varying β_l's, there is no uniformly most powerful test, and any test designed to have good power for a particular choice of $\{\beta_l^1\}_{l=1}^n$ necessarily leads to poor power properties under some alternative value of $\{\beta_l\}_{l=1}^n$.

This is a standard issue in the design of hypothesis tests, and one solution is to maximize a weighted average power criterion over the various possible paths of $\{\beta_l\}_{l=1}^n$, or equivalently, calls for the specification of a probability distribution over $\{\beta_l\}_{l=1}^n$ that serves as a weighting function that puts relatively more weights on paths of particular interest to the researcher. We use that approach here. We choose a probability distribution that reflects two characteristics of the evolution of $\{\beta_l\}_{l=1}^n$ under the alternative. First, we focus the test on persistent (i.e., low-frequency) variation in β_l. In the context of our zip-code level example, this focuses the test's power on variation across states or wide regions, and puts little weight on high-frequency zip-code to zip-code variation. Furthermore, in many applications it is reasonable to use a weighting function that is invariant to rotations of the locations, i.e., the probability distribution should be isometric. In the context of our example, the test puts equal weight on North-South, East-West, NNE-SSW, etc. variation. Lévy (1948)-Brownian motion provides a convenient weighting function that satisfies these requirements. Thus, we seek a test that maximizes power against

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

where $L(\cdot)$ is a Lévy-Brownian motion process and $\kappa > 0$ measures the size of instability.

Lévy-Brownian motion is a spatial generalization of Brownian motion. Specifically L is a continuous parameter mean-zero Gaussian process with almost surely continuous sample paths and covariance kernel equal to

$$k_L(s, r) = \mathbb{E}[L(s)L(r)] = \tfrac{1}{2}(||r|| + ||s|| - ||s - r||) \tag{5}$$

where $||\cdot||$ denotes the Euclidean norm. Notably, L induces a Wiener process along each line: For all $v_0, v_1 \in \mathbb{R}^d$, $L(v_0 + tv_1) - L(v_0) \sim \mathbb{W}(t)$, where $\mathbb{W}(t)$ with $t \in \mathbb{R}$ is a standard

scalar Wiener process. From (4) this implies that the variance of $(\beta_l - \beta_\ell)$ is proportional to the distance between the locations s_l and s_ℓ. Since a Wiener process is a martingale, the specification (4) generalizes Nyblom's (1989) assumption for parameter variation in time. We denote by Σ_L the $n \times n$ covariance matrix of $(L(s_1), \ldots, L(s_n))'$.[1]

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 the null 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 then shows that this locally best invariant test rejects for large values of the score statistic

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

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 $\xi^* = \hat{e}' D_x \bar{\Sigma}_L D_x \hat{e}$, 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,

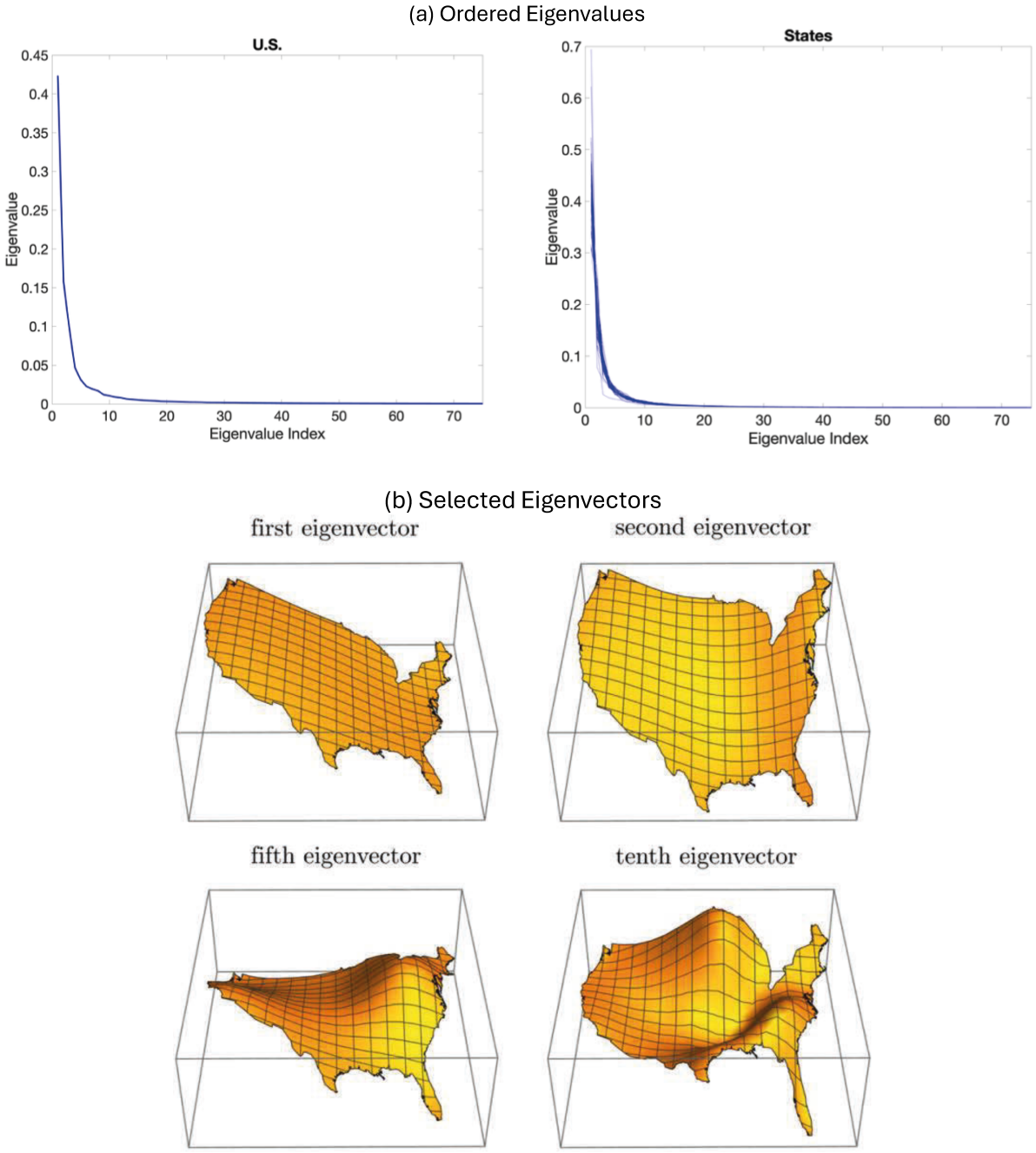
$$\xi^* = \sum_{j=1}^{n} \lambda_j \left(\sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l \right)^2 .$$

This representation shows 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.

Figure 2 shows eigenvalues, λ_j, and selected eigenvectors, r_j, of $\bar{\Sigma}_L$ for the zip code locations shown in Figure 1. The left figure in Panel (a) shows the eigenvalues of $\bar{\Sigma}_L$ for the locations across the entire U.S., and the right panel shows the eigenvalues for the locations in each of the 48 states (which are sufficiently similar that they are difficult to discern in the

[1] In geographic applications, where locations are on the surface of the earth, distances $||\cdot||$ computed using the great-circle formula also yield a positive definite covariance kernel (5), so Σ_L is guaranteed to be positive semi-definite. Conley (1999) discusses the use of "economic distance," rather than geographical distance, and provides several examples.

Figure 2: Eigenvalues and Eigenvectors of $\overline{\Sigma}_L$



Notes: The panels show eigenvalues and eigenvectors of $\overline{\Sigma}_L$ evaluated at the zip-code locations shown in Figure 1. The left panel in (a) shows eigenvalues across all locations in the U.S. The right panel shows eigenvalues for each of the 48 states (so there are 48 functions plotted). Panel (b) shows eigenvectors across the U.S.

plot). In both cases note that the eigenvalues decay quickly. Panel (b) shows selected eigenvectors over the entire U.S. Notice that the eigenvectors become less smooth as the index j increases. These figures highlight an important feature of Lévy-Brownian motion: the smooth weighted averages associated with the first few eigenvectors have markedly larger variances (given by the associated eigenvalues) than other weighted averages so that sample paths are dominated by low-frequency variability. The test that rejects for large values of ξ^* is the locally best test for such low-frequency variation in the regression coefficients.

The rapid decay of the eigenvalues evident in Panel (a) suggests that ξ^* can be approximated using a small number weighted averages, that is by

$$\xi_q = \sum_{j=1}^{q} \lambda_j \left(\sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l \right)^2 \tag{6}$$

for some fixed q, such as $q = 35$. As we show in the next section, this fixed-q approximation allows us to study the large-sample properties of the test statistic in settings that are much more general than those of the canonical model.

