Spatial Correlation Robust Inference*

Ulrich K. Müller and Mark W. Watson
Department of Economics, Princeton University
Princeton, NJ, 08544

First Draft: December 2020
This Draft: May 2022

Abstract

We propose a method for constructing confidence intervals that account for many forms of spatial correlation. The interval has the familiar 'estimator plus and minus a standard error times a critical value' form, but we propose new methods for constructing the standard error and the critical value. The standard error is constructed using population principal components from a given 'worst-case' spatial correlation model. The critical value is chosen to ensure coverage in a benchmark parametric model for the spatial correlations. The method is shown to control coverage in finite sample Gaussian settings in a restricted but nonparametric class of models and in large samples whenever the spatial correlation is weak, i.e., with average pairwise correlations that vanish as the sample size gets large. We also provide results on the efficiency of the method.

Key Words: Confidence interval, HAR, HAC, Random field

JEL: C12, C20

*We thank four referees, the editor, and participants at various workshops for helpful comments and suggestions. Müller acknowledges financial support from the National Science Foundation grant SES-191336.

1 Introduction

Prompted by advances in both data availability and theory in economic geography, international trade, urban economics, development and other fields, empirical work using spatial data has become commonplace in economics. These applications highlight the importance of econometric methods that appropriately account for spatial correlation in real-world settings. While important advances have been made, researchers arguably lack practical methods that allow for reliable inference about parameters estimated from spatial data for the wide-range spatial designs and correlation patterns encountered in applied work.[1] This paper takes a step forward in this regard.

Specifically, we consider the problem of constructing a confidence interval (or test of a hypothesized value) for the mean of a spatially-sampled random variable. We propose a confidence interval constructed in the usual way, i.e., as the sample mean plus and minus an estimate of its standard error multiplied by a critical value. The novelty is that the standard error and critical value are constructed so the resulting confidence interval has the desired coverage probability (say, 95%) for a relatively wide range of correlation patterns and spatial designs. The analysis is described for the mean, but the required modifications for regression coefficients or parameters in GMM settings follow from standard arguments.

To be more precise, suppose that a random variable y is associated with a location $s \in \mathcal{S}$, where $\mathcal{S} \subset \mathbb{R}^d$. Figure 1 provides two sets of examples. Panel (a) shows three one-dimensional ($d = 1$) spatial designs. It begins with the familiar case of regularly spaced locations, corresponding to the standard time series setting; the next two examples show irregularly spaced times series with randomly selected locations drawn from a density g, where g is either uniform or triangular. Panel (b) shows two geographic examples, so $d = 2$, for the U.S. state of Texas. In the left panel, locations are randomly selected from a uniform distribution, while in the right panel locations are more likely to be sampled from areas with high economic activity, here measured by light intensity as seen from space.[2] The goal of this paper is to construct confidence intervals with desired coverage, conditional on the observed locations, for a rich set of possible locations such as those shown in the figure.

[1] Ibragimov and Müller (2010), Sun and Kim (2012) and Bester, Conley, Hansen, and Vogelsang (2016), for instance, find nontrivial size distortions of modern methods even in arguably fairly benign designs, and Kelly (2019) reports very large distortions under spatial correlations calibrated to real-world data.

[2] The light data are from Henderson, Squires, Storeygard, and Weil (2018).

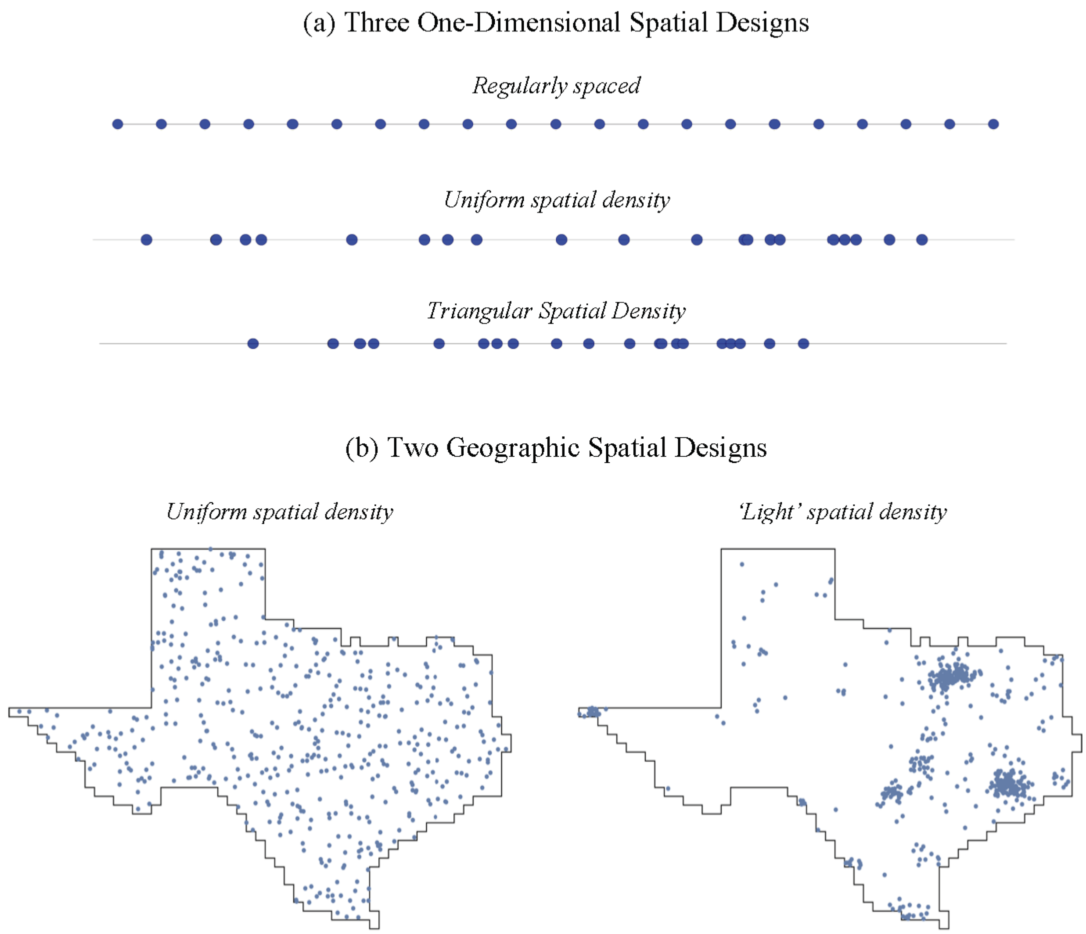

Figure 1: Examples of Spatial Designs

Adding some notation, suppose

$$y_l = \mu + u_l \text{ for } l = 1, ..., n \tag{1}$$

where y_l is associated with the observed spatial location s_l, μ is the mean of y_l, and conditional on the observed locations $\{s_l\}_{l=1}^n$, u_l is an unobserved mean-zero error that is covariance stationary, that is $\mathbb{E}[u_l u_\ell] = \sigma_u(s_l - s_\ell)$ for some covariance function $\sigma_u : \mathbb{R}^d \mapsto \mathbb{R}$. Let $\bar{y}$ denote the sample mean, and consider the usual t-statistic

$$\tau = \frac{\sqrt{n}(\bar{y} - \mu_0)}{\hat{\sigma}} \tag{2}$$

where $\hat{\sigma}^2$ is an estimator of σ^2, the variance of $\sqrt{n}(\bar{y} - \mu)$. Tests of the null hypothesis $H_0 : \mu = \mu_0$ reject when $|\tau| > \text{cv}$, where cv is the critical value, and the corresponding confidence interval for μ has endpoints $\bar{y} \pm \text{cv}\,\hat{\sigma}/\sqrt{n}$. Inference methods in this class differ in their choice of $\hat{\sigma}^2$ and critical value cv.

The case of regularly-spaced time series observations (the first example in Figure 1) is the most well-studied version of this problem. There, $\text{Var}(\sqrt{n}(\bar{y} - \mu))$ is the long-run variance of y, and traditional choices for $\hat{\sigma}^2$ are kernel-based consistent estimators such as those proposed in Newey and West (1987) and Andrews (1991), and inference uses standard normal critical values. A more recent literature initiated by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2005) accounts for the sampling uncertainty of kernel-based $\hat{\sigma}^2$ by considering "fixed-b" asymptotics where the bandwidth is a fixed fraction of the sample size, which leads to a corresponding upward adjustment of the critical value. Closely related are projection estimators of $\hat{\sigma}^2$ where the number of projections is treated as fixed in the asymptotics, as in Müller (2004), Phillips (2005), Sun (2013), and others, leading to Student-t critical values. These newer methods are found to markedly improve size control under moderate serial correlation compared to inference based on standard normal critical values. (For example, see the numerical results in Lazarus, Lewis, Stock, and Watson (2018).)

The econometrics literature on the derivation of spatial HAR inference is smaller, but has developed along similar lines: Conley (1999), Kelejian and Prucha (2007) and Kim and Sun (2011) derive consistent variance estimators, Bester, Conley, Hansen, and Vogelsang (2016) (also see Rho and Vogelsang (2019)) study the spatial analogue of the fixed-b kernel estimators, Sun and Kim (2012) suggest a spatial projection-based estimator, and Ibragimov and Müller (2010, 2015), Bester, Conley, and Hansen (2011) and Cao, Hansen, Kozbur, and Villacorta (2020) derive asymptotically justified spatial HAR inference based on a finite number of clusters.

This paper makes progress over this literature by developing a method that (i) accounts for sampling uncertainty in $\hat{\sigma}^2$; (ii) controls size under a restricted but nonparametric form of strongly correlated u_l; (iii) is asymptotically valid under generic weakly correlated u_l. The second property sets it apart from all previously mentioned methods; in a time series setting, Robinson (2005) and Müller (2014) derive inference under parametric forms of strong dependence, and Dou (2019) derives optimal inference under a non-parametric form of strong dependence under a simplifying Whittle-type approximation to the implied covariance matrices.

The remainder of the paper is organized as follows. Section 2 defines the new method. It uses a projection-type variance estimator, where the projection weights are *spatial correlation* *principal components* from a given 'worst case' benchmark correlation matrix. We correspondingly refer to the method as SCPC. Section 3 studies its small sample size control in Gaussian models. We derive a generic result about size control of t-statistics in a nonparametric class of covariance matrices, and apply it to study the robustness of SCPC under a large class of persistent processes defined in spectral terms. We note that both the basic idea of SCPC, as well as some of the results in Section 3 could potentially also be applied to settings other than (1), such as to HAR inference for data generated from spatial autoregressive models, or network data, but we do not pursue this further in this paper. Section 3 concludes with some numerical evidence on size control of SCPC under heteroskedasticity and mismeasured locations.

Section 4 studies the efficiency of the SCPC confidence interval. We compare its expected length to the length of confidence intervals derived from previously suggested spatial t-statistics, and to a lower bound that holds for all confidence intervals that, like SCPC, control size over a wide range of persistent spatial processes.

We turn to a large sample analysis in Section 5. We derive the asymptotic distribution of projection and fixed-b spatial t-statistics, including the SCPC t-statistic, and find that the density of the locations g plays a key role in their limiting distributions. This dependence is present even under weak correlation, that is, when the average correlation across observations shrinks to zero as $n \to \infty$. Notably, only when g is constant (that is, when the density is uniform) does the asymptotic distribution under weak correlation coincide with the asymptotic distribution induced by i.i.d. data. Thus, the usual suggestions for critical values, such as student-t critical values for projection t-statistics, are not generically valid under weak dependence for non-constant g. We suggest an alternative, easy-to-implement choice for the critical value that restores asymptotic validity under generic weak correlation, which is part of the definition of the SCPC method in Section 2.

Section 6 concludes with a brief discussion on how to apply SCPC in more general regression or GMM settings. Software for conducting SCPC inference for regression coefficients is available for STATA and Matlab.[3]

[3] The most recent implementation of the software is available at `https://www.princeton.edu/~mwatson/`.

2 Spatial Correlation Principal Components

This section provides details for computing the SCPC t-statistic, critical value and associated confidence interval. The definition of the SCPC t-test and critical value involves, among other things, various covariance matrices and probability calculations. We stress at the outset that these are used to describe the required calculations, and they are not assumptions about the probability distribution of the data under study. We study finite sample and asymptotic properties of the SCPC t-test under general conditions in Sections 3 and 5 below.

Let $\mathbf{y} = (y_1, y_2, ..., y_n)'$ and similarly for $\mathbf{s} = (s_1, s_2, ..., s_n)'$, $\mathbf{u} = (u_1, u_2, ..., u_n)'$ and the vector of residuals $\hat{\mathbf{u}} = (\hat{u}_1, \hat{u}_2, ..., \hat{u}_n)'$. Let $\mathbf{l}$ denote an $n \times 1$ vector of 1s, and $\mathbf{M} = \mathbf{I} - \mathbf{l}(\mathbf{l}'\mathbf{l})^{-1}\mathbf{l}'$. Consider a benchmark Gaussian 'exponential' covariance matrix for $\mathbf{u}$ with covariance function $\mathbb{E}[u_l u_\ell] = \exp(-c||s_l - s_\ell||)$ for $c > 0$. (Because the t-statistic is scale invariant, the assumption that $\mathbb{E}[u_l^2] = 1$ is without loss of generality.) Let $\mathbf{\Sigma}(c)$ denote the $n \times n$ covariance matrix of $\mathbf{u}$ in this model. Let c_0 denote a predetermined value of c that is meant to capture an upper bound on the spatial persistence in the data. (The choice of c_0 is discussed below). Let $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n$ denote the eigenvectors of $\mathbf{M\Sigma}(c_0)\mathbf{M}$ corresponding to the eigenvalues ordered from largest to smallest, and normalized so that $n^{-1}\mathbf{r}_j'\mathbf{r}_j = 1$ for all j. The scalar variable $n^{-1/2}\mathbf{r}_j'\hat{\mathbf{u}}$ has the interpretation as the jth population principle component of $\hat{\mathbf{u}}|\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{M\Sigma}(c_0)\mathbf{M})$. The SCPC estimator of σ^2 based on the first q of these principal components is

$$\hat{\sigma}^2_{\text{SCPC}}(q) = q^{-1} \sum_{j=1}^{q} (n^{-1/2}\mathbf{r}_j'\hat{\mathbf{u}})^2, \tag{3}$$

and the corresponding SCPC t-statistic is

$$\tau_{\text{SCPC}}(q) = \frac{\sqrt{n}(\bar{y} - \mu_0)}{\hat{\sigma}_{\text{SCPC}}(q)}. \tag{4}$$

The critical value $\text{cv}_{\text{SCPC}}(q)$ of the level-α SCPC test is chosen so that size is equal to α under the Gaussian benchmark model with $c \geq c_0$. That is, $\text{cv}_{\text{SCPC}}(q)$ satisfies

$$\sup_{c \geq c_0} \mathbb{P}^0_{\mathbf{\Sigma}(c)}(|\tau_{\text{SCPC}}(q)| > \text{cv}_{\text{SCPC}}(q)|\mathbf{s}) = \alpha, \tag{5}$$

where $\mathbb{P}^0_{\mathbf{\Sigma}}$ means that the probability is computed under the null hypothesis in the Gaussian model with covariance matrix $\mathbf{\Sigma}$, $\mathbf{y}|\mathbf{s} \sim \mathcal{N}(\mathbf{l}\mu_0, \mathbf{\Sigma}(c))$.

The final ingredient in the method is the choice of q. Let $\mathbb{E}_{\mathbf{\Sigma}=\mathbf{I}}[2\hat{\sigma}_{\text{SCPC}}(q)\,\text{cv}_{\text{SCPC}}(q)|\mathbf{s}]$ denote the expected length of the confidence interval constructed using $\tau_{\text{SCPC}}(q)$ under the

Gaussian i.i.d. model $\mathbf{y}|\mathbf{s} \sim \mathcal{N}(\mathbf{l}\mu, \mathbf{I})$. SCPC chooses q to make this expected length as small as possible, that is q_{SCPC} solves

$$\min_{q\geq 1} \mathbb{E}_{\mathbf{\Sigma}=\mathbf{I}}[2\hat{\sigma}_{\text{SCPC}}(q)\,\text{cv}_{\text{SCPC}}(q)|\mathbf{s}] = \min_{q\geq 1} \sqrt{8}n^{-1/2}q^{-1/2}\,\text{cv}_{\text{SCPC}}(q)\frac{\Gamma((q+1)/2)}{\Gamma(q/2)} \tag{6}$$

with the equality exploiting the fact that $q\hat{\sigma}^2_{\text{SCPC}}(q)|\mathbf{s} \sim \chi^2_q$ in the Gaussian i.i.d. model.

Remark 2.1. The primary concern in the construction of $\hat{\sigma}^2$ is downward bias. Recall that the eigenvector $\mathbf{r}_1$ maximizes $\mathbf{h}'\mathbf{M}\mathbf{\Sigma}(c_0)\mathbf{M}\mathbf{h}$ among all vectors $\mathbf{h}$ of the same length, the second eigenvector $\mathbf{r}_2$ maximizes $\mathbf{h}'\mathbf{M}\mathbf{\Sigma}(c_0)\mathbf{M}\mathbf{h}$ subject to $\mathbf{h}'\mathbf{r}_1 = 0$, and so forth, and for any $q \geq 1$, the $n \times q$ matrix $(\mathbf{r}_1, \ldots, \mathbf{r}_q)$ maximizes $\operatorname{tr} \mathbf{H}'\mathbf{M}\mathbf{\Sigma}(c_0)\mathbf{M}\mathbf{H}$ among all $n \times q$ matrices $\mathbf{H}$ with $n^{-1}\mathbf{H}'\mathbf{H} = \mathbf{I}_q$. Thus, the SCPC method selects the linear combinations of $\hat{\mathbf{u}}$ in the estimator of σ^2 that have the largest variance in the benchmark model with $c = c_0$, under the constraint of being unbiased in the i.i.d. model.

Remark 2.2. The choice of q trades off the downward bias in $\hat{\sigma}^2_{\text{SCPC}}(q)$ that occurs when q is large and its large variance when q is small. Both bias and variance lead to a large critical value, and (6) leads to a choice of q that optimally trades off these two effects to obtain the shortest possible expected confidence interval length in the i.i.d. model. In Section 4 we consider an alternative choice of q that minimizes expected length under $c = 2c_0$.

Remark 2.3. SCPC requires that the researcher chooses a value for c_0 which represents the highest degree of spatial correlation allowed by the method. One way to calibrate c_0 is via the average pairwise correlation of the spatial observations

$$\bar{\rho} = \frac{1}{n(n-1)} \sum_{l=1}^{n} \sum_{\ell \neq l} \text{Cor}\left(y_l, y_\ell \,|\mathbf{s}_n\right)$$

that is, c_0 is chosen so that it implies a given value $\bar{\rho}_0$ of $\bar{\rho}$. For example, $\bar{\rho}_0 = (0.003, 0.01, 0.03, 0.10)$ implies very weak, weak, strong and very strong correlation, respectively.

To put these values into perspective, recognize that the standard deviation of $\bar{y}$ relative to its value under i.i.d. sampling, say γ_n, satisfies $\gamma_n^2 = \text{Var}[\sqrt{n}\bar{y}]/\text{Var}[y_l] = 1 + (n-1)\bar{\rho}$, and therefore γ_n measures the increase in the length of the confidence interval with σ known relative to its i.i.d. counterpart. The parameter γ_n also governs the size distortion associated with using the standard t-statistic (i.e., based on i.i.d. sampling) when y is spatially correlated; for example, the rejection frequency for a nominal 5% level test is approximately $\mathbb{P}(|Z| > 1.96/\gamma_n)$ with $Z \sim \mathcal{N}(0,1)$. With $n = 500$, $\bar{\rho} = (0.003, 0.01, 0.03, 0.10)$ yields $\gamma_n = (1.6, 2.4, 4.0, 7.1)$ and approximate rejection frequencies of $(0.21, 0.42, 0.62, 0.78)$ using t-statistics constructed under an erroneous i.i.d. assumption.

Alternatively, in the equally spaced time series model, note that γ_n^2 is the long-run variance of the process in multiples of its variance. For an AR(1) process with coefficient ϕ_n, $\gamma_n^2 = (1+\phi_n)/(1-\phi_n)$ and $\phi_n \approx 1-(2/\bar{\rho})n^{-1}$ for large n. Using $n = 500$ and the four values of $\bar{\rho}$, $\phi_{500} = (0.43, 0.72, 0.88, 0.96)$. In their study of HAR inference in time series, Lazarus, Lewis, Stock, and Watson (2018) considered models with $n = 200$ and $\phi = 0.7$, corresponding to $\bar{\rho} \approx 0.03$.

Remark 2.4. In the regular spaced time series case, the SCPC eigenvectors, $\mathbf{r}_j$ are numerically close to the weights of the equal weighted cosine (EWC) projection estimator considered in Müller (2004, 2007), Lazarus, Lewis, Stock, and Watson (2018) and Dou (2019). This is not surprising, since the corresponding cosines are the limit of the eigenvectors of $\mathbf{M\Sigma}(c_0)\mathbf{M}$ as $c_0 \to 0$ (cf. Theorem 1 of Müller and Watson (2008)). What is more, the SCPC choice of q is also numerically close to the corresponding optimal choice of q in Dou (2019). So when applied to time series, SCPC comes close to replicating Dou's (2019) suggestion for optimal inference, with c_0 representing the upper bound for the degree of persistence. The same is true in a spatial design in $\mathbb{R}^d$, with arbitrary d, if the locations happens to fall on a line segment with approximately uniform empirical distribution.

