Spatial Correlation Robust Inference in Linear Regression and Panel Models*

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

This Draft: January 20, 2022

Abstract

We consider inference about a scalar coefficient in a linear regression with spatially correlated errors. Recent suggestions for more robust inference require stationarity of both regressors and dependent variables for their large sample validity. This rules out many empirically relevant applications, such as difference-in-difference designs. We develop a robustified version of the SCPC method of Müller and Watson (2021) that addresses this challenge. We find that the method has good size properties in a wide range of Monte Carlo designs that are calibrated to real world applications, both in a pure cross sectional setting, but also for spatially correlated panel data. We provide numerically efficient methods for computing the associated spatial-correlation robust test statistics, critical values and confidence intervals.

Keywords: HAC, HAR, Matérn process
JEL: C12, C20

*Müller acknowledges financial support from the National Science Foundation grant SES-191336.

1 Introduction

This paper studies inference about linear regression coefficients estimated from spatially correlated data. In a simple version of the model

$$y_l = x_l\beta + e_l, \tag{1}$$

where all variables are scalars, $\mathbb{E}[e_l|x_l] = 0$ and (y_l, x_l, e_l) are associated with the spatial location $s_l \in \mathbb{R}^d$. Spatial correlation invalidates inference about β using the usual heteroskesdasticity-robust t-statistic. Several spatial-correlation methods based on robustified versions of the t-statistic have been proposed, with the most well-known method proposed in Conley (1999). These approaches estimate $\sigma^2 = \text{Var}[n^{-1/2}\sum_{l=1}^n u_l]$ where $u_l = x_l e_l$ and n denotes the number of observations. In the usual time series model, the locations s_l are equidistant on a line, so $d = 1$, and σ^2 (or its limit as $n \to \infty$) is called the long-run variance of u. In that context, estimators for σ^2 are called HAC or HAR (Heteroskedastic and Autocorrelation Consistent/Robust). Spatial-correlation robust inference generalizes HAC and HAR time series methods to spatial settings.

As in the time series case, it is difficult to devise spatial correlation robust inference that continues to work well in small samples under empirically plausible degrees of spatial dependence; see, for instance, Ibragimov and Müller (2010), Sun and Kim (2012), Bester, Conley, Hansen, and Vogelsang (2016), and Kelly (2019, 2021) for corresponding simulation evidence. This is especially true for approaches, such as Conley (1999), Kelejian and Prucha (2007), and Kim and Sun (2011), that rely on the standard normal critical value for the t-statistic by invoking a consistency argument for the estimator σ^2. A more promising approach is to explicitly account for the sampling variability in the estimator of σ^2, and to correspondingly adjust the critical value, as in Bester, Conley, and Hansen (2011) and Sun and Kim (2012)—these are analogs of the fixed-b approach of Kiefer and Vogelsang (2005) and the projection approach of Müller (2004) to the spatial setting. However, the appropriate adjustment to the critical value is more complicated in the spatial setting: As demonstrated by Müller and Watson (2021), the distribution of the locations in space also matters, even in large samples under weak dependence. In particular, if the distribution of the locations is not uniform, then even under weak dependence, the asymptotic null distribution of spatial projection type t-statistics are not student-t, and fixed-b spatial t-statistics do not have the same asymptotic distribution

as under i.i.d. sampling, invalidating the usual suggestions for the critical value. Müller and Watson (2021) suggest the spatial correlation principal components (SCPC) method that addresses this challenge, and also provides explicit robustness under some empirically relevant forms of strong dependence.

In contrast to methods that rely on consistency of $\hat{\sigma}^2$, fixed-b type approaches (in time and space), including SCPC, rely on the *stationarity* of u_l for their large sample validity.[1] What is more, the methods require that the asymptotic properties of weighted averages of $x_l\hat{e}_l$, with $\hat{e}_l$ the OLS residuals, to behave like the weighted averages of the demeaned values of $u_l = x_l e_l$, which essentially requires (x_l, e_l) to be stationary and weakly dependent. Yet, in many empirical applications, the data may exhibit non-stationarities and/or strong dependence. A simple but important spatial example is when x_l is an indicator for a binary treatment, where treatment is more likely in one region (the 'north') than another region (the 'south'). This paper studies the finite-sample properties of spatial correlation inference procedures when x_l (and/or e_l) follow empirically relevant processes that may be non-stationary or strongly dependent. Our focus is on the SCPC method proposed in Müller and Watson (2021) because of its desirable theoretical properties. We use the performance of a version of the Conley (1999) method as a benchmark.