In particular, using the approximation in (6), it suffices to study the asymptotic properties of the $q \times 1$ vector Y_n and the diagonal matrix Λ_q where

$$Y_{n,j} = \sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l,\ Y_n = (Y_{n,1}, ..., Y_{n,q})', \text{ and } \Lambda_q = \text{diag}(\lambda_1, \ldots, \lambda_q). \tag{7}$$

Using this notation, the test statistic ξ_q in (6) becomes

$$\xi_q = Y_n' \Lambda_q Y_n. \tag{8}$$

3 Large-Sample Analysis

This section derives the asymptotic distribution of ξ_q under the null and local alternative hypotheses. The analysis proceeds in three steps. The first step focuses on 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 depend on the locations $\{s_l\}$, but not the data $\{y_l, w_l\}$; thus we

begin with an assumption about the distribution of the locations. The second step involves assumptions about the large-sample behavior of weighted averages of terms such as $x_l u_l$ and $x_l w_l$ that appear in Y_n; we postulate high-level central limit and law of large numbers results for these weighted averages. The third step combines the results from the previous steps to deduce the large-sample behavior of Y_n, and thus ξ_q.

3.1 Assumptions

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

We make the following assumption on the locations.

Condition 1. *For some $\nu_n \to \infty$, the locations s_l are generated via $s_l = \nu_n s_l^0$, with s_l^0 i.i.d. from a bounded and continuous density g (with c.d.f. denoted by G) with support in the compact set $\mathcal{S} \subset \mathbb{R}^d$, $d \geq 1$.*

Condition 1 follows Lahiri (2003) in modelling the (typically irregular) pattern of locations as random. The assumption rationalizes different concentrations of observed locations by allowing for non-uniform sampling with density g.

Since ν_n is assumed to diverge, Condition 1 assumes a form of increasing domain asymptotics, where the observed locations cover a larger and larger region $\nu_n \mathcal{S} = \{s : \nu_n^{-1} s \in \mathcal{S}\}$. This allows for the application of a central limit theorem and law of large numbers of random variables whose dependence decays as their distance increases. If $\nu_n / n^{1/d} \to 0$, then the increase in the overall size of the sampling region is sufficiently slow that the number of observations in a fixed subregion becomes relatively more dense, that is, Condition 1 also allows for a form of infill asymptotics. See Lahiri (1996, 2003) and Lahiri and Zhu (2006) for further discussion.

The i.i.d. modelling assumption for the locations plays an important role in the determination of the limiting variance in the central limit theorem below, but does not otherwise enter the analysis. All other random variables, such as $\{(w_l, u_l)\}_{l=1}^n$ are independent of $\{s_l\}_{l=1}^n$, and our analysis proceeds conditional on $\{s_l\}_{l=1}^n$ (which we do not emphasize in the notation to avoid clutter). In particular, the variability in the suggested test statistic is not driven by the randomness in Condition 1.

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 $\bar{L}(s) = L(s) - \int_{\mathcal{S}} L(r) dG(r)$, 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

$$\bar{k}_L(r,s) = \mathbb{E}[\bar{L}(s)\bar{L}(r)] = \sum_{j=1}^{\infty} \lambda_j^0 \varphi_j(s)\varphi_j(r) \tag{9}$$

with eigenvalues $\lambda_1^0 \geq \lambda_2^0 \geq \ldots$, and eigenfunctions $\varphi_j(s) = \lambda_j^{-1} \int_{\mathcal{S}} \varphi_j(r)\bar{k}_L(r,s)dG(r)$ that are continuous for $\lambda_j^0 > 0$, orthonormal $\int_{\mathcal{S}} \varphi_i(s)\varphi_j(s)dG(s) = \mathbf{1}[i=j]$ and orthogonal to a constant, $\int_{\mathcal{S}} \varphi_j(s)dG(s) = 0$, $j = 1, 2, \ldots$. Lemma 6 of Müller and Watson (2022) 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

$$\sup_{j \leq q} |\lambda_j - \lambda_j^0| \xrightarrow{p} 0 \tag{10}$$

$$\sup_{1 \leq l \leq n, j \leq q} |r_{j,l} - \varphi_j(s_l^0)| \xrightarrow{p} 0 \tag{11}$$

provided $\lambda_1^0 > \lambda_2^0 > \ldots > \lambda_{q-1}^0 > \lambda_q^0$.

3.1.2 Large-Sample Behavior of Weighted Averages of $\{x_l u_l\}$ and $\{x_l w_l\}$

We now turn to an appropriate high-level assumption about the behavior of weighted averages of $x_l u_l$ and $x_l w_l$. These are central limit and law of large numbers results that hold under weak dependence and second order stationarity.

Condition 2. *There exists a positive sequence a_n and $\omega \in [0,1]$ such that for any uniformly convergent sequence of functions $h_n : \mathcal{S} \mapsto \mathbb{R}$ with continuous limit $h : \mathcal{S} \mapsto \mathbb{R}$,*

$$a_n^{-1/2} \sum_{l=1}^{n} h_n(s_l^0) x_l u_l \Rightarrow \mathcal{N}\left(0, \int_{\mathcal{S}} h(s)^2 \left[\omega + (1-\omega)g(s)\right] g(s) ds\right). \tag{12}$$

This assumption has two parts: (i) the large-sample normality for the weighted averages, and (ii) the specific form for the limiting variance. Both parts are motivated by the central limit theorem in Lahiri (2003).

The particular form of the limiting variance plays an important role in spatial analysis. To gain intuition suppose that $\eta(\cdot)$ is a mean-zero stationary random field on $\nu_n \mathcal{S}$ with

$E[\eta(r)\eta(s)] = \sigma(r-s)$ for $\sigma : \mathbb{R}^d \mapsto \mathbb{R}$ and $\int_{\mathbb{R}^d} |\sigma(s)|ds < \infty$. Then

$$\begin{aligned}\text{Var}\left[\sum_{l=1}^{n} h(s_l^0)\eta(s_l)\right] &= \sum_l h(s_l^0)^2\sigma(0) + \sum_{l,\ell:l\neq\ell} h(s_l^0)h(s_\ell^0)\sigma(s_l - s_\ell) \\ &= \sum_l h(s_l^0)^2\sigma(0) + \sum_l h(s_l^0)\sum_{\ell:\ell\neq l} h(s_l^0 + \nu_n^{-1}\Delta_{l,\ell})\sigma(\Delta_{l,\ell})\end{aligned}$$

where $\Delta_{l,\ell} = s_l - s_\ell$. Since $\sigma(\Delta_{l,\ell}) \approx 0$ whenever $\nu_n^{-1}\Delta_{l,\ell}$ is non-negligible, one would expect

$$\sum_l h(s_l^0)\sum_{\ell\neq l} h(s_l^0 + \nu_n^{-1}\Delta_{l,\ell})\sigma(\Delta_{l,\ell}) \approx \sum_l h(s_l^0)^2 \sum_{\ell:\ell\neq l}\sigma(\Delta_{l,\ell}).$$

Furthermore for any r in the interior of $\mathcal{S}$, the distribution of locations s_l in a ball of any fixed radius centered at $\nu_n r$ is approximately uniform under Condition 1. The expected number of locations that fall into the ball is proportional to $g(r)n^{1/d}/\nu_n$, yielding

$$\frac{\nu_n}{n^{1/d}}\sum_l \sigma(s_l - r) \approx g(r)\int_{\mathbb{R}^d}\sigma(s)ds.$$

This suggests the form of the limiting variance in (12), with $na_n = \sigma(0) + \int_{\mathbb{R}^d}\sigma(s)ds \cdot n^{1/d}/\nu_n$ and $\omega \in [0,1]$ the limit of $\sigma(0)/(\sigma(0) + na_n)$.

These heuristics (as well as the convergence to a Gaussian limit) are made precise in Lahiri's (2003) Theorems 3.1 and 3.2 (and where Lemma 12 of Müller and Watson (2022) shows that Lahiri's result also applies to uniformly converging weight functions, as assumed in Condition 2). The upshot is that even if $x_l u_l$ is stationary and weakly dependent, as a general matter the limiting variance in (12) depends on the density g. This is an important departure of the spatial case from the usual time series model in which the "locations" correspond to the equally-spaced time indices $t = 1, 2, ...$, which yield a uniform spatial density.

The test statistic also involves weighted averages of $x_l w_l$. For these, we assume the following.