Remark 2.5. The SCPC method with c_0 calibrated by a choice of $\bar{\rho}_0$ is invariant to the scale of the locations $\{s_l\}_{l=1}^n \mapsto \{as_l\}_{l=1}^n$ for $a > 0$, and (in contrast to Sun and Kim's (2012) and Conley's (1999) suggestion) also to arbitrary distance preserving transformations, such as rotations.

Remark 2.6. We suggest determining q_{SCPC} for $\alpha = 5\%$, and then using the same q_{SCPC} at all other significance levels α, and for the computation of p-values. This avoids discontinuities that arise from the dependence of q_{SCPC} on α, and is computationally convenient. In the following we excusively focus on 5% level tests.

Remark 2.7. The discussion focuses on inference about the mean, but all the results extend to regression and GMM problems using standard arguments. For example, in the simple linear regression $w_l = x_l\beta + \varepsilon_l$, where β is the parameter of interest, $x_l\varepsilon_l$ replaces u_l in the analysis,

Table 1: SCPC for Different Choices of $\bar{\rho}_0$ in the U.S. States Spatial Designs

	$\bar{\rho}_0 = 0.003$	$\bar{\rho}_0 = 0.01$	$\bar{\rho}_0 = 0.03$	$\bar{\rho}_0 = 0.10$
	Uniform Spatial Designs $g = g_{\text{uniform}}$			
$\Delta_{1/2}(c_0)$ in %	$\langle 0.7, 1.0, 1.1\rangle$	$\langle 1.3, 1.8, 2.1\rangle$	$\langle 2.5, 3.4, 3.9\rangle$	$\langle 5.4, 7.0, 8.0\rangle$
q_{SCPC}	$\langle 38, 42, 46\rangle$	$\langle 11, 12, 13\rangle$	$\langle 8, 8, 9\rangle$	$\langle 5, 6, 6\rangle$
Expected length	$\langle 1.02, 1.02, 1.03\rangle$	$\langle 1.10, 1.12, 1.13\rangle$	$\langle 1.28, 1.30, 1.31\rangle$	$\langle 1.64, 1.68, 1.70\rangle$
	"Light" Spatial Designs $g = g_{\text{light}}$			
$\Delta_{1/2}(c_0)$ in %	$\langle 0.2, 0.4, 0.7\rangle$	$\langle 0.5, 0.9, 1.5\rangle$	$\langle 0.9, 1.9, 3.0\rangle$	$\langle 2.3, 4.6, 6.9\rangle$
q_{SCPC}	$\langle 44, 47, 50\rangle$	$\langle 14, 18, 20\rangle$	$\langle 6, 8, 9\rangle$	$\langle 4, 5, 6\rangle$
Expected length	$\langle 1.02, 1.02, 1.02\rangle$	$\langle 1.05, 1.06, 1.07\rangle$	$\langle 1.12, 1.16, 1.24\rangle$	$\langle 1.38, 1.51, 1.61\rangle$

Notes: Entries are 5th, 50th and 95th percentiles of the distribution of across the 240 location draws in the U.S. states spatial design. $\Delta_{1/2}(c_0)$ is the distance that leads to a correlation of 1/2 measured in multiples of largest distance in sample, and expected length is computed in the i.i.d. model and measured in multiples of the known-σ interval length $2 \cdot 1.96\sigma/\sqrt{n}$.

and the test can be constructed as described above using $y_l = \hat{\beta} + x_l\hat{\varepsilon}_l/(n^{-1}\sum_{i=1}^{n} x_i^2)$, where $\hat{\beta}$ is the OLS estimator and $\hat{\varepsilon}_l$ is the residual. Details are provided in Section 6.

U.S. states spatial designs: SCPC inference is conditioned on the value of the locations, $\mathbf{s}$, observed in the sample. To gauge how well the method is likely to perform in applications, we use 480 different values of $\mathbf{s}$. The values are generated by randomly drawing $n = 500$ locations within the boundaries of each of the 48 contiguous U.S. states. The density of locations g within each state is either uniform (g_{uniform}), or it is proportional to light measured from space (g_{light}) as a proxy for economic activity; the bottom panel of Figure 1 shows two values of $\mathbf{s}$ that were drawn using Texas. We draw five sets of $n = 500$ independent locations under each density $g \in \{g_{\text{uniform}}, g_{\text{light}}\}$ for a total of 240 (= 48 states × 5 location draws) sets of locations $\mathbf{s} = \{s_l\}_{l=1}^{500}$ using g_{uniform} and 240 using g_{light}.

Table 1 reports the 5th, 50th and 95th percentiles of selected SCPC properties across these 240 location draws for different values of $\bar{\rho}_0$ for each $g \in \{g_{\text{uniform}}, g_{\text{light}}\}$. In the table and throughout the paper, we use the notation $\langle\cdot,\cdot,\cdot\rangle$ to indicate these three quantiles of some statistic that describes each location $\mathbf{s}$. The first row of the table shows the quantiles of the 'half-life' distance $\Delta_{1/2}(c_0)$ satisfying $\exp(-c_0||r - s||) = 1/2$ whenever $||r - s|| = \Delta_{1/2}(c_0)$, measured in multiples of the largest distance $\Delta_{\max} = \max_{l,\ell}||s_l - s_\ell||$. For example, when $\bar{\rho}_0 = 0.03$, the median half-life distance is 1.9% of the maximum distance across the 240 values of $\mathbf{s}$ generated from the g_{light} density. The next row of the table shows the quantiles for the values of q_{SCPC} chosen by (6), and the final row shows the implied expected length of the SCPC confidence interval relative to the length of the known-σ interval with endpoints $\bar{y}\pm 1.96\sigma/\sqrt{n}$. The results shown in the table indicate, for example, that a researcher using the SCPC t-statistic chosen to accommodate spatial correlation as large as $\bar{\rho}_0 = 0.03$ will typically use $q \approx 8$ principal components and the resulting confidence interval will be, on average, roughly 20% to 30% longer than the known-σ confidence interval. This is slightly larger than the Student-t confidence interval using $q = 8$ principal components in an i.i.d. model because SCPC is "bias aware" and chooses the critical value to control size under $\bar{\rho}_0 = 0.03$.

Remark 2.8. The U.S. states spatial designs will be used throughout the paper to illustrate the properties of the SCPC t-statistic.

Remark 2.9. The supremum over $c \geq c_0$ in (5) plays an important role to guarantee asymptotic size control under weak correlations; see Section 5.4 below. At the same time, as one might intuit, in most designs, the condition binds at the smallest value $c = c_0$. In the U.S. states spatial designs, the null rejection probabilities of SCPC under c_0 have percentiles $\langle 5.0\%, 5.0\%, 5.0\%\rangle$ and $\langle 4.7\%, 5.0\%, 5.0\%\rangle$ for $g = g_{\text{uniform}}$ and $g = g_{\text{light}}$, respectively. The condition doesn't always bind at $c = c_0$ because $\bar{y}$ and $\hat{\sigma}_{\text{SCPC}}(q)$ are in general dependent, a feature that is discussed more in Section 4.

3 Finite-Sample Size Control in Gaussian Models

In this section, we study the size control of spatial t-statistics in Gaussian models where $\mathbf{y} \sim \mathcal{N}(\mathbf{l}\mu, \mathbf{\Sigma})$ for some $\mathbf{\Sigma}$. Conditioning on the locations $\mathbf{s}$ is implicit. While our main interest is on the SCPC t-statistic, many of our results apply more generally to t-statistics (2) with a quadratic form estimator of $\hat{\sigma}^2$,

$$\tau(\mathbf{W}\mathbf{W}') = \frac{\sqrt{n}(\bar{y}-\mu_0)}{\hat{\sigma}} = \frac{\mathbf{l}'(\mathbf{y}-\mu_0\mathbf{l})}{\sqrt{\mathbf{y}'\mathbf{W}\mathbf{W}'\mathbf{y}}} \qquad \hat{\sigma}^2 = n^{-1}\mathbf{y}'\mathbf{W}\mathbf{W}'\mathbf{y} = n^{-1}\mathbf{u}'\mathbf{W}\mathbf{W}'\mathbf{u} \tag{7}$$

for some $n \times q$ matrix $\mathbf{W}$, $1 \leq q \leq n-1$ satisfying $\mathbf{W}'\mathbf{l} = 0$. Note that for any positive semi-definite $n \times n$ matrix $\mathbf{Q}$, $\hat{\sigma}^2 = n^{-1}\hat{\mathbf{u}}'\mathbf{Q}'\hat{\mathbf{u}}$ can be represented in this way. For future reference, it will be useful to define the $n \times (q+1)$ matrix $\mathbf{W}^0 = [\mathbf{l}, \mathbf{W}]$.

By construction, SCPC controls size in exponential Gaussian models with $c \geq c_0$, that is in exponential models with spatial persistence less than the c_0-benchmark model. Our goal in this section is to investigate SCPC size control for covariance matrices outside of this exponential class. Thus, let $\mathcal{V}$ denote a set of covariance matrices. A test using the t-statistic

$\tau^2(\mathbf{WW}')$ with critical value cv and level α controls size under $\mathcal{V}$ if $\sup_{\mathbf{\Sigma}\in\mathcal{V}} \mathbb{P}^0_{\mathbf{\Sigma}}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2) \leq \alpha$.

For our purposes, the interesting set of covariance matrices $\mathcal{V}$ are those that exhibit less spatial persistence than $\mathbf{\Sigma}(c_0)$ (recall that c_0 was chosen to represent an upper bound on persistence in the data). In time series data, long-run persistence is intimately related to the slope of the spectrum near frequency zero. An analogous result holds for spatial persistence, and in Section 3.2 we use this to characterize a set of covariance matrices $\mathcal{V}$ with less spatial persistence than $\mathbf{\Sigma}(c_0)$. While the resulting $\mathcal{V}$ is a nonparametric set of covariance matrices, we show that elements in $\mathcal{V}$ can be represented as mixtures of covariance matrices in a parametric class, say $\mathbf{\Sigma}^p(\theta)$, $\theta \in \Theta$, that is $\mathcal{V} = \{\mathbf{\Sigma} : \mathbf{\Sigma} = \int_\Theta \mathbf{\Sigma}^p(\theta) d\Pi(\theta)$ for some probability distribution $\Pi\}$. This motivates studying size control over arbitrary mixtures of a set of parametric covariance matrices $\mathbf{\Sigma}^p(\theta)$—see Theorem 2 below.

The next two subsections carry out this analysis, and we find that SCPC controls size over a large class of processes that are less persistent than the $\mathbf{\Sigma}(c_0)$ worst-case benchmark in the U.S. states spatial designs. Section 3.3 briefly analyzes the null rejection properties of SCPC under heteroskedasticity and mismeasured locations.

3.1 Generic Results

The following is a useful result for computing the null rejection frequency of $\tau^2(\mathbf{WW}')$ for a given covariance matrix $\mathbf{\Sigma}$.

Lemma 1. *Assume* $\mathbf{y} \sim \mathcal{N}(\mathbf{l}\mu_0, \mathbf{\Sigma})$ *and let* $\mathbf{\Omega} = \mathbf{W}^{0\prime}\mathbf{\Sigma}\mathbf{W}^0$*. For* $\mathrm{cv} > 0$*, define* $\mathbf{D}(\mathrm{cv}) = \mathrm{diag}(1, -\mathrm{cv}^2\,\mathbf{I}_q)$ *and* $\mathbf{A} = \mathbf{D}(\mathrm{cv})\mathbf{\Omega}$*, and let* $(\omega_0, \omega_1, ..., \omega_q)$ *denote the eigenvalues of* $\mathbf{A}$ *ordered from largest to smallest. Then with* $(Z_0, Z_1, ..., Z_q) \sim \mathcal{N}(0, \mathbf{I}_{q+1})$*,*

(i) $\omega_0 > 0$*, and* $\omega_i \leq 0$ *for* $i = 1, ..., q$*;*

(ii) $\mathbb{P}^0_{\mathbf{\Sigma}}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2) = \mathbb{P}(\sum_{i=0}^q \omega_i Z_i^2 > 0) = \frac{1}{\pi}\int_0^1 x^{\frac{q-1}{2}}(1-x)^{1/2}\prod_{i=1}^q (x - (\omega_i/\omega_0))^{1/2} dx$*.*

Remark 3.1. Result (i) and the first equality in (ii) follow from standard calculations. The final equality in (ii) is shown in Bakirov and Székely (2005); this result makes it straightforward to compute the null rejection frequency by evaluating the integral via numerical quadrature.

We now turn to an analytic result about size control for a set $\mathcal{V}$ of covariance matrices with elements that are a mixture of covariance matrices from a parametric class. Specifically, suppose for a given $\mathbf{\Sigma}_0$, cv is such that $\mathbb{P}^0_{\mathbf{\Sigma}_0}(\tau^2(\mathbf{WW}') > \text{cv}^2) = \alpha$. Let $\mathbf{\Sigma}^p(\theta)$, $\theta \in \Theta$ be a parametric class of covariance matrices. We seek conditions under which

$$\mathbb{P}^0_{\mathbf{\Sigma}_1}(\tau^2(\mathbf{WW}') > \text{cv}^2) \leq \alpha \text{ for } \mathbf{\Sigma}_1 = \int_\Theta \mathbf{\Sigma}^p(\theta) d\Pi(\theta) \tag{8}$$

for a probability distribution Π. Let $\lambda_j(\cdot)$ denote the jth largest eigenvalue of some matrix.

Theorem 2. *Let $\mathbf{\Omega}_0 = \mathbf{W}^{0\prime}\mathbf{\Sigma}_0\mathbf{W}^0$, $\mathbf{\Omega}(\theta) = \mathbf{W}^{0\prime}\mathbf{\Sigma}^p(\theta)\mathbf{W}^0$, and assume $\mathbf{\Omega}_0$ and $\mathbf{\Omega}(\theta)$, $\theta \in \Theta$ are full rank. Suppose $\mathbf{A}_0 = \mathbf{D}(\text{cv})\mathbf{\Omega}_0$ is diagonalizable, and let $\mathbf{P}$ be its eigenvectors. Let $\mathbf{A}(\theta) = \mathbf{P}^{-1}\mathbf{D}(\text{cv})\mathbf{\Omega}(\theta)\mathbf{P}$ and $\bar{\mathbf{A}}(\theta) = \frac{1}{2}(\mathbf{A}(\theta) + \mathbf{A}(\theta)')$. Suppose $\mathbf{A}_0$ and $\mathbf{A}(\theta)$, $\theta \in \Theta$ are scale normalized such that $\lambda_1(\mathbf{A}_0) = \lambda_1(\mathbf{A}(\theta)) = 1$. Let*

$$\begin{aligned}
\nu_1(\theta) &= \lambda_q(-\bar{\mathbf{A}}(\theta)) - \lambda_1(\bar{\mathbf{A}}(\theta))\lambda_q(-\mathbf{A}_0) - (\lambda_1(\bar{\mathbf{A}}(\theta)) - 1) \\
\nu_i(\theta) &= \lambda_{q+1-i}(-\bar{\mathbf{A}}(\theta)) - \lambda_1(\bar{\mathbf{A}}(\theta))\lambda_{q+1-i}(-\mathbf{A}_0) \text{ for } i = 2, \dots, q.
\end{aligned}$$

If $\sum_{i=1}^j \nu_i(\theta) \geq 0$ for all $\theta \in \Theta$ and $1 \leq j \leq q$, then (8) holds for all Π.

Remark 3.2. The theorem is based on the following logic: First, as shown in Lemma 1, the eigenvalues of $\mathbf{A}_0$ and $\mathbf{A}(\theta)$ (or, equivalently, of $\mathbf{D}(\text{cv})\mathbf{\Omega}(\theta)$) govern the rejection probability of $\tau^2(\mathbf{WW}')$ under $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}^p(\theta)$. Given the scale normalization $\lambda_1(\mathbf{A}_0) = \lambda_1(\mathbf{A}(\theta)) = 1$, if $\lambda_j(\mathbf{A}(\theta)) \leq \lambda_j(\mathbf{A}_0)$ for all $j \geq 2$, then, using the notation in Lemma 1, $\omega_i(\mathbf{A}(\theta)) \leq \omega_i(\mathbf{A}_0)$ which yields $\mathbb{P}^0_{\mathbf{\Sigma}^p(\theta)}(\tau^2(\mathbf{WW}') > \text{cv}^2) \leq \mathbb{P}^0_{\mathbf{\Sigma}_0}(\tau^2(\mathbf{WW}') > \text{cv}^2)$. Second, the integral representation in part (ii) of Lemma 1 can be used to show that the null rejection probability of the t-statistic is Schur convex in these negative eigenvalues, so that the inequality holds whenever the negative eigenvalues of $\mathbf{A}(\theta)$ weakly majorize those of $\mathbf{A}_0$. Majorization inequalities about eigenvalues of sums of matrices and additional calculations then extend this further to the result in Theorem 2.

Remark 3.3. In the appendix we prove a more general result: If for some probability distribution Π on Θ,

$$\sum_{i=1}^j \int \nu_i(\theta) d\Pi(\theta) \geq 0 \text{ for all } 1 \leq j \leq q, \tag{9}$$

then (8) holds. The conditions stated in Theorem 2 guarantee that (9) holds for all Π.

Remark 3.4. If for some $\theta_0 \in \Theta$, $\mathbf{\Sigma}_0 = \mathbf{\Sigma}^p(\theta_0)$, then $\nu_i(\theta_0) = 0$ for $1 \leq j \leq q$, so the inequalities of the theorem have no 'minimal slack' and potentially apply also to parametric models with a covariance matrix $\mathbf{\Sigma}^p(\theta)$ that takes on values arbitrarily close to $\mathbf{\Sigma}_0$.

3.2 SCPC Size Control Under Alternative Forms of Persistence

The problem of estimating the variance of $\bar{y}$ is intimately linked to the properties of the spectral density close to zero. In the time series case, the long-run variance (that is, the variance of $\sqrt{n}\bar{y}$) converges to the spectral density at frequency zero, multiplied by 2π, as $n \to \infty$ for a large class of weakly dependent stationary processes. From this perspective, the aim of correlation robust inference is to extract information about the variance of $\bar{y}$ by extrapolating the observed variability of weighted averages that contain information about the spectrum close to the origin, such as low-frequency periodogram ordinates. Such an extrapolation can only be successful under some *a priori* smoothness of the spectral density close to zero (cf. Pötscher (2002)), so in this perspective, specification of a worst case benchmark model amounts to the specification of a bound on the smoothness of the spectral density close to zero. This motivates an application of Theorem 2 to a class of covariance matrices that is defined in terms of a class of underlying spectral densities.

3.2.1 Spatial Case

If the covariance function σ_u in (1) is isotropic, then its spectrum $\mathfrak{f} : \mathbb{R}^d \mapsto [0, \infty)$ at frequency $\boldsymbol{\omega} \in \mathbb{R}^d$ can be written as function of the scalar $\omega = ||\boldsymbol{\omega}||$, that is $\mathfrak{f}(\boldsymbol{\omega}) = f(\omega)$ for some $f : \mathbb{R} \mapsto [0, \infty)$. Since the null rejection probability of spatial t-statistics does not depend on the scale of σ_u, it is without loss of generality to normalize $f(0) = 1$. The spectrum of the benchmark covariance function $\exp(-c||s - r||)$ is

$$f_c^{\text{bnch}}(\omega) = \frac{c^3}{(c^2 + \omega^2)^{3/2}}. \tag{10}$$

By construction, SCPC controls size in the benchmark model with $c \geq c_0$, and $f_0 = f_{c_0}^{\text{bnch}}$ is the spectral density with the steepest decline at the origin in the benchmark model. A spectral density f would naturally be considered less persistent than f_0 if $r(\omega) = f(\omega)/f_0(\omega)$ is (weakly) monotonically increasing in $|\omega|$, since this implies that f has relatively more mass at higher frequencies.

Note that any symmetric function $r : \mathbb{R} \mapsto \mathbb{R}$ with $r(0) = 1$ that is increasing in $|\omega|$ with $\lim_{\omega \to \infty} r(\omega) = M \geq 1$ can be written in the form $r(\omega) = 1 + (M - 1)\Pi(|\omega|) = \Pi(|\omega|) + M(1 - \Pi(|\omega|))$ for some CDF Π on $[0, \infty)$. Since $\Pi(|\omega|) = \int \mathbf{1}[\theta \leq |\omega|] d\Pi(\theta)$ and $1 - \Pi(|\omega|) = \int \mathbf{1}[\theta > |\omega|] d\Pi(\theta)$, any such r can therefore be written as the mixture $r(\omega) = \int r_\theta^{\text{step}}(\omega) d\Pi(\theta)$ with $r_\theta^{\text{step}}(\omega) = \mathbf{1}[\theta \leq |\omega|] + M \cdot \mathbf{1}[\theta > |\omega|]$. Moreover, if we define

$r_0^{\text{step}}(\omega) = 1$, then by letting Π have some mass on $\theta = 0$, we can further induce any value for $\lim_{\omega\to\infty} \int r_\theta^{\text{step}}(\omega) d\Pi(\theta)$ smaller or equal to M.

Rewriting these representations in terms of $f(\omega) = r(\omega) f_0(\omega)$ thus yields that any f such that $f(\omega)/f_0(\omega)$ is (weakly) monotonically increasing in $|\omega|$ and $\lim_{\omega\to\infty} f(\omega)/f_0(\omega) \leq M$ can be written as a mixture $f(\omega) = \int f_\theta^{\text{step}}(\omega) d\Pi(\theta)$, where $f_0^{\text{step}} = f_0$ and for $\theta > 0$

$$f_\theta^{\text{step}}(\omega) = \mathbf{1}[|\omega| \leq \theta] f_0(\omega) + \mathbf{1}[|\omega| > \theta] M \cdot f_0(\omega). \tag{11}$$

Here, $f_\theta^{\text{step}}(\omega)$ is equal to the benchmark spectrum $f_0(\omega)$ for $\omega \leq \theta$ and jumps to $M \cdot f_0(\omega)$ for larger values of ω. Let $\mathbf{\Sigma}_\theta^{\text{step}}$ be the covariance matrices induced by f_θ^{step}, $\theta \geq 0$.

