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: February 2021

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 covariance 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 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 correct coverage in a restricted but nonparametric class of strong spatial correlations, as well as on the efficiency of the method. In a design calibrated to match economic activity in U.S. states the method outperforms previous suggestions for spatially robust inference about the population mean.

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

JEL: C12, C20

*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 large-sample 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 shows three one-dimensional ($d = 1$) spatial designs. Panel (a) shows the familiar case of regularly spaced locations, corresponding to the standard time series setting; panels (b) and (c) show randomly selected locations drawn from a density g, where g is uniform in panel (b) and triangular in panel (c). Figure 2 shows two geographic examples, so $d = 2$, for the U.S. state of Texas. In panel (a), the locations are randomly selected from a uniform distribution, while in panel (b) they are more likely to be sampled from areas with high economic activity, here measured by light intensity as seen from space.[2] In much of our analysis, we will assume that locations are i.i.d. draws from a distribution with density g, and so will encompass the irregularly spaced time series and Texas examples.

Adding some notation, suppose

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

[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).

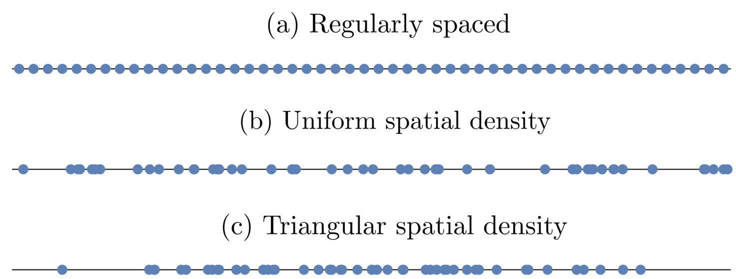

Figure 1: Three One-Dimensional Spatial Designs

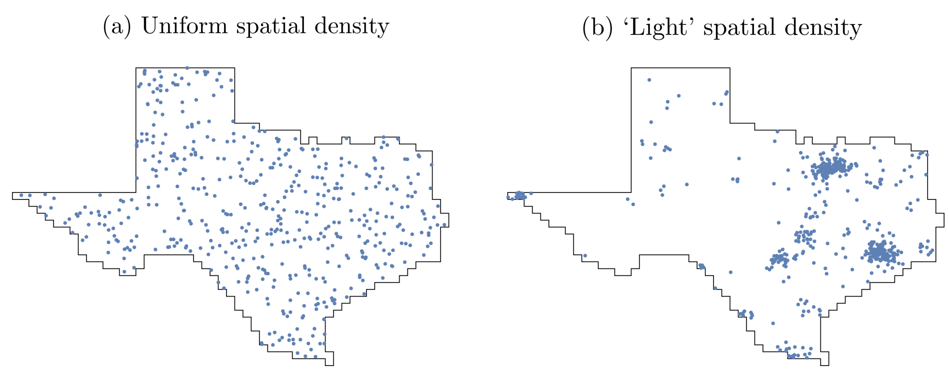

Figure 2: Two Geographic Spatial Designs

where y_l is associated with the spatial location s_l, μ is the mean of y_l, and u_l is an unobserved error, assumed to be covariance stationary with mean zero and covariance function $\mathbb{E}[u(r)u(s)] = \sigma_u(r-s)$. Let $\overline{y}$ denote the sample mean, and consider the usual t-statistic

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

where $\hat{\sigma}^2$ is an estimator for the variance of $\sqrt{n}(\overline{y}-\mu)$. 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 $\overline{y} \pm \text{cv}\,\hat{\sigma}/\sqrt{n}$. 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 (panel (a) of Figure 1) is the most well-studied version of this problem. Here $\text{Var}(\sqrt{n}(\overline{y}-\mu))$ is the long-run variance of y. Classic choices for $\hat{\sigma}^2$ are kernel-based consistent estimators such as those proposed in Newey and West (1987) and Andrews (1991), and associated 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, 2007), 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.

In the general spatial case, the variance of $\overline{y}$ depends on the correlation between all of the observations, and this in turn depends on two distinct features of the problem. The first is the correlation between observations at arbitrary locations (say r and s); this is given by the covariance function $\sigma_u(r-s)$. The second feature is which locations in $\mathcal{S}$ are likely to be sampled; this is given by the spatial density g. Only the first of these features is important in the regularly-spaced time series example because the locations do not vary from one application to the next.

Most existing suggestions for spatial inference are derived under the assumption that the locations are (asymptotically) uniformly distributed, corresponding to a constant density g: This includes the consistent kernel-based estimator in Conley (1999), the spatial analogue of the fixed-b kernel approach analyzed in Bester, Conley, Hansen, and Vogelsang (2016), as well as the spatial projection-based estimator put forward in Sun and Kim (2012). Exceptions include Kelejian and Prucha (2007), who derive a consistent kernel for $\hat{\sigma}^2$ under assumptions that can accommodate arbitrary locations s_l, and the cluster approach suggested by Ibragimov and Müller (2010, 2015) and Bester, Conley, and Hansen (2011) (also see Cao, Hansen, Kozbur, and Villacorta (2020)).

This paper makes progress over this literature by developing a method that (i) accounts for sampling uncertainty in $\hat{\sigma}^2$ in a spatial context while allowing for nonuniform spatial densities g; (ii) is valid under generic weakly correlated u_l; (iii) also controls size under a restricted but nonparametric form of strongly correlated u_l. The last 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.

Our method works as follows: First, a benchmark parametric model is specified for the covariance function, say $\sigma_u^0(\cdot) = \sigma_u^0(\cdot|c)$, where c is a persistence parameter with larger values indicating less dependence. For a given lower bound on the persistence parameter, say c_0, a hypothetical covariance matrix for $(y_1, ..., y_n)'$ is constructed using $\sigma_u^0(\cdot|c_0)$ evaluated at the actual sample locations $(s_1, ..., s_n)$. The eigenvectors of the demeaned version of this covariance matrix are the (population) principal components of the residuals $\hat{u}_l = y_l - \overline{y}$ under $\sigma_u^0(\cdot|c_0)$, and the sample variance of q of these principal components is the estimator $\hat{\sigma}^2$. The critical value is chosen to ensure coverage for all $c \geq c_0$. The number of principal components q is chosen to minimize the expect length of the confidence interval in the model where u_l is i.i.d. For shorthand, we refer to the method as *spatial correlation principal components*, abbreviated SCPC.

Intuitively, variance estimators $\hat{\sigma}^2$ that are quadratic forms in $\hat{u}$ are sums of squares of weighted averages of $\hat{u}$. Under spatial correlation, most weighted averages are less variable than $\overline{y}$, leading to a downward biased $\hat{\sigma}^2$. SCPC selects the linear combinations of $\hat{u}$ that are most variable, so that the bias is as small as possible in the benchmark model with parameter c_0.

The remainder of the paper studies this method. Section 2 provides the specifics for SCPC. These specifics raise a variety of issues that are the focus of the remaining sections of the paper. In particular, Section 3 lays out the analytic framework used to study the large-sample and finite-sample Gaussian properties of spatial t-statistics. We use the framework to analyze SCPC, but several of the results in Section 3 encompass other methods, notably "fixed-b" kernel-based methods, and general projection estimators with a fixed number of basis functions. We find that in contrast to the regularly spaced time series case, such t-statistics with analogously adjusted critical values are *not* generically valid under weak correlation as soon as the spatial density function is not uniform. We develop an alternative approach to the construction of critical values that restores validity, and this is used for SCPC inference. Section 4 thus shows that SCPC has the desired large-sample coverage probability under generic weak correlation. Moreover, Section 4 provides a set of (easily verifiable) sufficient conditions that guarantee coverage under arbitrary mixtures of a set of strong correlation patterns in a finite-sample Gaussian setting. Section 4 also investigates the finite-sample coverage probability of SCPC confidence sets when there is heteroskedasticity across locations or measurement errors in locations — two problems faced in some applications. Section 5 addresses the question of efficiency of SCPC by computing a lower bound on the expected length of confidence intervals for *any* inference method that controls coverage in a particular class of spatial correlations. Comparing the expected length of SCPC to this lower bound provides a measure of the efficiency of the method. Section 6 compares the properties of SCPC to other methods that have been proposed in the literature, and the results suggest that SCPC dominates these methods over the range of covariance functions and spatial designs considered. Section 7 discusses extensions and implementation issues. First, it discusses how the results developed in the body of the paper for inference about the population mean can be applied to inference problems about regression coefficients or parameters in GMM models. It then discusses two important computational issues involved in computing the critical value and computing the required eigenvectors for the construction of SCPC in very large-n applications. Finally, Section 7 provides a sketch of the generalization of the SCPC method to multivariate (F-test) settings. Proofs are collected in the appendix.

2 Spatial Correlation Principal Components

This section provides details for computing the SCPC t-statistic, critical value and associated confidence interval. The construction of SCPC raises a variety of questions about its properties, many of which are posed here and discussed in detail in the remaining sections of the paper.

The construction of the SCPC t-test and confidence interval 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. Those assumptions will be listed in Section 3 and, it will turn out, are significantly more general than what would follow from the description in this section.

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 model for u_l with a parametric covariance function $\text{Cov}(u(r), u(s)) = \sigma_u^0(r - s|c)$, where smaller values of the scalar parameter c indicate stronger correlations. In the following, we focus on the simple Gaussian exponential ('AR(1)') model where $\sigma_u^0(r - s|c) = \exp(-c||r - s||)$ for $c > 0$. Let $\mathbf{\Sigma}(c)$ denote the $n \times n$ covariance matrix with $\mathbf{\Sigma}(c)_{ij} = \exp(-c||s_i - s_j||)$, so that $\mathbf{\Sigma}(c)$ is the covariance matrix of $u(s)$ evaluated at the sample locations $\mathbf{s}$. Let c_0 denote a pre-determined 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 $\mathbf{M}\mathbf{\Sigma}(c_0)\mathbf{M}$ 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 $\hat{\mathbf{u}}|\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{M}\mathbf{\Sigma}(c_0)\mathbf{M})$. The

SCPC estimator of σ^2 based on the first q of these principal components is

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

and the corresponding SCPC t-statistic is

$$\tau_{\mathrm{SCPC}}(q) = \frac{\sqrt{n}(\overline{y} - \mu_0)}{\hat{\sigma}_{\mathrm{SCPC}}(q)}. \tag{3}$$

The critical value $\mathrm{cv}_{\mathrm{SCPC}}(q)$ 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, $\mathrm{cv}_{\mathrm{SCPC}}(q)$ satisfies

$$\sup_{c \geq c_0} \mathbb{P}^0_{\mathbf{\Sigma}(c)}(|\tau_{\mathrm{SCPC}}(q)| > \mathrm{cv}_{\mathrm{SCPC}}(q)|\mathbf{s}) = \alpha, \tag{4}$$

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

The final ingredient in the method is the choice of q. Let $\mathbb{E}^1[2\hat{\sigma}_{\mathrm{SCPC}}(q)\,\mathrm{cv}_{\mathrm{SCPC}}(q)|\mathbf{s}]$ denote the expected length of the confidence interval constructed using $\tau_{\mathrm{SCPC}}(q)$ under the Gaussian i.i.d. model $\mathbf{y}|\mathbf{s} \sim \mathcal{N}(\mathbf{l}\mu, \mathbf{I})$. (The superscript "1" on $\mathbb{E}$ differentiates this from the benchmark model with covariance matrix $\mathbf{\Sigma}(c)$.) SCPC chooses q_{SCPC} to make this length as small as possible, that is q solves

$$\min_{q \geq 1} \mathbb{E}^1[2\hat{\sigma}_{\mathrm{SCPC}}(q)\,\mathrm{cv}_{\mathrm{SCPC}}(q)|\mathbf{s}] = \min_{q \geq 1} \sqrt{8} n^{-1/2} q^{-1/2}\,\mathrm{cv}_{\mathrm{SCPC}}(q) \frac{\Gamma((q+1)/2)}{\Gamma(q/2)} \tag{5}$$

with the equality exploiting that $q\hat{\sigma}^2_{\mathrm{SCPC}}(q)|\mathbf{s} \sim \chi^2_q$ 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 $\mathrm{tr}\,\mathbf{H}'\mathbf{M}\mathbf{\Sigma}(c_0)\mathbf{M}\mathbf{H}$ 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 induce the smallest bias in the benchmark model with $c = c_0$, under the constraint of being unbiased in the i.i.d. model.

The choice of q trades off the downward bias in $\hat{\sigma}^2_{\mathrm{SCPC}}(q)$ 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 (5) 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.

Remark 2.2. By construction, SCPC confidence intervals have correct coverage in Gaussian models with a spatial exponential covariance function ('AR(1)' models) with spatial persistence level less than or equal to the model with $c = c_0$. Lemma 1 in Section 3 provides a central limit result that rationalizes the normality assumption. Theorem 7 provides conditions on the choice of c_0 so that the SCPC t-test controls size in large samples not just in the exponential model, but under generic 'weak correlation', as defined in Section 3. Theorem 8 provides easily verifiable sufficient conditions for size control under mixtures of parametric small sample Gaussian models.

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

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

that is, we set c_0 so that it implies a given value $\overline{\rho}_0$ of $\overline{\rho}$. For example, $\overline{\rho}_0 = 0.001$ implies very weak correlation, $\overline{\rho}_0 = 0.02$ stronger correlation, and $\overline{\rho}_0 = 0.10$ very strong correlation. In our examples, we calibrate c_0 to these three values of $\overline{\rho}$.

Remark 2.4. The SCPC method with c_0 calibrated in this way 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) suggestion) also to arbitrary distance preserving transformations, such as rotations.