Condition 3. *There exists a $q \times 1$ vector $\Sigma_{xw} = (\sigma_x^2, \Sigma_{xz}')'$ with $\sigma_x^2 > 0$ such that for any uniformly convergent sequence of functions $h_n : \mathcal{S} \mapsto \mathbb{R}$ with limit h,*

$$n^{-1}\sum_{l=1}^{n} h_n(s_l^0)x_l w_l \xrightarrow{p} \int_{\mathcal{S}} h(s)g(s)ds \cdot \Sigma_{xw}. \tag{13}$$

Condition 3 assumes a law of large numbers for weighted averages of $x_l w_l$. One could again apply Lahiri's (2003) results, or invoke the mixing conditions in Jenish and Prucha (2009) to obtain sufficient conditions. Condition 3 is compatible with some form of nonstationarities: If x_l is centered and weakly dependent, then (weighted) averages with any smooth function will converge to zero. Thus one would expect (13) to hold with Σ_{xw} having zero elements in the rows that correspond to region dummies or (nonlinear) controls for latitude/longitude, for example.

3.1.3 Asymptotic properties of $\hat{\delta}$

Recall that $\hat{\delta}$ is the OLS estimator of the entire coefficient vector $\delta = (\beta, \alpha')'$. We make the following high level assumption.

Condition 4. $\hat{\delta} - \delta = O_p(a_n^{1/2} n^{-1})$.

If one were to strengthen Conditions 2 and 3 to hold for averages $\{w_l u_l\}$ and $\{w_l w_l'\}$, then Condition 4 would immediately follow from $a_n^{-1/2} n(\hat{\delta} - \delta) = \left(n^{-1} \sum_l w_l w_l'\right)^{-1} a_n^{-1/2} \sum_l w_l u_l$, so weak dependence and stationarity are sufficient. They are not necessary, however, as Condition 4 is also compatible with, for example, controls such as region dummies and other forms of non-stationarity.

3.1.4 Evolution of β_l under the alternative

We consider two local alternatives for β_l. The first assumes that β_l follows a deterministic piecewise continuous function (and thus allows for discrete breaks in β), and the second assumes that β_l evolves as a scaled Lévy-Brownian motion. The assumptions are:

Condition 5. *(a)* $\beta_l - \beta = a_n^{1/2} n^{-1} b(s_l^0)$, $l = 1, \ldots, n$, *where* $b : \mathcal{S} \mapsto \mathbb{R}$ *is sample size independent piecewise continuous function (formally, for a finite partition* $\{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ *of* $\mathcal{S}$, $b : \mathcal{S}_j \mapsto \mathbb{R}$ *is uniformly continuous, and the boundary* $\partial \mathcal{S}_j$ *has Lebesgue measure zero, for all* $j = 1, \ldots, m$*); or*

(b) $\beta_l - \beta = \kappa_n L(s_l^0)$ *with* $\kappa_n = a_n^{1/2} n^{-1} \gamma$ *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 $O_p(a_n^{1/2} n^{-1})$, a rate which is related to the rate of convergence of the CLT in Condition 2 and the "usual"

n^{-1} rate in the LLN of Condition 3. 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.

3.2 Limiting Distribution of Y_n and ξ_q

Combining these high level assumptions yields the following result.

Theorem 1. *(a) Under Conditions 1-4 and 5(a),* $a_n^{-1/2} Y_n \Rightarrow Y_0 + B$, *where* $Y_0 \sim \mathcal{N}(0, V_0)$, *with* $V_{0,i,j} = \int_{\mathcal{S}} \varphi_i(s)\varphi_j(s)[\omega + (1-\omega)g(s)]g(s)ds$ *and* $B_i = \sigma_x^2 \int_{\mathcal{S}} \varphi_i(s)b(s)dG(s)$ *for* $i, j = 1, \ldots, q$. *Furthermore,*

$$a_n^{-1}\xi_q \Rightarrow (Y_0 + B)'\Lambda_q^0(Y_0 + B) \tag{14}$$

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

(b) Under Conditions 1-4 and 5(b), $a_n^{-1/2} Y_n \Rightarrow Y_0 + \gamma\sigma_x^2 Y_1$ *where* $Y_1 \sim \mathcal{N}(0, \Lambda_q^0)$. *Furthermore,*

$$a_n^{-1}\xi_q \Rightarrow (Y_0 + \gamma Y_1)'\Lambda_q^0(Y_0 + \gamma Y_1). \tag{15}$$

Note that these limits are those of a location model where x_t is a constant and there are no additional controls z_t. This equivalence is essentially due to Condition 3, which ensures that second moments are approximately constant over $\mathcal{S}$. We exploit this equivalence below when we derive critical values.

To get a better understanding of the result in part (b), it is instructive to consider the case where $\omega = \sigma_x^2 = 1$. 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)

$$a_n^{-1}\xi_q \Rightarrow Z'(\Lambda_q^0 + \gamma^2(\Lambda_q^0)^2)Z = \sum_{j=1}^{q}(\lambda_j^0 + (\gamma\lambda_j^0)^2)Z_j^2 \quad \text{with } Z_j \sim iid\mathcal{N}(0,1), \tag{16}$$

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.

4 Spatial Correlation-Robust Inference for Varying Coefficients

In this section we discuss how to use the asymptotic results in Theorem 1 for inference. A first subsection concerns the test of the null hypothesis of parameter stability. A second subsection provides a suggestion for the choice of q, and the final subsection discusses estimation of the magnitude of parameter instability κ in the Lévy model (4).

4.1 Test Statistic and Critical Value

The construction of an asymptotically valid test based on the results in Theorem 1 faces two challenges: First, the normalization sequence $\{a_n\}$ for ξ_q in (14) and (15) is unknown, and second, the scalar $\omega \in [0,1]$ that appears in the limiting variance V_0 in Theorem 1 is an unknown nuisance parameter. We handle these in turn.

To deal with the first issue, we suggest basing inference on the self-normalized statistic

$$\xi_q^s = \frac{Y_n' \Lambda_q Y_n}{Y_n' Y_n} \Rightarrow_{H_0} \frac{Y_0' \Lambda_q^0 Y_0}{Y_0' Y_0} \tag{17}$$

where the convergence follows straightforwardly from Theorem 1 and the continuous mapping theorem.

The presence of ω in the expression for V_0 in $Y_0 \sim \mathcal{N}(0, V_0)$, $V_{0,i,j} = \int_{\mathcal{S}} \varphi_i(s)\varphi_j(s)[\omega + (1-\omega)g(s)]g(s)ds$, is more difficult to handle. A direct approach would require the estimation of the density g of the locations s_l^0. This would be cumbersome. We instead follow Müller and Watson (2022, 2024) and argue that an asymptotically valid critical value for ξ_q^s is obtained by considering the small sample critical value that is appropriate in a simple one-parameter model of dependence: Define the Ornstein-Uhlenbeck process on $\mathcal{S}$ as the zero-mean Gaussian processes η with covariance kernel $\mathbb{E}[\eta(r)\eta(s)] = k_c(r,s) = \exp(-c||r-s||)$. We denote the process by $\eta \sim \mathcal{G}_c$. For a sample size independent value of c, $\eta(s_l^0)$ is strongly dependent, as $\mathbb{E}[\eta(r)\eta(s)]$ is non-zero for all $r \neq s$. For $c = c_n$ with $c_n \to \infty$, this model induces weaker dependence, with the exact degree of dependence governed by how quickly $c_n \to \infty$. Lahiri's (2003) result in Condition 2 also applies to this simple model, and different sequences $c_n \to \infty$ induce all of the possible ω-indexed limit variances V_0. Furthermore, recall that the

asymptotic distribution in Theorem 1 is also obtained in a location model. Thus, if we define $\tilde{Y}_{n,j} = \sum_{l=1}^{n} r_{j,l}\eta(s_l^0)$, $\tilde{Y}_n = (\tilde{Y}_{n,1}, \ldots, \tilde{Y}_{n,q})'$ and $\tilde{\xi}_q^s = \tilde{Y}_n'\Lambda_q\tilde{Y}_n/\tilde{Y}_n'\tilde{Y}_n$, and let $\text{cv}_{q,n}$ satisfy

$$\sup_{c \geq c_n^*} \mathbb{P}(\tilde{\xi}_q^s > \text{cv}_{q,n}) = \alpha \tag{18}$$

for some c_n^* satisfying $\limsup_n c_n^* < \infty$, then $\text{cv}_{q,n}$ will be asymptotically valid in the general model under Conditions 1–4 (see the proof of Theorem 8 in Müller and Watson (2024) for a precise argument).

To make this approach feasible, we need to choose c_n^*. We find it useful to parameterize c in terms of the implied average correlation between locations. Thus, let $c_{\bar{\rho},n}$ 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 smaller c. We suggest using an average correlation of $\overline{\rho} = 0.01$ and thus use $c_n^* = c_{0.01,n}$. Under Condition 1, $c_{0.01,n} \to c_{0.01}$, where $c_{0.01}$ solves $\int_{\mathcal{S}}\int_{\mathcal{S}} k_{c_{0.01}}(r,s)g(s)ds \cdot g(r)dr = 0.01$, so $c_{0.01,n}$ remains bounded for all n (irrespective of ν_n). Thus, (18) not only yields asymptotic size control under all forms of weak dependence that induce the null distribution in Theorem 1, but the test also controls size by construction under a particular form of strong dependence in the location model (for reference, a time series AR(1) with $n = 500$ observations induces an average correlation of 0.01 for an AR(1) coefficient equal to 0.72, for example). While this choice of critical value induces conservativeness in the test, at least for large enough n, we consider it attractive to have some absolute sense of the degree of correlations for which the test is designed to control size.