Since SCPC controls size at $\mathbf{\Sigma}_0 = \mathbf{\Sigma}(c_0)$, one can apply Theorem 2 to the SCPC t-statistic with $\bar{\rho} = 0.03$ to the parametric class $\mathbf{\Sigma}_\theta^{\text{step}}$ in the U.S. states spatial designs. Numerical experimentation shows that for $g = g_{\text{uniform}}$, we may choose $M = 10$ for all 240 locations $\mathbf{s}$. Thus, in those designs, SCPC controls size under all isotropic spectral densities $\mathfrak{f}(\boldsymbol{\omega}) = f(||\boldsymbol{\omega}||)$ such that $f(\omega)/f_0(\omega)$ is monotonically increasing in $|\omega|$ with $\lim_{\omega\to\infty} f(\omega)/f_0(\omega) \leq 10$.

It turns out that for some location draws generated under $g = g_{\text{light}}$, some of the $\nu_j(\theta)$ defined in Theorem 2 are negative. So instead, we let f_0 in (11) be flatter than $f_{c_0}^{\text{bnch}}$, weakening the claim about size control. In particular, we let $f_0 = f_{\tilde{c}_0}^{\text{bnch}}$, with $\tilde{c}_0 > c_0$, and determine for what kind of values of $\tilde{c}_0$ the claim holds again for $M = 10$. Across the 240 locations generated under $g = g_{\text{light}}$, the percentiles of the ratios $\tilde{c}_0/c_0$ are $\langle 1.00, 1.04, 1.18 \rangle$, so SCPC controls size for a large class of spectral densities that are nearly as steep as $f_{c_0}^{\text{bnch}}$ close to the origin.

3.2.2 Regularly-Spaced Time Series Case

A particularly interesting application of these ideas is the familiar time series case with $s_l = l/n \in \mathcal{S} = [0, 1]$. The benchmark model then simply becomes an AR(1) process with coefficient $\phi_n = e^{-c/n}$. Due to aliasing, the spectral density, h, of a stationary regularly-spaced time series is usefully defined on the interval $[-\pi, \pi]$, $h : [-\pi, \pi] \mapsto \mathbb{R}$. We again normalize $h(0) = 1$. The corresponding benchmark spectral density is proportional to

$$h_c^{\text{bnch}}(\lambda) \propto \frac{1}{(1 + \phi_n^2 - 2\phi_n \cos(\lambda))}, \quad \lambda \in [-\pi, \pi].$$

As in the spatial case, a spectral density h would naturally be considered less persistent than h_0 if $h(\lambda)/h_0(\lambda)$ is (weakly) monotonically increasing in $|\lambda|$, motivating the consideration of mixtures of $h_\theta^{\text{step}}(\lambda) = \mathbf{1}[|\lambda| \leq \theta] h_0(\omega) + \mathbf{1}[|\lambda| > \theta] M \cdot h_0(\omega)$.

Numerical experimentation using the expressions in Theorem 2 now shows that the SCPC t-statistic applied to the time series case controls size for these mixtures for $M = 5$ and $h_0 = h_{\tilde{c}_0}^{\text{bnch}}$ with $\tilde{c}_0 = 1.03c_0$ for $n \in \{50, 100, 200, 500\}$.

Remark 3.5. Taking limits as $n \to \infty$ yields a corresponding asymptotic robustness statement: The function $f_0(\omega) = \lim_{n\to\infty} h_0(\omega/n) = \tilde{c}_0^2/(\omega^2 + \tilde{c}_0^2)$ is the 'local-to-zero' spectral density (cf. Müller and Watson (2016, 2017)) of a local-to-unity process with parameter $\tilde{c}_0$. Consider any process with spectral density $h = h_n$ whose local-to-zero spectral density $f(\omega) = \lim_{n\to\infty} h_n(\omega/n)$ is such that $f(\omega)/f_0(\omega)$ is monotonically increasing in $|\omega|$ with $\lim_{\omega\to\infty} f(\omega)/f_0(\omega) \leq 5$ and that satisfies the CLT in Müller and Watson (2016, 2017). Application of Theorem 2 then implies that the SCPC t-test controls asymptotic size for all such processes.

3.3 SCPC Size Control Under Heteroskedasticity and Mismeasured Locations

We now briefly study size control of SCPC in the U.S. states spatial designs if either the variance of u_l is a function of the location, or the locations are mismeasured.

The first experiment is a heteroskedastic model where $u_l = \psi(s_l)\tilde{u}_l$, with $\tilde{u}_l$ following the benchmark model with $c = c_0$. We let $\log\psi$ increase or decrease linearly from $\log\psi(s) = 0$ to $\log\psi(s) = \log 3$ moving from the most westward to the most eastward location, or from north to south. The largest of the four rejection frequencies of SCPC has percentiles $\langle 4.6\%, 4.9\%, 5.3\%\rangle$ and $\langle 5.1\%, 6.4\%, 8.7\%\rangle$ under g_{uniform} and g_{light}, respectively. We conclude that heteroskedasticity does not seem to be a major driver of size distortions.

The second experiment investigates location measurement error of a form studied in Conley and Molinari (2007). Specifically for each location, $s_l^* = s_l + e_l$ where s_l^* is the measured location, s_l is the true location and e_l is the measurement error. The error term is $e_l = (e_{1,l}, e_{2,l})$ with $e_{1,l}$ the north-south and $e_{2,l}$ the east-west coordinate and $e_{i,l}$ i.i.d.$\mathcal{U}(-\delta, \delta)$ over i and l, and $\delta = 0.0375H$ with H the length of the smallest square that encompasses all locations, corresponding to medium "level 4" errors in Conley and Molinari's (2007) classification. The null rejection frequencies of SCPC have percentiles $\langle 5.3\%, 5.6\%, 6.1\%\rangle$ and $\langle 5.1\%, 7.3\%, 17.5\%\rangle$ under g_{uniform} and g_{light}, respectively. Evidently, measurement error of this sort has little effect on the size of SCPC under uniformly distributed locations, but can lead to substantial size distortions for some highly concentrated spatial distributions.

4 Efficiency of SCPC

In this section we study the average length of SCPC intervals. We again focus exclusively on a Gaussian finite sample framework, so we adopt the notation of the last section. We consider two comparisons. First, we compare SCPC to previously proposed spatial t-statistics. Second, we assess absolute efficiency by computing a lower bound on the average length of a length-optimal confidence interval.

For the latter comparison, we consider confidence intervals $\text{CI}(\mathbf{y}) \subset \mathbb{R}$ of the form

$$\text{CI}(\mathbf{y}) = [\bar{y} - \delta(\hat{\mathbf{u}}), \bar{y} + \delta(\hat{\mathbf{u}})] \tag{12}$$

with a margin-of-error estimator $\delta : \mathbb{R}^n \mapsto [0, \infty)$ that is a scale equivariant function of the residuals $\hat{\mathbf{u}}$, $\delta(\lambda\hat{\mathbf{u}}) = \lambda\delta(\hat{\mathbf{u}})$ for all $\lambda > 0$, but is otherwise unrestricted. We want to compare the SCPC interval with a version of $\text{CI}(\mathbf{y})$ that, like SCPC, has good coverage $\mathbb{P}^0_{\mathbf{\Sigma}}(\mu \in \text{CI}(\mathbf{y}))$ over a range of potential spatial correlation patterns $\mathbf{\Sigma} \in \mathcal{V}$. The metric for measuring efficiency is the expected length $\mathbb{E}_{\mathbf{\Sigma}(c_1)}[\int \mathbf{1}[x \in \text{CI}(\mathbf{y})]dx]$ in the SCPC benchmark model $\mathbf{y} \sim \mathcal{N}(\mathbf{l}\mu, \mathbf{\Sigma}(c_1))$ for a given $c_1 > c_0$, or expected length in the i.i.d. model, $c_1 \to \infty$. We compare these expected lengths in the U.S. states spatial designs, using $c_1 = 2c_0$, $c_1 = 5c_0$, or $c_1 \to \infty$ (i.e., the i.i.d. model).

As in Section 3, we take a spectral perspective to guide our choice of $\mathcal{V}$: Intuitively, for a method that seeks to minimize expected length under c_1, it is hardest to control size if the spectral density is proportional to $f = f_1 = f^{\text{bnch}}_{c_1}$ for high frequencies, but steeper for lower frequencies, so that the variance of $\bar{y}$ is larger than one would expect based on an extrapolation using high frequency variation. As discussed in the last section, the choice of $\bar{\rho}_0$ and hence c_0 of SCPC is usefully thought of as specifying the the worst-case steepest spectral density $f = f_0 = f^{\text{bnch}}_{c_0}$. This motivates a choice of a putative "least favorable" continuous spectral density of the form

$$f^{\text{kink}}_\theta(\omega) = \mathbf{1}[|\omega| \leq \theta] f_0(\omega) + \mathbf{1}[|\omega| > \theta] \frac{f_0(\theta)}{f_1(\theta)} f_1(\omega)$$

so that f^{kink}_θ coincides with f_0 over low frequencies, has a kink at $\omega = \theta$, after which it coincides with a scaled version of f_1. Let $\mathbf{\Sigma}^{\text{kink}}_\theta$ denote the implied covariance matrix, and set $\mathcal{V} = \mathcal{V}^{\text{kink}} = \{\mathbf{\Sigma}^{\text{kink}}_\theta : \theta \geq 0\}$.

This construction is not applicable to the i.i.d. case, since setting f_1 equal to a constant does not yield an integrable spectral density. Instead, define $f_\Delta(\omega) = \mathbf{1}[|\omega| \leq \theta](f_0(\omega) - f_0(\theta))$, and let $f_R(\omega) = f_0(\omega) - f_\Delta(\omega)$, so that $f_0(\omega) = f_\Delta(\omega) + f_R(\omega)$. In obvious notation, the corresponding covariance matrices satisfy $\mathbf{\Sigma}(c_0) = \mathbf{\Sigma}_\Delta(\theta) + \mathbf{\Sigma}_R(\theta)$. Since $f_R(\omega)$ is a continuous density that is flat for $|\omega| \leq \theta$, and that follows the same decline as $f_0(\omega)$ for $|\omega| > \theta$, it also contributes to the overall persistence of $\mathbf{\Sigma}(c_0)$. Thus, replacing $\mathbf{\Sigma}_R(\theta)$ by $\lambda_1(\mathbf{\Sigma}_R(\theta))\mathbf{I}_n$ reduces overall persistence, motivating the construction of $\mathbf{\Sigma}_\theta^{\text{kink}}$ in the i.i.d. case as $\mathbf{\Sigma}_\theta^{\text{kink}} = \mathbf{\Sigma}_\Delta(\theta) + \lambda_1(\mathbf{\Sigma}_R(\theta))\mathbf{I}_n$.

As one would expect given the results of Section 3, SCPC controls size in the U.S. states spatial designs under $\mathcal{V}^{\text{kink}}$, or at least nearly so: With $\alpha_{\text{SCPC}}(\theta) = \mathbb{P}^0_{\mathbf{\Sigma}_\theta^{\text{kink}}}(\tau^2_{\text{SCPC}} > \text{cv}^2_{\text{SCPC}})$, the distribution of $\sup_{\theta \geq 0} \alpha_{\text{SCPC}}(\theta)$ has 95th percentile smaller than 5.2% under $g = g_{\text{uniform}}$ for all considered values of c_1, and smaller than 7.3% under $g = g_{\text{light}}$. To keep things on an equal footing, we allow CI the same degree of undercoverage, that is we consider the problem

$$\inf_\delta \mathbb{E}_{\mathbf{\Sigma}(c_1)}[\int \mathbf{1}[x \in \text{CI}(\mathbf{y})]dx] \text{ s.t. } \mathbb{P}_{\mathbf{\Sigma}_\theta^{\text{kink}}}(\mu \notin \text{CI}(\mathbf{y})) \leq \max(\alpha_{\text{SCPC}}(\theta), \alpha) \text{ for all } \theta \geq 0. \quad (13)$$

In words, we seek the confidence interval with the shortest expected length in the $\mathbf{\Sigma}(c_1)$ model among all confidence intervals of the form (12) that are as robust as the SCPC interval under $\mathbf{\Sigma}_\theta^{\text{kink}}$, $\theta \geq 0$.

Since θ is one-dimensional, one can apply the numerical techniques of Elliott, Müller, and Watson (2015) and Müller and Wang (2019) to obtain an informative lower bound on the objective $\inf_\delta \mathbb{E}_{\mathbf{\Sigma}(c_1)}[\int \mathbf{1}[x \in \text{CI}(\mathbf{y})]dx]$ that holds for *any* $\text{CI}(\mathbf{y})$ of the form (12) that satisfies the constraint in (13).

We compare these lower bounds on expected lengths with five confidence intervals based on spatial t-statistics: (i) the SCPC t-statistic as defined in Section 2; (ii) an alternative version of the SCPC t-statistic that chooses q to minimize expected length in the $\mathbf{\Sigma}(c_1)$ model with $c_1 = 2c_0$ (alt-SCPC); (iii) a t-statistic based on a Bartlett-type kernel variance estimator with bandwidth equal to 0.3 of the largest distance of all observations, $\Delta_{\max} = \max_{l,\ell} ||s_l - s_\ell||$, that is $k(s_l, s_\ell) = \max(1 - 0.3||s_l - s_\ell||/\Delta_{\max}, 0)$ (Bartlett Kernel); (iv) Sun and Kim's (2012) projection t-statistic with $k_1 = 1$, $k_2 = 2$ Fourier weights for a total of $q = 2(k_1 + k_2 + k_1 k_2) = 10$ weighted averages (Fourier Projection); (v) Ibragimov and Müller's (2010) cluster t-statistic with $q = 9$ equal-sized clusters (Cluster).[4] All five methods use a critical value so that size is

[4]The assignment of locations to clusters is performed sequentially, where at each step, we minimize (across yet unassigned locations) the maximal distance over clusters (among those that have not yet been assigned n/q locations). Cluster distances are computed from the northwest, northeast, southeast and southwest corners of the location circumscribing rectangle, and in the $q = 9$ case, also from the mid-points of the four sides of this rectangle, and its center.

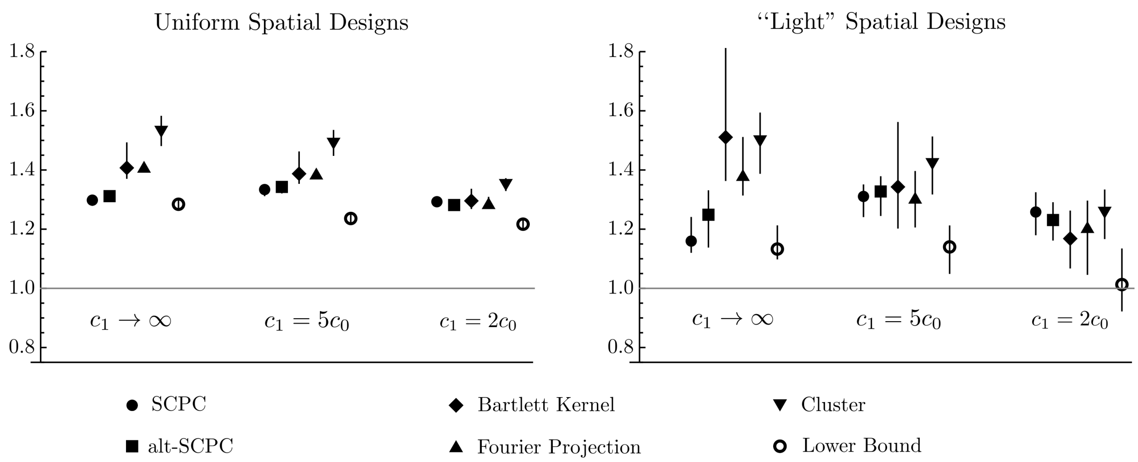

Figure 2: 5th, 50th and 95th Percentiles of Average Confidence Interval Lengths in U.S. States designs

controlled in the benchmark model with $c \geq c_0$. (The following results are nearly unchanged if in addition, one also imposes the coverage constraint in (13).)

Figure 2 reports the 5th, 50th and 95th percentiles of the distribution of expected lengths under $c_1 \to \infty$ (the i.i.d. case), $c_1 = 2c_0$ and $c_1 = 5c_0$, in multiples of the length of the known-σ interval with endpoints $\bar{y} \pm 1.96\sigma/\sqrt{n}$. (In the g_{uniform} designs, the expected lengths are often not very variable, so the 5th and 95th percentiles are sometimes hidden by the median marker in Figure 2.) In the uniform spatial designs, and with $c_1 \to \infty$ in the light designs, the SCPC interval comes reasonably close to being as short as the lower bound, and performs better than the alternative confidence intervals. The differences between SCPC and alt-SCPC are small throughout, motivating our choice of q_{SCPC} to minimize length in the i.i.d. model. In the light design with $c_1 = 2c_0$, SCPC performs somewhat worse than the other confidence intervals, and all intervals are much longer than the lower bound. The latter effect is due to $\bar{y}$ being far from the efficient estimator of μ when c_1 is small and the location distribution is not uniform: the population R^2 of a regression of $\bar{y}$ on $\hat{\mathbf{u}}$ has percentiles of $\langle 26\%, 47\%, 59\% \rangle$ under $g = g_{\text{light}}$ and $c_1 = 2c_0$. Thus, there exist margin-of-error functions $\delta(\hat{\mathbf{u}})$ in (12) that exploit this partial information about the realization of $\bar{y}$, leading to small lower bounds that are even below the length of the known-σ interval in some cases.

Remark 4.1. These efficiency results imply a limit on the possibility of using data-dependent methods to learn about the value of the worst-case correlation c_0: For example, consider an approach that pre-tests whether there is any spatial correlation (that is, whether c_0 can be chosen arbitrarily large), and that conservatively reverts to a very wide interval if it detects any correlation. If one could devise a pre-test that reliably indicates the presence or absence of spatial correlation, then one could easily construct a function δ that (i) controls size under $\mathcal{V}^{\text{kink}}$; and (ii) is nearly as efficient as the oracle interval in the i.i.d. case. But given our lower bound results, such a function cannot exist. The same argument applies to pre-tests that seek to determine whether, say, c_0 can safely be chosen five times as large as a given value, while still trying to control size if it cannot.

More generally, any attempt to estimate $\mathbf{\Sigma}$ from the data and to use this value for inference about μ must either yield confidence intervals that are not much shorter than SCPC, at least in the uniform designs and the i.i.d. light designs; or fail to control size under $\mathcal{V}^{\text{kink}}$. For example, consider a plug-in estimator $\hat{\sigma}^2_{\text{PI}}$ of $\sigma^2 = \mathbf{l}'\mathbf{\Sigma}\mathbf{l}/n$ with $\mathbf{\Sigma}$ in the Matérn class, so that the spectral density is proportional to $(c^2+\omega^2)^{-1-\nu}$, $\nu, c > 0$. Suppose we estimate the Matérn scale parameter, $c > 0$ and $\nu \in \{1/2, 3/2, 5/2\}$ by maximizing the Gaussian likelihood (this grid of values for ν is computationally convenient, since it yields simple expression for the covariance function). We find that in the U.S. states spatial designs with $g = g_{\text{uniform}}$, the confidence interval with endpoints $\bar{y} \pm 1.96\hat{\sigma}_{\text{PI}}/\sqrt{n}$ induces non-coverage probabilities with percentiles $\langle 21\%, 22\%, 23\% \rangle$ in the $\mathcal{V}^{\text{kink}}$ class with $c_1 = 5c_0$.

5 Large-Sample Analysis of Spatial t-Statistics

This section extends the results for finite-sample Gaussian models to large-sample non-Gaussian settings. The discussion is facilitated using notation that emphasizes the sample size, and we do by appending a subscript n to many of the variables defined previously. For example, the t-statistic defined in (2) will be denoted τ_n, and so forth for other variables.

The large-sample distribution of τ_n depends on two characteristics of the model. The first is the covariance function of the u_l process, that is, the covariance between u_l and u_ℓ at locations $s_l, s_\ell \in \mathcal{S}$. The second is the distribution of locations s that are sampled. The first sub-section provides a large-n framework for characterizing these two features of the model. With this framework in hand, the following subsections discuss the large sample normality of the linear functions $\mathbf{W}_n^{0\prime}\mathbf{u}_n$ that determine the null distribution of τ_n, extensions for kernel-based t-statistics, the implications of these results for size control of SCPC-based inference, and the key role that the density g of s plays in these results even under weak correlation.

5.1 Sampling and Large-n Framework

This subsection provides assumptions on sampling of the spatial locations, the spatial correlation properties of u conditional on the locations, and the set of weight functions used to determine the weighted averages of $\mathbf{u}_n$ that enter the t-statistic τ_n. We discuss these in turn.

Spatial locations: The spatial locations s_l are chosen from $\mathcal{S}$, a compact subset of $\mathbb{R}^d$. Sample locations are selected as i.i.d. draws from a distribution G with density g, which is continuous and positive on $\mathcal{S}$.