Remark 2.5. The weights $\mathbf{r}_j$ used to construct the principal components and $\hat{\sigma}^2_{\mathrm{SCPC}}(q)$ depend on $\mathbf{s}$, the sample values of the spatial locations. Because the spatial locations are randomly drawn, the $\mathbf{r}_j$ weights are random. But as shown in Section 3, the weights have well-defined limits in terms of appropriately defined nonrandom eigenfunctions. Figure 3 plots selected eigenfunctions for two one-dimensional spatial designs and Figure 4 shows the associated plots for the Texas example, where in both cases $\overline{\rho}_0 = 0.02$. With uniform spatial densities (panel (a) in both figures), the eigenfunctions are much like the weighting functions used for low-frequency projection methods for regularly spaced time series (e.g., Müller (2004), Phillips (2005), Sun (2013)) or its spatial analogue (e.g., Sun and Kim (2012)). In contrast, the non-uniform densities (panel (b) in the figures) produce weights that are distorted versions of their uniform counterparts, with most of the variation concentrated in high-density areas.

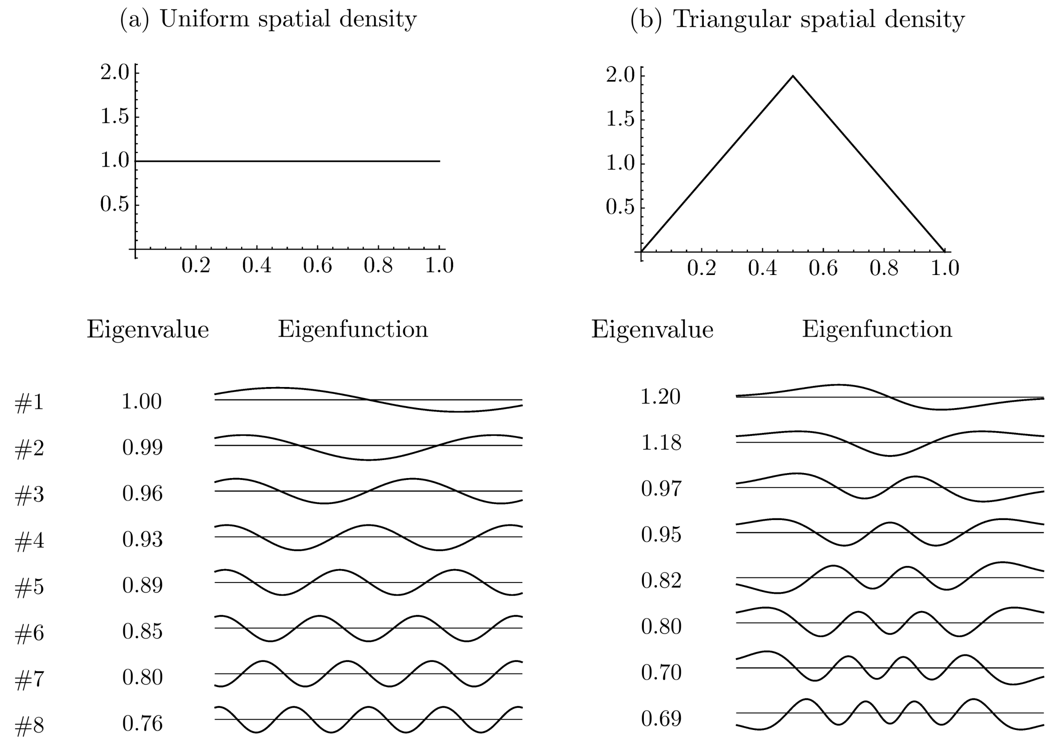

Figure 3: Eigenfunctions for Two One-Dimensional Spatial Designs

The figures also show the associated normalized eigenvalues, that is the variance of the principal components under the assumed exponential model, relative to the variance of $\sqrt{n}(\bar{y}-\mu)$. When the density is uniform, these relative variances are slightly below 1.0 for small j, and decline monotonically with j. This leads to the familiar downward bias of $\hat{\sigma}^2$ in projection methods. When the spatial density is not uniform, the relative variance of the principal components can be larger than unity, mitigating this downward bias.

Remark 2.6. In the regular spaced time series case, the eigenvectors of SCPC for $\bar{\rho}_0 \in \{0.02, 0.10\}$ are numerically close to the type-II cosine transforms considered in Müller (2004, 2007), Lazarus, Lewis, Stock, and Watson (2018) and Dou (2019). 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 with locations that happen to fall on a line with approximately uniform empirical distribution.

U.S. states spatial correlation designs. Before making two additional remarks about the

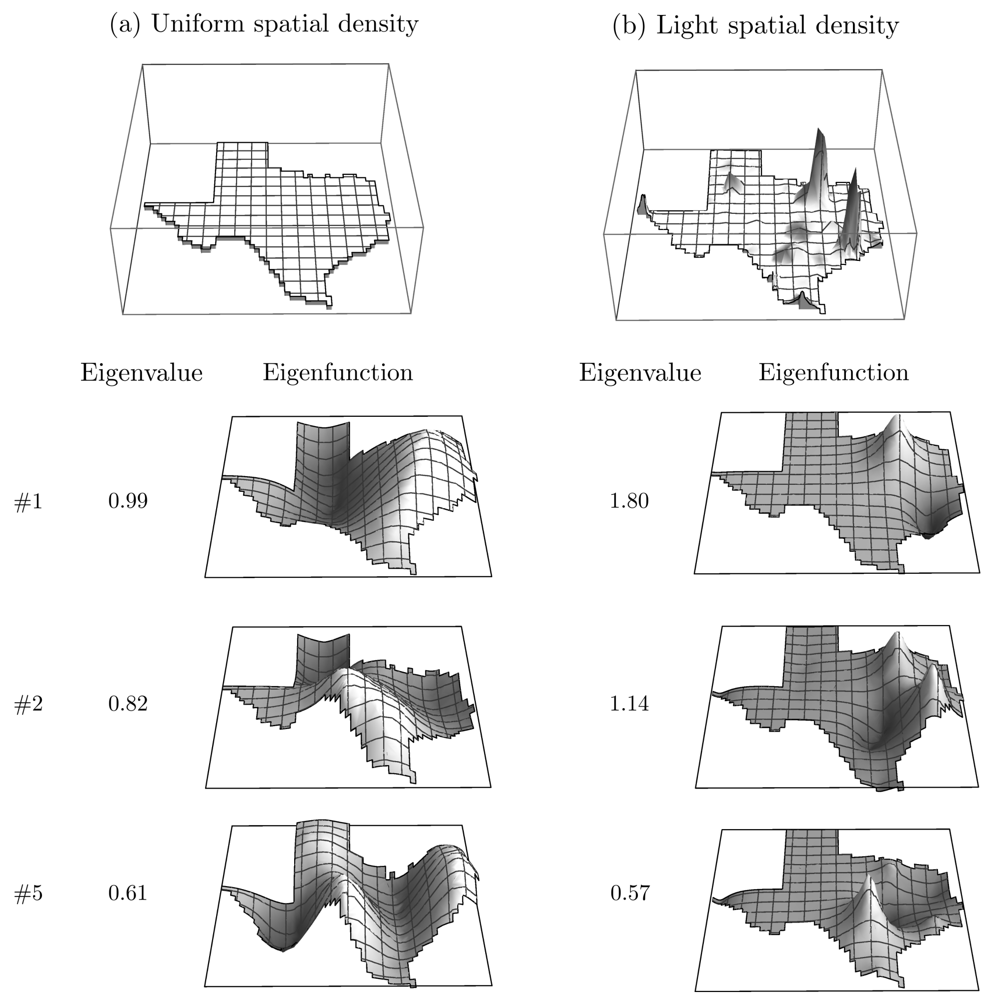

Figure 4: Eigenfunctions for Two Geographic Spatial Designs

SCPC method, we introduce a set of spatial correlation designs that will be used throughout the paper. The idea is to consider a set of real world designs to learn about the usefulness of the SCPC and other methods in practice. In particular, we randomly draw $n = 500$ locations within the boundaries of the 48 contiguous states of the U.S. (we also considered $n = 1000$ draws, and found nearly identical results in all exercises). 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. We draw five sets of 500 independent locations under each density $g \in \{g_{\text{uniform}}, g_{\text{light}}\}$ and $\bar{\rho}_0 \in \{0.02, 0.10\}$ for each state, for a total of 240 (= 48 states $\times$ 5 location draws) sets of locations $\{s_l\}_{l=1}^{500}$ and associated covariances under each of the four $(g, \bar{\rho}_0)$ pairs.

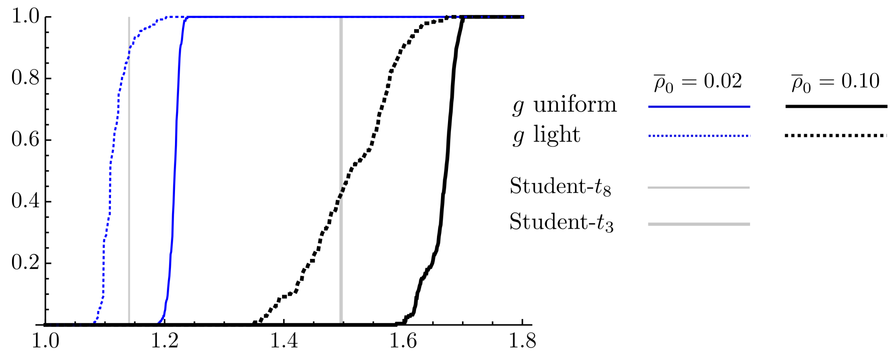

Figure 5: CDFs of Expected Length of SCPC Confidence Interval Relative to Known Variance Interval

Remark 2.7. The critical value of the SCPC t-statistic reflects randomness in both $\bar{y}$ and $\hat{\sigma}^2_{\mathrm{SCPC}}$. This is analogous to inference in small-sample Gaussian models using critical values from the Student-t distribution. Figure 5 shows the effect of the uncertainty in σ^2 on the expected length of 95% confidence intervals in the U.S. states spatial correlation designs, by comparing the expected length of the SCPC confidence interval in the i.i.d. model to the the length with σ^2 known: this relative length is $\mathbb{E}^1[(\mathrm{cv}\,/1.96)(\hat{\sigma}_{\mathrm{SCPC}}/\sigma)|\mathbf{s}]$, where 1.96 is the standard normal critical value. The figure plots the CDF of these relative lengths over the 240 draws under each $(g, \bar{\rho}_0)$ pair. For example, the left-most CDF (dashed blue, for $g = g_{\mathrm{light}}$ and $\bar{\rho}_0 = 0.02$) shows that the relative expected length ranges from roughly 1.08 to 1.18 across the 240 draws. The figure indicates that the expected lengths are higher under g_{uniform} than under the g_{light} design and are higher under $\bar{\rho}_0 = 0.10$ than $\bar{\rho}_0 = 0.02$. For comparison the figure also shows the relative expected lengths of Student-t confidence intervals with 8 and 3 degrees of freedom, in multiples of the length of the known variance z-interval. Evidently, when $\bar{\rho}_0 = 0.02$, the increase in expected length of the SCPC confidence interval relative to an oracle endowed with the value of σ^2 is roughly like learning about the value of σ^2 from 8 i.i.d.$\mathcal{N}(0, \sigma^2)$ observations. When $\bar{\rho}_0 = 0.10$, relative lengths increase to approximately what would obtain from Student-t_3 inference.

Remark 2.8. Consider the related question about the efficiency of SCPC relative to other methods that do not assume that the value of σ^2 is known. This question can be answered in two ways. The first is to compare SCPC to methods that have previously been proposed. This is done in Section 6. A more ambitious approach compares SCPC to the most efficient method constructed for any particular spatial density that, like SCPC, produces confidence intervals with the desired coverage over a wide range of covariance functions. This is done in Section 5 which computes a lower bound on the expected length of confidence intervals for any such method.

3 Large-sample analysis of spatial t-statistics

This section outlines a large-sample framework used to study SCPC and other spatial t-statistics. The first two subsections introduce notation and the asymptotic sampling framework. With these in hand, the remainder of the section summarizes the large-sample distribution of various statistics including the SCPC and kernel-based t-statistics. Proofs are provided in the appendix.

3.1 Notation

Some of this notation has been introduced earlier, but is repeated here for easy reference.

The sample mean is denoted by $\overline{y}_n$, where here and elsewhere we append the subscript n for clarity in the asymptotic analysis. The residual is $\hat{u}_l = y_l - \overline{y}_n$. Let $\mathbf{y}_n = (y_1, ..., y_n)'$, and similarly for $\mathbf{u}_n$, $\hat{\mathbf{u}}_n$ and $\mathbf{s}_n$. The vector $\mathbf{l}_n$ is a $n \times 1$ vector of 1s, and $\mathbf{M}_n = \mathbf{I}_n - \mathbf{l}_n(\mathbf{l}_n'\mathbf{l}_n)^{-1}\mathbf{l}_n$, so that $\hat{\mathbf{u}}_n = \mathbf{M}_n\mathbf{u}_n$.

Generically, we consider estimators $\hat{\sigma}_n^2$ that are quadratic forms in $\hat{\mathbf{u}}_n$. Let $\mathbf{Q}_n$ be a positive semidefinite matrix with $\mathbf{Q}_n\mathbf{l}_n = \mathbf{0}$. We consider estimators of the form

$$\hat{\sigma}_n^2(\mathbf{Q}_n) = n^{-1}\hat{\mathbf{u}}_n'\mathbf{Q}_n\hat{\mathbf{u}}_n = n^{-1}\mathbf{u}_n'\mathbf{Q}_n\mathbf{u}_n \tag{6}$$

where the final equality follows from $\mathbf{Q}_n\mathbf{l}_n = \mathbf{0}$.