Because $\eta \sim \mathcal{G}_c$ is Gaussian, the critical value $\text{cv}_{q,n}$ can be determined quickly and accurately by applying Imhof's (1961) procedure to compute the c.d.f. of quadratic forms in Gaussian variables. We provide details in the appendix, and corresponding Matlab code in the replication files.

4.2 Determining q

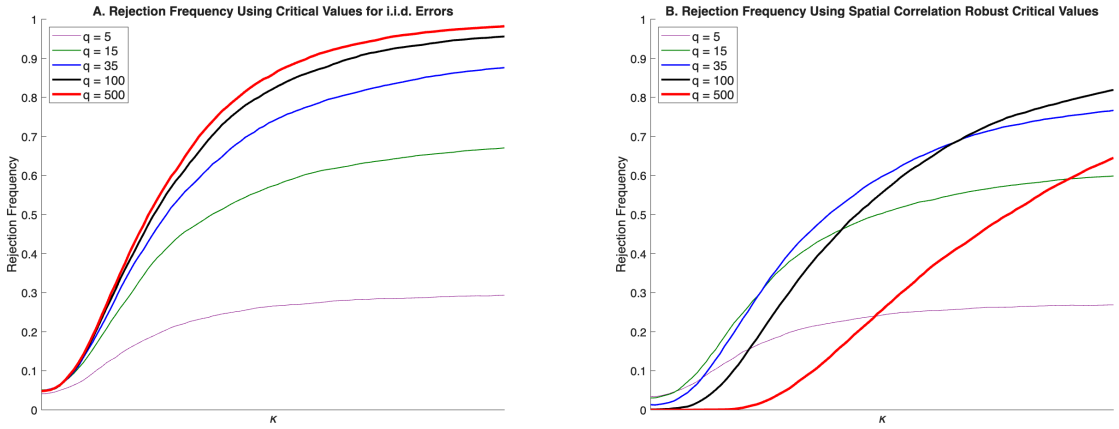
An additional advantage of this construction of the critical value is that it naturally lends itself to a method for choosing q. Intuitively, a small q leads to low power because too few weighted averages do not contain enough information to reliably distinguish between the null and alternative distributions. On the other hand, a very large q also leads to low power, as the "higher frequency" eigenvectors corresponding to smaller eigenvalues capture the correlation

implied by $c_n^* = c_{0.01}$, which necessitates choosing a large value of cv_n in (18). Thus, one does not want q to be too small or too large, and this trade-off leads to a simple rule for a choice of q that maximizes power.

These issues are illustrated in Figure 3. The figure shows the size and power of the ξ_q^s test in the location version of the canonical model of Section 2, that is, the model with $x_l = 1$, no additional regressors z, $u_l \sim iid\mathcal{N}(0,1)$, using the U.S. spatial design with n =21,194 zip codes. The left panel in Figure 3 shows rejection frequencies using critical values that induce 5% rejection probability under the i.i.d. model for u_l. There, larger values of q yield more powerful tests. Indeed, as $q = q_n \to \infty$, the test based on ξ_q^s is asymptotically equivalent to the locally best invariant test ξ^* in the canonical model, and the power of ξ_{500}^s is indistinguishable to the power of ξ^* (not shown in the figure). However, because the left panel uses critical values computed under the i.i.d. model, the associated tests are invalid when the errors are spatially correlated. For example, if the errors are generated by the $\mathcal{G}_c$ model with $c = c_{0.01}$, the null rejection probability of nominal 5% tests using i.i.d. critical values is 6% for $q = 5$, increases to 13% for $q = 35$, to 30% for $q = 100$, and is greater than 80% for $q = 500$. The right panel corrects this deficiency by using the spatial-correlation robust critical values from (18) with $c_n^* = c_{0.01,n}$. This figure shows the trade-off alluded to in the previous subsection: the spatial-correlation robust critical values are necessarily larger than their i.i.d. counterparts, this leads to a conservative test when the errors are i.i.d., and an associated reduction in power. Importantly this conservativeness increases with q, as does the reduction in power.

We utilize the trade-off illustrated in the right panel to choose q. As in the figure, consider power in the location model under i.i.d. Gaussian errors. Let $\check{Y}_{n,j} = \sum_{l=1}^n r_{j,l}(\varepsilon_l + \kappa L(s_l))$, where $\varepsilon_l \sim iid\mathcal{N}(0,1)$, L is a Lévy-Brownian motion, and $\kappa > 0$ indicates the distance of the alternative from the null hypothesis. Denote by $\check{\xi}_q^s = \check{Y}_n'\Lambda_q\check{Y}_n/\check{Y}_n'\check{Y}_n$ the test statistic applied to this location model with $\check{Y}_n = (\check{Y}_{n,1}, \ldots, \check{Y}_{n,q})'$. As shown in the figure, we would expect that for any critical value $\text{cv} > 0$, the rejection probability $\mathbb{P}_\kappa(\check{\xi}_q^s > \text{cv})$ is increasing in κ. Let κ_q solve $\mathbb{P}_{\kappa_q}(\check{\xi}_q^s > \text{cv}_{q,n}) = 0.5$ (and set $\kappa_q = \infty$ if no solution exists), that is, for the critical value determined in (18), κ_q indicates the magnitude of the alternative in the location model that is required to induce 50% power. Our suggestion is to choose q such that κ_q is minimized. Among the tests depicted in Figure 3, this leads to using $q = 35$, as the power curve of this test reaches 50% for the smallest value of κ. In practice, we consider all values in the interval $2 \leq q \leq 50$, where the upper bound of 50 avoids relying on the multivariate

Figure 3: Rejection Frequencis Using *i.i.d.* and Spatial-Correlation Robust Critical Values



Notes: Rejection frequencies for the ξ_q^s-test using the U.S. zip code spatial design. Panel (a) uses critical values from the canonical model with i.i.d. errors. Panel (b) uses spatial correlation robust critical values from (18).

Gaussian implication of Theorem 1 for a very high dimensional Y_n. For the U.S. zip code design, this yields the optimal value $q = 32$, which is close to the $q = 35$ value shown in the figure.

4.3 Estimating the Variation in Varying Coefficients

We now discuss a method that takes advantage of the result $a_n^{-1/2} Y_n \Rightarrow Y_0 + \gamma\sigma_x^2 Y_1$ in Theorem 1 (b) to construct an asymptotically approximately median unbiased estimator of the magnitude $\kappa_n = a_n^{1/2} n^{-1}\gamma$ of parameter variation in the model (4), that is, under the assumption that $\beta_l - \beta = \kappa_n L(s_l)$, $l = 1, \ldots, n$ with L equal to Lévy Brownian motion. (See Stock and Watson (1998) for a related estimator for time series regressions.) Denote the estimator by $\hat{\kappa}(Y_n)$, where $\hat{\kappa} : \mathbb{R}^q \mapsto \mathbb{R}$. The construction of $\hat{\kappa}$ faces the same two problems already encountered for testing the null hypothesis of parameter stability in Section 4.1 above, namely the sequence a_n is unknown, and the presence of ω in the distribution of $Y_0 \sim \mathcal{N}(0, V_0)$.

Write $Y = Y_0 + \gamma\sigma_x^2 Y_1 = Y_0 + \psi Y_1$, where $\psi = \gamma\sigma_x^2$ is the scale associated with $Y_1 \sim \mathcal{N}(0, \Lambda_q^0)$. Let $\hat{\psi} : \mathbb{R}^q \mapsto \mathbb{R}$ be an estimator of ψ that satisfies

$$\sup_{\psi \geq 0, \omega \in [0,1]} \mathbb{P}(\hat{\psi}(Y) > \psi) \leq 1/2. \tag{19}$$

In words, the estimator $\hat{\psi}$ overestimates ψ with probability of at most 1/2, for all values of ψ and $\omega \in [0,1]$ (or, equivalently, the interval $[\hat{\psi}(Y), \infty)$ forms a 50% level one-sided confidence interval for ψ). Now suppose $\hat{\psi}$ is scale equivariant, that is, $\hat{\psi}(ay) = a \cdot \hat{\psi}(y)$, for all $a > 0$ and $y \in \mathbb{R}^q$. Then setting $\hat{\kappa}(Y_n) = \hat{\psi}(Y_n)/(n \cdot \hat{\sigma}_x^2)$ with $\hat{\sigma}_x^2 = n^{-1} \sum_{l=1}^n x_l^2$ yields

$$\begin{aligned} \mathbb{P}(\hat{\kappa}(Y_n) > \kappa_n) &= \mathbb{P}(\hat{\psi}(a_n^{-1/2} Y_n) > n a_n^{-1/2} \hat{\sigma}_x^2 \kappa_n) \\ &\to \mathbb{P}(\hat{\psi}(Y) > \gamma \sigma_x^2) = \mathbb{P}(\hat{\psi}(Y) > \psi) \leq 1/2 \end{aligned}$$

where the convergence follows from Condition 3, Theorem 1 (b) and the continuous mapping theorem (provided $\hat{\psi}$ is continuous almost everywhere and the c.d.f. of $\hat{\psi}(Y)$ is continuous at ψ).