Correlation properties of $\mathbf{u}_n|\mathbf{s}_n$*:* The average pairwise correlation of u_l, conditional on the sample locations $\mathbf{s}_n$, is $\bar{\rho}_n = \frac{1}{n(n-1)} \sum_{l=1}^n \sum_{\ell \neq l} \text{Cor}\,(u_l, u_\ell\,|\mathbf{s}_n)$. When $\bar{\rho}_n = 0$, $\mathbf{u}_n\,|\mathbf{s}_n$ is white noise. When $\bar{\rho}_n = O_p(1)$ (and not $o_p(1)$), we will say the process exhibits *strong correlation*. When $\bar{\rho}_n = O_p(1/c_n^d)$ where c_n is a sequence of constants with $c_n \to \infty$, we follow Lahiri (2003) and say the process exhibits *weak correlation*. As shown in the next section, the large-sample distribution of τ_n is different under weak and strong correlation.

Distribution of $\mathbf{u}_n|\mathbf{s}_n$*:* The following asymptotic framework, adapted from Lahiri (2003), is useful for modelling weak and strong correlation. Let B be a zero-mean stationary random field on $\mathbb{R}^d$ with continuous covariance function $\mathbb{E}[B(s)B(r)] = \sigma_B\,(s - r)$, and B and $\mathbf{s}_n$ are independent. To avoid pathological cases, we assume $\int \sigma_B(s)ds > 0$ and B is nonsingular in the sense that $\inf_{||f||=1} \int\int f(r)f(s)\sigma_B(s-r)dG(r)dG(s) > 0$ with $||f||^2 = \int f^2(s)dG(s)$. Let c_n denote a sequence of constants with either $c_n \to \infty$ or $c_n = c > 0$. We consider a triangular-array framework with $u_l = B(c_n s_l)$ for $s_l \in \mathcal{S}$, so that $\sigma_u(s) = \sigma_B(c_n s)$. A calculation shows that $\bar{\rho}_n = O_p(1/c_n^d)$, so the sequence c_n characterizes weak and strong correlation as described above.

The sequence c_n determines the 'infill' and 'outfill' nature of the asymptotics. To see this, note that the volume of the relevant domain for the random field B is $c_n^d\,\text{vol}(\mathcal{S})$, where $\text{vol}(\mathcal{S})$ is the volume of $\mathcal{S}$. The average number of sample points per unit of volume is then $n/(c_n^d\,\text{vol}(\mathcal{S}))$. If $c_n^d \propto n$, the volume of the domain is increasing, while the number of points per unit of volume is not; this is the usual outfill asymptotic sampling scheme. On the other hand, when $c_n = c$, a constant, the volume of the domain is fixed, and the number of points per unit of volume is proportional to n; this is the usual infill sampling. Finally, when $c_n \to \infty$ with $c_n^d = o(n)$ the sampling scheme features both infill and outfill asymptotics.

Weight Functions: Finally, we specify a set of weighting functions. Specifically, for $j = 1, \ldots, q$, let $w_j : \mathcal{S} \mapsto \mathbb{R}$ denote a set of continuous functions that satisfy $\int w_j(s)dG(s) = 0$. We introduce the following notation involving these functions: $\mathbf{w}(s)$ is a $q \times 1$ vector-valued continuous function with $\mathbf{w}(s) = (w_1(s), ..., w_q(s))'$; $\mathbf{w}^0(s) = (1, \mathbf{w}(s)')'$; $\mathbf{W}_n$ is a $n \times q$ matrix with lth row given by $\mathbf{w}(s_l)'$, and $\mathbf{W}_n^0$ is a $n \times (q+1)$ matrix with lth row given by $\mathbf{w}^0(s_l)'$ so that $\mathbf{W}_n^0 = [\mathbf{l}_n, \mathbf{W}_n]$.

Remark 5.1. In our framework, locations s_l are sampled within $\mathcal{S}$ for a fixed and given $\mathcal{S}$. But nothing changes in our derivations if instead we treated the observations y_l as being indexed by $c_n s_l \in c_n \mathcal{S}$, as in Lahiri (2003), or any other one-to-one transformation of s_l. The essential characteristic is the dependence pattern over the spatial domain of the observations which is governed by c_n and B.

5.2 Large-Sample Behavior of Weighted Averages

As is evident from equation (7), the t-statistic is a function of weighted averages of the elements of $\mathbf{u}_n$. This subsection discusses the large-sample distribution of such weighted averages. These results involve weak convergence (i.e., convergence in distribution) where our interest lies in these limits conditional on the locations $\mathbf{s}_n$. With this in mind, for $\mathbf{X}_n$ and $\mathbf{X}$ p-dimensional random vectors, we use the notation $\mathbf{X}_n|\mathbf{s}_n \Rightarrow_p \mathbf{X}$ to denote $\mathbb{E}[h(\mathbf{X}_n)|\mathbf{s}_n] \xrightarrow{p} \mathbb{E}[h(\mathbf{X})]$ for any bounded continuous function $h : \mathbb{R}^p \mapsto \mathbb{R}$. This notion of weak convergence in probability is weaker than almost sure weak convergence of conditional distributions, but nevertheless ensures that the limiting distribution is not induced by the randomness in the locations $\mathbf{s}_n$.

Lemma 3. *(i) (strong correlation) Suppose $c_n = c > 0$ and B is a Gaussian process. Then*

$$n^{-1}\mathbf{W}_n^{0\prime}\mathbf{u}_n|\mathbf{s}_n \Rightarrow_p \mathbf{X} \sim \mathcal{N}(0, \mathbf{\Omega}_{sc})$$

with

$$\mathbf{\Omega}_{sc} = \int\int \mathbf{w}^0(r)\mathbf{w}^0(s)'\sigma_B(c(r-s))dG(r)dG(s).$$

(ii) (weak correlation) Let $a_n = c_n^d/n$. Suppose $c_n \to \infty$, $a_n \to a \in [0, \infty)$, and the assumptions of Lahiri's (2003) Theorem 3.2 hold. Then

$$a_n^{1/2}n^{-1/2}\mathbf{W}_n^{0\prime}\mathbf{u}_n|\mathbf{s}_n \Rightarrow_p \mathbf{X} \sim \mathcal{N}(0, \mathbf{\Omega}_{wc})$$

with

$$\mathbf{\Omega}_{wc} = a\sigma_B(0)\mathbf{V}_1 + \left(\int \sigma_B(s)ds\right)\mathbf{V}_2$$

where

$$\mathbf{V}_1 = \int \mathbf{w}^0(s)\mathbf{w}^0(s)' g(s) ds \text{ and } \mathbf{V}_2 = \int \mathbf{w}^0(s)\mathbf{w}^0(s)' g(s)^2 ds.$$

Remark 5.2. Note that the variance of $\sum_{l=1}^{n} \mathbf{w}^0(s_l) u_l$ conditional on $\mathbf{s}_n$ is

$$\begin{aligned}\mathrm{Var}\left[\sum_{i=1}^{n} \mathbf{w}^0(s_l) u_l \,|\mathbf{s}_n\right] &= \sum_{l}\sum_{\ell} \mathbf{w}^0(s_l)\mathbf{w}^0(s_\ell)' \sigma_u(s_l - s_\ell) \\ &= \sum_{l}\sum_{\ell} \mathbf{w}^0(s_l)\mathbf{w}^0(s_\ell)' \sigma_B(c_n\,(s_l - s_\ell)). \qquad (14)\end{aligned}$$

The strong-correlation covariance matrix, $\mathbf{\Omega}_{sc}$, is recognized as the large-n analogue of this expression after appropriate normalization and averaging over the locations. The weak-correlation covariance matrix, $\mathbf{\Omega}_{wc}$, differs from $\mathbf{\Omega}_{sc}$ in two ways. First, because $c_n \to \infty$ in the weak-correlation case, and $\sigma_B(r)$ vanishes for large $|r|$, the second term in $\mathbf{\Omega}_{wc}$ is recognized as the limit of $\mathbf{\Omega}_{sc}$ as the double integral concentrates entirely on 'the diagonal' where $r \approx s$. Second, as outfill becomes more important (that is, $a_n = c_n^d/n$ gets larger), variances become more important relative to covariances; this explains the first term in $\mathbf{\Omega}_{wc}$.

Remark 5.3. In the strong-correlation case, normality is assumed. That said, CLTs have been established also for strongly correlated models when $d = 1$ (i.e., the time series case), such as Taqqu (1975), Phillips (1987) or Chan and Wei (1987), and to a lesser extent also for $d > 1$, as in Wang (2014) or Lahiri and Robinson (2016). For the weak correlation case, large-sample normality follows from Theorem 3.2 in Lahiri (2003), which imposes mixing and moment conditions on B.

Remark 5.4. The regularly-spaced time series analogue of part (i) of Lemma 3 is the convergence $n^{-1}\mathbf{W}_n^{0\prime}\mathbf{u}_n|\mathbf{s}_n \Rightarrow \mathbf{X} = \int_0^1 \mathbf{w}^0(s) B(cs) ds$. The result in part (ii) has no such analogue, as the complications arise precisely under non-uniformly distributed locations.

Remark 5.5. The factor $\int \sigma_B(s) ds$ in front of $\mathbf{V}_2$ is the spatial analogue of the long-run variance of the process B. In this integral, the distances are weighted as if s was uniform on $\mathbb{R}^d$. This is a consequence of the i.i.d. sampling assumption on s_l: Under weak correlation, only observations very close to each other are meaningfully correlated, and with g continuous, the density of the locations s_l is locally flat in a small enough neighborhood around any given point $s \in \mathcal{S}$. This asymptotic approximation hence requires that the observed $\mathbf{s}_n$ is such that the empirical distribution of $s_l - s_\ell$ is approximately uniform conditional on $||s_l - s_\ell||$ being small. If B is assumed isotropic, a sufficient condition is that $\Delta_{l,\ell} = ||s_l - s_\ell||$ has an empirical distribution that is reasonably well approximated by a density proportional to Δ^{d-1} close to the origin.

Remark 5.6. The form of $\mathbf{V}_2$ is recognized as the limit covariance matrix in a model where the observations are independent, with variance proportional to $g(s_l)$. Thus, $\mathbf{V}_2$ is what one would obtain for the limit covariance matrix under a specific form of non-stationarity. Intuitively, a high density area does not only yield many observations, but under spatial correlation, the variance contribution is further amplified by the resulting high average correlation.

5.3 Large-Sample Distribution of Spatial t-Statistics

5.3.1 Projection Variance Estimators

Lemmas 1 and 3 lead to the following representation for the limiting distribution of $\tau_n^2(\mathbf{W}_n\mathbf{W}_n')$.

Theorem 4. *With $\mathbf{\Omega} \in \{\mathbf{\Omega}_{sc}, \mathbf{\Omega}_{wc}\}$, ω_i defined in Lemma 1, and $(Z_0, Z_1, ..., Z_q)' \sim \mathcal{N}(0, \mathbf{I}_{q+1})$, under the assumptions of Lemma 3, $\mathbb{P}\left(\tau_n^2(\mathbf{W}_n\mathbf{W}_n') > \mathrm{cv}^2 \,|\mathbf{s}_n\right) \xrightarrow{p} \mathbb{P}\left(Z_0^2 > \sum_{i=1}^q(-\frac{\omega_i}{\omega_0})Z_i^2\right)$ under the null hypothesis.*

Remark 5.7. In the general weak correlation case with arbitrary spatial density g, $\mathbf{\Omega}_{wc} = a\sigma_B(0)\mathbf{V}_1 + \left(\int \sigma_B(s)ds\right)\mathbf{V}_2$. Because τ_n^2 is a scale-invariant function of $\mathbf{u}_n$, it is without loss of generality to normalize the scale of $\sigma_B(\cdot)$ so that $a\sigma_B(0) + \int \sigma_B(s)ds = 1$. Under this normalization

$$\mathbf{\Omega}_{wc} = \kappa\mathbf{V}_1 + (1-\kappa)\mathbf{V}_2 \tag{15}$$

where κ is scalar with $0 \leq \kappa < 1$. Thus, the limit distribution of τ_n^2 is seen to depend on σ_B only through the scalar κ; the matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ are functions of the weights $\mathbf{w}^0$ and the spatial density g. The scalar κ thus completely summarizes the large sample effect of alternative underlying random fields B and weak correlation sequences $c_n \to \infty$.

Remark 5.8. When g is constant, so the spatial distribution is uniform, $\mathbf{V}_1 \propto \mathbf{V}_2$ and $\mathbf{\Omega}_{wc} \propto \int \mathbf{w}^0(s)\mathbf{w}^0(s)'ds$. In a leading case with orthogonal w_j of length $1/\sqrt{q}$, $\int w_j(s)w_i(s)dG(s) = q^{-1}\mathbf{1}[i=j]$, $\mathbf{\Omega}_{wc} \propto \mathrm{diag}(1, q^{-1}\mathbf{I}_q)$. Thus the asymptotic rejection probability becomes the corresponding quantile of the $F_{1,q}$ distribution, a result familiar from the limiting distribution of projection based squared t-statistics in the regularly spaced time series case. Importantly, while this result holds under constant g, it does *not* hold for other spatial distributions, so that the typical HAR results about inconsistent variance estimators for regularly spaced time series under weak dependence do not carry over to the spatial case.

For example, consider Sun and Kim (2012) inference in the U.S. states spatial design with $g = g_{\text{light}}$ and $n \to \infty$. Suppose we use $k_1 = 1$ and $k_2 = 2$ Fourier weights, so that the total number of weighted averages is $q = 2(k_1 + k_2 + k_1 k_2) = 10$, and Sun and Kim (2012) suggest using the critical value from a student-t distribution with 10 degrees of freedom (corresponding to computing the critical value under $\kappa \to 1$, or equivalently, under i.i.d. sampling). Under a weak-correlation sequence with $\kappa = 0$, so that $\mathbf{\Omega}_{wc} = \mathbf{V}_2$, a direct calculation shows that these nominal 5% level tests have asymptotic null rejection probabilities with percentiles $\langle 6.2\%, 10.3\%, 30.0\% \rangle$ across the 48 U.S. states.

In contrast, for the Ibragimov and Müller (2010) cluster t-statistic with clusters defined by a partition of $\mathcal{S}$ into q subregions, the special structure of the corresponding weighting functions $\mathbf{w}$ implies that the lower right $q \times q$ block of $\mathbf{\Omega}_{wc}$ is diagonal irrespective of g, which guarantees asymptotic validity of the student-t q critical value by virtue of Bakirov and Székely's (2005) result about the small sample validity of the usual t-test with heteroskedastic observations at conventional significance levels (cf. Remark 5.6).

Remark 5.9. For SCPC and other estimators, the weights $\mathbf{w}(s)$ are estimated using the sample locations $\mathbf{s}_n$. Lemma 12 in the appendix provides conditions under which the result in Theorem 4 continues to hold for estimated weights $\hat{\mathbf{w}}(s)$.

5.3.2 Kernel Variance Estimators

This subsection discusses how these results can be generalized so they apply to kernel-based variance estimators, $\hat{\sigma}_n^2(\mathbf{M}_n\mathbf{K}_n\mathbf{M}_n)$ and associated t-statistics $\tau_n^2(\mathbf{M}_n\mathbf{K}_n\mathbf{M}_n)$, where the $n \times n$ matrix $\mathbf{K}_n$ has (l, ℓ) element equal to $k(s_l, s_\ell)$ for a positive semidefinite continuous kernel $k : \mathcal{S} \times \mathcal{S} \mapsto \mathbb{R}$. Since in our framework, $s_l \in \mathcal{S}$ for a fixed sampling region $\mathcal{S}$, and k does not depend on n, these kernel estimators are spatial analogues of fixed-b time series long-run variance estimators considered by Kiefer and Vogelsang (2005), as also investigated by Bester, Conley, Hansen, and Vogelsang (2016).

Let $\hat{\mathbf{K}}_n = \mathbf{M}_n\mathbf{K}_n\mathbf{M}_n$, and note that the (l, ℓ) element of $\hat{\mathbf{K}}_n$ is $\hat{k}_n(s_l, s_\ell)$ with

$$\hat{k}_n(r, s) = k(r, s) - n^{-1}\sum_{l=1}^{n} k(s_l, s) - n^{-1}\sum_{\ell=1}^{n} k(r, s_\ell) + n^{-2}\sum_{l=1}^{n}\sum_{\ell=1}^{n} k(s_l, s_\ell). \tag{16}$$

To begin, consider a simpler problem using a kernel that replaces the sample means in (16) with populations means

$$\bar{k}(r,s) = k(r,s) - \int k(u,s)dG(u) - \int k(r,u)dG(u) + \int\int k(u,t)dG(u)dG(t). \quad (17)$$

By Mercer's Theorem, $\bar{k}(r,s)$ has the representation

$$\bar{k}(s,r) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(s)\varphi_i(r) \quad (18)$$

where (λ_i, φ_i) are the eigenvalues and eigenfunctions of $\bar{k}$, with eigenvalues ordered from largest to smallest, normalized so that $\int \varphi_i(s)\varphi_j(s)dG(s) = \mathbf{1}[i=j]$. By definition of an eigenfunction, for $\lambda_i > 0$, $\varphi_i(\cdot) = \lambda_i^{-1}\int \bar{k}(\cdot,s)\varphi_i(s)dG(s)$, so φ_i is continuous, and $\int \varphi_i(s)dG(s) = 0$.

Consider the problem with a truncated version of $\bar{k}$,

$$\bar{k}_q(s,r) = \sum_{i=1}^{q} \lambda_i \varphi_i(s)\varphi_i(r).$$

We can directly apply Theorem 4 using $w_j(s) = \lambda_j^{1/2}\varphi_j(s)$. Specifically, let $\bar{\mathbf{K}}_{n,q}$ be an $n \times n$ matrix with (l,ℓ) element equal to $\bar{k}_q(s_l, s_\ell)$. Then $\mathbf{u}_n'\bar{\mathbf{K}}_{n,q}\mathbf{u}_n = \mathbf{u}_n'\mathbf{W}_n\mathbf{W}_n'\mathbf{u}_n$ so that $\tau_n^2(\bar{\mathbf{K}}_{n,q}) = \tau_n^2(\mathbf{W}_n\mathbf{W}_n')$, and $\mathbb{P}\left(\tau_n^2(\bar{\mathbf{K}}_{n,q}) > \mathrm{cv}^2 \,|\mathbf{s}_n\right) \xrightarrow{p} \mathbb{P}\left(Z_0^2 > \sum_{i=1}^{q}(-\frac{\omega_i}{\omega_0})Z_i^2\right)$ by Theorem 4.

To extend this result to the original problem, it is useful to reformulate it in terms of eigenvalues of linear operators. Specifically, denote by $\mathcal{L}_G^2$ the Hilbert space of functions $\mathcal{S} \mapsto \mathbb{R}$ with inner product $\langle f_1, f_2 \rangle = \int f_1(s)f_2(s)dG(s)$. Normalize $\mathbf{\Omega}_{wc} = \kappa\mathbf{V}_1 + (1-\kappa)\mathbf{V}_2$, as in (15). A tedious but straightforward calculation (see (29) in the appendix) shows that the eigenvalues ω_i of $\mathbf{A} = \mathbf{D}(\mathrm{cv})\mathbf{\Omega}$ with $\mathbf{\Omega} = \{\mathbf{\Omega}_{sc}, \mathbf{\Omega}_{wc}\}$ are also the eigenvalues of finite rank self-adjoint linear operators $\mathcal{L}_G^2 \mapsto \mathcal{L}_G^2$, namely $R_{sc}T_qR_{sc}$ and $R_{wc}T_qR_{wc}$ in the strong and weak correlation case, respectively, where

$$\begin{aligned} R_{sc}^2(f)(s) &= \int \sigma_B(c(s-r))f(r)dG(r) \\ R_{wc}^2(f)(s) &= (\kappa + (1-\kappa)g(s))f(s) \\ T_q(f)(s) &= \int \left(1 - \mathrm{cv}^2\, \bar{k}_q(s,r)\right) f(r)dG(r). \end{aligned}$$

This suggests that the limiting rejection probability for the original non-truncated $\bar{k}$ might be characterized by the (potentially infinite) number of eigenvalues of the operators RTR :

$\mathcal{L}_G^2 \mapsto \mathcal{L}_G^2$ with $R \in \{R_{wc}, R_{sc}\}$, where

$$T(f)(s) = \int \left(1 - \mathrm{cv}^2\, \bar{k}(s,r)\right) f(r)dG(r).$$

The following theorem shows this to be the case, and it also includes the generalization to sample demeaned kernels (16) instead of (17).

Theorem 5. *Let ω_0 denote the largest eigenvalue, and $\omega_i, i \geq 1$ the remaining eigenvalues of RTR for $R \in \{R_{wc}, R_{sc}\}$. Then under the assumptions of Lemma 3, $\omega_0 > 0$ and $\omega_i \leq 0$ for $i \geq 1$, and $\mathbb{P}(\tau_n^2(\hat{\mathbf{K}}_n) > \mathrm{cv}^2\,|\mathbf{s}_n) \xrightarrow{p} \mathbb{P}(Z_0^2 > \sum_{i=1}^{\infty}(-\omega_i/\omega_0)Z_i^2)$.*

Remark 5.10. Under weak correlation the limit distribution of kernel-based spatial t-statistics depends on the spatial density g, since the eigenvalues of $R_{wc}TR_{wc}$ are a function of g. This is analogous to the results for projection estimators discussed above. Thus, in both cases, using a critical value that is appropriate for i.i.d. data does not, in general, lead to valid inference under weak correlation.