Two leading examples of estimators in this class are kernel-based estimators and orthogonal-projections estimators. For kernel-based estimators, let $k(r, s)$ denote a positive semi-definite kernel, $k : \mathcal{S} \times \mathcal{S} \mapsto \mathbb{R}$. Let $\mathbf{K}_n$ denote an $n \times n$ matrix with (l, ℓ) element equal to $k(s_l, s_\ell)$ and let $\mathbf{Q}_n = \mathbf{M}_n\mathbf{K}_n\mathbf{M}_n$. Then $\hat{\sigma}_n^2 = n^{-1}\sum_l\sum_\ell k(s_l, s_\ell)\hat{u}_l\hat{u}_\ell = n^{-1}\hat{\mathbf{u}}_n'\mathbf{Q}_n\hat{\mathbf{u}}_n$. For orthogonal-projection estimators, let $\hat{\mathbf{W}}_n$ be an $n \times q$ matrix with jth column given by $\hat{\mathbf{w}}_j$ satisfying $n^{-1}\hat{\mathbf{W}}_n'\hat{\mathbf{W}}_n = q^{-1}\mathbf{I}_q$ and $\hat{\mathbf{W}}_n'\mathbf{l}_n = \mathbf{0}$ (the 'hat' notation is a reminder that $\hat{\mathbf{W}}$ depends on the locations $\mathbf{s}_n$, which are random). With $\mathbf{Q}_n = \hat{\mathbf{W}}_n\hat{\mathbf{W}}_n'$, the orthogonal projection estimator is $\hat{\sigma}_n^2 = \sum_{j=1}^q(n^{-1/2}\hat{\mathbf{w}}_j'\hat{\mathbf{u}}_n)^2 = n^{-1}\hat{\mathbf{u}}_n'\mathbf{Q}_n\hat{\mathbf{u}}_n$. The SCPC estimator is an orthogonal-projection estimator using the first q eigenvectors of $\mathbf{M}_n\mathbf{\Sigma}(c_0)\mathbf{M}_n$, scaled to have length $1/\sqrt{q}$, as the columns of $\hat{\mathbf{W}}_n$.

For quadratic form estimators $\hat{\sigma}_n^2(\mathbf{Q}_n)$, under the null hypothesis the squared t-statistic is a ratio of quadratic forms in $\mathbf{u}_n$

$$\tau_n^2(\mathbf{Q}_n) = \frac{\left(\sqrt{n}(\overline{y}_n - \mu_0)\right)^2}{\hat{\sigma}_n^2(\mathbf{Q}_n)} = \frac{\mathbf{u}_n'\mathbf{l}_n\mathbf{l}_n'\mathbf{u}_n}{\mathbf{u}_n'\mathbf{Q}_n\mathbf{u}_n}. \tag{7}$$

3.2 Sampling and large-n framework

The spatial locations s 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, where $g(s)$ is continuous and positive for all $s \in \mathcal{S}$.

The average pairwise correlation of y, conditional on the sample locations is $\bar{\rho}_n = \frac{1}{n(n-1)} \sum_{l=1}^{n} \sum_{\ell \neq l} \mathrm{Cor}\,(y_l, y_\ell \,|\mathbf{s}_n)$. When $\overline{\rho}_n = 0$, $\mathbf{y}_n \,|\mathbf{s}_n$ is white noise. When $\overline{\rho}_n = O_p(1)$ (and not $o_p(1)$), we will say the process exhibits *strong correlation*. When $\overline{\rho}_n = O_p(1/c_n^d)$ where c_n is a sequence of constants with $c_n \to \infty$, we follow Lahiri (2003) and say the process exhibits *weak correlation*.

The following asymptotic framework, adapted from Lahiri (2003), is useful for representing 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 $\{s_l\}_{l=1}^n$ are independent. To avoid pathological cases, we further assume $\int \sigma_B(s)ds > 0$ and that 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 $\sigma_u(s) = \sigma_B(c_n s)$. 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 $c_n^d \operatorname{vol}(\mathcal{S})$, where $\operatorname{vol}(\mathcal{S})$ is the volume of $\mathcal{S}$. The average number of sample points per unit of volume is then $n/(c_n^d \operatorname{vol}(\mathcal{S}))$. 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. A calculation shows that $\overline{\rho}_n = O_p(1/c_n^d)$, so the sequence c_n characterizes weak and strong correlation as described above. With this background, let $a_n = c_n^d/n$; we will assume that $a_n \to a \in [0, \infty)$.

Finally, we specify a set of weighting functions. To simplify the problem, we initially consider weights that are nonrandom. For $j = 1, \ldots, q$, let $w_j : \mathcal{S} \mapsto \mathbb{R}$ denote a set of continuous functions that satisfy $\int w_j(s)dG(s) = 0$ and $\int w_j^2(s)dG(s) > 0$. 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 3.1. 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, governed by c_n and B.

With this background, we now present the large-sample analysis.

3.3 Large-sample behavior of weighted averages

As is evident from equation (7) the squared t-statistic is a ratio of squares of weighted average of the elements of $\mathbf{u}_n$. This subsection discusses the large-sample distribution of such weighted averages. These results involve weak converge (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 slightly weaker than almost sure weak convergence of conditional distributions, but still ensures that the limiting distribution is not induced by the randomness in the locations $\mathbf{s}_n$.

Lemma 1 characterizes the large-sample behavior of sums of the form $\sum_{l=1}^n \mathbf{w}^0(s_l)u(s_l)$. For the weak correlation result, we invoke the mixing and moment assumptions of Lahiri (2003) on B that underlie his Theorem 3.2.

Lemma 1. *(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) Suppose $c_n \to \infty$, 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

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

Remark 3.2. Note that the variance of $\sum_{l=1}^n \mathbf{w}^0(s_l)u(s_l)$ conditional on $\mathbf{s}_n$ is

$$\begin{aligned}\operatorname{Var}\left[\sum_{l=1}^n \mathbf{w}^0(s_l)u(s_l)\,|\mathbf{s}_n\right] &= \sum_l \sum_\ell \mathbf{w}^0(s_l)\mathbf{w}^0(s_\ell)'\sigma_u(s_l - s_\ell) \\ &= \sum_l \sum_\ell \mathbf{w}^0(s_l)\mathbf{w}^0(s_\ell)'\sigma_B(c_n(s_l - s_\ell)). \end{aligned} \tag{8}$$

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, $a_n = c_n^n/n$ gets larger), variances become more important relative to covariances; this explains the first term in $\mathbf{\Omega}_{wc}$.

Remark 3.3. The form of $\mathbf{V}_2$ is further 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.

Remark 3.4. 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).

Remark 3.5. When $g(s)$ is constant, so the spatial distribution is uniform, $\mathbf{V}_1 \propto \mathbf{V}_2$ and $\mathbf{\Omega}_{wc} \propto \int \mathbf{w}^0(s)\mathbf{w}^0(s)'ds$. Thus, in a leading case with orthogonal w_j of length $1/\sqrt{q}$, $\int w_j(s)w_i(s)dG(s) = q^{-1}\mathbf{1}[i = j]$, $\mathbf{\Omega}_{wc} \propto \operatorname{diag}(1, q^{-1}\mathbf{I}_q)$, a familiar result from the literature on HAR inference in time series with regularly spaced observations. Importantly, while this result holds under constant $g(s)$, 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.

3.4 Large-sample null rejection probability of spatial t-tests

This section presents a useful representation for the limiting distribution of $\tau_n^2(\mathbf{W}_n\mathbf{W}_n')$ under the assumptions of Lemma 1.

Theorem 2. *For* $\mathrm{cv} > 0$*, let* $\mathbf{D}(\mathrm{cv}) = \mathrm{diag}(1, -\mathrm{cv}^2\,\mathbf{I}_q)$*,* $\mathbf{A} = \mathbf{D}(\mathrm{cv})\boldsymbol{\Omega}$ *with* $\boldsymbol{\Omega} \in \{\boldsymbol{\Omega}_{sc}, \boldsymbol{\Omega}_{wc}\}$*, and let* $(\omega_0, \omega_1, ..., \omega_q)$ *denote the eigenvalues of* $\mathbf{A}$ *ordered from largest to smallest. Then under the assumptions of Lemma 1, under the null hypothesis and 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* $i \geq 1$*;*

(ii) $\mathbb{P}\left(\tau_n^2(\mathbf{W}_n\mathbf{W}_n') > \mathrm{cv}^2\,|\mathbf{s}_n\right) \xrightarrow{p} \mathbb{P}\left(Z_0^2 > \sum_{i=1}^q (-\frac{\omega_i}{\omega_0}) Z_i^2\right)$.

Remark 3.6. In the weak-correlation case with constant spatial density $g(s)$ and orthogonal w_j of length $1/\sqrt{q}$, $\boldsymbol{\Omega} = \boldsymbol{\Omega}_{wc} \propto \mathrm{diag}(1, q^{-1}\mathbf{I}_q)$. Thus $-\omega_i/\omega_0 = \mathrm{cv}^2/q$, and 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.

Remark 3.7. In the general weak correlation case with arbitrary spatial density g, $\boldsymbol{\Omega}_{wc} = a\sigma_B(0)\mathbf{V}_1 + \left(\int \sigma_B(s)ds\right)\mathbf{V}_2$. 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

$$\boldsymbol{\Omega}_{wc} = \kappa\mathbf{V}_1 + (1-\kappa)\mathbf{V}_2 \tag{9}$$

where κ is scalar with $0 \leq \kappa < 1$. Thus, the limiting CDF 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 $c_n \to \infty$.

3.4.1 Extensions for estimated weights

For SCPC and other estimators, the weights in $\mathbf{w}(s)$ are estimated using the sample locations $\mathbf{s}_n$. The conditions under which Lemma 1 continues to hold for such estimated weights is given in the following theorem.

Theorem 3. *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*

$$\sup_{s \in \mathcal{S}} ||\hat{\mathbf{w}}^0(s) - \mathbf{w}^0(s)|| \xrightarrow{p} 0. \tag{10}$$

Then Lemma 1 and Theorem 2 continue to hold with $\hat{\mathbf{W}}_n^0$ in place of $\mathbf{W}_n^0$, where the lth row of $\hat{\mathbf{W}}_n^0$ is equal to $(1, \hat{\mathbf{w}}(s_l)')$.

Remark 3.8. The theorem also accommodates location dependent convergent critical values $\mathrm{cv}_n \xrightarrow{p} \mathrm{cv}$ by setting $\hat{\mathbf{w}}^0(s) = (\mathrm{cv}_n \,/\, \mathrm{cv})\mathbf{w}^0(s)$.

3.4.2 Extension for 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

$$\hat{k}_n(r, s) = k(r, s) - n^{-1}\sum_{l=1}^{n} k(s_l, s) - n^{-1}\sum_{l=1}^{n} k(r, s_l) - n^{-2}\sum_{l=1}^{n}\sum_{\ell=1}^{n} k(s_l, s_\ell). \tag{11}$$

To begin, consider a simpler problem using a kernel that replaces the sample means in (11) with populations means

$$\overline{k}(r, s) = k(r, s) - \int k(u, s)dG(u) - \int k(r, u)dG(u) + \int\int k(u, t)dG(u)dG(t). \tag{12}$$

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

$$\overline{k}(s, r) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(s)\varphi_i(r) \tag{13}$$

where $\{\lambda_i, \varphi_i\}$ are the eigenvalues and eigenfunctions of $\overline{k}$, with eigenvalues ordered from largest to smallest, $\int \varphi_i(s)dG(s) = 0$ and $\int \varphi_i(s)\varphi_j(s)dG(s) = \mathbf{1}[i = j]$.

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

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

We can directly apply Theorem 2 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 $\overline{k}_q(s_l, s_\ell)$. 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 $\mathbb{P}\left(\tau_n^2(\bar{\mathbf{K}}_{n,q}) > \text{cv}^2\,|\mathbf{s}_n\right) \xrightarrow{p} \mathbb{P}\left(Z_0^2 > \sum_{i=1}^{q}(-\frac{\omega_i}{\omega_0})Z_i^2\right)$ by Theorem 2.

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 (9). A tedious but straightforward calculation (see (27) in the appendix) shows that the eigenvalues ω_i of $\mathbf{A} = \mathbf{D}(\text{cv})\mathbf{\Omega}$ 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

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

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

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

The following theorem shows this to be the case, and it also includes the generalization to sample demeaned kernels (11) instead of (12).

Theorem 4. *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 1, $\omega_0 > 0$ and $\omega_i \leq 0$ for $i \geq 1$, and $\mathbb{P}(\tau_n^2(\hat{\mathbf{K}}_n) > \text{cv}^2\,|\mathbf{s}_n) \xrightarrow{p} \mathbb{P}(Z_0^2 > \sum_{i=1}^{\infty}(-\omega_i/\omega_0)Z_i^2)$.*

Remark 3.9. 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 (that is, setting $\kappa = 1$) does not, in general, lead to valid inference under weak correlation.

Remark 3.10. The theorem is also applicable to projection estimators using basis functions ϕ_i that are orthogonalized using the sample locations (such as those suggested in Sun and Kim (2012)) by setting $k(r,s) = q^{-1}\sum_{i=1}^{q}\phi_i(r)\phi_i(s)$.

Remark 3.11. The framework of Theorem 4 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 (9) 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 $\int \phi_i(s)^2dG(s) = 1$ 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 4 involves showing that in large samples, the difference between the eigenfunctions of the sample demeaned kernel (11) and the population demeaned kernel (12) becomes small. The following lemma extends and adapts previous results by Rosasco, Belkin, and Vito (2010) to the case of sample demeaned kernels.

Lemma 5. *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*

$$\hat{\varphi}_i(\cdot) = n^{-1}\hat{\lambda}_i^{-1}\sum_{l=1}^{n}\hat{v}_{i,l}\hat{k}_n(\cdot, s_l). \tag{14}$$

Let $\lambda_{(j)}$, $j = 1, \ldots$ be the unique positive values of λ_i, ordered descendingly, 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 $\hat{\boldsymbol{\varphi}} = (\hat{\varphi}_1, \ldots, \hat{\varphi}_q)'$,*

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

(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.

3.4.3 SCPC t-statistic

Beyond its use in the proof of Theorem 4, Lemma 5 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 5, we are interested in the eigenfunctions of the covariance kernel $k^0(r, s) = \sigma_u^0(r - s|c_0)$ 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 $k^0(r, s) = \sigma_u^0(r - s|c_0)$ and $\bar{k}^0$ is the demeaned version of k^0 in analogy to (12). Lemma 5 and a slightly extended version of Theorem 3 (see Lemma 10 in the appendix) then yields the following corollary.