We are thus left to construct a scale equivariant estimator $\hat{\psi}$ satisfying (19). However, both g and Λ_q^0 are not directly observed. So consider instead the construction of a scale equivariant estimator $\hat{\psi}_n(\tilde{Y}_n)$ in the Gaussian location model where $\tilde{Y}_{n,j} = \sum_{l=1}^n r_{j,l}(\eta(s_l) + \psi L(s_l)/n)$, $j = 1, \ldots, q$, and $\eta \sim \mathcal{G}_c$. Then by the same logic as applied in the construction of the critical value in Section 4.1, if $\hat{\psi}_n$ satisfies

$$\sup_{\psi \geq 0, c \geq c_{0.01}} \mathbb{P}(\hat{\psi}_n(\tilde{Y}_n) > \psi) \leq 1/2 \tag{20}$$

for all n, then $\sup_{\psi \geq 0, \omega \in [0,1]} \limsup_{n \to \infty} \mathbb{P}(\hat{\psi}_n(Y) > \psi) \leq 1/2$ follows under Conditions 1–4 and 5(b).

In practice, there are many equivariant estimators $\hat{\psi}_n$ satisfying (20), including highly uninformative estimators such as $\hat{\psi}_n(Y_n) = 0$. So in addition to the constraint (20) it makes sense to posit an objective that maximizes informativeness. We choose to minimize the weighted average excess length of the 50% confidence interval $[\hat{\psi}_n(\tilde{Y}_n), \infty)$, that is,

$$\int \mathbb{E}\left[(\psi - \hat{\psi}_n(\tilde{Y}_n)) \mathbf{1}[\psi > \hat{\psi}_n(\tilde{Y}_n)\right] dW(c, \psi) \tag{21}$$

where $W(c, \psi)$ is a suitably chosen weighting function. We then rely on the numerical algorithm developed in Müller and Wang (2019) to determine $\hat{\psi}_n$ that approximately minimizes (21) subject to (20). See the Appendix for details.

5 Monte Carlo Results and Empirical Analysis

Section 1 introduced locations across zip codes in the continental United States. In this section we use data on a wide range of socioeconomic variables measured at these locations to investigate two questions: How well do the methods developed in the earlier sections perform in realistic environments calibrated to these data? And how stable or unstable are the bivariate relationships between these socioeconomic variables across the U.S. and within states?

We begin with a description of the data.

5.1 Data Description

The data are from the American Community Survey, 5-year estimates from 2018-2022, for the zip codes regions (ZIP Code Tabulation Areas (ZCTA)) making up the contiguous 48 states and the District of Columbia. The dataset contains sixty-one variables measuring 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.

The raw data were transformed in three ways. First, in most cases, the variables were scaled by the relevant population in the region. 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 across the 21,194 zip codes. Third, the resulting variables typically exhibit strong spatial persistence as indicated by the spatial unit roots test from Müller and Watson (2024). Because strong spatial persistence can lead to spurious regressions, all of the variables were "spatially differenced" using the LBM-GLS transformation discussed in Müller and Watson (2024). This transformation eliminates spatial low-frequency trends in the variables, analogous to the first differencing transformation of persistent time series.

We are interested in bivariate regressions involving these variables. By selecting pairs of the 61 variables from the dataset it is possible to construct 3,660 bivariate regressions. Some of these involve variables that are closely related by construction (e.g., the fraction of married adults and the fraction of divorced adults), and as described in the online appendix, we used simple rules to eliminate pairs of closely-related variables. This yielded the 3,458 bivariate regressions that are used in the exercises reported below.

5.2 Size and Power

The first set of exercises investigate the size and power properties of the ξ^s test. It focuses on two questions. First, how does spatial correlation and (potentially) spatial heteroskedasticity affect size and power? Second, 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?

5.2.1 Monte Carlo Design

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 1. The experiments involve three ingredients: (i) the spatial locations, (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.

The spatial locations are given by the entire set of locations shown in Figure 1 (U.S. locations), or each of the individual states (state locations). The U.S. locations include all 48 states and the District of Columbia. The state locations include the 45 states with 100 or more zip code locations.[2]

The data generating processes (DGP) for $\{x_l, u_l\}$ involve zero-mean spatial Ornstein-Uhlenbeck (OU) processes $\eta \sim \mathcal{G}_c$ with covariance kernel $\mathbb{E}[\eta(r)\eta(s)] = k_c(r,s) = \exp(-c||r-s||)$ discussed previously in Section 4.1. As described there, $c_{\overline{\rho}}$ denotes the value of c that yields an average correlation of $\eta(s_l)$ of $\overline{\rho}$ over the sample locations. We use four values of $\overline{\rho}$, (0.0, 0.001, 0.01, 0.03), and chose critical values for the tests that are designed to control size for values $\overline{\rho} \leq 0.01$ as described in Section 4.1.

[2]This excludes the District of Columbia, Delaware, Rhode Island and Wyoming.

We use four data generating processes for $\{x_l, u_l\}$:

- DGP1: $x_l = 1$ and $u_l \sim \mathcal{G}_c$.
- DGP2: x_l and u_l are generated by independent $\mathcal{G}_{c/2}$ processes.
- DGP3: x_l is randomly selected from the 61 standardized variables in our dataset and u_l follows a $\mathcal{G}_{c/2}$ process.
- DGP4: $\{y_l^{acs}, x_l^{acs}\}$ are a pair of series from the list of 3,458 bivariate regressions in our dataset. Then $x_l = x_l^{acs}\eta_{x,l}$ and $u_l = y_l^{acs}\eta_{u,l}$ where $\eta_{x,l}$ and $\eta_{u,l}$ follow independent $\mathcal{G}_{c/2}$ processes; this DGP is a version of the "spatially dependent wild bootstrap" in the spirit of Shao (2010), Conley, Gonçalves, Kim, and Perron (2023) and Kurisu, Kato, and Shao (2024).

In DGP1, the regressor is constant and errors follow the spatial process used to determine the critical value. Thus, by construction, the size constraint will be satisfied for values of $\overline{\rho} \leq 0.01$. For DGP2, x_l and u_l are independent $\mathcal{G}_{c/2}$ processes, so the product $x_l u_l$ has the same second moments as a $\mathcal{G}_c$ process, but different higher-order moments. In DGP3, the spatial autocovariances of $x_l u_l$ are determined in part by the empirical data, and similarly for DGP4. Note also that any spatial heteroskedasticity in x_l will be inherited by $x_l u_l$ in DGP3 and DGP4.

The final ingredient is the data generating process for β_l. We consider three.

- Constant β: This is the null model with $\beta_l = \beta$ (and where invariance makes the value of β irrelevant).
- Random Walk: β_l is generated 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 experiments, where the scale of the regressors and sample sizes differ across states, we normalize the maximum distance in each experiment to be unity so that $\max_{l,\ell} \sigma(\beta_l - \beta_\ell) = \kappa$, use the scaling from Section 4.3 with $\kappa = \gamma/(\sigma_x n^{1/2})$, and use five values of γ, $(0, 5, 10, 20, 40)$.
- Discrete Break: In this model, β_l takes on different values in different regions of the country. We define regions in three ways: (i) by the 48 states plus the District of

Table 1: Size of nominal 5% tests

$\overline{\rho}$	DGP1	DGP2	DGP3	DGP4
0	0.011 (0.001, 0.046)	0.011 (0.001, 0.046)	0.013 (0.002, 0.046)	0.015 (0.002, 0.074)
0.001	0.035 (0.018, 0.036)	0.029 (0.012, 0.038)	0.011 (0.002, 0.043)	0.023 (0.002, 0.073)
0.01	0.050 (0.050, 0.050)	0.036 (0.021, 0.048)	0.011 (0.001, 0.040)	0.030 (0.003, 0.074)
0.03	0.111 (0.096, 0.154)	0.072 (0.044, 0.100)	0.010 (0.001, 0.037)	0.056 (0.004, 0.080)

Notes: See text for descriptions of DGP1-DGP4. The entries in the table are the rejection rates using the U.S. spatial design, followed by the minimum and maximum rejections rates across the states designs shown in parentheses.