Remark 5.11. The framework of Theorem 5 also sheds light on the asymptotic bias of kernel-based and orthogonal projection estimators under weak correlation. The estimand σ^2 is the limiting variance of $a_n^{1/2} n^{-1/2} \sum_{l=1}^n u_l$, which under the normalization (15) is equal to the (single) eigenvalue of the operator $R_{wc}T_{\sigma^2}R_{wc}$ with $T_{\sigma^2}(f)(s) = \int f(r)dG(r)$, that is $\int(\kappa + (1-\kappa)g(s))dG(s)$. The expectation of $a_n\hat{\sigma}_n^2(\hat{\mathbf{K}}_n)$ converges to the trace of the operator $R_{wc}T_{\bar{k}}R_{wc}$ with $T_{\bar{k}}(f)(s) = \int \bar{k}(s,r)f(r)dG(r)$, that is $\int(\kappa + (1-\kappa)g(s))\bar{k}(s,s)dG(s)$. Thus, the estimator is asymptotically unbiased for all g if and only if $\bar{k}(s,s) = 1$. For standard choices of k, $k(s,s) = 1$, so the only source of asymptotic bias is the demeaning (and if the estimator $\hat{\sigma}_n^2$ uses the null value $\mathbf{y}_n - \mu_0\mathbf{l}_n$ instead of the residuals $\hat{\mathbf{u}}_n$, the asymptotic bias is zero under the null hypothesis). Moreover, if $k(r,s)$ concentrates around the 'diagonal' where $r \approx s$, corresponding to a fixed-b kernel estimator with small b, the demeaning effect is small, as is the asymptotic variability of $a_n\hat{\sigma}_n^2(\hat{\mathbf{K}}_n)$. Thus, fixed-$b$ kernel estimators with standard kernel choices and small b yield nearly valid and efficient inference under weak correlation.

In contrast, orthogonal projection estimators where $\bar{k}(r,s) = q^{-1}\sum_{i=1}^q \phi_i(r)\phi_i(s)$ do not share this approximate unbiasedness property, even for q large, since $\int \phi_i(s)^2 dG(s) = 1$ does not, in general, imply that $\bar{k}(s,s) = q^{-1}\sum_{i=1}^q \phi_i(s)^2 \approx 1$.

The proof of Theorem 5 involves showing that in large samples, the difference between the eigenfunctions of the sample demeaned kernel (16) and the population demeaned kernel

(17) becomes small. The following lemma extends and adapts previous results by Rosasco, Belkin, and Vito (2010) to the case of sample demeaned kernels.

Lemma 6. *Let* $(\hat{\mathbf{v}}_i, \hat{\lambda}_i)$ *with* $\hat{\mathbf{v}}_i = (\hat{v}_{i,1}, \ldots, \hat{v}_{i,n})'$ *be the eigenvector-eigenvalue pairs of* $n^{-1}\hat{\mathbf{K}}_n$ *with* $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_n$ *and* $n^{-1}\hat{\mathbf{v}}_i'\hat{\mathbf{v}}_i = 1$*. For all* i *with* $\hat{\lambda}_i > 0$*, define the* $\mathcal{S} \mapsto \mathbb{R}$ *functions*

$$\hat{\varphi}_i(\cdot) = n^{-1}\hat{\lambda}_i^{-1} \sum_{l=1}^{n} \hat{v}_{i,l}\hat{k}_n(\cdot, s_l). \tag{19}$$

Let $\lambda_{(j)}$*,* $j = 1, \ldots$ *be the unique positive values of* λ_i *in descending order, and suppose* $\lambda_{(j)}$ *has multiplicity* $m_j \geq 1$*. Then for any* p *such that* $\lambda_{(p)} > 0$*,*

(a) there exist rotation matrices $\hat{\mathbf{O}}_{(j)}$ *of dimension* $m_j \times m_j$*,* $j = 1, \ldots, p$ *such that with* $q = \sum_{j=1}^{p} m_j$*,* $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_q)'$ *and* $\hat{\boldsymbol{\varphi}} = (\hat{\varphi}_1, \ldots, \hat{\varphi}_q)'$*,*

$$\sup_{s \in \mathcal{S}} ||\boldsymbol{\varphi}(s) - \mathrm{diag}(\hat{\mathbf{O}}_{(1)}, \ldots, \hat{\mathbf{O}}_{(p)})\hat{\boldsymbol{\varphi}}(s)|| = O_p(n^{-1/2});$$

(b) $\sum_{i=1}^{q}(\hat{\lambda}_i - \lambda_i)^2 = O_p(n^{-1})$*.*

Part (a) shows convergence of the eigenspace corresponding to unique eigenvalues, and part (b) shows convergence of the eigenvalues.

5.3.3 SCPC t-Statistic

Beyond its use in the proof of Theorem 5, Lemma 6 can be used to establish the large sample distribution of the SCPC t-statistic for nonrandom q and critical value cv. Note that in this application of Lemma 6, we are interested in the eigenfunctions of the demeaned covariance kernel $k^0(r, s) = \exp(-c_0||r - s||)$ of the benchmark model, rather than the eigenfunctions of a kernel that defines a kernel-based variance estimator.

Recall from Section 2 that $\mathbf{r}_i$ is the eigenvector of $\mathbf{M}_n\boldsymbol{\Sigma}_n(c_0)\mathbf{M}_n$ corresponding to the ith largest eigenvalue, normalized to satisfy $n^{-1}\mathbf{r}_i'\mathbf{r}_i = 1$. Let φ_i^0 be the eigenfunction of the kernel $\bar{k}^0(r, s)$ corresponding to the ith largest eigenvalue λ_i^0, where $\bar{k}^0$ is the demeaned version of k^0 in analogy to (17). Combining Lemma 6 with a result (Lemma 13 of the appendix) that suitably accounts for estimated weights yields the following corollary.

Corollary 7. *Suppose* $\lambda_q^0 > \lambda_{q+1}^0$ *and the assumptions of Lemma 3 hold. Then the convergence in Theorem 4 holds for* $\tau_{SCPC}^2(q) = \tau_n^2(q^{-1}\sum_{i=1}^{q} \mathbf{r}_i\mathbf{r}_i')$ *with* $\mathbf{w}(s) = (\varphi_1^0(s), \ldots, \varphi_q^0(s))'/\sqrt{q}$*.*

5.4 Asymptotic Size Control under Weak Correlation

As discussed above (see equation (15)), under weak correlation, the asymptotic rejection probability of τ_n for finite q can be studied via $\mathbf{\Omega}_{wc}(\kappa) = \kappa\mathbf{V}_1 + (1-\kappa)\mathbf{V}_2$, where the covariance function of B and the sequence c_n affects the large-sample distribution of τ_n only through the scalar $\kappa \in [0,1)$. Thus, if $\overline{\text{cv}}$ is such that $\sup_{0\leq\kappa<1}\mathbb{P}\left(\sum_{i=0}^{q}\omega_i(\kappa,\overline{\text{cv}})Z_i^2 > 0\right) = \alpha$, where $\{\omega_i(\kappa,\overline{\text{cv}})\}_{i=0}^{q}$ are the eigenvalues of $\mathbf{A}(\kappa,\overline{\text{cv}}) = \mathbf{D}(\overline{\text{cv}})\mathbf{\Omega}_{wc}(\kappa)$, then setting $\text{cv}_n \geq \overline{\text{cv}}$ for all n yields inference that is asymptotically robust under all forms of weak correlation covered by Theorem 3 (ii). In the case of a kernel-based variance estimator, the same holds as long as $\overline{\text{cv}}$ satisfies $\sup_{0\leq\kappa<1}\mathbb{P}\left(\sum_{i=0}^{\infty}\omega_i(\kappa,\overline{\text{cv}})Z_i^2 > 0\right) = \alpha$ where $\{\omega_i(\kappa,\overline{\text{cv}})\}_{i=0}^{\infty}$ are the eigenvalues of the linear operator $L(f)(s) = \int\sqrt{\kappa+(1-\kappa)g(s)}\left(1-\overline{\text{cv}}^2\bar{k}(s,r)\right)\sqrt{\kappa+(1-\kappa)g(r)}f(r)dG(r)$.

The value $\overline{\text{cv}}$ depends on the spatial density g, which can be seen directly by inspecting the form of $\mathbf{\Omega}_{wc}$ and the operator L. In principle, one could use these expressions to estimate $\overline{\text{cv}}$ directly. But this would involve estimates of the spatial density g, which leads to difficult bandwidth and other choices. We now discuss a simpler approach.

Consider a benchmark model B^0 that satisfies the assumptions of Lemma 3 (ii), such as the Gaussian exponential model introduced in Section 2. Let σ_B^0 denote the covariance kernel of B^0, and suppose $c_{n,0}$, is chosen so that $a_{n,0} = c_{n,0}^d/n \to a_0 = 0$. For instance, $c_{n,0} = c_0 > 0$ satisfies this condition, as does $c_{n,0} = n^{1/d}/\log(n)$. Note that for this model $\kappa = 0$. Suppose $\text{cv}_n = \text{cv}_n(\mathbf{s}_n)$ satisfies

$$\sup_{c\geq c_{n,0}}\mathbb{P}^0_{\mathbf{\Sigma}(c)}(\tau_n^2 \geq \text{cv}_n^2\,|\mathbf{s}_n) \leq \alpha \tag{20}$$

where $\mathbb{P}^0_{\mathbf{\Sigma}(c)}$ is computed under the benchmark model, that is under $\mathbf{u}_n|\mathbf{s}_n \sim \mathcal{N}(0,\mathbf{\Sigma}(c))$ with $\mathbf{\Sigma}(c)$ the covariance matrix of $(B^0(cs_1),...,B^0(cs_n))'$.

Theorem 8. *Let* cv_n^2 *satisfy (20). Under arbitrary weak correlation in the sense of Lemma 3 (ii), for the SCPC t-statistic and t-statistics covered by Theorems 4 and 5,* $\max(\overline{\text{cv}}^2 - \text{cv}_n^2, 0) \xrightarrow{p} 0$. *Consequently, for any* $\epsilon > 0$, $\limsup_n \mathbb{P}(\mathbb{P}(\tau_n^2 > \text{cv}_n^2\,|\mathbf{s}_n) > \alpha+\epsilon) \to 0$, *so that* $\limsup_n \mathbb{P}(\tau_n^2 \geq \text{cv}_n^2) \leq \alpha$.

The intuition for Theorem 8 is as follows. The critical value cv_n in (20) is valid in the benchmark model for all $c \geq c_{n,0}$ and n. Thus, it is also valid along arbitrary sequences $c_n \geq c_{n,0}$. Since the $c_{n,0}$ model has $\kappa = 0$, there exists sequences $c_n \geq c_{n,0}$ that induce any $\kappa \in [0,1)$ in the benchmark model; different sequences c_n in the benchmark model therefore trace out all possible limit distributions under generic weak correlation, so that size control in the benchmark model for all $c \geq c_{n,0}$ translates into size control under generic weak correlation.

For SCPC, the benchmark covariance kernel for B^0 is exponential $\sigma_B^0(r,s) = \exp(-c||r - s||)$ and (from equation (5)) the critical value is chosen to satisfy (20) with equality. Thus, with a fixed value of c_0 (or a fixed value of $\bar{\rho}_0$), the SCPC t-test $\tau_{\mathrm{SCPC}}(q)$ controls size in large samples under generic weak correlation.[5]

6 Extensions to Regression and GMM

The extension of these results to regression and GMM problems follows from standard arguments. For example, consider the linear regression problem

$$w_l = x_l\beta + \mathbf{z}_l'\delta + \varepsilon_l \text{ for } l = 1, ..., n \tag{21}$$

where β is the (scalar) parameter of interest, $\mathbf{z}_l$ are additional controls in the regression, and $(w_l, x_l, \mathbf{z}_l)$ are associated with location s_l. Let $\tilde{x}_l = x_l - \mathbf{S}_{xz}\mathbf{S}_{zz}^{-1}\mathbf{z}_l$ denote the residual from regressing x_l on $\mathbf{z}_l$, where we use the notation $\mathbf{S}_{ab} = n^{-1}\sum_{l=1}^n \mathbf{a}_l\mathbf{b}_l'$ for any vectors $\mathbf{a}_l$ and $\mathbf{b}_l$. Suppose $\mathbf{S}_{\tilde{x}\tilde{x}} \xrightarrow{p} \sigma_{\tilde{x}\tilde{x}}^2 > 0$ and $n^{-1/2}\sum_{l=1}^n \tilde{x}_l\varepsilon_l|\mathbf{s} \Rightarrow_p \mathcal{N}(0, \sigma_{\tilde{x}\varepsilon}^2)$. Then

$$\sqrt{n}(\hat{\beta} - \beta)|\mathbf{s} \Rightarrow_p \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \sigma_{\tilde{x}\varepsilon}^2/\sigma_{\tilde{x}\tilde{x}}^4$. Spatial correlation affects inference in this model through $\sigma_{\tilde{x}\varepsilon}^2$ which incorporates potential correlation between $\tilde{x}_l\varepsilon_l$ and $\tilde{x}_\ell\varepsilon_\ell$ at spatial locations s_l and s_ℓ.

Thus, suppose that $\tilde{x}_l\varepsilon_l$ satisfies the assumptions previously made for u_l. Then a straightforward calculation shows that setting

$$y_l = \hat{\beta} + \frac{\tilde{x}_l\hat{\varepsilon}_l}{n^{-1}\sum_{l=1}^n \tilde{x}_l^2}$$

in the analysis of the previous sections leads to analogous results with β replacing μ as the parameter of interest. The extension to GMM inference, potentially with clustering, is analogous; see, for instance, Section 4.4 of Müller (2020).

As usual, these extensions require that $\tilde{x}_l\varepsilon_l$ (or its GMM analogue) is stationary. This may be implausible is some applications. Müller and Watson (2022) investigate the performance of SCPC inference in regression models with a range of non-stationary processes for x_l and or

[5] Technically, the SCPC choice of q in (6) is also a function of the locations of $\mathbf{s}_n$, so q_{SCPC} is random. However, the argument that establishes Theorem 8 can be extended under this complication as long as $q_{\mathrm{SCPC}} \leq q_{\max}$ almost surely for some finite and fixed $q_{\max}$. See Theorem 14 in the appendix for a formal statement.

e_l. That paper also takes up the problem of computing the SCPC test statistic in applications with very large n.

A Appendix

Proof of Lemma 1: with $\mathbf{X} = \mathbf{W}^{0\prime}\mathbf{u} = (X_0, \mathbf{X}'_{1:q})'$ and $\mathbf{Z} = (Z_0, Z_1, ..., Z_q)'$ we have

$$\begin{aligned}
\mathbb{P}\left(\tau^2(\mathbf{W}\mathbf{W}') > \mathrm{cv}^2\right) &= \mathbb{P}\left(\frac{X_0^2}{\mathbf{X}'_{1:q}\mathbf{X}_{1:q}} > \mathrm{cv}^2\right) = \mathbb{P}\left(X_0^2 - \mathrm{cv}^2\,\mathbf{X}'_{1:q}\mathbf{X}_{1:q} > 0\right) \\
&= \mathbb{P}\left(\mathbf{X}'\mathbf{D}(\mathrm{cv})\mathbf{X} > 0\right) = \mathbb{P}(\mathbf{Z}'\mathbf{\Omega}^{1/2}\mathbf{D}(\mathrm{cv})\mathbf{\Omega}^{1/2}\mathbf{Z} > 0) \\
&= \mathbb{P}\left(\sum_{i=0}^{q}\omega_i Z_i^2 > 0\right)
\end{aligned}$$

where the last equality follows by similarity of the matrices $\mathbf{\Omega}^{1/2}\mathbf{D}(\mathrm{cv})\mathbf{\Omega}^{1/2}$ and $\mathbf{D}(\mathrm{cv})\mathbf{\Omega}$. The claim about the sign of the eigenvalues follows from Lemma 10 below. ■

The proof of Theorem 2 relies on some preliminary results.

Lemma 9. *For any two $q \times q$ positive semi-definite matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ and vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^q$, and all $p \in [0, 1]$,*

$$\begin{aligned}
\varsigma(p) = (p\mathbf{v}_1 + (1-p)\mathbf{v}_2)'(\mathbf{I}_q + p\mathbf{B}_1 + (1-p)\mathbf{B}_2)^{-1}(p\mathbf{v}_1 + (1-p)\mathbf{v}_2) \\
- p\mathbf{v}_1'(\mathbf{I}_q + \mathbf{B}_1)^{-1}\mathbf{v}_1 - (1-p)\mathbf{v}_2'(\mathbf{I}_q + \mathbf{B}_2)^{-1}\mathbf{v}_2 \leq 0.
\end{aligned}$$

Proof. We first show that $\varsigma(p)$ is convex. Write $\mathbf{G}(p) = \mathbf{I}_q + p\mathbf{B}_1 + (1-p)\mathbf{B}_2$. The first derivative of the nonlinear part of $\frac{1}{2}\varsigma(p)$ is given by

$$(\mathbf{v}_1 - \mathbf{v}_2)'\mathbf{G}(p)^{-1}(p\mathbf{v}_1 + (1-p)\mathbf{v}_2) - \tfrac{1}{2}(p\mathbf{v}_1 + (1-p)\mathbf{v}_2)'\mathbf{G}(p)^{-1}(\mathbf{B}_1 - \mathbf{B}_2)\mathbf{G}(p)^{-1}(p\mathbf{v}_1 + (1-p)\mathbf{v}_2)$$

so that the second derivative of $\frac{1}{2}\varsigma(p)$ equals

$$\begin{aligned}
(\mathbf{v}_1 - \mathbf{v}_2)'\mathbf{G}(p)^{-1}(\mathbf{v}_1 - \mathbf{v}_2) - 2(\mathbf{v}_1 - \mathbf{v}_2)'\mathbf{G}(p)^{-1}(\mathbf{B}_1 - \mathbf{B}_2)\mathbf{G}(p)^{-1}(p\mathbf{v}_1 + (1-p)\mathbf{v}_2) \\
+ (p\mathbf{v}_1 + (1-p)\mathbf{v}_2)'\mathbf{G}(p)^{-1}(\mathbf{B}_1 - \mathbf{B}_2)\mathbf{G}(p)^{-1}(\mathbf{B}_1 - \mathbf{B}_2)\mathbf{G}(p)^{-1}(p\mathbf{v}_1 + (1-p)\mathbf{v}_2).
\end{aligned}$$

With $\mathbf{\Delta}(p) = \mathbf{G}(p)^{-1/2}(\mathbf{v}_1 - \mathbf{v}_2)$ and $\mathbf{r}(p) = -\mathbf{G}(p)^{-1/2}(\mathbf{B}_1 - \mathbf{B}_2)\mathbf{G}(p)^{-1}(p\mathbf{v}_1 + (1-p)\mathbf{v}_2)$, the second derivative may be rewritten as

$$\begin{pmatrix} \mathbf{\Delta}(p) \\ \mathbf{r}(p) \end{pmatrix}' \begin{pmatrix} \mathbf{I}_q & \mathbf{I}_q \\ \mathbf{I}_q & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{\Delta}(p) \\ \mathbf{r}(p) \end{pmatrix} \geq 0$$

and convexity follows. Thus $\max_{p\in[0,1]}\varsigma(p) \leq \max(\varsigma(1), \varsigma(0)) = 0$. □

Lemma 10. *Let* $\mathbf{A}_1 = \int \mathbf{P}^{-1}\mathbf{D}(\text{cv})\boldsymbol{\Omega}(\theta)\mathbf{P}dF(\theta)$*. The* $q+1$ *eigenvalues of* $\mathbf{A}_1$ *are real, and only one is positive, and the same holds for* $\mathbf{A}(\theta)$*,* $\theta \in \Theta$*. Furthermore,* $\lambda_1(\mathbf{A}_1) \geq 1$*.*

Proof. By similarity, the eigenvalues of $\mathbf{A}_1$ are equal to those of $\mathbf{P}\mathbf{A}_1\mathbf{P}^{-1}$, which in turn is similar to the symmetric matrix

$$\begin{pmatrix} \mathbf{l}'\boldsymbol{\Sigma}_1\mathbf{l} & \mathbf{l}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}} \\ \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\mathbf{l} & \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}} \end{pmatrix}^{1/2} \begin{pmatrix} 1 & 0 \\ 0 & -\mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{l}'\boldsymbol{\Sigma}_1\mathbf{l} & \mathbf{l}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}} \\ \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\mathbf{l} & \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}} \end{pmatrix}^{1/2}$$

with $\tilde{\mathbf{W}} = (\mathbf{l}, \mathbf{W}/\,\text{cv})$, and the first claim follows for $\mathbf{A}_1$. The claim for $\mathbf{A}(\theta)$ follows from the same argument.