Corollary 6. *Suppose $\lambda_q^0 > \lambda_{q+1}^0$. Then Theorem 2 holds for $\tau_{SCPC}^2(q) = \tau_n^2(q^{-1}\sum_{i=1}^{q}\mathbf{r}_i\mathbf{r}_i')$ with $\mathbf{w}(s) = (\varphi_1^0(s), \ldots, \varphi_q^0(s))'/\sqrt{q}$.*

4 Size control of spatial t-statistics

This section presents two results on size control of spatial t-statistics, the first asymptotic and the second a finite-sample result, and applies these to SCPC.

4.1 Asymptotic size control under weak correlation

As discussed above (see equation (9)), under weak correlation, the asymptotic rejection probability of τ_n for finite q can be studied via $\boldsymbol{\Omega}_{wc}(\kappa) = \kappa\mathbf{V}_1 + (1 - \kappa)\mathbf{V}_2$, where the covariance function of u 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 1 (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 $L(f)(s) = \int \sqrt{\kappa + (1-\kappa)g(s)}\left(1 - \overline{\text{cv}}^2 \overline{k}(s, r)\right)\sqrt{\kappa + (1-\kappa)g(r)} f(r) dG(r)$.

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 an other choices. We now discuss a simpler approach.

Consider a benchmark model B^0 that satisfies the assumptions of Theorem 1 (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

$$\sup_{c \geq c_{n,0}} \mathbb{P}^0_{\mathbf{\Sigma}(c)}(\tau_n^2 \geq \text{cv}_n^2 \,|\mathbf{s}_n) \leq \alpha \tag{15}$$

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 7. *Let cv_n^2 satisfy (15). Under weak correlation in the sense of Lemma 1 (ii), for t-statistics covered by Theorems 2, 3, 4 and Corollary 6,* $\max(\overline{\text{cv}}^2 - \text{cv}_n^2, 0) \xrightarrow{p} 0$. *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 7 is as follows. The critical value cv_n in (15) 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; thus different sequences c_n in the benchmark model recreate any possible limit distribution 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.

4.1.1 Implications for SCPC

For SCPC, the benchmark covariance kernel for B^0 is exponential $\sigma_B^0(r, s) = \exp(-||r - s||)$ and (from equation (4)) the critical value is chosen to satisfy (15) with equality. Thus, with a fixed value of c_0, the SCPC t-test $\tau_{\mathrm{SCPC}}(q)$ controls size in large samples under generic weak correlation.[3]

In addition and by construction, the SCPC critical value is chosen to satisfy the size constraint for all values of $c \geq c_0$ in the benchmark model. Thus, size is controlled by construction also in strong-correlation models with exponential covariance kernels for all $c \geq c_0$.

4.2 Finite sample size control in the Gaussian model

The asymptotic results of the last subsection are comforting, but in finite samples, the robustness of a spatial t-statistic with critical value chosen according to (15) still depends on the choice of $c_{n,0}$ and the benchmark model. This motivates investigating size control in finite samples, which potentially includes 'strong' correlation cases.

We restrict attention to Gaussian models where $\mathbf{y} \sim \mathcal{N}(\mathbf{l}\mu, \mathbf{\Sigma})$ for some $\mathbf{\Sigma}$ and implicitly condition on $\mathbf{s}$, and we also omit the dependence on n to ease notation. In this finite sample conditional framework, the distinction between $\mathbf{W}$ and $\hat{\mathbf{W}}$ is immaterial, so for simplicity, we write $\tau^2(\mathbf{WW}')$ for the t-statistic.[4]

Let $\mathcal{V}$ denote a set of covariance matrices. A test using the t-statistic $\tau^2(\mathbf{WW}')$ with critical value cv is robust for $\mathcal{V}$ if $\sup_{\mathbf{\Sigma}\in\mathcal{V}} \mathbb{P}_{\mathbf{\Sigma}}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2) \leq \alpha$. For a finite or parametric set of $\mathcal{V}$, $\sup_{\mathbf{\Sigma}\in\mathcal{V}} \mathbb{P}_{\mathbf{\Sigma}}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2)$ can be established numerically. We therefore focus on an analytical robustness result for a non-parametric class $\mathcal{V}$.

Specifically, we establish a set of readily verifiable sufficient conditions to check robustness for sets $\mathcal{V}$ that are composed of mixtures of parametric covariance matrices $\mathbf{\Sigma}^p(\theta)$ for $\theta \in \Theta$. We then apply this result to a set of Matérn covariance matrices with parameter θ and investigate the robustness of SCPC over arbitrary mixtures of these Matérn models. In addition, we use the result to study the robustness of a popular projection based t-test in a regularly spaced time series setting.

Consider a benchmark model with $\mathbf{\Sigma} = \mathbf{\Sigma}_0$, and suppose that cv has been chosen so that

[3] Technically, the SCPC choice of q in (5) is also a function of the locations of $\mathbf{s}_n$, so q_{SCPC} is random. However, the argument that establishes Theorem 7 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 11 in the appendix for a formal statement.

[4] This also covers kernel variance estimators by setting $q = T - 1$ and using the Choleksy decomposition $\mathbf{MKM} = \mathbf{WW}'$.

$\mathbb{P}_{\mathbf{\Sigma}_0}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2) = \alpha$. We are interested in conditions under which

$$\mathbb{P}_{\mathbf{\Sigma}_1}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2) \leq \alpha \text{ for } \mathbf{\Sigma}_1 = \int_{\Theta} \mathbf{\Sigma}^p(\theta) dF(\theta) \tag{16}$$

for a probability distribution F.

Let $\lambda_j(\cdot)$ denote the jth largest eigenvalue of some matrix.

Theorem 8. *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 $\mathbf{A}_0 = \mathbf{D}(\mathrm{cv})\mathbf{\Omega}_0$ is diagonalizable, and let $\mathbf{P}$ be its eigenvectors. Let $\mathbf{A}(\theta) = \mathbf{P}^{-1}\mathbf{D}(\mathrm{cv})\mathbf{\Omega}(\theta)\mathbf{P}$ 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*

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

If for some probability distribution F on Θ, $\sum_{i=1}^{j} \int \nu_i(\theta) dF(\theta) \geq 0$ for all $1 \leq j \leq q$, then (16) holds.

Remark 4.1. If $\sum_{i=1}^{j} \nu_i(\theta) \geq 0$ for all $\theta \in \Theta$ and $1 \leq j \leq q$, then the theorem implies that $\mathbb{P}_{\mathbf{\Sigma}_1}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2) \leq \alpha$ for $\mathbf{\Sigma}_1$ an arbitrary mixture of $\mathbf{\Sigma}^p(\theta)$.

Remark 4.2. Note that for $\mathbf{\Sigma}^p(\theta_0) = \mathbf{\Sigma}_0$, $\nu_i(\theta_0) = 0$ 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 $\mathbf{\Sigma}_0$.

Remark 4.3. As shown in Theorem 2, the eigenvalues of $\mathbf{A}_0$ and $\mathbf{A}(\theta)$ (or, equivalently, of $\mathbf{D}(\mathrm{cv})\mathbf{\Omega}(\theta)$) 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 $-\lambda_j(\mathbf{A}(\theta)) \geq -\lambda_j(\mathbf{A}_0)$ for all $j \geq 2$, then the result there implies that $\mathbb{P}_{\mathbf{\Sigma}^p(\theta)}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2) \leq \mathbb{P}_{\mathbf{\Sigma}_0}(\tau^2(\mathbf{WW}') > \mathrm{cv}^2)$. It follows from an integral representation (cf. equation (20) below) 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 from Marshall, Olkin, and Arnold (2011) and additional calculations then extend this further to the result in Theorem 8.

Remark 4.4. The conditions of Theorem 8 implicitly depend on the locations $\mathbf{s}$, so the implications are specific to the application. In the spatial case, the practical importance of the theorem is that the conditions are straightforward to check numerically for a given parametric family $\mathbf{\Sigma}^p(\theta)$. This can establish a range of robustness of a spatial t-test in a given application and is illustrated in the next subsection with the SCPC t-test and the Matérn class of spatial correlations. The theorem also provides insights for inference in the regularly-spaced time series case, where the spatial design is fixed across applications. This is illustrated in the subsequent subsection for a projection-based t-statistic for mixtures of AR(1) processes and processes that are 'less persistent' than a benchmark AR(1) model.

4.2.1 Implications for SCPC

The critical value for the SCPC t-test is chosen to control size in exponential models with $c \geq c_0$, where c_0 is calibrated to a value $\overline{\rho}_0$. Because $\overline{\rho}$ is monotone in c, the resulting SCPC t-test controls size for all $\overline{\rho} \leq \overline{\rho}_0$ in the exponential model by construction.

Let $\mathbf{\Sigma}^p(\theta)$ denote the covariance matrix associated with a parameter θ, with average pairwise correlation $\overline{\rho}(\theta)$. Let $\Theta_{\overline{\rho}_L,\overline{\rho}_U} = \{\theta | \overline{\rho}_L \leq \overline{\rho}(\theta) \leq \overline{\rho}_U\}$ denote the set of values of θ that induce correlations between $\overline{\rho}_L$ and $\overline{\rho}_U$. If the inequalities in Theorem 8 are satisfied for all values of $\theta \in \Theta_{\overline{\rho}_L,\overline{\rho}_U}$, then the SCPC t-test controls size for all mixtures of $\mathbf{\Sigma}^p(\theta)$ in this set.

In this section we consider $\mathbf{\Sigma}^p(\theta)$ computed from Matérn processes with parameter $\theta = (\nu, c)$, where ν and c are positive constants. If u follows a Matérn process, its covariance function $\sigma_u(r - s)$ depends on the locations only through $d = ||r - s||$. For $\nu \in \{1/2, 3/2, 5/2, \infty\}$, the Matérn covariance functions are

- $\nu = 1/2$: $\sigma_u(d) \propto \exp[-cd]$
- $\nu = 3/2$: $\sigma_u(d) \propto \left(1 + \sqrt{3}dc\right) \exp[-\sqrt{3}cd]$
- $\nu = 5/2$: $\sigma_u(d) \propto \left(1 + \sqrt{5}dc + (5/2)d^2c^2\right) \exp[-\sqrt{5}cd]$
- $\nu = \infty$: $\sigma_u(d) \propto \exp[-c^2d^2/2]$.

For any $\mathbf{\Sigma}(c_0)$ it is straightforward to compute the bounds $\overline{\rho}_L$ and $\overline{\rho}_U$ such that the inequalities in Theorem 8 are satisfied for all values of $\theta \in \Theta_{\overline{\rho}_L,\overline{\rho}_U}$ with $\nu \in \{1/2, 3/2, 5/2, \infty\}$ and $c > 0$. We carried out this exercise for the U.S. states spatial correlation designs of Section 2 (the calculations for one set of locations take less than a second). We find $\overline{\rho}_L \leq 0.001$ and $\overline{\rho}_U = \overline{\rho}_0 \in \{0.02, 0.10\}$, with very few minor exceptions.

We conclude that SCPC controls size in finite Gaussian samples for a wide range of Matérn process mixtures that imply $\overline{\rho} \leq \overline{\rho}_0$, at least for this set of spatial designs.

4.2.2 Implications for regularly-spaced time series

The spatial design is fixed for regularly-spaced time series, so the theorem can provide general robustness results. Consider, for instance, the equal weighted cosine (EWC) projection estimator of Müller (2004, 2007), Lazarus, Lewis, Stock, and Watson (2018) and Dou (2019) where $\mathbf{w}(s) = \sqrt{2/q}(\cos \pi s, \cos(2\pi s), \ldots, \cos(q\pi s))$. Suppose the critical value cv_n is chosen so that size is controlled in a Gaussian AR(1) with coefficient $\exp(-c_0/n)$, and q is chosen to minimize expected length in the i.i.d. model. For $c_0 = 10$, $c_0 = 25$ and $c_0 = 50$, we obtain $q = 5, 7$ and 10, respectively, for all $n \in \{50, 100, 500\}$. Call this test the EWC(c_0) t-test.

Calculations based on Theorem 8 for these values of c_0 and n show that the EWC(c_0) t-test controls size for arbitrary mixtures of AR(1) processes with coefficients $\exp(-c/n)$, $c \geq c_0$. By taking the limit in n and using standard local-to-unity weak convergence results (as in Müller (2014)), one can further apply Theorem 1 to the limiting covariance matrices $\mathbf{\Omega}_0$ and $\mathbf{\Omega}(\theta)$ to study asymptotic robustness of the EWC(c_0) t-test with an asymptotically justified critical value (which are equal to $\mathrm{cv} = 3.53$, 2.71, 2.40 for $c_0 = 10$, 25, 50, respectively). Another numerical calculation based on Theorem 8 then shows that these EWC(c_0) t-tests control asymptotic size for underlying processes that are arbitrary mixtures of local-to-unity models with parameters $c \geq c_0$.

Moreover, let $f_{n,0} : [-\pi, \pi] \mapsto [0, \infty)$ be the spectral density of an AR(1) process with coefficient $\exp(-c_0/n)$, so $f_{n,0}(\omega) \propto (1 - 2e^{-c_0/n} \cos \omega + e^{-2c_0/n})^{-1}$. A spectral density $f_{n,1}$ would naturally be considered less persistent than $f_{n,0}$ if $f_{n,1}(\omega)/f_{n,0}(\omega)$ is (weakly) monotonically increasing in $|\omega|$. Denote all such functions by $\mathcal{F}_n$. Define

$$M = \frac{f_{n,1}(\pi)/f_{n,0}(\pi)}{f_{n,1}(0)/f_{n,0}(0)},$$

so M measures by how much $f_{n,1}(\omega)/f_{n,0}(\omega)$ increases over $[0, \pi]$, and denote by $\mathcal{F}_n^{\bar{M}}$ all functions in $\mathcal{F}_n$ with $M \leq \bar{M}$ for some $\bar{M} > 1$. Then for any $f_{n,1} \in \mathcal{F}_n^{\bar{M}}$, there exists a CDF H on $[0, \pi]$ such that

$$\begin{aligned} f_{1,n}(\omega) &\propto f_{n,0}(\omega) + (M-1)H(|\omega|)f_{n,0}(\omega) \\ &= \frac{\bar{M} - M}{\bar{M} - 1} f_{n,0}(\omega) + \frac{M-1}{\bar{M}-1} \int [f_{n,0}(\omega) + (\bar{M}-1)\mathbf{1}[|\omega| \geq \theta] f_{n,0}(\omega)] dH(\theta) \end{aligned}$$

so $f_{n,1}$ has a representation as a scale mixture of $f_{n,0}(\omega) + (\bar{M} - 1)\mathbf{1}[|\omega| \geq \theta] f_{n,0}(\omega)$, $0 \leq \theta \leq \pi$. After translating this back into a corresponding mixture of covariance matrices $\mathbf{\Sigma}^p(\theta)$, an application of Theorem 8 shows that the EWC(c_0) t-test also controls size in this class, for $(c_0, \bar{M}) \in \{(10,10),(25,10),(50,5)\}$ and all $n \in \{50,100,500\}$. These results refine corresponding results in Dou (2019) that are based on a Whittle-type diagonal approximation to $\mathbf{\Sigma}$.