Columbia, (ii) by the nine Census divisions, and (iii) by the eastern and western regions of the U.S.[3] In these experiments the values of β are $iid\mathcal{N}(0, \sigma^2_{\kappa/\sqrt{2}})$ for each region, so κ is the standard deviation of the difference in β across regions.

In the experiments with spatially varying coefficients, x_l is generated by DGP1 and DGP3 and u_l are independent $iid\mathcal{N}(0,1)$ random variables.

5.2.2 Results

Table 1 summarizes the results for the size of the ξ^s test for the four DGPs across each of the spatial designs. Each table cell shows the rejection frequency for the U.S. spatial design followed by parentheses showing the minimum and maximum rejection frequencies across the 45 state designs. DGP1 corresponds to the large-sample model where the regressor is a constant and the errors follow spatial OU processes; this model is used to construct the test's critical value, so that the rejection frequency is 0.05 by construction for $\overline{\rho} = 0.01$, and is less than 0.05 for smaller values of $\overline{\rho}$. DGP2 has the same spatial autocorrelations for $x_l u_l$ as DGP1, but has different higher order moments. The rejection frequencies for DGP2 are slightly smaller than for DGP1. DGP3 differs from DGP2 by using the ACS data for x. Recall that these data have been spatially differenced, and the resulting series turn out to be less spatially correlated than the regressors in DGP2. This results in smaller rejection frequencies. In DGP4, the regressors and errors inherit the spatial heteroskedasticity of the ACS data and this results in somewhat large rejection frequencies, with values that (marginally) exceed 0.05 for a handful of state designs.

[3]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 districts, the number of observations range from 1,273 to 3,649; for the East-West regions there are 8,192 zip codes in the west and 13,002 in the east.

Table 2: Power of nominal 5% tests

γ	RandomWalk	East-West	Districts	States
	(a) $x = 1$ (as in DGP1)			
0	0.01	0.01	0.01	0.01
5	0.12	0.24	0.07	0.02
10	0.36	0.52	0.21	0.05
20	0.64	0.74	0.39	0.09
40	0.80	1.0	0.48	0.12
	(b) $x = x^{acs}$ (as in DGP3)			
0	0.01	0.01	0.01	0.01
5	0.12	0.23	0.07	0.03
10	0.35	0.51	0.21	0.06
20	0.62	0.73	0.35	0.11
40	0.77	0.88	0.43	0.14

Table 2 summarizes the power results. Two results stand out. First, while the test was designed to have good power against random walk variation in the regression coefficients, it has similarly good power against discrete breaks in the coefficients as long as these discrete breaks are persistent. Thus, power is high for i.i.d. breaks across 2 regions labeled "East-West", falls somewhat for 9 regions ("Districts"), and falls substantially for i.i.d. breaks across the 49 "States".

The bottom line from these experiments is that the test appears to have good size control and good power against alternatives with spatially persistent variation of the regression coefficient, as intended.

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

We now turn to the empirical analysis of the 3,458 bivariate spatial regressions involving the 61 socioeconomic variables in the dataset. Of course, these bivariate regressions are not structural (or causal), and this raises the question of how should think about spatial stability or instability. The most straightforward answer is that the tests provide evidence on the spatial stationarity of second moments across these socioeconomic variables. Recall that all variables were spatially differenced, so rejections are not easily explained by the presence of spatial trends in the original level variables. Similarly, the omission of stationary control

Table 3: Empirical results for 3,458 regressions

	Percentiles						
	0.05	0.10	0.25	0.50	0.75	0.90	0.95
$\lvert\hat{\beta}\rvert$ U.S.	0.01	0.02	0.05	0.11	0.23	0.36	0.46
$\lvert\hat{\beta}\rvert$ States	0.01	0.02	0.06	0.13	0.25	0.41	0.52
ξ^s p-value U.S.	0.01	0.01	0.06	0.19	0.42	0.69	0.83
ξ^s p-value States	0.06	0.12	0.29	0.56	0.81	0.93	0.97
$\sigma_{\Delta\beta,1000km}(\hat{\kappa})$ U.S.	0.001	0.009	0.021	0.044	0.072	0.105	0.130
$\sigma_{\Delta\beta,1000km}(\hat{\kappa})$ States	0.003	0.004	0.005	0.011	0.224	0.530	0.789

variables also cannot account for such a finding. The presence or absence of stability is thus of interest not only for prediction, but potentially also to guide structural econometric analysis.

The specifics of the analysis is as follows: For each of the regressions we computed three statistics: (i) $\hat{\beta}$, the OLS estimate of β, (ii) the p-value for the ξ^s test, and (iii) the value of the standard deviation of the change in β_l over 1000km in the random walk model that is implied by $\hat{\kappa}$, the estimator of κ from Section 4.3. These statistics were computed over the entire continental U.S. and for each of the 45 states with 100 or more zip codes. Table 3 shows the percentiles across the $3,458$ regressions for the U.S. and the 45×3458 regressions for the states.

Recall that the variables are measured in percentiles across zip codes in the U.S., so a value of $\beta = 0.10$ implies that changing x from, say, the 50th to the 60th percentile predicts a change of y of 1 percentile; the first two rows of the table suggest that estimating β over the U.S. or across the states yield little difference. In contrast, the following rows suggest that the evidence of instability in the β coefficients is much stronger across the U.S. than within states. For example, the null of stability is rejected at the 5% level in nearly 25% of the U.S. regressions, but in less than 5% of the state-wide regressions. And the final two rows of the table suggest that this is not simply a matter of power – across the U.S. the median value of the standard deviation of $\Delta\beta$ is estimated to be 0.04 across 1000km, but within states, the corresponding value is only 0.01. Evidently, the unique features of each state have important effects on the reduced-form relationship between the variables in our dataset.

6 Concluding Remarks

This paper has developed a test for spatial variation in a spatial regression coefficient. The test is designed to control size when data are spatially correlated and to have good power when the coefficient evolves in a persistent manner over the spatial domain. Numerical experiments suggest that the test succeeds both in controlling size and detecting persistent spatial variation in regression coefficients in realistic designs.

The extension to IV regressions estimated by 2SLS is straightforward. Let $\hat{w}_l$ denote the fitted value from the first stage regression, that is, the regression of w_l onto the instruments. Let $\hat{x}_l$ denote the first element of $\hat{w}_l$ and $\hat{e}_l = y_l - w_l'\hat{\delta}_{2SLS}$ denote the IV residual. An IV version of the test statistic replaces $x_l\hat{e}_l$ with $\hat{x}_l\hat{e}_l$, and after making this change the test proceeds as described in Sections 2–4.

The paper has focused on a test for variation in a scalar coefficient, with the coefficients on additional regressors assumed to be constant. Extending the analysis to consider spatial variation for a vector-valued coefficient is conceptually straightforward—calculations involve a vector-vector score $x_l\hat{e}_l$—although the details associated with constructing a critical value that controls for multivariate spatial correlation are less straightforward. We leave this to future work.

A Proofs and Description of Numerical Calculations

Proof of Theorem 1:

From $\hat{e}_l = e_l - w_l(\hat{\delta} - \delta)$, we obtain

$$a_n^{-1/2} \sum_{l=1}^{n} r_{j,l} x_l \hat{e}_l = a_n^{-1/2} \sum_{l=1}^{n} r_{j,l} x_l e_l - \left(n^{-1} \sum_{l=1}^{n} r_{j,l} x_l w_l' \right) a_n^{-1/2} n(\hat{\delta} - \delta).$$

By (11) and Condition 3, $n^{-1} \sum_{l=1}^{n} r_{j,l} x_l w_l' \xrightarrow{p} \int \varphi_j(s) dG(s) \cdot \Sigma_{xw} = 0$, and by Condition 4, $a_n^{-1/2} n(\hat{\delta} - \delta) = O_p(1)$, so it suffices to show the claim for $a_n^{-1/2} \sum_{l=1}^{n} r_{j,l} x_l e_l$.