For the last claim, let $\bar{h} : \mathbb{R} \mapsto \mathbb{R}$

$$\bar{h}(t) = 1 - t\mathbf{l}'\boldsymbol{\Sigma}_1\mathbf{l} + t^2\mathbf{l}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}}(\mathbf{I}_q + t\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\mathbf{l}.$$

Note that $\bar{h}(t)$ is weakly decreasing in $t > 0$, since with $\tilde{\mathbf{H}} = -t\tilde{\mathbf{W}}(\mathbf{I}_q + t\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\mathbf{l}$

$$\bar{h}'(t) = -\begin{pmatrix} \mathbf{l} \\ \tilde{\mathbf{H}} \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1 \end{pmatrix} \begin{pmatrix} \mathbf{l} \\ \tilde{\mathbf{H}} \end{pmatrix} < 0.$$

The characteristic polynomial of $\mathbf{A}_1$ is given by

$$\begin{aligned} &\det\begin{pmatrix} s - \mathbf{l}'\boldsymbol{\Sigma}_1\mathbf{l} & \mathbf{l}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}} \\ -\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\mathbf{l} & s\mathbf{I}_q + \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}} \end{pmatrix} \\ = \;& (s - \mathbf{l}'\boldsymbol{\Sigma}_1\mathbf{l} + \mathbf{l}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}}(s\mathbf{I}_q + \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\mathbf{l})\det(s\mathbf{I}_q + \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}}) \\ = \;& s\bar{h}(s^{-1})\det(s\mathbf{I}_q + \tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}}) \end{aligned}$$

so that $\lambda_1(\mathbf{A}_1)$ satisfies $\bar{h}(1/\lambda_1(\mathbf{A}_1)) = 0$. Similarly, $1/\lambda_1(\mathbf{A}(\theta)) = 1$ is a root of

$$h_\theta(t) = 1 - t\mathbf{l}'\boldsymbol{\Sigma}(\theta)\mathbf{l} + t^2\mathbf{l}'\boldsymbol{\Sigma}(\theta)\tilde{\mathbf{W}}(\mathbf{I}_q + t\tilde{\mathbf{W}}'\boldsymbol{\Sigma}(\theta)\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\boldsymbol{\Sigma}(\theta)\mathbf{l}.$$

By Lemma 9, for any $t > 0$,

$$\begin{aligned} &\mathbf{l}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}}(\mathbf{I}_q + t\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\boldsymbol{\Sigma}_1\mathbf{l} \\ = \;& \left(\int \tilde{\mathbf{W}}'\boldsymbol{\Sigma}(\theta)\mathbf{l}dF(\theta)\right)'\left(\mathbf{I}_q + t\int \tilde{\mathbf{W}}'\boldsymbol{\Sigma}(\theta)\tilde{\mathbf{W}}dF(\theta)\right)^{-1}\left(\int \tilde{\mathbf{W}}'\boldsymbol{\Sigma}(\theta)\mathbf{l}dF(\theta)\right) \\ \leq \;& \int \mathbf{l}'\boldsymbol{\Sigma}(\theta)\tilde{\mathbf{W}}(\mathbf{I}_q + t\tilde{\mathbf{W}}'\boldsymbol{\Sigma}(\theta)\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}'\boldsymbol{\Sigma}(\theta)\mathbf{l}dF(\theta). \end{aligned}$$

Thus, $\bar{h}(t) \leq \int h_\theta(t)dF(\theta)$, and from $h_\theta(1) = 0$ for all θ, $\bar{h}(1) \leq 0$. Since h is decreasing, its root $1/\lambda_1(\mathbf{A}_1)$ must thus be smaller than unity, and the conclusion follows. □

Proof of Theorem 2: Proceeding as in the proof of Lemma 1, $\mathbb{P}_{\mathbf{\Sigma}_1}(\tau^2(\mathbf{WW}') > \text{cv}^2) = \mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \bar{\eta}_i Z_i^2\right)$ with $\bar{\eta}_i = \lambda_i\left(-\mathbf{A}_1\right)/\lambda_1(\mathbf{A}_1)$. By Lemma 10, $\bar{\eta}_i \geq 0$ for $i = 1, \ldots, q$. For future reference, note that $\mathbb{P}_{\mathbf{\Sigma}_0}(\tau^2(\mathbf{WW}') > \text{cv}^2) = \alpha$ yields $\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \eta_i Z_i^2\right) \leq \alpha$ for $\eta_i = \lambda_i\left(-\mathbf{A}_0\right)$.

In the following, we write $\mathbf{a} \prec \mathbf{b}$ for two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$ to indicate that $\mathbf{b}$ majorizes $\mathbf{a}$, that is, with the elements of a_i and b_i sorted in descending order, $\sum_{i=1}^j a_i \leq \sum_{i=1}^j b_i$ for all $j = 1, \ldots, k$, and $\sum_{i=1}^k a_i = \sum_{i=1}^k b_i$. Let $\bar{\mathbf{A}}_1 = \frac{1}{2}(\mathbf{A}_1 + \mathbf{A}_1')$. From Theorems 9.F.1 and 9.G.1 in Marshall, Olkin, and Arnold (2011)

$$
\begin{aligned}
(\lambda_1(-\mathbf{A}_1), \ldots, \lambda_{q+1}(-\mathbf{A}_1)) &\prec (\lambda_1(-\bar{\mathbf{A}}_1), \ldots, \lambda_{q+1}(-\bar{\mathbf{A}}_1)) \\
&\prec \left(\int \lambda_1(-\bar{\mathbf{A}}(\theta))dF(\theta), \ldots, \right. \\
&\qquad \left. \int \lambda_q(-\bar{\mathbf{A}}(\theta))dF(\theta), \int \lambda_{q+1}(-\bar{\mathbf{A}}(\theta))dF(\theta)\right).
\end{aligned} \tag{22}
$$

Since $\int \lambda_{q+1}(-\bar{\mathbf{A}}(\theta))dF(\theta) = -\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta)$ and $\lambda_{q+1}(-\mathbf{A}_1) = -\lambda_1(\mathbf{A}_1)$, we have

$$
-\lambda_1(\mathbf{A}_1) + \sum_{j=1}^q \lambda_j(-\mathbf{A}_1) = -\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) + \sum_{j=1}^q \int \lambda_j(-\bar{\mathbf{A}}(\theta))dF(\theta).
$$

The majorization result (22) further implies

$$
\lambda_1(\mathbf{A}_1) \leq \lambda_1(\bar{\mathbf{A}}_1) \leq \int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) \tag{23}
$$

so that also

$$
\begin{aligned}
(\lambda_1(-\mathbf{A}_1), \ldots, \lambda_q(-\mathbf{A}_1)) \prec &\left(\int \lambda_1(-\bar{\mathbf{A}}(\theta))dF(\theta), \ldots, \right. \\
&\left. \int \lambda_{q-1}(-\bar{\mathbf{A}}(\theta))dF(\theta), \int \lambda_q(-\bar{\mathbf{A}}(\theta))dF(\theta) - \left(\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta)) - \lambda_1(\mathbf{A}_1)\right)\right).
\end{aligned}
$$

with the elements still sorted in descending order. Thus, with $\tilde{\eta}_i = \int \lambda_i(-\bar{\mathbf{A}}(\theta))dF(\theta)/\lambda_1(\mathbf{A}_1)$ for $i = 1, \ldots, q-1$ and

$$
\tilde{\eta}_q = \frac{\int \lambda_q(-\bar{\mathbf{A}}(\theta))dF(\theta) - (\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta)) - \lambda_1(\mathbf{A}_1))}{\lambda_1(\mathbf{A}_1)}
$$

we have $(\bar{\eta}_1, \ldots, \bar{\eta}_q) \prec (\tilde{\eta}_1, \ldots, \tilde{\eta}_q)$. From the integral representation of Lemma 1 (ii), the application of the Schur-Ostrowski criterion (Theorem 3.A.4 in Marshall, Olkin, and

Arnold (2011)) shows that $\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q a_i Z_i^2\right)$ is Schur convex in $(a_1, \ldots, a_q)$, so that $\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \bar{\eta}_i Z_i^2\right) \leq \mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \tilde{\eta}_i Z_i^2\right)$.

Now applying (23), $\tilde{\eta}_i^* = \int \lambda_i(-\bar{\mathbf{A}}(\theta))dF(\theta) / \int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) \leq \tilde{\eta}_i$ for $i = 1, \ldots, q-1$, and since from Lemma 10, $\lambda_1(\mathbf{A}_1) \geq 1$, also

$$\tilde{\eta}_q^* = \frac{\int \lambda_q(-\bar{\mathbf{A}}(\theta))dF(\theta) - (\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) - 1)}{\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta)} \leq \tilde{\eta}_q$$

provided

$$\int \lambda_q(-\bar{\mathbf{A}}(\theta))dF(\theta) - \left(\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) - 1\right) \geq 0. \tag{24}$$

Since $\mathbb{P}(Z_0^2 \geq \sum_{i=1}^q \tilde{\eta}_i Z_i^2)$ is a decreasing function in $\tilde{\eta}_i$, $\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \tilde{\eta}_i Z_i^2\right) \leq \mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \tilde{\eta}_i^* Z_i^2\right)$. By Theorem 3.A.8 of Marshall, Olkin, and Arnold (2011), the Schur-convexity of $\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q a_i Z_i^2\right)$ in $(a_1, \ldots, a_q)$ and $\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \eta_i Z_i^2\right) \leq \alpha$, it now suffices to show that $\sum_{i=1}^j \tilde{\eta}_{q+1-i}^* \geq \sum_{i=1}^j \eta_{q+1-i}$ for all $1 \leq j \leq q$, and since $\eta_q \geq 0$, this also ensures that (24) holds. This latter condition may be rewritten as $\sum_{i=1}^j \int \nu_i(\theta) d\Pi(\theta) \geq 0$, and the result follows. ■

Proof of Lemma 3: (i) Since B is Gaussian, $n^{-1}\mathbf{W}_n^{0\prime}\mathbf{u}_n|\mathbf{s}_n \sim \mathcal{N}(0, \mathbf{\Omega}_n)$ with $\mathbf{\Omega}_n = n^{-2}\sum_{l,\ell} \mathbf{w}^0(s_l)\mathbf{w}^0(s_\ell)'\sigma_B(c\,(s_l - s_\ell))$. It thus suffices to show that $\mathbf{\Omega}_n \xrightarrow{p} \mathbf{\Omega}_{sc}$.

We have $\mathbf{\Omega}_n = \sigma_B(0)n^{-2}\sum_l \mathbf{w}^0(s_l)\mathbf{w}^0(s_l)' + n^{-2}\sum_{l\neq\ell} \mathbf{w}^0(s_l)\mathbf{w}^0(s_\ell)'\sigma_B(c\,(s_l - s_\ell))$, and $||n^{-2}\sum_l \mathbf{w}^0(s_l)\mathbf{w}^0(s_l)'|| \leq n^{-1}\sup_{s\in\mathcal{S}} ||\mathbf{w}^0(s)||^2 \to 0$. Furthermore,

$$\mathbb{E}\left[\frac{1}{n(n-1)}\sum_{l\neq\ell} \mathbf{w}^0(s_l)\mathbf{w}^0(s_\ell)'\sigma_B(c\,(s_l - s_\ell))\right] = \mathbb{E}[\mathbf{w}^0(s_1)\mathbf{w}^0(s_2)'\sigma_B(c\,(s_1 - s_2))] = \mathbf{\Omega}_{sc}$$

and with $w_i^0(s)$ the ith element of $\mathbf{w}^0(s)$,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{n(n-1)}\sum_{l\neq\ell} w_i^0(s_l)w_j^0(s_\ell)'\sigma_B(c\,(s_l - s_\ell))\right)^2\right] \\
&\quad = \frac{(n-2)(n-3)}{n(n-1)}\mathbb{E}[w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1 - s_2))]\mathbb{E}[w_i^0(s_3)w_j^0(s_4)'\sigma_B(c\,(s_3 - s_4))] \\
&\qquad + \frac{4(n-2)}{n(n-1)}\mathbb{E}[w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1 - s_2))w_i^0(s_1)w_j^0(s_3)'\sigma_B(c\,(s_1 - s_3))] \\
&\qquad\quad + \frac{2}{n(n-1)}\mathbb{E}[w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1 - s_2))w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1 - s_2))]
\end{aligned}$$

so that $\mathrm{Var}[\frac{1}{n(n-1)}\sum_{l\neq\ell} w_i^0(s_l)w_j^0(s_\ell)'\sigma_B(c\,(s_l - s_\ell))] = O(n^{-1})$, and therefore $\mathbf{\Omega}_n \xrightarrow{p} \mathbf{\Omega}_{sc}$.

(ii) By the Cramér-Wold device, it suffices to obtain the desired convergence for fixed linear combinations. Thus, for $\boldsymbol{\upsilon} \in \mathbb{R}^{q+1}$, define $w_\upsilon : \mathcal{S} \mapsto \mathbb{R}$ via $w_\upsilon(s) = \boldsymbol{\upsilon}'\mathbf{w}^0(s)$, a continuous function on a compact set $\mathcal{S}$. We apply Lahiri's (2003) Theorem 3.2, in the notation $\lambda_n \triangleq c_n$, $\mathbf{X}_i \triangleq s_l$, $\omega_n(\lambda_n^{-1}\mathbf{x}) \triangleq w_\upsilon(s)$ and $s_{1n}^2 \triangleq \int w_\upsilon(s)^2 g(s) ds$. For any $h \in \mathbb{R}^d$

$$\frac{\int w_\upsilon(s + c_n^{-1}h) w_\upsilon(s) g(s)^2 ds}{\int w_\upsilon(s)^2 g(s) ds} \to \frac{\int w_\upsilon(s)^2 g(s)^2 ds}{\int w_\upsilon(s)^2 g(s) ds} \triangleq Q_1$$

since the density g is also continuous, so Lahiri's Condition (S.1) holds. If $a_n \to a > 0$, then $n/c_n^d \to a^{-1} \triangleq C_1$, and the result follows from Lahiri's equation (3.4). If $a_n \to 0$, then $n/c_n^d \to \infty$ and $a_n^{1/2} n^{-1/2} = (c_n^d/n^2)^{1/2}$, so the result follows from Lahiri's equation (3.5). ■

Proof of Theorem 4: In the notation of Lemma 3, with $\mathbf{X} = (X_0, \mathbf{X}'_{1:q})'$ and $\mathbf{Z} = (Z_0, \ldots, Z_q)'$ we have $\mathbb{P}\left(\tau_n^2(\mathbf{W}_n\mathbf{W}'_n) > \text{cv}^2 \,|\mathbf{s}_n\right) \xrightarrow{p} \mathbb{P}\left(X_0^2/(\mathbf{X}'_{1:q}\mathbf{X}_{1:q}) > \text{cv}^2\right)$ from Lemma 3 and the continuous mapping theorem, so the result follows as in the proof of Lemma 1. ■

The proof of Theorem 5 requires a number of technical preliminaries.

Lemma 11. *If $\mathbf{X}_n|\mathbf{s}_n \Rightarrow_p \mathbf{X}$ and $\mathbf{Y}_n \xrightarrow{p} 0$, then $(\mathbf{X}_n + \mathbf{Y}_n)|\mathbf{s}_n \Rightarrow_p \mathbf{X}$.*

Proof. Let BL be the space of Lipschitz continuous functions $\mathbb{R}^p \mapsto \mathbb{R}$ bounded by one with unit Lipschitz constant. By Berti, Pratelli, and Rigo (2006), page 93, $\mathbf{X}_n|\mathbf{s}_n \Rightarrow_p \mathbf{X}$ is equivalent to $\sup_{h\in\text{BL}} |\mathbb{E}[h(\mathbf{X}_n) - h(\mathbf{X})|\mathbf{s}_n]| \xrightarrow{p} 0$, so it suffices to show that $\sup_{h\in\text{BL}} |\mathbb{E}[h(\mathbf{X}_n + \mathbf{Y}_n) - h(\mathbf{X})|\mathbf{s}_n]| \xrightarrow{p} 0$. Let $\mathbf{Y}_n^* = \mathbf{Y}_n \mathbf{1}[||\mathbf{Y}_n|| \leq 1]$, so that

$$\sup_{h\in\text{BL}} |\mathbb{E}[h(\mathbf{X}_n + \mathbf{Y}_n) - h(\mathbf{X})|\mathbf{s}_n]| \leq \sup_{h\in\text{BL}} |\mathbb{E}[h(\mathbf{X}_n + \mathbf{Y}_n^*) - h(\mathbf{X})|\mathbf{s}_n]| + 2\mathbb{P}(||\mathbf{Y}_n^*|| > 1|\mathbf{s}_n).$$

Note that with $\Delta_n(h) = h(\mathbf{X}_n + \mathbf{Y}_n^*) - h(\mathbf{X}_n)$, $|\Delta_n(h)| \leq ||\mathbf{Y}_n^*||$ a.s. for all $h \in \text{BL}$, so that

$$\begin{aligned}
\sup_{h\in\text{BL}} |\mathbb{E}[h(\mathbf{X}_n + \mathbf{Y}_n^*) - h(\mathbf{X})|\mathbf{s}_n]| &= \sup_{h\in\text{BL}} |\mathbb{E}[\Delta_n(h) + h(\mathbf{X}_n) - h(\mathbf{X})|\mathbf{s}_n]| \\
&\leq \sup_{h\in\text{BL}} \left(|\mathbb{E}[\Delta_n(h)|\mathbf{s}_n]| + |\mathbb{E}[h(\mathbf{X}_n) - h(\mathbf{X})|\mathbf{s}_n]|\right) \\
&\leq \mathbb{E}[||\mathbf{Y}_n^*|||\mathbf{s}_n] + \sup_{h\in\text{BL}} |\mathbb{E}[h(\mathbf{X}_n) - h(\mathbf{X})|\mathbf{s}_n]|.
\end{aligned}$$

We are left to show that $\mathbf{Y}_n \xrightarrow{p} 0$ implies $\mathbb{P}(||\mathbf{Y}_n^*|| > 1|\mathbf{s}_n) \xrightarrow{p} 0$ and $\mathbb{E}[||\mathbf{Y}_n^*|||\mathbf{s}_n] \xrightarrow{p} 0$.

Consider the latter claim. Suppose otherwise. Then for some $\varepsilon > 0$, and some subsequence n' of n, $\lim_{n'\to\infty} \mathbb{P}(\mathbb{E}[||\mathbf{Y}_{n'}^*|||\mathbf{s}_{n'}] > \varepsilon) > \varepsilon$, so that $\liminf_{n'\to\infty} \mathbb{E}[||\mathbf{Y}_{n'}^*||] > \varepsilon^2$. But since $\mathbf{Y}_n^*$ is bounded, $\mathbf{Y}_n \xrightarrow{p} 0$ implies $\lim_{n\to\infty} \mathbb{E}[||\mathbf{Y}_n^*||] = 0$, a contradiction. A similar argument yields $\mathbb{E}[||\mathbf{Y}_n^*|||\mathbf{s}_n] \xrightarrow{p} 0$, concluding the proof. □

Lemma 12. *Suppose the mapping* $\hat{\mathbf{w}}^0 : \mathcal{S} \mapsto \mathbb{R}^{q+1}$ *is a function of* $\mathbf{s}_n$ *(but not of* B*), and*

$$\sup_{s\in\mathcal{S}} ||\hat{\mathbf{w}}^0(s) - \mathbf{w}^0(s)|| \xrightarrow{p} 0. \tag{25}$$

Then Lemma 3 and Theorem 4 continue to hold with $\hat{\mathbf{W}}_n^0$ *in place of* $\mathbf{W}_n^0$*, where the* i*th row of* $\hat{\mathbf{W}}_n^0$ *is equal to* $(1, \hat{\mathbf{w}}(s_i)')$*.*

Proof. We show that Lemma 3 (i) and (ii) continue to hold with $\mathbf{w}^0$ replaced by $\hat{\mathbf{w}}^0$. We have

$$\mathbb{E}\left[\left(\sum_{l=1}^{n}(\hat{w}_i^0(s_l) - w_i^0(s_l))u(s_l)\right)^2 |\mathbf{s}_n\right] \leq \sup_{s\in\mathcal{S}} |\hat{w}_i^0(s) - w_i^0(s)|^2 \sum_{l,\ell} |\sigma_B(c_n(s_l - s_\ell))|$$

almost surely. Proceeding as in the proof of Lemma 3 (i) now shows that $\mathbb{E}[n^{-2}\sum_{l,\ell}|\sigma_B(c(s_l - s_\ell))|] = \int\int |\sigma_B(c(r-s))|g(r)g(s)drds$, so $n^{-2}\sum_{l,\ell}|\sigma_B(c(s_l - s_\ell))| = O_p(1)$. Similarly, under the assumptions of part (ii) of Lemma 3, proceeding as in the proof of Lemma 5.2 of Lahiri (2003) yields $\mathbb{E}[a_n n^{-1}\sum_{l,\ell}|\sigma_B(c_n(s_l - s_\ell))|] \to a\sigma_u^2 + \int_{\mathbb{R}^d}|\sigma_B(s)|ds \int g(s)^2 ds$. The result thus follows from (25) and Lemma 11. □

Lemma 13. *In the notation of Lemma 6, suppose* $\hat{\mathbf{W}} = \hat{\mathbf{L}}\hat{\mathbf{\Phi}}$*, where the* i*th column of the* $n\times q$ *matrix* $\hat{\mathbf{\Phi}}$ *is* $\hat{\mathbf{v}}_i = (\hat{\varphi}_i(s_1), \ldots, \hat{\varphi}_i(s_n))'$ *and* $\hat{\mathbf{L}} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_q)$*. Under the assumptions of Lemma 3,* $c_n^d n^{-2}(\mathbf{u}'\hat{\mathbf{W}}\hat{\mathbf{W}}'\mathbf{u} - \mathbf{u}'\mathbf{W}\mathbf{W}'\mathbf{u})|\mathbf{s}_n \xrightarrow{p} 0$*, where* $\mathbf{W} = \mathbf{L}\mathbf{\Phi}$*,* $\mathbf{L} = \mathrm{diag}(\lambda_1\mathbf{l}_{m_1}, \ldots, \lambda_p\mathbf{l}_{m_p})$ *and the* i*th column of* $\mathbf{\Phi}$ *is equal to* $(\varphi_i(s_1), \ldots, \varphi_i(s_n))'$*.*

Proof. With $\hat{\mathbf{O}} = \mathrm{diag}(\hat{\mathbf{O}}_{(1)}, \ldots, \hat{\mathbf{O}}_{(p)})$,

$$\begin{aligned} c_n^d n^{-2}\mathbf{u}'\hat{\mathbf{\Phi}}\hat{\mathbf{L}}^2\hat{\mathbf{\Phi}}'\mathbf{u} &= c_n^d n^{-2}\mathbf{u}'\hat{\mathbf{\Phi}}\hat{\mathbf{O}}\hat{\mathbf{O}}'\hat{\mathbf{L}}^2\hat{\mathbf{O}}'\hat{\mathbf{O}}\hat{\mathbf{\Phi}}'\mathbf{u} = c_n^d n^{-2}\mathbf{u}'\mathbf{\Phi}\hat{\mathbf{O}}'\hat{\mathbf{L}}^2\hat{\mathbf{O}}'\mathbf{\Phi}'\mathbf{u} + o_p(1) \\ &= c_n^d n^{-2}\mathbf{u}'\mathbf{\Phi}\hat{\mathbf{O}}'\mathbf{L}^2\hat{\mathbf{O}}'\mathbf{\Phi}'\mathbf{u} + o_p(1) = c_n^d n^{-2}\mathbf{u}'\mathbf{\Phi}\mathbf{L}^2\mathbf{\Phi}'\mathbf{u} + o_p(1) \end{aligned}$$

where the first equality follows from $\hat{\mathbf{O}}'\hat{\mathbf{O}} = \mathbf{I}_q$, the second from Lemma 6 (a) and (b) and the reasoning in the proof of Lemma 12, the third from Lemma 6 (b) and $||c_n^{d/2}n^{-1}\hat{\mathbf{O}}'\mathbf{\Phi}'\mathbf{u}|| \leq ||\hat{\mathbf{O}}|| \cdot ||c_n^{d/2}n^{-1}\mathbf{\Phi}'\mathbf{u}|| = O_p(1)$ using Lemma 3, and the fourth from $\hat{\mathbf{O}}'\mathbf{L}^2\hat{\mathbf{O}}' = \mathbf{L}^2$ a.s. The result now follows from Lemma 11. □

Proof of Theorem 5: For the first claim, by Theorem 4.4.6 of Harkrishan (2017), $\omega_0 = \sup_{||f||=1}\langle f, RTRf\rangle$, so it suffices to show that for some $f \in \mathcal{L}_G^2$, $\langle f, RTRf\rangle > 0$. In the weak correlation case, this holds for $f(s) = (\kappa + (1-\kappa)g(s))^{-1/2}$, since $\langle f, R_{wc}TR_{wc}f\rangle =$

$\langle 1, T1\rangle = \int\int(1-\bar{k}(r,s))dG(r)dG(s) = 1$. In the strong correlation case, the same conclusion holds by setting f such that $R_{sc}f = 1$. Such an f exists, because the kernel of R_{sc}^2 is equal to $\{0\}$ by assumption about σ_B, so the range of R_{sc} is $\mathcal{L}_G^2\backslash\{0\}$ by Theorem 3.5.8 of Harkrishan (2017).