Taking limits as $n \to \infty$ yields a corresponding asymptotic robustness statement: The function $f_0 : \mathbb{R} \mapsto [0,\infty)$ with $f_0(\omega) = (\omega^2 + c_0^2)^{-1}$ is proportional to the 'local-to-zero' spectral density (cf. Müller and Watson (2016, 2017)) of a local-to-unity process with parameter c_0. Consider any process whose local-to-zero spectral density f_1 is such that $f_1(\omega)/f_0(\omega)$ is monotonically increasing in $|\omega|$ with $\lim_{\omega\to\infty} f_1(\omega)/f_0(\omega) \leq \bar{M} f_1(0)/f_0(0)$ and that satisfies the CLT in Müller and Watson (2016, 2017). A numerical calculation based on Theorem 8 then shows that the EWC(c_0) t-tests for $(c_0, \bar{M}) \in \{(10,10),(25,10),(50,5)\}$ controls size in large samples under all such processes.

4.3 Size properties of SCPC under heteroskedasticity and mismeasured locations

The SCPC t-test is not robust to heteroskedasticity or measurement error in locations by construction. For example, suppose that $u(s) = h(s)\tilde{u}(s)$, where $\tilde{u}$ is homoskedastic and satisfies the assumptions outlined above for u, and $h : \mathcal{S} \mapsto \mathbb{R}$ is a non-random function that induces heteroskedasticity in the u process. The linear combinations of u studied in Lemma 2 are now $\sum_{l=1}^n \mathbf{w}^0(s_l)u(s_l) = \sum_{l=1}^n \mathbf{w}_h^0(s_l)\tilde{u}(s_l)$ where $\mathbf{w}_h^0(s) = \mathbf{w}^0(s)h(s)$. The results of the lemma and subsequent theorems then follow with $\mathbf{w}_h^0$ replacing $\mathbf{w}^0$. But, the test statistic and critical value is computed using $\mathbf{w}^0$, not $\mathbf{w}_h^0$, so that size control is not guaranteed, even in large samples. An analogous problem arises when the locations s_i are measured with error.

In both cases, the particulars of the size distortion depend on the distribution of spatial locations, g, the weights $\mathbf{w}^0$ (which in turn depend on the value of $\overline{\rho}_0$ used to calibrate c_0), the function h in the heteroskedastic model and the distribution of the measurement error for the locations.

We summarize two experiments that illustrate and quantify the size distortions in the U.S. states spatial correlation designs. The first experiment is a heteroskedastic model with $\log h$ increasing or decreasing linearly from $\log h(s) = 0$ to $\log h(s) = \log 3$ moving from the most westward to the most eastward location, the experiment is repeated with h increasing or decreasing moving north to south, and we record the largest of the four rejection frequencies. Panel (a) of Figure 6 plots the CDF of rejection frequencies for nominal 5% SCPC tests for each $(\overline{\rho}_0, g)$ pair. For these designs, the resulting size distortions are not large, except for a

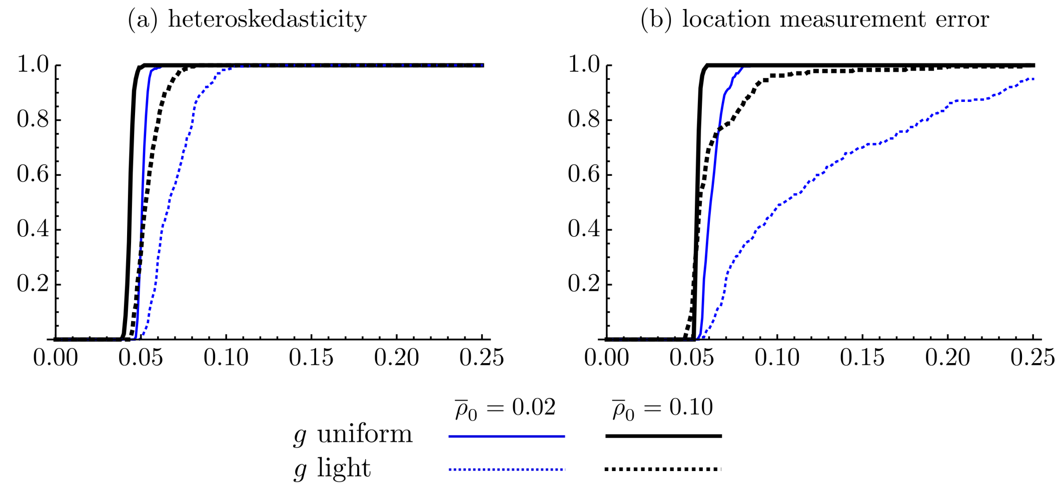

Figure 6: CDFs of Size under Heteroskedasticity and Location Measurement Error

few states with $\overline{\rho}_0 = 0.02$ and the light spatial density g, where rejection frequencies approach 10%.

The second experiment investigates location measurement error of a form studied in Conley and Molinari (2007). Specifically for each location, $s_i^* = s_i + e_i$ where s_i^* is the measured location, s_i is the true location and e_i is the measurement error. The error term is $e_i = (e_{1,i}, e_{2,i})$ with $e_{1,i}$ the north-south and $e_{2,i}$ the east-west coordinate and $e_{j,i}$ i.i.d.$\mathcal{U}(-\delta, \delta)$ over j and i, and $\delta = 0.0375H$ with H the length of the smallest square that encompasses all locations, corresponding to "level 4" errors in Conley and Molinari's (2007) classification. The CDFs for the rejection frequencies are shown in panel (b) of Figure 6. Evidently, measurement error of this sort has little effect on the size of SCPC under uniformly distributed locations, but can have a substantial effect for highly concentrated spatial distributions, especially when $\overline{\rho}_0 = 0.02$.

5 Efficiency of SCPC

Figure 5 showed the expected length of the SCPC confidence interval relative to the length of an oracle confidence interval that uses the true value of $\text{Var}(\sqrt{n}(\overline{y} - \mu))$ conditional on the observed locations $\mathbf{s}$. (As before, in this subsection we keep the conditioning on $\mathbf{s}$ and the dependence on n implicit.) For studying efficiency, a more relevant comparison involves the expected length of the SCPC confidence interval relative to a confidence interval that, like SCPC, does not depend on the true (unknown) value of $\text{Var}(\sqrt{n}(\overline{y} - \mu))$. Ideally, such a comparison would involve SCPC and the most efficient method for constructing a confidence interval. We undertake such a comparison here.

To be specific, let $\mathrm{CS}(\mathbf{y}) \subset \mathbb{R}$ denote a confidence set for μ constructed from $\mathbf{y}$. We restrict attention to location and scale equivariant confidence sets, that is CS satisfies $\mathrm{CS}(a_\mu + a_\sigma \mathbf{y}) = \{\mu_0 : (\mu_0 - a_\mu)/a_\sigma \in \mathrm{CS}(\mathbf{y})\}$ for all $\mathbf{y}$, $a_\mu \in \mathbb{R}$ and $a_\sigma > 0$. As in Section 4.2, we focus on the Gaussian model $\mathbf{y} \sim \mathcal{N}(\mathbf{l}\mu, \boldsymbol{\Sigma})$. We want to compare the SCPC interval with a confidence interval that, like SCPC, has good coverage $\mathbb{P}_{\boldsymbol{\Sigma}}(\mu \in \mathrm{CS}(\mathbf{y}))$ over a range of potential spatial correlation patterns $\boldsymbol{\Sigma} \in \mathcal{V}$. The metric for measuring efficiency is the expected length $\mathbb{E}^1[\int \mathbf{1}[x \in \mathrm{CS}(\mathbf{y})]dx]$ in the i.i.d. model $\mathbf{y} \sim \mathcal{N}(\mathbf{l}\mu, \mathbf{I})$.

Our choice of $\mathcal{V}$ is motivated by the structure of the SCPC benchmark covariance matrix $\boldsymbol{\Sigma}(c_0)$. The idea is to include in $\mathcal{V}$ covariance matrices that are weakly less persistent than $\boldsymbol{\Sigma}(c_0)$, and that cannot be easily distinguished from the i.i.d. model. To characterize these covariance matrices, note that $\boldsymbol{\Sigma}(c_0)$ is generated from u, an isotropic random field with covariance function $\sigma_u(s, r) = \exp(-c_0||s - r||)$. Isotropy implies that the spectrum of this random field $F_0 : \mathbb{R}^d \mapsto [0, \infty)$ at frequency $\boldsymbol{\omega} \in \mathbb{R}^d$ can be written as function of the scalar $\omega = ||\boldsymbol{\omega}||$, that is $F_0(\boldsymbol{\omega}) = f_0(\omega)$ for some $f_0 : \mathbb{R} \mapsto [0, \infty)$. As is well known, the exponential covariance model for $d = 2$ corresponds to a spectral density function f_0 proportional to $(c_0 + \omega^2)^{-3/2}$. By scale invariance of both CS and the SCPC interval, it is without loss of generality to set f_0 equal to

$$f_0(\omega) = \frac{1}{(c_0 + \omega^2)^{3/2}}.$$

For some $\bar{\omega} > 0$, define $f_\Delta(\omega) = \mathbf{1}[|\omega| \le \bar{\omega}](f_0(\omega) - f_\Delta(\bar{\omega}))$, and let $f_R(\omega) = f_0(\omega) - f_\Delta(\omega)$, so that

$$f_0(\omega) = f_\Delta(\omega) + f_R(\omega).$$

For $0 \le |\omega| \le \bar{\omega}$, the density f_Δ is equal to $f_0(\omega) - f_0(\bar{\omega})$, so that the remainder $f_R(\omega)$ is a continuous density that is flat for $|\omega| \le \bar{\omega}$, and that follows the same decline as f_0 for $|\omega| > \bar{\omega}$. Since both $f_\Delta(\omega)$ and $f_R(\omega)$ are non-negative, we have the corresponding identity in covariance matrices

$$\boldsymbol{\Sigma}(c_0) = \boldsymbol{\Sigma}_\Delta(\bar{\omega}) + \boldsymbol{\Sigma}_R(\bar{\omega}) \tag{17}$$

where $\boldsymbol{\Sigma}_\Delta(\bar{\omega})$ and $\boldsymbol{\Sigma}_R(\bar{\omega})$ are induced by the isotropic random fields with spectral densities $F_\Delta(\boldsymbol{\omega}) = f_\Delta(||\boldsymbol{\omega}||)$ and $F_R(\boldsymbol{\omega}) = f_R(||\boldsymbol{\omega}||)$, respectively.

Now consider the covariance matrix

$$\bar{\boldsymbol{\Sigma}}(\bar{\omega}) = \boldsymbol{\Sigma}_\Delta(\bar{\omega}) + \lambda_1(\boldsymbol{\Sigma}_R(\bar{\omega}))\mathbf{I}_n$$

where $\lambda_1(\mathbf{\Sigma}_R(\bar{\omega}))$ is the largest eigenvalue of $\mathbf{\Sigma}_R(\bar{\omega})$. Since $f_R(\omega)$ is monotonically decreasing in $|\omega|$, also $\mathbf{\Sigma}_R(\bar{\omega})$ contributes to the persistence of $\mathbf{\Sigma}(c_0)$ in (17), so replacing it with white noise of weakly larger variance should make inference about μ under $\bar{\mathbf{\Sigma}}(\bar{\omega})$ no harder than under $\mathbf{\Sigma}(c_0)$.[5] Said differently, a method that is robust under correlation patterns weakly less persistent than $\mathbf{\Sigma}(c_0)$ should continue to have good coverage after replacing medium and high frequency variation in $\mathbf{y}$ by white noise, that is, under $\bar{\mathbf{\Sigma}}(\bar{\omega})$. This motivates the set $\mathcal{V} = \{\bar{\mathbf{\Sigma}}(\bar{\omega})|\bar{\omega} > 0\}$.

A calculation shows that in the U.S. states spatial correlation designs, the SCPC interval has good coverage properties under this $\mathcal{V}$. With $\alpha_{\text{SCPC}}(\bar{\omega}) = \mathbb{P}_{\bar{\mathbf{\Sigma}}(\bar{\omega})}(\tau^2_{\text{SCPC}} > \text{cv}^2_{\text{SCPC}})$ for the nominal 5% level SCPC test, for most designs, $\sup_{\bar{\omega}\geq 0} \alpha_{\text{SCPC}}(\bar{\omega})$ is equal or very close to 5%, and it never exceeds 8%. To keep things on an equal footing, we allow CS the same degree of undercoverage, that is we consider the problem

$$\inf_{\text{CS}} \mathbb{E}^1[\int \mathbf{1}[x \in \text{CS}(\mathbf{y})]dx] \text{ s.t. } \mathbb{P}_{\bar{\mathbf{\Sigma}}(\bar{\omega})}(\mu \notin \text{CS}(\mathbf{y})) \leq \max(\alpha_{\text{SCPC}}(\bar{\omega}), \alpha) \text{ for all } \bar{\omega} > 0. \tag{18}$$

In words, we seek the invariant confidence set with the shortest expected length in the i.i.d. location model among all confidence sets that are as robust as the SCPC interval under $\bar{\mathbf{\Sigma}}(\bar{\omega})$, $\bar{\omega} > 0$.