(a) For arbitrary $v_1, \ldots, v_q \in \mathbb{R}$, we have

$$a_n^{-1/2} \sum_{j=1}^{q} v_j Y_{n,j} = a_n^{-1/2} \sum_{l=1}^{n} r_{v,l} x_l u_l + n^{-1} \sum_{l=1}^{n} r_{v,l} x_l^2 b(s_l)$$

with $r_{v,l} = \sum_{j=1}^{q} v_j r_{j,l}$. Given (11) and Condition 3, it follows that $\sup_l |r_{v,l} - \varphi_v(s_l)| \xrightarrow{p} 0$, where $\varphi_v(s) = \sum_{j=1}^{q} v_j \varphi_j(s)$. We thus have $a_n^{-1/2} \sum_{l=1}^{n} r_{v,l} x_l u_l \Rightarrow \mathcal{N}(0, \int \varphi_v(s)^2 [\omega + (1-\omega) g(s)] g(s) ds$ by Condition 2. Furthermore, $n^{-1} \sum_{l=1}^{n} r_{v,l} x_l^2 b(s_l) = \sum_{j=1}^{m} n^{-1} \sum_{l=1}^{n} r_{v,l} x_l^2 \mathbf{1}[s_l \in \mathcal{S}_j] b(s_l)$. Now for each j, since $b : \mathcal{S}_j \mapsto \mathbb{R}$ is uniformly continuous, there exists an extension b_j of b that is continuous on the closure $\bar{\mathcal{S}}_j$ of $\mathcal{S}_j$, and by the Tietze extension theorem (Theorem 4.4. of Lang (1993)), it can further be extended to be continuous on all of $\mathcal{S}$. Furthermore, since $\partial \mathcal{S}_j$ has Lebesgue measure zero and g is bounded, for every $\varepsilon > 0$ there exists an open set $\mathcal{B}_j \supset \partial \mathcal{S}_j$ such that $\int_{\mathcal{B}_j} g(s) ds < \varepsilon$. The sets $\mathcal{S} \backslash (\mathcal{S}_j \cup \mathcal{B}_j)$ and $\bar{\mathcal{S}}_j$ are closed and disjoint, so by Urysohn's Lemma (Lemma 4.2. of Lang (1993)), there exists a continuous function $\bar{h}_j : \mathcal{S} \mapsto [0,1]$ with $\bar{h}_j(s) = 1$ for $s \in \bar{\mathcal{S}}_j$ and $\bar{h}_j(s) = 0$ for $s \in \mathcal{S} \backslash (\mathcal{S}_j \cup \mathcal{B}_j)$. Similarly, the sets $\partial \mathcal{S}_j \cup (\mathcal{S} \backslash \mathcal{S}_j)$ and $\mathcal{S}_j \backslash \mathcal{B}_j$ are closed and disjoint, so there exists a continuous function $\underline{h}_j : \mathcal{S} \mapsto [0,1]$ with $\underline{h}_j(s) = 1$ for $s \in \mathcal{S}_j \backslash \mathcal{B}_j$ and $\underline{h}_j(s) = 0$ for $s \in \partial \mathcal{S}_j \cup (\mathcal{S} \backslash \mathcal{S}_j)$. Applying Condition 3 yields

$$n^{-1} \sum_{l=1}^{n} r_{v,l} x_l^2 b_j(s_l) \bar{h}_j(s_l) \xrightarrow{p} \sigma_x^2 \int \varphi_v(s) b_j(s) \bar{h}_j(s) g(s) ds$$

and

$$\begin{aligned} \left| \int \varphi_v(s) b_j(s) \bar{h}_j(s) g(s) ds - \int_{\mathcal{S}_j} \varphi_v(s) b(s) g(s) ds \right| &\leq \int |\varphi_v(s) b_j(s)| (\bar{h}_j(s) - \mathbf{1}[s \in \mathcal{S}_j]) g(s) ds \\ &\leq \varepsilon \sup_s |\varphi_v(s) b_j(s)| \end{aligned}$$

Furthermore, again applying Condition 3,

$$\begin{aligned}
& \left| n^{-1}\sum_{l=1}^{n} r_{v,l}x_l^2 b_j(s_l)\bar{h}_j(s_l) - n^{-1}\sum_{l=1}^{n} r_{v,l}x_l^2 \mathbf{1}\left[s_l \in \mathcal{S}_j\right] b(s_l) \right| \\
\leq \; & n^{-1}\sum_{l=1}^{n} x_l^2 |r_{v,l}b_j(s_l)| \left(\bar{h}_j(s_l) - \mathbf{1}\left[s_l \in \mathcal{S}_j\right]\right) \\
\leq \; & n^{-1}\sum_{l=1}^{n} x_l^2 |r_{v,l}b_j(s_l)| \left(\bar{h}_j(s_l) - \underline{h}_j(s_l)\right) \\
& \xrightarrow{p} \sigma_x^2 \int |\varphi_v(s)b_j(s)| \left(\bar{h}_j(s_l) - \underline{h}_j(s_l)\right) g(s)ds \\
\leq \; & \varepsilon \cdot \sigma_x^2 \sup_s |\varphi_v(s)b_j(s)|.
\end{aligned}$$

But ε was arbitrary, so $n^{-1}\sum_{l=1}^{n} r_{v,l}x_l^2 \mathbf{1}\left[s_l \in \mathcal{S}_j\right] b(s_l) \xrightarrow{p} \int_{\mathcal{S}_j} \varphi_v(s)b(s)g(s)ds$, and thus $n^{-1}\sum_{l=1}^{n} r_{v,l}x_l^2 b(s_l) \xrightarrow{p} \sum_{j=1}^{m}\int_{\mathcal{S}_j} \varphi_v(s)b(s)g(s)ds = \int \varphi_v(s)b(s)g(s)ds$. The convergence $a_n^{-1/2}Y_n \Rightarrow Y_0 + B$ now follows from the Cramér-Wold device, and (14) is a consequence of the continuous mapping theorem.

(b) Let $Q_n(L) = n^{-1}\sum_{l=1} r_{v,l}x_l^2 L(s_l)$, where $r_{v,l}$ is defined in the proof of part (a). We now show that $Q_n(L) \xrightarrow{p} Q_0(L) = \sigma_x^2 \int \varphi_v(s)L(s)dG(s)$. 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 Condition 3, a contradiction.) We conclude that $\mathbf{1}[L \in \mathcal{C}_\varepsilon]Q_n(L) \xrightarrow{p} \mathbf{1}[L \in \mathcal{C}_\varepsilon]Q_0(L)$, and since $\mathbb{P}(L \in \mathcal{C}_\varepsilon) > 1 - \varepsilon$ for arbitrary $\varepsilon > 0$, also $Q_n(L) \xrightarrow{p} Q_0(L)$. □

Numerical Determination of $\mathrm{cv}_{n,q}$ in (18):

Let Ω_q be the $q \times q$ covariance matrix of $\tilde{Y}_n \sim \mathcal{N}(0, \Omega_q)$ in the $\mathcal{G}_c$ model, that is, $\Omega_q = R_q'\Sigma_c R_q$ with $R_q = (r_1, \ldots, r_q)$ and Σ_c the $n \times n$ matrix with l, ℓ element equal to $\exp(-c||s_l - s_\ell||)$. Clearly

$$\begin{aligned}
\mathbb{P}(\tilde{\xi}_q^s > \mathrm{cv}) &= \mathbb{P}(\tilde{Y}_n'(\Lambda_n - \mathrm{cv}\, I_q)\tilde{Y}_n > 0) \\
&= \mathbb{P}(Z'\Omega_q^{1/2}(\Lambda_n - \mathrm{cv}\, I_q)\Omega_q^{1/2}Z > 0)
\end{aligned}$$

$$= \mathbb{P}(\sum_{j=1}^{q} \varkappa_j Z_j^2 > 0)$$

where $\varkappa_j$ are the (real) eigenvalues of $(\Lambda_n - \mathrm{cv}\, I_q)\Omega_q$, scale normalized such that $\max_j |\varkappa_j| = 1$. Using the results of Imhof (1961), we find

$$\mathbb{P}(\sum_{j=1}^{q} \varkappa_j Z_j^2 > 0) = \tfrac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin\left(\frac{1}{2}\sum_{j=1}^{q} \arctan(\varkappa_j t)\right)}{t\prod_{j=1}^{q}(1+t^2\varkappa_j^2)^{1/4}}dt.$$

We approximate this indefinite integral by the definite integral with an upper bound of 100 and Gaussian quadrature with 500 points. Note that $0 < \mathbb{P}(\tilde{\xi}_q^s > \mathrm{cv}) < 1$ requires $\lambda_q < \mathrm{cv} < \lambda_1$, so the critical value can be obtained by simple bisection for a given c. We approximate the maximization over $c \geq c_{0.01}$ by searching over the grid $c_{0.01}\exp(5.0j/24)$, $j = 0, \ldots, 24$.