Under the null hypothesis, $\mathbb{P}(\tau_n^2(\overline{\mathbf{K}}_n) > \text{cv}^2\,|\mathbf{s}_n) = \mathbb{P}(\hat{\xi}_n > 0|\mathbf{s}_n)$, where $\hat{\xi}_n = c_n^d n^{-2}\sum_{l,\ell} u_l u_\ell(1 - \text{cv}^2\,\hat{k}_n(s_l, s_\ell))$. By construction of $\hat{\lambda}_i$ and $\hat{\varphi}_i(\cdot)$ in Lemma 6, for all $1 \leq l, \ell \leq n$, $\hat{k}_n(s_l, s_\ell) = \sum_{i=1}^n \hat{\lambda}_i\hat{\varphi}_i(s_l)\hat{\varphi}_i(s_\ell)$. For a given q satisfying the assumption of Lemma 6, and all $n > q$, let $\hat{k}_{n,q}(r,s) = \sum_{i=1}^q \hat{\lambda}_i\hat{\varphi}_i(r)\hat{\varphi}_i(s)$ and $\hat{\xi}_n^q = c_n^d n^{-2}\sum_{l,\ell} u_l u_\ell(1 - \text{cv}^2\,\hat{k}_{n,q}(s_l, s_\ell))$. We now show the last claim, that is $\mathbb{P}(\hat{\xi}_n > 0|\mathbf{s}_n) \xrightarrow{p} \mathbb{P}(\sum_{i=0}^\infty \omega_i Z_i^2 > 0)$, which is implied by the following three claims

$$\text{(i) for any } \varepsilon > 0 \quad \lim_{q\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\hat{\xi}_n - \hat{\xi}_n^q| > \varepsilon) = 0 \tag{26}$$

$$\text{(ii) for any fixed } q,\ \mathbb{P}(\hat{\xi}_n^q > 0|\mathbf{s}_n) \xrightarrow{p} \mathbb{P}\left(\sum_{i=0}^q \omega_{q,i}Z_i^2 > 0\right) \tag{27}$$

$$\text{(iii)} \quad \lim_{q\to\infty}\mathbb{P}\left(\sum_{i=0}^q \omega_{q,i}Z_i^2 > 0\right) = \mathbb{P}\left(\sum_{i=0}^\infty \omega_i Z_i^2 > 0\right) \tag{28}$$

for some double array of real numbers $\omega_{q,i}$ by invoking Lemma 11.

For claim (i), note that for all $n > q$, $\hat{\xi}_n \leq \hat{\xi}_n^q$ a.s., and

$$\begin{aligned}\mathbb{E}[\hat{\xi}_n^q - \hat{\xi}_n|\mathbf{s}_n] &= c_n^d n^{-2}\sum_{l,\ell}\sigma_B(c_n(s_l - s_\ell))\left(\sum_{i=q+1}^n \hat{\lambda}_i\hat{\varphi}_i(s_l)\hat{\varphi}_i(s_\ell)\right) \\ &\leq \hat{\lambda}_{q+1}c_n^d n^{-2}\sum_{l,\ell}\sigma_B(c_n(s_l - s_\ell))\end{aligned}$$

where the inequality follows from $\text{tr}(\mathbf{AB}) \leq \lambda_1(\mathbf{A})\,\text{tr}\,\mathbf{B}$ for positive semidefinite matrices $\mathbf{A}, \mathbf{B}$ and $\lambda_1(\mathbf{A})$ the largest eigenvalue of $\mathbf{A}$. By the same reasoning as employed in Theorem 12, $c_n^d n^{-2}\sum_{l,\ell}\sigma_B(c_n(s_l - s_\ell)) = O_p(1)$. Furthermore, by Lemma 6 (b), $|\hat{\lambda}_{q+1} - \lambda_{q+1}| = O_q(n^{-1/2})$, and $\lim_{q\to\infty}\lambda_q = 0$. Thus (26) follows.

For claim (ii), let $\varphi_0(s) = 1$ and $\lambda_0 = 1$. By Lemma 6 (a), Lemma 13 and Theorem 1, claim (27) holds, where $\omega_{q,i}$ are the eigenvalues of $\mathbf{D}(\text{cv})\mathbf{\Omega}$ for $\mathbf{\Omega} \in \{\mathbf{\Omega}_{sc}, \mathbf{\Omega}_{wc}\}$, and the $(i+1), (j+1)$ element of $\mathbf{\Omega}$ is equal to $\sqrt{\lambda_i\lambda_j}\int\int\varphi_i(s)\sigma_B(c(r-s))\varphi_j(r)dG(s)dG(r)$ and $\sqrt{\lambda_i\lambda_j}\int\varphi_i(s)\varphi_j(s)(\kappa + (1-\kappa)g(s))ds$ under strong and weak correlation, respectively.

For claim (iii), we first show that these $\omega_{q,i}$ are also the eigenvalues of the finite rank self-adjoint linear operators RT_qR, $R \in \{R_{sc}, R_{wc}\}$. To this end, let $\varphi_i^*(s) = \sqrt{\lambda_i}R\varphi_i(s)$. With

$d_0 = 1$ and $d_i = -\,\mathrm{cv}^2$, we have

$$RT_qR(f)(s) = \int \left(\sum_{i=0}^{q} d_i \varphi_i^*(s)\varphi_i^*(r) \right) f(r)dG(r)$$

and the $(i+1),(j+1)$ element of $\mathbf{\Omega}$ stated above is equal to $\sqrt{\lambda_i\lambda_j}\langle \varphi_i, R^2\varphi_j\rangle = \sqrt{\lambda_i\lambda_j}\langle R\varphi_i, R\varphi_j\rangle = \int \varphi_i^*(s)\varphi_j^*(s)dG(s)$. Let $\mathbf{v} = (v_0,\ldots,v_q)'$ be an eigenvector of $\mathbf{D}(\mathrm{cv})\mathbf{\Omega}$ corresponding to eigenvalue ω, $\mathbf{D}(\mathrm{cv})\mathbf{\Omega}\mathbf{v} = \omega\mathbf{v}$. Then $\mathbf{D}(\mathrm{cv})\mathbf{\Omega}\mathbf{v} = \omega\mathbf{v}$ implies

$$\int \begin{pmatrix} \varphi_0^*(r)\varphi_0^*(r) & \cdots & \varphi_q^*(r)\varphi_0^*(r) \\ -\,\mathrm{cv}^2\,\varphi_0^*(r)\varphi_1^*(r) & \cdots & -\,\mathrm{cv}^2\,\varphi_q^*(r)\varphi_1^*(r) \\ \vdots & \ddots & \vdots \\ -\,\mathrm{cv}^2\,\varphi_0^*(r)\varphi_q^*(r) & \cdots & -\,\mathrm{cv}^2\,\varphi_q^*(r)\varphi_q^*(r) \end{pmatrix} dG(r)\mathbf{v} = \omega\mathbf{v}.$$

Premultiplying both sides of this equation by $(\varphi_0^*(s),\ldots,\varphi_q^*(s))$ yields

$$\begin{aligned} \sum_{j=0}^{q}\sum_{i=0}^{q} v_j\varphi_i^*(s)\int d_i\varphi_j^*(r)\varphi_i^*(r)dG(r) &= \omega\sum_{j=0}^{q} v_j\varphi_j^*(s) \\ \int \left(\sum_{i=0}^{q} d_i\varphi_i^*(s)\varphi_i^*(r)\right)\left(\sum_{j=0}^{q} v_j\varphi_j^*(r)\right) dG(r) &= \omega\sum_{j=0}^{q} v_j\varphi_j^*(s) \qquad (29) \end{aligned}$$

so $\sum_{j=0}^{q} v_j\varphi_j^*(r)$ is an eigenvector of RT_qR with eigenvalue ω, and since the kernel of RT_qR contains all functions that are orthogonal to $\{\varphi_i^*\}_{i=0}^{q}$, these are the only nonzero eigenvalues.

Now let $\omega_{q,i}^{\Delta}$ be the eigenvalues of the self-adjoint linear operator $R(T-T_q)R$. By Kato (1987) (also see the development on page 911 of Rosasco, Belkin, and Vito (2010)), there is an enumeration of the eigenvalues $\omega_{q,i}$ such that

$$\sum_{i=0}^{\infty}(\omega_{q,i}-\omega_i)^2 \le \sum_{i=0}^{\infty}(\omega_{q,i}^{\Delta})^2 = ||R(T-T_q)R||_{HS} \qquad (30)$$

where $||R(T-T_q)R||_{HS}$ is the Hilbert-Schmidt norm on the operator $R(T-T_q)R : \mathcal{L}_G^2 \mapsto \mathcal{L}_G^2$ induced by the norm $\sqrt{\langle f,f\rangle}$. Now $||R(T-T_q)R||_{HS} \le ||R||^2 \cdot ||T-T_q||_{HS}$ (cf. (34) below), and since $T-T_q$ is an integral operator, $||T-T_q||_{HS} = \int\int \left(\sum_{i=q+1}^{\infty}\lambda_i\varphi_i(s)\varphi_j(s)\right)^2 dG(s)dG(r)$. By Mercer's Theorem, this converges to zero as $q\to\infty$, so that

$$\lim_{q\to\infty}\sum_{i=0}^{\infty}(\omega_{q,i}-\omega_i)^2 = 0. \qquad (31)$$

Thus using the same order of eigenvalues as in (30), we also have $\mathrm{Var}[\sum_{i=0}^{q}\omega_{q,i}Z_i^2 - \sum_{i=0}^{\infty}\omega_i Z_i^2] \leq 2\sum_{i=0}^{\infty}(\omega_{q,i}-\omega_i)^2$, with the right-hand side converging to zero as $q \to \infty$ by (31). But mean-square convergence implies convergence in distribution, and (28) follows.

For the second claim of the theorem, by Lemma 3, $\omega_{q,i} \leq 0$ for $i \geq 1$, which in conjunction with (31) implies $\omega_i \leq 0$ for $i \geq 1$. ■

Proof of Lemma 6: We initially show a weaker claim than part (a), namely that there exists a sequence of $q \times q$ rotation matrices $\hat{\mathbf{O}}_n = \hat{\mathbf{O}}_n(\mathbf{s}_n)$ with elements $\hat{O}_{n,ij}$ such that

$$\max_{i \leq q} \sup_{s \in \mathcal{S}} \left| \varphi_i(s) - \sum_{j=1}^{q} \hat{O}_{n,ij}\hat{\varphi}_i(s) \right| = O_p(n^{-1/2}). \tag{32}$$

The proof follows closely the development in Rosasco, Belkin, and Vito (2010), denoted RBV in the following. Let $k_0(r,s) = \bar{k}(r,s) + 1$. Conditional on $\mathbf{s}_n$, define the linear operators $\mathcal{L}_G^2 \mapsto \mathcal{L}_G^2$ $M(f)(s) = f(s) - \int f(r)dG(r)$, $M_n(f)(s) = f(s) - \int f(r)dG_n(r)$, $L(f)(s) = \int k_0(r,s)f(r)dG(r)$ and $L_n(f)(s) = \int k_0(r,s)f(r)dG_n(r)$ and the derived operators $\bar{L} = MLM$, $\bar{L}_n = ML_nM$ and $\hat{L}_n = M_nL_nM_n$, so that $\bar{L}(f)(s) = \int f(r)\bar{k}(r,s)dG(r)$, $\bar{L}_n(f)(s) = \int \bar{k}(r,s)f(r)dG_n(r)$ and $\hat{L}_n(f)(s) = \int \hat{k}_n(r,s)f(r)dG_n(r)$, where G_n is the empirical distribution of $\{s_l\}_{l=1}^n$.

Let $\mathcal{H} \subset \mathcal{L}_G^2$ be the (separable) Reproducing Kernel Hilbert Space (RKHS) of functions $f : \mathcal{S} \mapsto \mathbb{R}$ with kernel k_0 and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ satisfying $\langle f, k_0(\cdot, r)\rangle_{\mathcal{H}} = f(r)$ and associated norm $||f||_{\mathcal{H}}$. Let $K = \sup_{s \in \mathcal{S}} k_0(s,s)$. Define $\bar{\mathcal{H}}$ as the RKHS of functions $f : \mathcal{S} \mapsto \mathbb{R}$ with kernel $\bar{k}$, and $\mathcal{H}_1$ as the RKHS of functions $f : \mathcal{S} \mapsto \mathbb{R}$ with kernel equal to 1, which only consists of the constant function. Since $k_0 = \bar{k} + 1$, $\mathcal{H}$ contains all functions that can be written as linear combinations of $\bar{\mathcal{H}}$ and $\mathcal{H}_1$ (see, for instance, Theorem 2.16 in Saitoh and Sawano (2016)). Thus $\mathcal{H}$ contains the constant function, and $||1||_{\mathcal{H}} < \infty$. Furthermore, since for any $f \in \mathcal{H}$, $|f(r)| = |\langle f(\cdot), k_0(\cdot, r)\rangle_{\mathcal{H}}| \leq ||f||_{\mathcal{H}} \cdot ||k_0(\cdot, r)||_{\mathcal{H}} \leq \sqrt{K}||f||_{\mathcal{H}}$, we have

$$\sup_{r \in \mathcal{S}} |f(r)| \leq \sqrt{K} \cdot ||f||_{\mathcal{H}}. \tag{33}$$

As in RBV, view the operators above as operators on $\mathcal{H} \mapsto \mathcal{H}$. The operator norm $||A||$ of the operator $A : \mathcal{H} \mapsto \mathcal{H}$ is defined as $\sup_{||f||_{\mathcal{H}}=1} ||Af||_{\mathcal{H}}$, and A is called bounded if $||A|| < \infty$. A bounded operator A is Hilbert-Schmidt if $\sum_{j=1}^{\infty} ||Ae_j|| < \infty$ for some (any) orthonormal basis e_j. The space of Hilbert-Schmidt operators is a Hilbert space endowed with the norm $||A||_{HS} = \sqrt{\sum_{j=1}^{\infty}\langle Ae_j, Ae_j\rangle_{\mathcal{H}}}$, and for any Hilbert-Schmidt operator A and bounded operator

B,

$$||AB||_{HS} \leq ||A||_{HS}||B||, \; ||BA||_{HS} \leq ||B|| \cdot ||A||_{HS}. \tag{34}$$

By Theorem 7 of RBV, L and L_n are Hilbert-Schmidt.

Furthermore, for any $f \in \mathcal{H}$,

$$||Mf||_{\mathcal{H}} = ||f - \int f(r)dG(r)||_{\mathcal{H}} \leq ||f||_{\mathcal{H}} + ||1||_{\mathcal{H}} \int f(r)dG(r) \leq ||f||_{\mathcal{H}} + ||1||_{\mathcal{H}} \sup_{r \in \mathcal{S}} |f(r)|$$

so that (33) implies that $||M||$ is a bounded operator. By the same argument, so is M_n (almost surely). Thus, from (34), also $\bar{L}$, $\bar{L}_n$ and $\hat{L}_n$ are Hilbert-Schmidt for almost all $\mathbf{s}_n$.

Conditioning on $\mathbf{s}_n$ throughout, we have the almost sure inequalities $||\hat{L}_n - \bar{L}||_{HS} \leq ||\hat{L}_n - \bar{L}_n||_{HS} + ||\bar{L}_n - \bar{L}||_{HS}$ and, using (34),

$$\begin{aligned} ||\hat{L}_n - \bar{L}_n||_{HS} &\leq ||(M_n - M)L_n M_n||_{HS} + ||ML_n(M_n - M)||_{HS} \\ &\leq ||M_n - M|| \cdot ||M_n|| \cdot ||L_n||_{HS} + ||M_n - M|| \cdot ||M|| \cdot ||L_n||_{HS} \end{aligned}$$

as well as

$$\begin{aligned} ||(M_n - M)f||_{\mathcal{H}} &= \left\| \int f(r)dG_n(r) - \int f(r)dG(r) \right\|_{\mathcal{H}} \\ &= ||1||_{\mathcal{H}} \left| \int f(r)dG_n(r) - \int f(r)dG(r) \right| \\ &= ||1||_{\mathcal{H}} \left| \left\langle f, n^{-1} \sum_{l=1}^{n} \zeta_l \right\rangle_{\mathcal{H}} \right| \leq ||1||_{\mathcal{H}} ||f||_{\mathcal{H}} \left\| n^{-1} \sum_{l=1}^{n} \zeta_l \right\|_{\mathcal{H}} \end{aligned}$$

with $\zeta_l = k_0(\cdot, s_l) - \int k_0(\cdot, r)dG(r) \in \mathcal{H}$. Since s_l is i.i.d. with distribution G, $\mathbb{E}[\zeta_l] = 0$ and $||\zeta_l||_{\mathcal{H}} \leq 2\sqrt{K}$ a.s. By Hoeffding's inequality for random elements that take values in separable Hilbert spaces (cf. equation (3) in RBV), $\|n^{-1} \sum_{l=1}^{n} \zeta_l\|_{\mathcal{H}} \leq 2\sqrt{K\delta} n^{-1/2}$ with probability of at least $1 - 2e^{-\delta}$. We conclude that $||M_n - M|| = \sup_{||f||_{\mathcal{H}}=1} ||(M_n - M)f||_{\mathcal{H}} = O_p(n^{-1/2})$.

Furthermore, applying the same reasoning as in the proof of Theorem 7 of RBV, $||\bar{L}_n - \bar{L}||_{HS} = O_p(n^{-1/2})$. Thus, $||\hat{L}_n - \bar{L}||_{HS} = O_p(n^{-1/2})$.

The conclusion now follows from similar arguments as employed in Proposition 10 and 12 of RBV. In particular, note that $\varphi_i \in \mathcal{H}$ for all i. Furthermore, $\int \varphi_i(s)dG(s) = \lambda_i^{-1} \int \varphi_i(r)\bar{k}(r,s)dG(r)dG(s) = 0$. Thus, with $e_i = \sqrt{\lambda_i}\varphi_i \in \mathcal{H}$, $Me_i = e_i$, and $\langle e_i, e_i \rangle_{\mathcal{H}} = \langle e_i(\cdot), \lambda_i^{-1} \int \bar{k}(r, \cdot)e_i(r)dG(r) \rangle_{\mathcal{H}} = \lambda_i^{-1} \langle e_i, \bar{L}e_i \rangle_{\mathcal{H}} = \lambda_i^{-1} \langle e_i, Le_i \rangle_{\mathcal{H}} =$

$\lambda_i^{-1}\int\langle e_i(\cdot),k_0(r,\cdot)\rangle_{\mathcal{H}}e_i(r)dG(r) = \lambda_i^{-1}\int e_i^2(r)dG(r) = 1$, so that e_i are normalized eigenvectors of $\bar{L}:\mathcal{H}\mapsto\mathcal{H}$. Since $\mathcal{H}\subset\mathcal{L}_G^2$, these are the only eigenfunctions of $\bar{L}:\mathcal{H}\mapsto\mathcal{H}$ with positive eigenvalue, so that the spectrum of $\bar{L}$ is equal to $\{\lambda_i\}_{i=1}^{\infty}$ (cf. Proposition 8 of RBV).