Since $\bar{\omega}$ is one-dimensional, one can apply the numerical techniques of Elliott, Müller, and Watson (2015) and Müller and Norets (2016) (also see Müller and Watson (in preparation)) to obtain an informative lower bound on the objective $\inf_{\text{CS}} \mathbb{E}^1[\int \mathbf{1}[x \in \text{CS}(\mathbf{y})]dx]$ that holds for *any* equivariant $\text{CS}(\mathbf{y})$ that satisfies the constraint in (18).

We compute such lower bounds in the U.S. states spatial correlation designs. Panel (a) of Figure 7 shows the CDFs of the length of SCPC confidence intervals relative to the lower bounds for the 240 designs in each $(\overline{\rho}_0, g)$ pair. The expected lengths of SCPC are within 7% of the efficiency bound for all designs when $\overline{\rho}_0 = 0.02$. When $\overline{\rho}_0 = 0.10$, so that spatial correlation is high, and the spatial locations are highly concentrated as under the light design, the expected length of the SCPC confidence interval can be more that 15% longer than the efficiency bound. In part, this is because the implied efficient confidence sets are complicated and rather uninterpretable functions of $\mathbf{y}$ in this case. We thus repeat the exercise for confidence sets constrained to be symmetric around $\bar{y}$ by imposing $\text{CS}(a_\mu + a_\sigma \mathbf{y}) = \{\mu_0 : (\mu_0 - a_\mu)/a_\sigma \in \text{CS}(\mathbf{y})\}$

[5] In the regularly-spaced time series setting, white noise amounts to a flat spectrum, so $\mathbf{\Sigma}_0(\bar{\omega})$ corresponds to an underlying spectral density equal to $f_\Delta(\omega) + f_0(\bar{\omega})$, which is the "kinked" spectral density considered by Dou (2019). For arbitrary locations, however, the domain of the spectrum doesn't fold onto the interval $[-\pi, \pi]$, so that white noise cannot mathematically be represented by a flat spectrum.

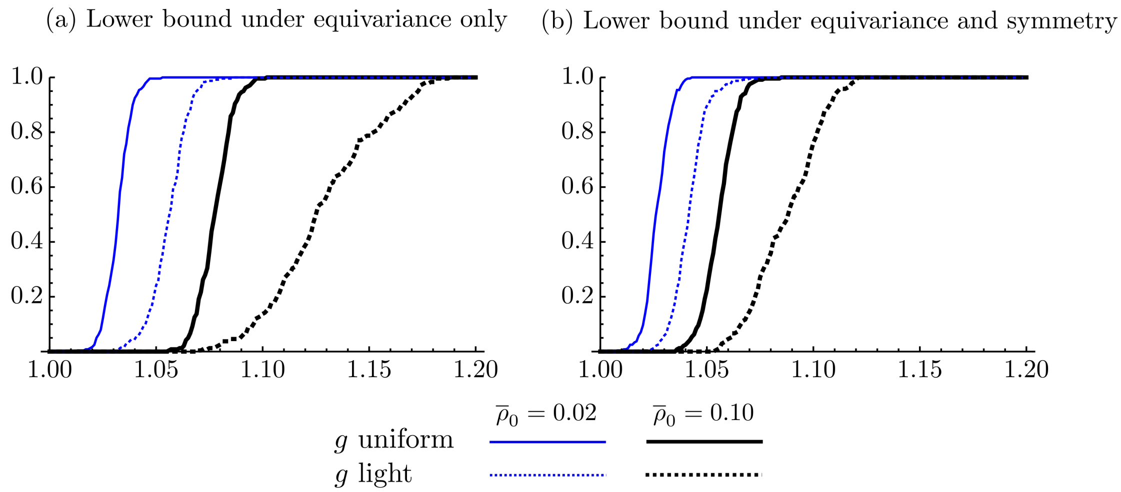

Figure 7: CDFs of Expected Length of SCPC Relative to Lower Bound on Expected Length

for all $\mathbf{y}$, $a_\mu \in \mathbb{R}$ and $a_\sigma \neq 0$. The results are summarized in panel (b), and we can see that SCPC comes closer to the resulting higher bound on confidence interval length.

Remark 5.1. These efficiency results also provide a limit on the possibility of using data-dependent methods to learn about the value of the worst-case correlation c_0: Since the i.i.d. model corresponds to $c \to \infty$, if it was possible to learn the value of c from the data, one would be able to conduct much more efficient inference than what is reported in Figure 7. The results here thus provide a rationalization for treating c_0 as given.[6]

6 Comparison with other methods

This section compares SCPC with other methods that have been proposed, focusing on size and expected length of confidence intervals in the benchmark Gaussian model with exponential covariance kernel and parameter c_0 (calibrated by $\bar{\rho}_0$). We consider two kernel-based methods, two versions of a cluster method, and one projection method. All these methods are t-statistic based tests of the form considered in Section 3.

The kernel based methods use a Bartlett kernel, $k(s,r) = k_{\text{Bartlett}}(||s-r||/b)$. The methods differ in their choice of bandwidth b and critical value. The first method uses a standard normal critical value with b chosen so the resulting test has size as close as possible to 5%. This is a version of the method proposed by Conley (1999), but with an oracle choice for the

[6]Also see Dou (2019) for a related discussion and associated impossibility results.

bandwidth. The second method sets $b = \max_{l,\ell} ||s_l - s_\ell||$ and chooses the critical value to obtain exact coverage under $\mathbf{\Sigma} = \mathbf{I}$. This is the spatial analogue of the method suggested by Kiefer, Vogelsang, and Bunzel (2000) (KVB) for regularly spaced time series. The cluster methods follow the approach of Ibragimov and Müller (2010) (IM) with student-t_q critical values and is implemented with $q = 4$ and $q = 9$ equal-sized clusters.[7] The projection method follows Sun and Kim (2012). It uses a student-t_q critical value and q low-frequency Fourier weights orthogonalized using the sample locations, where q is chosen as a function of the exponential model parameter c_0 using the formula in their equation (8). The first and last method are thus tailored to the true value c_0, just like SCPC.

We analyze these methods in the U.S. states spatial correlation designs, augmented to also include the value $\overline{\rho}_0 = 0.001$ for the average pairwise correlation to investigate performance under 'weak' spatial correlations. Figure 8 summarizes the results for size control and expected lengths by plotting the CDFs for each $(\overline{\rho}_0, g)$ pair. The first column shows the null rejection frequency for each method; by construction, the rejection frequency for SCPC is at most 5% in all designs. The expected lengths in the second and third column use size-corrected critical values to ensure 95% coverage under $\mathbf{\Sigma}(c_0)$, and are given in multiples of the expected length of the (non-adjusted) SCPC method. The second column reports these relative expected lengths under $\mathbf{\Sigma} = \mathbf{I}$, and the third column under $\mathbf{\Sigma}(c_0)$.

Looking at the first column, the kernel and cluster methods have null rejection probabilities close to 5% when $\overline{\rho}_0 = 0.001$, but exhibit significant size distortions for $\overline{\rho}_0 = 0.02$ or 0.10. Evidently, the kernel and cluster methods substantially underestimate the variance of $\overline{y}$ for the latter two values of $\overline{\rho}_0$. In contrast, the Fourier projection method has relatively small size distortions under $g = g_{\text{uniform}}$ but can have substantial size distortions under $g = g_{\text{light}}$, even when $\overline{\rho}_0 = 0.001$. This is consistent with the implications of Theorem 2: the student-t critical value for the projection method is appropriate when $\mathbf{\Omega} \propto \mathbf{I}$, which it is under weak-correlation with g uniform, but not otherwise, even for large q (cf. Remark 3.11).

The relative lengths shown in the second column are above unity, sometimes by a wide margin, indicating that SCPC is closer to the efficiency bound computed in Section 5 than these alternative methods, at least for the designs considered here. The third column shows

[7] 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.

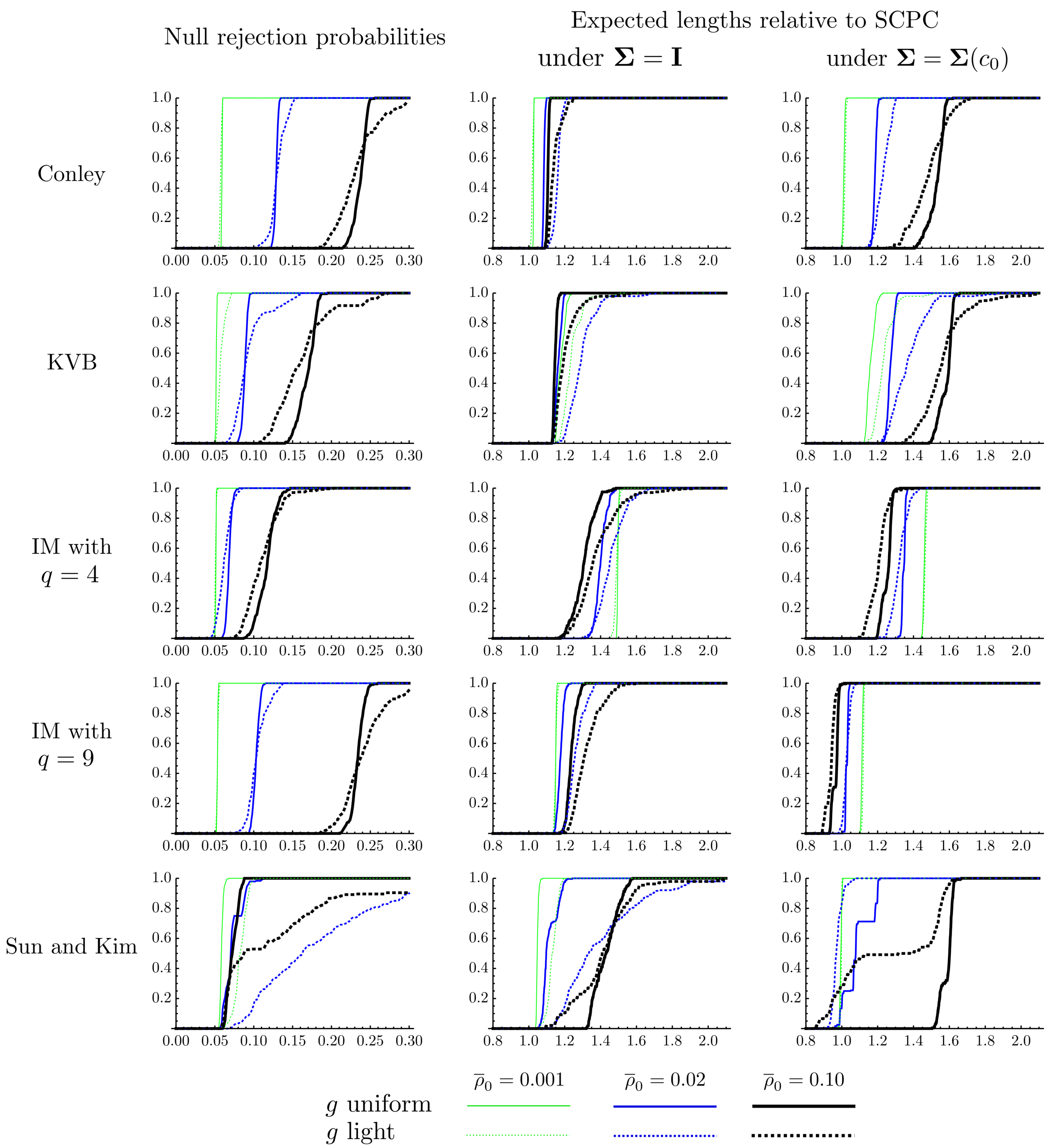

Figure 8: CDFs of Null Rejection Probability and Relative Expected Length of Alternative Methods

that this continues hold for lengths computed under $\mathbf{\Sigma}(c_0)$ with a few exceptions. Notably, the expected length of the size-adjusted 9-cluster method is smaller than SCPC when $\bar{\rho}_0 = 0.10$. This apparent good performance comes at the cost of substantially longer confidence intervals in the i.i.d. model.

7 Extensions and computational issues

This section discusses extensions of the method to regression and GMM models, some computational issues, and the multivariate extension of SCPC.

7.1 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

$$w_l = x_l\beta + \mathbf{z}_l'\delta + \varepsilon_l \text{ for } l = 1, ..., n \tag{19}$$

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 is analogous; see, for instance, Section 4.4 of Müller (2020).

7.2 Computational issues

We highlight two computational issues. The first involves the calculation of the SCPC critical value, and the second involves the problem of computing the eigenvectors $\mathbf{r}_j$ of $\mathbf{M}\boldsymbol{\Sigma}(c_0)\mathbf{M}$ when n is very large.

The critical value $\mathrm{cv} = \mathrm{cv}_{\mathrm{SCPC}}(q)$ solves $\sup_{c\geq c_0} \mathbb{P}_{\boldsymbol{\Sigma}(c)}(\tau^2(q^{-1}\sum_{j=1}^q \mathbf{r}_j\mathbf{r}_j') > \mathrm{cv}^2) = \alpha$ or equivalently (from Theorem 2) $\sup_{c\geq c_0} \mathbb{P}\,(Z_0^2 > \sum_{i=1}^q \eta_i Z_i^2) = \alpha$ where $\eta_i = -\omega_i/\omega_0$, ω_i are the eigenvalues of $\hat{\mathbf{W}}^{0\prime}\boldsymbol{\Sigma}(c)\hat{\mathbf{W}}^0\mathbf{D}(\mathrm{cv})$ with $\hat{\mathbf{W}}^0 = [\mathbf{l}, \mathbf{r}_1/\sqrt{q}, \ldots, \mathbf{r}_q/\sqrt{q}]$ and $Z_j \sim$i.i.d. $\mathcal{N}(0,1)$. Bakirov and Székely (2005) show that

$$\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \eta_i Z_i^2\right) = \frac{1}{\pi}\int_0^1 \frac{x^{\frac{q-1}{2}}}{\sqrt{(1-x)\prod_{i=1}^q (x+\eta_i)}}dx, \tag{20}$$

which is readily evaluated by numerical quadrature. Thus $\mathrm{cv}_{\mathrm{SCPC}}(q)$ can be obtained by combining a root-finder with a grid search over $c \geq c_0$.