Numerical Determination of $\hat{\psi}_n$ of Section 4.3:

To ease notation, we drop the subscript n on $\hat{\psi}_n$ and $\tilde{Y}_n$ in the following. Note that any scale equivariant estimator $\hat{\psi}$ can be written in the form $\hat{\psi}(y) = \sqrt{y'y}\Psi(y/\sqrt{y'y})$, where $\Psi : S_q \mapsto \mathbb{R}$ and S_q is the surface of a q dimensional unit sphere. Define $Y^s = \tilde{Y}/\sqrt{\tilde{Y}'\tilde{Y}}$ and $U = \sqrt{\tilde{Y}'\tilde{Y}}$. Let $\theta_i = (c_i, \psi_i)$, $i = 1, \ldots, m$ be a discretization of the parameter space, with associated weights w_i. We use a grid of 5 values for c_i that approximate a uniform distribution of $\ln c$ on $[\ln c_{0.01}, 5 + \ln c_{0.01}]$ and a grid of 20 values of ψ_i that approximate a uniform distribution on $[1, 50]$, so $w_i = 1/m$ and $m = 100$. Here $\bar{\Sigma}_L$ is scale normalized to have largest eigenvalue equal to $1/\pi^2$, matching the value one would obtain for a demeaned Wiener process on the unit interval, and thus the scale of parameter variation in the time series case. With this normalization, our test has power of approximately 50% against $\psi = 15$, and the upper bound of $\psi = 50$ induces near 100% rejections, so our grid for ψ includes all degrees of spatial variation whose presence isn't entirely obvious.

We seek to solve the program

$$\min_{\delta} \sum_i w_i E_{\theta_i}[(\psi_i - \Psi(Y^s)U)_+] \quad \text{s.t.} \;\; P_{\theta_i}(\Psi(Y^s)U > \psi_i) \leq 1/2, \;\; i = 1, \ldots, m$$

where $(x)_+ = \max(x, 0)$. Write $f_\theta(y^s)$ for the density of Y^s relative to ν, the uniform measure on S_q, and $f_\theta(u|y^s)$ for the conditional density of U given Y^s. Under $\tilde{Y} \sim \mathcal{N}(0, \Omega_\theta)$, the results in Appendix B of Müller and Watson (2016) imply that

$$\begin{aligned} f_\theta(y^s) &= \tfrac{1}{2}\pi^{-q/2}|\Omega_\theta|^{-1/2}\Gamma(q/2)(y^{s\prime}\Omega_\theta^{-1}y^s)^{-q/2} \\ f_\theta(u|y^s) &= \frac{2^{-q/2+1}}{\Gamma(q/2)}(y^{s\prime}\Omega_\theta^{-1}y^s)^{q/2}u^{q-1}\exp[-\tfrac{1}{2}u^2 y^{s\prime}\Omega_\theta^{-1}y^s]. \end{aligned}$$

From a direct calculation, for $x > 0$

$$\begin{aligned} F_\theta(x, y^s) &= \int_0^x f_\theta(u|y^s)du = \frac{\gamma(q/2, x^2 y^{s\prime}\Omega_\theta^{-1} y^s/2)}{\Gamma(q/2)} \\ M_\theta(x, y^s) &= \int_0^x u f_\theta(u|y^s)du = \frac{\sqrt{2}}{\sqrt{y^{s\prime}\Sigma^{-1}y^s}} \frac{\gamma(q/2+1/2, x^2 y^{s\prime}\Omega_\theta^{-1} y^s/2)}{\Gamma(q/2)} \\ A_\theta(x, y^s) &= \int_0^x (x-u) f_\theta(u|y^s)du \\ &= x F_\theta(x, y^s) - M_\theta(x, y^s) \end{aligned}$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function. The Lagrangian corresponding to the above program equals $L_0 = \int L(y^s)d\nu(y^s)$, where

$$\begin{aligned} L(y^s) &= \sum_{i=1}^m f_{\theta_i}(y^s) \left\{ w_i \int f_{\theta_i}(u|y^s)(\psi_i - \Psi(y^s)u)_+ du + \lambda_i \left[\tfrac{1}{2} - \int f_{\theta_i}(u|y^s)\mathbf{1}[\Psi(y^s)u < \psi_i]du \right] \right\} \\ &= \sum_{i=1}^m f_{\theta_i}(y^s) \left\{ w_i \Psi(y^s) A_{\theta_i}\left(\frac{\psi_i}{\Psi(y^s)}, y^s\right) + \lambda_i \left[\tfrac{1}{2} - F_{\theta_i}\left(\frac{\psi_i}{\Psi(y^s)}, y^s\right)\right]\right\}. \end{aligned}$$

For given Lagrange multipliers $\{\lambda_i\}_{i=1}^n$, the optimal $\Psi(y^s)$ is hence obtained by minimizing the "profile function"

$$\text{prof}(d|y_s) = \sum_{i=1}^m f_{\theta_i}(y^s) \left\{ w_i d A_{\theta_i}\left(\frac{\psi_i}{d}, y^s\right) + \lambda_i \left[\tfrac{1}{2} - F_{\theta_i}\left(\frac{\psi_i}{d}, y^s\right)\right]\right\}.$$

The minimizer $\Psi(y^s)$ is approximated by evaluating $\text{prof}(d|y_s)$ on a grid of values for d, and then choosing $\Psi(y^s)$ by minimizing the quadratic approximation to $\text{prof}(d|y_s)$ around the minimizer on the grid. We use an equal spaced grid for d from $0, 1, \ldots, 50$.

The determination of the Lagrange multipliers now proceeds by computing the (approximately) best $\Psi(y^s)$ for a given set of multipliers $\{\lambda_i\}_{i=1}^m$. Then evaluate the m constraints

$$P_{\theta_i}(\Psi(Y^s)U > \psi_i) = \tfrac{1}{2} - E_{\theta_i}\left[F_{\theta_i}\left(\frac{\psi_i}{\Psi(Y^s)}, Y^s\right)\right] \leq 1/2,\ i = 1, \ldots, m$$

for that choice of $\Psi(y^s)$, slightly adjust the multipliers as a function of whether the constraint is violated or not (so that if a constraint never binds, the corresponding multiplier converges to zero), and iterate to convergence. See Appendix B of Müller and Wang (2019) for details. To speed up computations, we further approximate $E_{\theta_i}\left[F_{\theta_i}\left(\frac{\psi_i}{d}, Y^s\right)\right]$ for $d = \Psi(Y^s)$ by a linear interpolation between the two precomputed values on the grid for d. Furthermore, we approximate the expectations over Y^s by an importance sampling estimator, that is for N i.i.d. draws of Y_j^s, $j = 1, \ldots, N$

from the equal probability mixture $f_p(y^s) = m^{-1} \sum_{i=1}^{m} f_{\theta_i}(y^s)$, we approximate

$$E_{\theta_i}\left[F_{\theta_i}\left(\frac{\psi_i}{d}, Y^s\right)\right] \approx \frac{1}{N}\sum_{j=1}^{N} \frac{f_{\theta_i}(Y_j^s)}{f_p(Y_j^s)} F_{\theta_i}\left(\frac{\psi_i}{d}, Y_j^s\right)$$

with $N = 100,000$.

The numerical determination of the Lagrange multipliers for a given set of locations takes less than a minute on a modern workstation in a Fortran implementation, and the evaluation of $\hat{\psi}_n$ given the multipliers is practically instantaneous.

References

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.

Chow, G. C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605.

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

Conley, T. G., S. Gonçalves, M. S. Kim, and B. Perron (2023): "Bootstrap inference under cross-sectional dependence," *Quantitative Economics*, 14(2), 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.

Imhof, J. P. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

Jenish, N., and I. R. Prucha (2009): "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays and Random Fields," *Journal of Econometrics*, 150(1), 86–98.

Kurisu, D., K. Kato, and X. Shao (2024): "Gaussian Approximation and Spatially Dependent Wild Bootstrap for High-Dimensional Spatial Data," *Journal of the American Statistical Association*, 119(547), 1820–1832.

Lahiri, S. N. (1996): "On inconsistency of estimators based on spatial data under infill asymptotics," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 403–417.

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

Lahiri, S. N., and J. Zhu (2006): "Resampling methods for spatial regression models under a class of stochastic designs," *Annals of Statistics*, 34, 1774–1813.

Lang, S. (1993): *Real and Functional Analysis*. Springer, 3rd edn.

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*, 30(8), 1622–1643.

Müller, U. K., and Y. Wang (2019): "Nearly Weighted Risk Minimal Unbiased Estimation," *Journal of Econometrics*, 209, 18–34.

Müller, U. K., and M. W. Watson (2016): "Measuring Uncertainty about Long-Run Predictions," *Review of Economic Studies*, 83, 1711–1740.

——— (2022): "Spatial Correlation Robust Inference," *Econometrica*, 90, 2901–2935.

——— (2024): "Spatial Unit Roots and Spurious Regression," *Econometrica*, 92(5), 1661–1695.

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

Quandt, R. E. (1960): "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–330.

Shao, X. (2010): "The dependent wild bootstrap," *Journal of the American Statistical Association*, 105, 218–235.

Stock, J. H., and M. W. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

——— (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.