Also, $\hat{\varphi}_i\in\mathcal{H}$, and since $\hat{\mathbf{v}}_i$ is the eigenvector of $n^{-1}\hat{\mathbf{K}}_n$ with eigenvalue $\hat{\lambda}_i$, $n^{-1}\hat{\mathbf{K}}_n\hat{\mathbf{v}}_i=\hat{\lambda}_i\hat{\mathbf{v}}_i$, we obtain for $\hat{\lambda}_i>0$ that

$$\begin{aligned}\hat{L}_n(\hat{\varphi}_i)(\cdot) &= \int\hat{k}_n(r,\cdot)\hat{\varphi}_i(r)dG_n(r)=n^{-1}\sum_{j=1}^n\hat{k}_n(\cdot,s_j)\hat{\varphi}_i(s_j)\\ &= n^{-2}\hat{\lambda}_i^{-1}\sum_{j=1}^n\hat{k}_n(\cdot,s_j)\sum_{l=1}^n\hat{v}_{i,l}\hat{k}_n(s_j,s_l)=n^{-1}\sum_{j=1}^n\hat{k}_n(\cdot,s_j)\hat{v}_{i,j}=\hat{\lambda}_i\hat{\varphi}_i(\cdot)\end{aligned}$$

and

$$\int\hat{\varphi}_i(r)^2dG_n(r)=n^{-3}\hat{\lambda}_i^{-2}\sum_{j=1}^n\sum_{\ell=1}^n\sum_{t=1}^n\hat{v}_{i,j}\hat{k}_n(s_j,s_\ell)\hat{k}_n(s_\ell,s_t)\hat{v}_{i,t}=1.$$

Furthermore, from $\sum_{l=1}^n\hat{v}_{i,l}=0$, also $\int\hat{\varphi}_i(s)dG_n(s)=0$, so that $M_n\hat{e}_i=\hat{e}_i$. Thus, with $\hat{e}_i=\sqrt{\hat{\lambda}_i}\hat{\varphi}_i\in\mathcal{H}$, $\langle\hat{e}_i,\hat{e}_i\rangle_{\mathcal{H}}=\langle\hat{e}_i(\cdot),\hat{\lambda}_i^{-1}\int\hat{k}_n(r,\cdot)\hat{e}_i(r)dG_n(r)\rangle_{\mathcal{H}}=\hat{\lambda}_i^{-1}\langle\hat{e}_i,\hat{L}_n\hat{e}_i\rangle_{\mathcal{H}}=\hat{\lambda}_i^{-1}\langle\hat{e}_i,L_n\hat{e}_i\rangle_{\mathcal{H}}=\hat{\lambda}_i^{-1}\int\langle\hat{e}_i(\cdot),k_0(r,\cdot)\rangle_{\mathcal{H}}\hat{e}_i(r)dG_n(r)=\hat{\lambda}_i^{-1}\int\hat{e}_i(r)^2dG_n(r)=1$. Therefore $\hat{e}_i$ are normalized eigenfunctions of $\hat{L}_n:\mathcal{H}\mapsto\mathcal{H}$, and since all $f\in\mathcal{H}$ that are orthogonal to $\hat{e}_i$, $i=1,\ldots,n$ are in the kernel of $\hat{L}_n$, these are the only eigenfunctions of $\bar{L}:\mathcal{H}\mapsto\mathcal{H}$ with positive eigenvalue, so the spectrum of $\hat{L}_n:\mathcal{H}\mapsto\mathcal{H}$ is equal to $\{\hat{\lambda}_i\}_{i=1}^n$ (cf. Proposition 9 of RBV).

Part (b) of the lemma now follows from $||\hat{L}_n-\bar{L}||_{HS}^2=O_p(n^{-1})$ and the development on page 911 of RBV.

To establish (32), note that with the projection operators $P^q:\mathcal{H}\mapsto\mathcal{H}$ and $\hat{P}^q:\mathcal{H}\mapsto\mathcal{H}$ defined via $P^q(f)(\cdot)=\sum_{i=1}^q\langle f,e_i\rangle_{\mathcal{H}}e_i(\cdot)$ and $\hat{P}^q(f)(\cdot)=\sum_{i=1}^q\langle f,\hat{e}_i\rangle_{\mathcal{H}}\hat{e}_i(\cdot)$, by Proposition 6 of RBV, $||\hat{P}^q-P^q||_{HS}\leq 2(\lambda_q-\lambda_{q+1})^{-1}||\hat{L}_n-\bar{L}||_{HS}+o_p(n^{-1/2})=O_p(n^{-1/2})$. Define the $q\times q$ matrix $\tilde{\mathbf{O}}_n$ with i,jth element $\tilde{O}_{n,ij}=\langle\hat{e}_i,e_j\rangle_{\mathcal{H}}$. Then the j,tth element of $\tilde{\mathbf{O}}_n'\tilde{\mathbf{O}}_n$ is given by $\sum_{i=1}^q\tilde{O}_{n,ij}\tilde{O}_{n,it}=\sum_{i=1}^q\langle\hat{e}_i,e_j\rangle_{\mathcal{H}}\langle\hat{e}_i,e_t\rangle_{\mathcal{H}}=\langle e_j,\hat{P}^q(e_t)\rangle_{\mathcal{H}}$, and $\mathbf{1}[j=t]=\langle e_j,P^q(e_t)\rangle_{\mathcal{H}}$, so that by the Cauchy-Schwarz inequality

$$\begin{aligned}\left|\sum_{i=1}^q\tilde{O}_{n,ij}\tilde{O}_{n,it}-\mathbf{1}[j=t]\right| &= \left|\langle e_j,(\hat{P}^q-P^q)e_t\rangle_{\mathcal{H}}\right|\\ &\leq ||\hat{P}^q-P^q||_{HS}=O_p(n^{-1/2}).\end{aligned}$$

Thus $||\tilde{\mathbf{O}}_n'\tilde{\mathbf{O}}_n-\mathbf{I}_q||=O_p(n^{-1/2})$, and with $\hat{\mathbf{O}}_n=(\tilde{\mathbf{O}}_n'\tilde{\mathbf{O}}_n)^{-1/2}\tilde{\mathbf{O}}_n$, also $||\hat{\mathbf{O}}_n-\tilde{\mathbf{O}}_n||=O_p(n^{-1/2})$.

Furthermore, with $\hat{r}_i^2 = \lambda_i/\hat{\lambda}_i \xrightarrow{p} 1$ using part (b) of the lemma,

$$\begin{aligned}
\sqrt{\lambda_i}||\sum_{j=1}^{q}\hat{O}_{n,ij}\hat{\varphi}_j - \varphi_i||_{\mathcal{H}} &= ||\hat{r}_i\sum_{j=1}^{q}\hat{O}_{n,ij}\hat{e}_j - e_i||_{\mathcal{H}} \\
&\leq ||\sum_{j=1}^{q}\tilde{O}_{n,ij}\hat{e}_j - e_i||_{\mathcal{H}} + ||\sum_{j=1}^{q}(\hat{r}_i\hat{O}_{n,ij} - \tilde{O}_{n,ij})\hat{e}_j||_{\mathcal{H}} \\
&\leq ||(\hat{P}^q - P^q)e_i||_{\mathcal{H}} + \sum_{j=1}^{q}|\hat{r}_i\hat{O}_{n,ij} - \tilde{O}_{n,ij}| \\
&\leq ||\hat{P}^q - P^q||_{HS} + \sum_{j=1}^{q}|\hat{r}_i\hat{O}_{n,ij} - \tilde{O}_{n,ij}| = O_p(n^{-1/2})
\end{aligned}$$

so (32) follows from (33).

The claim in part (a) of the lemma now follows by induction from (32): For $p = 1$, this follows directly. Suppose the result holds for $p-1$, and let $\hat{\mathbf{O}}_B = \text{diag}(\hat{\mathbf{O}}_{(1)}, \ldots, \hat{\mathbf{O}}_{(p-1)})$, so that

$$\sup_{s\in\mathcal{S}}||\hat{\mathbf{O}}_B\hat{\boldsymbol{\varphi}}_B(s) - \boldsymbol{\varphi}_B(s)|| = O_p(n^{-1/2}), \tag{35}$$

with $\boldsymbol{\varphi}_B$ and $\hat{\boldsymbol{\varphi}}_B$ the vector of the first $\sum_{j=1}^{p-1} m_j$ eigenfunctions. Now let

$$\hat{\mathbf{O}}_I = \begin{pmatrix} \hat{\mathbf{O}}_{11} & \hat{\mathbf{O}}_{12} \\ \hat{\mathbf{O}}_{21} & \hat{\mathbf{O}}_{22} \end{pmatrix}$$

be the $(\sum_{j=1}^{p} m_j) \times (\sum_{j=1}^{p} m_j)$ matrix $\hat{\mathbf{O}}_n$ of (32) applied with $q = \sum_{j=1}^{p} m_j$, with $\hat{\mathbf{O}}_{11}$ of the same dimensions as $\hat{\mathbf{O}}_B$. Let $\boldsymbol{\varphi}_{I-B}$ and $\hat{\boldsymbol{\varphi}}_{I-B}$ be the $m_p \times 1$ vectors of eigenfunctions with indices $\sum_{j=1}^{p-1} m_j + 1, \ldots, \sum_{j=1}^{p} m_j$, so that by the conclusion of (32), $\sup_{s\in\mathcal{S}}||\hat{\mathbf{O}}_{11}\hat{\boldsymbol{\varphi}}_B(s) + \hat{\mathbf{O}}_{12}\hat{\boldsymbol{\varphi}}_{I-B}(s) - \boldsymbol{\varphi}_B(s)|| = O_p(n^{-1/2})$ and $\sup_{s\in\mathcal{S}}||\hat{\mathbf{O}}_{21}\hat{\boldsymbol{\varphi}}_B(s) + \hat{\mathbf{O}}_{22}\hat{\boldsymbol{\varphi}}_{I-B}(s) - \boldsymbol{\varphi}_{I-B}(s)|| = O_p(n^{-1/2})$. In conjunction with (35), the former yields $\sup_{s\in\mathcal{S}}||(\hat{\mathbf{O}}_{11} - \hat{\mathbf{O}}_B)\hat{\boldsymbol{\varphi}}_B(s) + \hat{\mathbf{O}}_{12}\hat{\boldsymbol{\varphi}}_{I-B}(s)|| = O_p(n^{-1/2})$, which implies in light of (32) and the linear independence of eigenvectors that both $||\hat{\mathbf{O}}_{11} - \hat{\mathbf{O}}_B|| = O_p(n^{-1/2})$ and $||\hat{\mathbf{O}}_{12}|| = O_p(n^{-1/2})$. Since $\hat{\mathbf{O}}_I$ and $\hat{\mathbf{O}}_B$ are rotation matrices, $\hat{\mathbf{O}}_B'\hat{\mathbf{O}}_B = \hat{\mathbf{O}}_{11}'\hat{\mathbf{O}}_{11} + \hat{\mathbf{O}}_{21}'\hat{\mathbf{O}}_{21} = \mathbf{I}$, so that $||\hat{\mathbf{O}}_{11} - \hat{\mathbf{O}}_B|| = O_p(n^{-1/2})$ further implies $||\hat{\mathbf{O}}_{21}|| = O_p(n^{-1/2})$. We conclude that also $\sup_{s\in\mathcal{S}}||\hat{\mathbf{O}}_{22}\hat{\boldsymbol{\varphi}}_{I-B}(s) - \boldsymbol{\varphi}_{I-B}(s)|| = O_p(n^{-1/2})$, so that the result for p holds with $\hat{\mathbf{O}}_{(p)} = \hat{\mathbf{O}}_{22}$, which concludes the proof. ■

Proof of Theorem 8: Suppose $\max(\overline{\text{cv}}^2 - \text{cv}_n^2, 0) \xrightarrow{p} 0$ does not hold. Then there exists $\delta > 0$ such that $\limsup_{n\to\infty}\mathbb{P}(\overline{\text{cv}}^2 - \text{cv}_n^2 > \delta) > \delta$. Define $\varkappa(\kappa, \overline{\text{cv}}^2) =$

$\mathbb{P}\left(\sum_{i=0}^{\infty}\omega_i(\kappa,\overline{\mathrm{cv}})Z_i^2>0\right)$, so that $\sup_{0\leq\kappa<1}\varkappa(\kappa,\overline{\mathrm{cv}}^2)=\alpha$ by definition of $\overline{\mathrm{cv}}$. By continuity of $\varkappa$, there exists $0\leq\kappa_0<1$ and $\overline{\mathrm{cv}}^2-\delta/2\leq\overline{\mathrm{cv}}_0^2\leq\overline{\mathrm{cv}}^2$ such that $\varkappa(\kappa_0,\overline{\mathrm{cv}}_0^2)=\alpha$. If $\kappa_0=0$, set $c_{n,1}=c_{n,0}$. Otherwise, let $c_{n,1}\to\infty$ be such that the corresponding $a_{n,1}=c_{n,1}^d/n\to a_1$ satisfies $a_1\sigma_B^0(0)/(a_1\sigma_B^0(0)+\int\sigma_B^0(s)ds)=\kappa_0$. Now let $\mathrm{cv}_{n,1}^2$ solve $\mathbb{P}^0_{\mathbf{\Sigma}(c_{n,1})}(\tau_n^2\geq\mathrm{cv}_{n,1}^2\,|\mathbf{s}_n)=\alpha$ a.s., so that clearly, $\mathrm{cv}_{n,1}^2\leq\mathrm{cv}_n^2$ a.s. for all large enough n. Thus, with $\mathcal{A}_n$ the event that $\mathbf{s}_n$ takes on a value such that $\overline{\mathrm{cv}}^2-\mathrm{cv}_{n',1}^2>\delta$, we also have $\limsup_{n\to\infty}\mathbb{P}(\mathcal{A}_n)>\delta$, and there exists a subsequence $n'\to\infty$ of n such that $\mathbb{P}(\mathcal{A}_{n'})>\delta$ for all n'.

For all such n',

$$\alpha=\mathbb{P}^0_{\mathbf{\Sigma}(c_{n',1})}(\tau_{n'}^2\geq\mathrm{cv}_{n',1}^2\,|\mathcal{A}_{n'})\geq\mathbb{P}^0_{\mathbf{\Sigma}(c_{n',1})}(\tau_{n'}^2\geq\overline{\mathrm{cv}}^2-\delta|\mathcal{A}_{n'})\text{ a.s.}\tag{36}$$

and by Theorem 5, $\mathbb{P}^0_{\mathbf{\Sigma}(c_{n',1})}(\tau_{n'}^2\geq\overline{\mathrm{cv}}^2-\delta|\mathcal{A}_{n'})\to\varkappa(\kappa_0,\overline{\mathrm{cv}}^2-\delta)>\alpha$. This contradicts (36), and the result follows. ■

Theorem 14. *Let $\hat{q}_n$ be an arbitrary function of $\mathbf{s}_n$ taking values in $\mathcal{Q}=\{1,2,\ldots,q_{\max}\}$ for some sample size independent finite and nonrandom $q_{\max}$. Then for a t-statistic $\tau_n(q)$ that satisfies the conditions of Theorem 8 for all $q\in\mathcal{Q}$ with critical value $\mathrm{cv}_n(q)$ as in (20), for any $\epsilon>0$, $\limsup_{n\to\infty}\mathbb{P}(\mathbb{P}(\tau_n^2(\hat{q}_n)>\mathrm{cv}_n(\hat{q}_n)^2|\mathbf{s}_n)>\alpha+\epsilon)=0$.*

Proof. Suppose otherwise. Then there exists $\epsilon>0$ and a subsequence $n'\to\infty$ such that with $\mathcal{B}_n=\{\mathbf{s}_n:\mathbb{P}(\tau_n^2(\hat{q})>\mathrm{cv}_n(\hat{q})^2|\mathbf{s}_n)>\alpha+\epsilon\}\subset\mathcal{S}$, $\lim_{n'\to\infty}\mathbb{P}(\mathbf{s}_{n'}\in\mathcal{B}_{n'})>\epsilon$. Let $\mathcal{A}_{n,i}=\{\mathbf{s}_n:\hat{q}_n=i\}$, so that $\lim_{n'\to\infty}\sum_{i=1}^{q_{\max}}\mathbb{P}(\mathbf{s}_{n'}\in\mathcal{B}_{n'}\cap\mathcal{A}_{n',i})>\epsilon$. There hence exists some $1\leq q\leq q_{\max}$ and a further subsequence n'' of n' such that $\lim_{n''}\mathbb{P}(\mathbf{s}_{n''}\in\mathcal{B}_{n''}\cap\mathcal{A}_{n'',q})>\epsilon/q_{\max}$. But along this subsequence, q is fixed, so Theorem 8 applies and yields $\lim_{n''\to\infty}\mathbb{P}(\mathbf{s}_{n''}\in\mathcal{B}_{n''}\cap\mathcal{A}_{n'',q})\to 0$, yielding the desired contradiction. □

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Bakirov, N. K., and G. J. Székely (2005): "Student's T-Test for Gaussian Scale Mixtures," *Zapiski Nauchnyh Seminarov POMI*, 328, 5–19.

Berti, P., L. Pratelli, and P. Rigo (2006): "Almost sure weak convergence of random probability measures," *Stochastics: An International Journal of Probability and Stochastic Processes*, 78(2), 91–97.

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

Bester, C. A., T. G. Conley, and C. B. Hansen (2011): "Inference with Dependent Data Using Cluster Covariance Estimators," *Journal of Econometrics*, 165, 137–151.

Cao, J., C. Hansen, D. Kozbur, and L. Villacorta (2020): "Inference for Dependent Data with Learned Clusters," *Working paper.*

Chan, N. H., and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

Conley, T., and F. Molinari (2007): "Spatial Correlation Robust Inference with Errors in Location or Distance," *Journal of Econometrics*, 140, 76–96.

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

Dou, L. (2019): "Optimal HAR Inference," *Working Paper, Princeton University.*

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

Harkrishan, L. V. (2017): *Elements of Hilbert Spaces and Operator Theory.* Springer.

Henderson, J., T. Squires, A. Storeygard, and D. Weil (2018): "The Global Distribution of Economic Activity: Nature, History, and the Role of Trade," *Quarterly Journal of Economics*, 133(1), 357–406.

Ibragimov, R., and U. K. Müller (2010): "T-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, 453–468.

——— (2016): "Inference with Few Heterogeneous Clusters," *Review of Economics and Statistics*, 98(1), 83–96.

Kato, T. (1987): "Variation of discrete spectra," *Communications in Mathematical Physics*, 111(3), 501–504.

Kelejian, H. H., and I. R. Prucha (2007): "HAC estimation in a spatial framework," *Journal of Econometrics*, 140(1), 131 – 154.

Kelly, M. (2019): "The standard errors of persistence," *University College Dublin WP19/13.*

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

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

Kim, M., and Y. Sun (2011): "Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix," *Journal of Econometrics*, 160, 349–371.

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

Lahiri, S., and P. M. Robinson (2016): "Central limit theorems for long range dependent spatial linear processes," *Bernoulli*, 22(1), 345–375.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR Inference: Recommendations for Practice," *Journal of Business and Economic Statistics*, 36(4), 541–559.

Marshall, A. W., I. Olkin, and B. C. Arnold (2011): *Inequalities: theory of majorization and its applications*. Springer Series in Statistics, New York.

Müller, U. K. (2004): "A Theory of Robust Long-Run Variance Estimation," *Working paper, Princeton University*.

——— (2014): "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, 32, 311–322.

Müller, U. K. (2020): "A More Robust t-Test," *arXiv:2007.07065*.

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 (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016.

——— (2016): "Measuring Uncertainty about Long-Run Predictions," *Review of Economic Studies*, 83.

——— (2017): "Low-Frequency Econometrics," in *Advances in Economics: Eleventh World Congress of the Econometric Society*, ed. by B. Honoré, and L. Samuelson, vol. II, pp. 63–94. Cambridge University Press.

——— (2022): "Spatial Correlation Robust Inference in Linear Regression and Panel Models," *Manuscript*.

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

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.

Pötscher, B. M. (2002): "Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters," *Econometrica*, 70, 1035–1065.

Rho, S., and T. Vogelsang (2019): "Heteroskedasticity Autocorrelation Robust Inference in Time Series Regressions with Missing Data," *Econometric Theory*, 35, 601–629.

Robinson, P. M. (2005): "Robust Covariance Matrix Estimation: HAC Estimates with Long Memory/Antipersistence Correction," *Econometric Theory*, 21, 171–180.

Rosasco, L., M. Belkin, and E. D. Vito (2010): "On Learning with Integral Operators," *Journal of Machine Learning Research*, 11(30), 905–934.

Saitoh, S., and Y. Sawano (2016): *Theory of Reproducing Kernels and Applications*. Springer, New York.

Sun, Y. (2013): "Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator," *The Econometrics Journal*, 16, 1–26.

Sun, Y., and M. Kim (2012): "Asymptotic F-Test in a GMM Framework with Cross-Sectional Dependence," *Review of Economics and Statistics*, 91(1), 210–233.

Taqqu, M. (1975): "Weak convergence to fractional Brownian motion and to the Rosenblatt process," *Advances in Applied Probability*, 7(2), 249.

Wang, Y. (2014): "An Invariance Principle for Fractional Brownian Sheets," *Journal of Theoretical Probability*, 27(4), 1124–1139.