The second problem involves computing the eigenvectors $\mathbf{r}_j = (r_{j,1}, \ldots, r_{j,n})'$ of the $n \times n$ matrix $\mathbf{M}\boldsymbol{\Sigma}(c_0)\mathbf{M}$ when n is very large (say, larger than $n = 2000$). Here we can leverage the eigenfunction convergence result in Lemma 5 as discussed in Section 3.4.3: In the notation defined there, we seek to approximate $\mathbf{r}_j = (\hat{\varphi}_j^0(s_1), \ldots, \hat{\varphi}_j^0(s_n))'$. Consider a random subset of size $\tilde{n} < n$ of the observed locations $\{\tilde{s}_l\}_{l=1}^{\tilde{n}} \subset \{s_l\}_{l=1}^n$, and let $\tilde{\boldsymbol{\Sigma}}(c_0)$ be the implied $\tilde{n} \times \tilde{n}$ covariance matrix of $(u(\tilde{s}_1), \ldots, u(\tilde{s}_n))'$ using the benchmark covariance function $\sigma_u^0(r-s|c_0) = \exp[-c_0||r-s||]$. Let the eigenvector corresponding to the jth largest eigenvalue $\tilde{\lambda}_j$ of $\tilde{\boldsymbol{\Sigma}}(c_0)$ be $\tilde{\mathbf{r}}_j = (\tilde{r}_{1,j}, \ldots, \tilde{r}_{\tilde{n},j})'$ with $\tilde{n}^{-1}\tilde{\mathbf{r}}_j'\tilde{\mathbf{r}}_j = 1$. As long as $\tilde{n} \to \infty$ and $\lambda_{q+1} > \lambda_q$, Lemma 5 implies that the span of the $\mathcal{S} \mapsto \mathbb{R}$ functions

$$\tilde{\varphi}_j^0(s) = \tilde{n}^{-1}\tilde{\lambda}_j^{-1}\sum_{l=1}^{\tilde{n}} \tilde{r}_{j,l}\left(\exp[-c_0||s-\tilde{s}_l||] - \tilde{n}^{-1}\sum_{\ell=1}^{\tilde{n}}\exp[-c_0||\tilde{s}_l - \tilde{s}_\ell||]\right), \; j = 1, \ldots, q$$

converges to the eigenspace spanned by φ_j^0, $j = 1, \ldots, q$, just like the full sample estimators $\hat{\varphi}_j^0$. Thus, it is formally justified to approximate the value of $\hat{\varphi}_j^0$ at locations $\{s_l\}_{l=1}^n \ni s_\ell \notin \{\tilde{s}_l\}_{l=1}^{\tilde{n}}$ via $r_{j,\ell} = \hat{\varphi}_j^0(s_\ell) \approx \tilde{\varphi}_j^0(s_\ell)$—this is a version of the so-called Nyström method (see, for instance, Rasmussen and Williams (2005) for discussion and references).

In practice, such approximations can be carried out for several random subsets of $\tilde{n}$ locations, followed by a (sample) principle component analysis to extract the best approximation to the space spanned by the first q eigenvectors. The resulting algorithm has $O(n)$ running time (in contrast to the $O(n^2)$ running time of a basic implementation of Conley (1999)-type kernel estimators). We provide corresponding STATA and Matlab code in the replication files.

7.3 Extension to F-tests

Consider the case where $\mathbf{y}_l = \boldsymbol{\mu} + \mathbf{u}_l$ with $\mathbf{y}_l$, $\boldsymbol{\mu}$ and $\mathbf{u}_l$ $m \times 1$ vectors, and we seek to test the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. Suppose the observations conditional on $\mathbf{s}$ are generated by the model

$$\mathbf{u}(s_l) = \mathbf{B}(c_n s_l),\ l = 1, \ldots, n$$

where $\mathbf{B}(s)$ is an $\mathbb{R}^m$-valued mean-zero stationary random field on $\mathbb{R}^d$ with covariance function $\mathbb{E}[\mathbf{B}(s)\mathbf{B}(r)'] = \boldsymbol{\sigma}_B(r-s)$. Let $\mathbf{Y}$ and $\mathbf{U}$ be the $n \times m$ matrices of observations and innovations, respectively, and $\bar{\mathbf{y}} = n^{-1}\sum_{l=1}^{n} \mathbf{y}_l$ the sample mean. The natural analogue to the t-statistic $\tau^2(\hat{\mathbf{W}}\hat{\mathbf{W}}')$ is Hotelling's-T^2 statistic

$$T^2(\hat{\mathbf{W}}\hat{\mathbf{W}}') = n(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)' \left(\mathbf{Y}'\hat{\mathbf{W}}\hat{\mathbf{W}}'\mathbf{Y}\right)^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}_0). \tag{21}$$

One would expect that under mixing and moment conditions similar to those of Lemma 1 (ii)

$$\operatorname{vec}(\mathbf{W}^{0\prime}\mathbf{U})|\mathbf{s} \Rightarrow_p \mathcal{N}\left(0, a\boldsymbol{\sigma}_B(r-s) \otimes \mathbf{V}_1 + \left[\int \boldsymbol{\sigma}_B(s)ds\right] \otimes \mathbf{V}_2\right). \tag{22}$$

Note that $T^2(\hat{\mathbf{W}}\hat{\mathbf{W}}')$ is invariant to the transformation $\mathbf{Y} \to \mathbf{YH}$ for nonsingular $\mathbf{H}$. For the purposes of studying the limit distribution of $T^2(q)$ under weak correlation, it is thus without loss of generality to normalize $\boldsymbol{\sigma}_B(\cdot)$ such that the limit covariance matrix in (22) becomes

$$\operatorname{diag}(\boldsymbol{\kappa}) \otimes \mathbf{V}_1 + (\mathbf{I}_m - \operatorname{diag}(\boldsymbol{\kappa})) \otimes \mathbf{V}_2 \tag{23}$$

where $\boldsymbol{\kappa}$ is a $m \times 1$ vector with elements in $[0, 1)$.

For the extension of the SCPC method, consider a benchmark model indexed by $\mathbf{c} = (c_1, \ldots, c_m)$ where $\operatorname{vec}(\mathbf{Y})|\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu} \otimes \mathbf{l}_n, \boldsymbol{\Sigma}(\mathbf{c}))$ with $\boldsymbol{\Sigma}(\mathbf{c}) = \operatorname{diag}(\boldsymbol{\Sigma}(c_1), \ldots, \boldsymbol{\Sigma}(c_m))$, and $\boldsymbol{\Sigma}(c)$ is as in Section 2. Let $\mathbf{c}_0 = c_0\mathbf{l}_m$, a $m \times 1$ vector of identical elements c_0. The SCPC test statistic $T^2_{\text{SCPC}}(q)$ is a special case of (21) with the columns of $\hat{\mathbf{W}}$ equal to the first q eigenvectors of $\boldsymbol{\Sigma}(c_0)$, scaled to have length $1/\sqrt{q}$, and with critical value $\text{cv}^T_{\text{SCPC}}$ chosen to satisfy

$$\sup_{\mathbf{c} \geq \mathbf{c}_0} \mathbb{P}^0_{\boldsymbol{\Sigma}(\mathbf{c})}(T^2_{\text{SCPC}}(q) > \text{cv}^T_{\text{SCPC}}(q)|\mathbf{s}) = \alpha,$$

under the null hypothesis, where $\mathbf{c} \geq \mathbf{c}_0$ is understood as an elementwise inequality. The value of q that minimizes the expected volume of the confidence ellipsoid under $\text{vec}(\mathbf{Y})|\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu} \otimes \mathbf{l}, \mathbf{I}_m \otimes \mathbf{I}_n)$ is

$$\min_{q \geq m} \mathbb{E}[\text{vol}\{\mathbf{m} : \mathbf{m}'(q^{-1}\mathbf{S}_q)^{-1}\mathbf{m} \leq n^{-1} \operatorname{cv}^T_{\text{SCPC}}(q)\}] = \min_{q \geq m} \frac{(2\pi \operatorname{cv}^T_{\text{SCPC}}(q)/n)^{m/2}\Gamma((q+1)/2)}{\sqrt{q}\Gamma((q-m+1)/2)\Gamma(m/2+1)}$$

where $\mathbf{S}_q$ is distributed Wishart with q degrees of freedom, and the equality follows from Bartlett's decomposition of a Wishart random matrix, and the formulas for the expectation of a χ random variable and the volume of an m dimensional ellipsoid.

Since appropriate choices of $c_{j,n} \to \infty$, $j = 1, \ldots, m$ in the benchmark model can replicate the normalized limit distributions (23) for all $\boldsymbol{\kappa}$, by the same arguments that lead to Theorem 7, $T^2_{\text{SCPC}}(q)$ controls size under all weak correlation patterns that induce (22). And as in Section 7.1, it is straightforward to adapt $T^2_{\text{SCPC}}(q)$ to test m restrictions in linear regression and GMM problems. We omit details for brevity. Generalizing the results about the small sample robustness of τ_{SCPC} under potentially strong correlations in Theorem 8 to T^2_{SCPC} is interesting but challenging, and beyond the scope of this paper.

A Appendix

Lemma 9. *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 $\sup_{h\in\mathrm{BL}} |\mathbb{E}[h(\mathbf{X}_n) - h(\mathbf{X})|\mathbf{s}_n]| \xrightarrow{p} 0$, so it suffices to show that $\sup_{h\in\mathrm{BL}} |\mathbb{E}[h(\mathbf{X}_n + \mathbf{Y}_n) - h(\mathbf{X})|\mathbf{s}_n]| \xrightarrow{p} 0$. Let $\mathbf{Y}_n^* = \mathbf{Y}_n \mathbf{1}[||\mathbf{Y}_n|| \leq 1]$, so that

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

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 $h \in \mathrm{BL}$, so that

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

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. □

Proof of Lemma 1: (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 $\mathbf{w}^0(s)$,

$$\mathbb{E}\left[\left(\frac{1}{n(n-1)}\sum_{l\neq\ell} w_i^0(s_l)w_j^0(s_\ell)'\sigma_B(c\,(s_l - s_\ell))\right)^2\right]$$

$$= \frac{(n-2)(n-3)}{n(n-1)}\mathbb{E}[w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1-s_2))]\mathbb{E}[w_i^0(s_3)w_j^0(s_4)'\sigma_B(c\,(s_3-s_4))]$$
$$+\frac{4(n-2)}{n(n-1)}\mathbb{E}[w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1-s_2))w_i^0(s_1)w_j^0(s_3)'\sigma_B(c\,(s_1-s_3))]$$
$$+\frac{2}{n(n-1)}\mathbb{E}[w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1-s_2))w_i^0(s_1)w_j^0(s_2)'\sigma_B(c\,(s_1-s_2))]$$

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) Follows from Theorem 3.2 in Lahiri (2003) and the Cramér-Wold device. ■

Proof of Theorem 2: In the notation of Lemma 1, with $\mathbf{X} = (X_0, \mathbf{X}_{1:q}')'$ and $\mathbf{Z} = (Z_0,\ldots,Z_q)'$ we have

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

where the convergence follows from Lemma 1 and the continuous mapping theorem, and 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 14 below. ■

Proof of Theorem 3: We show that Lemma 1 (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 1 (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 1, proceeding as in the proof of Lemma 5.2 of Lahiri (2003) yields $\mathbb{E}[a_n n^{-1}\sum_{l,\ell}|\sigma_B(c_n\,(s_l-s_\ell))|] \to a\sigma_u^2 + \int_{\mathbb{R}^d}|\sigma_B(s)|ds\int g(s)^2ds$. The result thus follows from (10) and Lemma 9.

The proof of Theorem 4 requires a slightly more general version of Theorem 3.

Lemma 10. *In the notation of Lemma 5, suppose $\hat{\mathbf{W}} = \hat{\mathbf{L}}\hat{\mathbf{\Phi}}$, where the ith 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 $\hat{\mathbf{L}} = \operatorname{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_q)$. Under the assumptions of Lemma 1, $c_n^d n^{-2}(\mathbf{u}'\hat{\mathbf{W}}\hat{\mathbf{W}}'\mathbf{u} - \mathbf{u}'\mathbf{W}\mathbf{W}'\mathbf{u})|\mathbf{s}_n \xrightarrow{p} 0$, where $\mathbf{W} = \mathbf{L}\mathbf{\Phi}$, $\mathbf{L} = \operatorname{diag}(\lambda_1 \mathbf{l}_{m_1}, \ldots, \lambda_p \mathbf{l}_{m_p})$ and the ith column of $\mathbf{\Phi}$ is equal to $(\varphi_i(s_1), \ldots, \varphi_i(s_n))'$.*

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

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

where the first line follows from $\hat{\mathbf{O}}'\hat{\mathbf{O}} = \mathbf{I}_q$, the second from Lemma 5 (a) and (b) and the reasoning in the proof of Theorem 3, the third from Lemma 5 (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 1, and the fourth from $\hat{\mathbf{O}}'\mathbf{L}^2\hat{\mathbf{O}}' = \mathbf{L}^2$ a.s. The result now follows from Lemma 9. □

Proof of Theorem 4: 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 5, 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 5, 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

$$\text{(i) for any } \varepsilon > 0 \quad \lim_{q\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\hat{\xi}_n - \hat{\xi}_n^q| > \varepsilon) = 0 \tag{24}$$

$$\text{(ii) for any fixed } q,\ \mathbb{P}(\hat{\xi}_n^q > 0|\mathbf{s}_n) \xrightarrow{p} \mathbb{P}\left(\sum_{i=0}^{q} \omega_{q,i} Z_i^2 > 0\right) \tag{25}$$

$$\text{(iii)}\ \lim_{q\to\infty} \mathbb{P}\left(\sum_{i=0}^{q} \omega_{q,i} Z_i^2 > 0\right) = \mathbb{P}\left(\sum_{i=0}^{\infty} \omega_i Z_i^2 > 0\right) \tag{26}$$

for some double array of real numbers $\omega_{q,i}$ by invoking Lemma 9.

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

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

where the inequality follows from $\operatorname{tr}(\mathbf{AB}) \le \lambda_1(\mathbf{A}) \operatorname{tr} \mathbf{B}$ 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 3, $c_n^d n^{-2} \sum_{l,\ell} \sigma_B(c_n(s_l - s_\ell)) = O_p(1)$. Furthermore, by Lemma 5 (b), $|\hat{\lambda}_{q+1} - \lambda_{q+1}| = O_q(n^{-1/2})$, and $\lim_{q\to\infty} \lambda_q = 0$. Thus (24) follows.

For claim (ii), let $\varphi_0(s) = 1$ and $\lambda_0 = 1$. By Lemma 5 (a), Lemma 10 and Theorem 2, claim (25) holds, where $\omega_{q,i}$ are the eigenvalues of $\mathbf{D}(\mathrm{cv})\mathbf{\Omega}$ 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

$$\begin{aligned}\sum_{j=0}^{q}\sum_{i=0}^{q} v_j \varphi_i^*(s) \int d_i \varphi_j^*(r)\varphi_i^*(r)dG(r) &= \omega \sum_{j=0}^{q} v_j \varphi_j^*(s) \\ \int \left(\sum_{i=0}^{q} d_i \varphi_i^*(s)\varphi_i^*(r)\right)\left(\sum_{j=0}^{q} v_j \varphi_j^*(r)\right) dG(r) &= \omega \sum_{j=0}^{q} v_j \varphi_j^*(s) \qquad (27)\end{aligned}$$

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

$$\sum_{i=0}^{\infty}(\omega_{q,i}-\omega_i)^2 \leq \sum_{i=0}^{\infty}(\omega_{q,i}^{\Delta})^2 = ||R(T-T_q)R||_{HS} \qquad (28)$$

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 $||R(T-T_q)R||_{HS} \leq ||R||^2 \cdot ||T-T_q||_{HS}$ (cf. (32) 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

$$\lim_{q\to\infty}\sum_{i=0}^{\infty}(\omega_{q,i}-\omega_i)^2 = 0. \qquad (29)$$

Thus using the same order of eigenvalues as in (28), 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 (29). But mean-square convergence implies convergence in distribution, and (26) follows.

For the second claim of the theorem, by Lemma 1, $\omega_{q,i} \leq 0$ for $i \geq 1$, which in conjunction with (29) implies $\omega_i \leq 0$ for $i \geq 1$. ■

Proof of Lemma 5: 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

$$\max_{i\leq q}\sup_{s\in\mathcal{S}}\left|\varphi_i(s) - \sum_{j=1}^{q}\hat{O}_{n,ij}\hat{\varphi}_i(s)\right| = O_p(n^{-1/2}). \qquad (30)$$

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

$\mathcal{L}_G^2 \mapsto \mathcal{L}_G^2$

$$\begin{aligned} M(f)(s) &= f(s) - \int f(r)dG(r) \\ M_n(f)(s) &= f(s) - \int f(r)dG_n(r) \\ L(f)(s) &= \int k_0(r,s)f(r)dG(r) \\ L_n(f)(s) &= \int k_0(r,s)f(r)dG_n(r) \end{aligned}$$

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 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 $f \in \mathcal{H}$, $|f(r)| = \langle f(\cdot), k_0(\cdot, r)\rangle_{\mathcal{H}} \le ||f||_{\mathcal{H}} \cdot ||k_0(\cdot, r)||_{\mathcal{H}} \le \sqrt{K}||f||_{\mathcal{H}}$, we have

$$\sup_{r\in\mathcal{S}} |f(r)| \le \sqrt{K} \cdot ||f||_{\mathcal{H}}. \tag{31}$$

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,

$$||AB||_{HS} \le ||A||_{HS}||B||,\ ||BA||_{HS} \le ||B|| \cdot ||A||_{HS}. \tag{32}$$

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

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

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

so that (31) implies that $||M||$ is a bounded operator. By the same argument, so is M_n (almost surely). Thus, from (32), 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 (32)

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

and

$$\begin{aligned} ||(M_n - M)f||_{\mathcal{H}} &= \left\| \int f(r)dG_n(r) - \int f(r)dG(r) \right\|_{\mathcal{H}} \\ &= ||1||_{\mathcal{H}} \left| \int f(r)dG_n(r) - \int f(r)dG(r) \right|. \end{aligned}$$

Now consider the sequence of real independent random variables $f(s_l)$, which have mean $\mathbb{E}[f(s_l)] = \int f(r)dG(r)$, and, by (31), are almost surely bounded. Since $\int f(r)(dG_n(r) - dG(r)) = n^{-1}\sum_{l=1}^n f(s_l) - \mathbb{E}[f(s_1)]$, so that by Hoeffding's inequality, with probability of at least $1 - 2e^{-\delta}$

$$\left| \int f(r)(dG_n(r) - dG(r)) \right| \leq \sqrt{2\delta} n^{-1/2} \sup_{r \in \mathcal{S}} |f(r)|$$

for all $\delta \geq 0$. This holds for all $f \in \mathcal{H}$, so we conclude that $||M_n - M|| = O_p(n^{-1/2})$.

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 (30), 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 (30) follows from (31).

The claim in part (a) of the lemma now follows by induction from (30): For $p=1$, this follows directly. Suppose the result holds for $p-1$, and let $\hat{\mathbf{O}}_B=\operatorname{diag}(\hat{\mathbf{O}}_{(1)},\ldots,\hat{\mathbf{O}}_{(p-1)})$, so that

$$\sup_{s\in\mathcal{S}}||\hat{\mathbf{O}}_B\hat{\boldsymbol{\varphi}}_B(s)-\boldsymbol{\varphi}_B(s)||=O_p(n^{-1/2}), \tag{33}$$

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 (30) 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 (30), $\sup_{s\in\mathcal{S}}||\hat{\mathbf{O}}_{11}\hat{\boldsymbol{\varphi}}_B(s)+\hat{\mathbf{O}}_{12}\hat{\boldsymbol{\varphi}}_{I-B}(s)-\boldsymbol{\varphi}_B(s)||=O_p(n^{-1/2})$ 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 (33), 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 (30) 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 7: Suppose $\max(\overline{\mathrm{cv}}^2 - \mathrm{cv}_n^2, 0) \xrightarrow{p} 0$ does not hold. Then there exists $\delta > 0$ such that $\limsup_{n\to\infty} \mathbb{P}(\overline{\mathrm{cv}}^2 - \mathrm{cv}_n^2 > \delta) > \delta$. Define $\varkappa(\kappa, \overline{\mathrm{cv}}^2) = \mathbb{P}\left(\sum_{i=0}^{\infty} \omega_i(\kappa, \overline{\mathrm{cv}}) Z_i^2 > 0\right)$, 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

$$\mathbb{P}^0_{\mathbf{\Sigma}(c_{n,1})}(\tau_n^2 \geq \mathrm{cv}_{n,1}^2 \,|\mathbf{s}_n) = \alpha \quad \text{a.s.},$$

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',

$$\alpha = \mathbb{P}^0_{\mathbf{\Sigma}(c_{n',1})}(\tau_{n'}^2 \geq \mathrm{cv}_{n',1}^2 \,|\mathcal{A}_{n'}) \geq \mathbb{P}^0_{\mathbf{\Sigma}(c_{n',1})}(\tau_{n'}^2 \geq \overline{\mathrm{cv}}^2 - \delta|\mathcal{A}_{n'}) \text{ a.s.} \tag{34}$$

and by Theorem 4, $\mathbb{P}^0_{\mathbf{\Sigma}(c_{n',1})}(\tau_{n'}^2 \geq \overline{\mathrm{cv}}^2 - \delta|\mathcal{A}_{n'}) \to \varkappa(\kappa_0, \overline{\mathrm{cv}}^2 - \delta) > \alpha$. This contradicts (34), and the result follows. ■

Theorem 11. *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 7 for all $q \in \mathcal{Q}$ with critical value $\mathrm{cv}_n(q)$ as in (15), 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 7 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. □

The proof of Theorem 8 relies on some preliminary results.

Lemma 12. *The $\mathbb{R}^q \mapsto \mathbb{R}$ function*

$$J(\boldsymbol{\eta}) = \frac{1}{\pi}\int_0^1 \frac{x^{\frac{q-1}{2}}}{\sqrt{(1-x)\prod_{i=1}^q (x+\eta_i)}}dx$$

with $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_q)$ is Schur convex.

Proof. By the Schur-Ostrowski criterion (Theorem 3.A.4 in Marshall, Olkin, and Arnold (2011)), J is Schur convex if (and only if)

$$(\eta_i - \eta_j)\left(\frac{\partial J}{\partial \eta_i} - \frac{\partial J}{\partial \eta_j}\right) \geq 0 \text{ for all } 1 \leq i, j \leq q.$$

With $\tilde{J} = (x+\eta_i)^{-1/2}(x+\eta_j)^{-1/2}$, by a direct calculation,

$$(\eta_i - \eta_j)\left(\frac{\partial \tilde{J}}{\partial \eta_i} - \frac{\partial \tilde{J}}{\partial \eta_j}\right) = \frac{(\eta_i - \eta_j)^2}{2(x+\eta_i)^{3/2}(x+\eta_j)^{3/2}} \geq 0$$

so the result follows. □

Lemma 13. *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

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

With $\boldsymbol{\Delta}(p) = \mathbf{G}(p)^{-1/2}(\mathbf{v}_1 - \mathbf{v}_2)$ 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

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

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

Lemma 14. *Let $\mathbf{A}_1 = \int \mathbf{P}^{-1}\mathbf{D}(\mathrm{cv})\mathbf{\Omega}(\theta)\mathbf{P}dF(\theta)$. 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

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

with $\tilde{\mathbf{W}} = (\mathbf{l}, \mathbf{W}/\,\mathrm{cv})$, 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 $\bar{h} : \mathbb{R} \mapsto \mathbb{R}$

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

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

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

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

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

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

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

By Lemma 13, for any $t > 0$,

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

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. $\square$

Proof of Theorem 8: Proceeding as in the proof of Theorem 2, $\mathbb{P}_{\mathbf{\Sigma}_1}(\tau^2(\mathbf{WW}') > \text{cv}^2) = \mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \bar{\eta}_i Z_i^2\right)$ with $\bar{\eta}_i = \lambda_i\left(-\mathbf{A}_1\right)/\lambda_1(\mathbf{A}_1)$. By Lemma 14, $\bar{\eta}_i \geq 0$ for $i = 1, \ldots, q$. For future reference, note that $\mathbb{P}_{\mathbf{\Sigma}_0}(\tau^2(\mathbf{WW}') > \text{cv}^2) = \alpha$ yields

$$\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \eta_i Z_i^2\right) \leq \alpha. \tag{35}$$

for $\eta_i = \lambda_i\left(-\mathbf{A}_0\right)$.

In the following, we write $\mathbf{a} \prec \mathbf{b}$ for two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^q$ to indicate that $\mathbf{b}$ majorizes $\mathbf{a}$, that is, with the elements of a_i and b_i sorted in descending order,

$$\sum_{i=1}^j a_i \leq \sum_{i=1}^j b_i \text{ for all } j = 1, \ldots, q$$

and $\sum_{i=1}^q a_i = \sum_{i=1}^q b_i$. 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)

$$\begin{aligned}
(\lambda_1(-\mathbf{A}_1), \ldots, \lambda_{q+1}(-\mathbf{A}_1)) &\prec (\lambda_1(-\bar{\mathbf{A}}_1), \ldots, \lambda_{q+1}(-\bar{\mathbf{A}}_1)) \\
&\prec \left(\int \lambda_1(-\bar{\mathbf{A}}(\theta))dF(\theta), \ldots,\right. \\
&\qquad \left.\int \lambda_q(-\bar{\mathbf{A}}(\theta))dF(\theta), \int \lambda_{q+1}(-\bar{\mathbf{A}}(\theta))dF(\theta)\right).
\end{aligned} \tag{36}$$

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 (36) further implies

$$\lambda_1(\mathbf{A}_1) \leq \lambda_1(\bar{\mathbf{A}}_1) \leq \int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) \tag{37}$$

so that also

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

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 $(\bar{\eta}_1, \ldots, \bar{\eta}_q) \prec (\tilde{\eta}_1, \ldots, \tilde{\eta}_q)$, so that by (20) and Lemma 12, $\mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \bar{\eta}_i Z_i^2\right) \leq \mathbb{P}\left(Z_0^2 \geq \sum_{i=1}^q \tilde{\eta}_i Z_i^2\right)$.

Now applying (37)

$$\tilde{\eta}_i^* = \int \lambda_i(-\bar{\mathbf{A}}(\theta))dF(\theta) / \int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) \leq \tilde{\eta}_i$$

for $i = 1, \ldots, q-1$, and since from Lemma 14, $\lambda_1(\mathbf{A}_1) \geq 1$, 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

$$\int \lambda_q(-\bar{\mathbf{A}}(\theta))dF(\theta) - \left(\int \lambda_1(\bar{\mathbf{A}}(\theta))dF(\theta) - 1\right) \geq 0. \tag{38}$$

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), Lemma 12, and (35), it now suffices to show that

$$\sum_{i=1}^{j} \tilde{\eta}_{q+1-i}^* \geq \sum_{i=1}^{j} \eta_{q+1-i} \tag{39}$$

for all $1 \leq j \leq q$, and since $\eta_q \geq 0$, this also ensures that (38) holds. Condition (39) may be rewritten as $\sum_{i=1}^j \int \nu_i(\theta)dF(\theta) \geq 0$, and the result follows. ■

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.

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*.

——— (2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141, 1331–1352.

——— (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 A. Norets (2016): "Credibility of Confidence Sets in Nonstandard Econometric Problems," *Econometrica*, 84, 2183–2213.

Müller, U. K., and M. W. Watson (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.

——— (in preparation): "Low-Frequency Analysis of Economic Time Series," in *Handbook of Econometrics*. Elsevier.

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.

Rasmussen, C. E., and C. K. I. Williams (2005): *Gaussian Processes for Machine Learning*. The MIT Press.

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.