We set the stage for the analysis in this paper by presenting finite-sample results for four data generation processes (DGPs). Each of the models uses a stationary Gaussian process for e_l, but differ in the DGP for x_l. Specifically, e_l is generated by a Gaussian process with covariance function $\text{Cov}(e_l, e_\ell) = \exp(-c||s_l - s_\ell||)$, where s_l and s_ℓ denote the spatial locations of e_l and e_ℓ, and $c > 0$ is a parameter that governs the strength of the spatial correlation. This process is the spatial analogue of the time series AR(1) model, with larger values of c implying less spatial correlation. This process will be used in many places in the paper, and from now on we use the shorthand $e_l \sim \mathcal{G}_{\exp}(c)$ to denote a Gaussian process with this exponential covariance function. We calibrate the value of c to induce a specific average

[1] A notable exception is the cluster-based method suggested by Ibragimov and Müller (2010), which remains asymptotically valid under many forms of variance heterogeneity. However, small sample simulations show that SCPC performs better under stationarity, and SCPC also avoids the issue of how to form the clusters (cf. Cao, Hansen, Kozbur, and Villacorta (2020)).

Table 1: Rejection frequency of nominal 5% level tests

Model	HR	KERNEL	SCPC
1: $x_l = 1$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	0.51	0.11	0.05
2: $x_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$	0.52	0.14	0.08
3: $x_l \sim$ step function, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	0.52	0.20	0.15
4: $x_l \sim$ random walk, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	0.50	0.14	0.08

Notes: Null rejection frequencies using heteroskedastic robust (HR), spatial kernel (KERNEL), and SCPC inference about β in regression model $y_l = x_l\beta + e_l$ with $n = 250$ and $s_l \sim$ i.i.d. $U(0,1)$.

pairwise correlation, say $\bar{\rho}$, at the sample spatial locations, that is, $c = c_{\bar{\rho}}$ solves

$$\frac{1}{n(n-1)} \sum_{l,\ell \neq l} \exp(-c_{\bar{\rho}}||s_l - s_\ell||) = \bar{\rho}.$$

Thus, if $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$, then the average correlation of e_l and e_ℓ is 0.03 over the sample locations. To put this value in perspective, note that for a discretely sampled AR(1) time series model with $n = 250$, an AR coefficient of 0.79 generates an average pairwise correlation of $\bar{\rho} = 0.03$.

In each of the four models, spatial locations are randomly selected from the unit interval, so $d = 1$ and $s_l \sim$ i.i.d. $U(0,1)$. Thus, this can be viewed as a time series setting with irregular (random) sampling. We consider more interesting and empirically relevant spatial designs in Section 3, but this simple design serves as a useful introduction. The sample size is $n = 250$. We construct t-statistics using three estimators for σ. The first uses the standard Eicker-Huber-White heteroskedastic robust estimator (HR) that ignores spatial correlation. The second is a Bartlett kernel estimator (KERNEL)—this is Conley's (1999) spatial generalization of the well-known Newey-West (1987) HAC standard error. The third is the SCPC estimator proposed in Müller and Watson (2021) which is analogous to projection based estimators used in time series regressions (e.g., Müller (2004), Phillips (2005), and Sun (2013)) but tailored to the particular spatial distribution of the data being studied. Standard normal critical values are used for HR and KERNEL; an 'oracle' bandwidth is used in the KERNEL method so the test's rejection frequency is as close as possible to its nominal level. The critical value for SCPC depends on $\bar{\rho}$ and the spatial locations in the sample; details are provided in Müller and Watson (2021) and reviewed in Section 2 below.

Table 2: Quantiles of conditional null rejection frequencies of nominal 5% level SCPC tests

	Quantile				
Model	0.05	0.25	0.50	0.75	0.95
1: $x_l = 1$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	0.05	0.05	0.05	0.05	0.05
2: $x_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$	0.05	0.06	0.07	0.09	0.11
3: $x_l \sim$ step function, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	0.11	0.13	0.14	0.17	0.21
4: $x_l \sim$ random walk, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	0.05	0.06	0.08	0.09	0.12

Notes: Quantiles of conditional null rejection frequencies of SCPC given $\{x_l, s_l\}_{l=1}^n$ for inference on β in regression model $y_l = x_l\beta + e_l$ with $n = 250$ and $s_l \sim$ i.i.d. $U(0,1)$.

Table 1 shows null rejection frequencies for tests with 5% nominal level for each of the four models and three methods. In the first model, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$ and $x_l = 1$. Here, and for the other models as well, HR exhibits a large size distortion, which is expected given the spatial correlation in the data. The Kernel method has a rejection frequency of 11%, which is a substantial improvement over HR, but is still far from its 5% nominal value, even though the best possible bandwidth choice was made. The SCPC method is designed to have small sample validity in this setting, so the null rejection probability is exactly 5%.

In Model 2, x_l and e_l are independent and follow a $\mathcal{G}_{\exp}(c_{0.03}/2)$ processes. Here $u_l = x_l e_l$ is non-Gaussian but with the same covariance function as $\mathcal{G}_{\exp}(c_{0.03})$ process. Yet the rejection frequencies of both Kernel and SCPC increase by 3%, which is primarily driven by the difference between $x_l\hat{e}_l$ and the demeaned version of $x_l e_l$.

In Models 3 and 4, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$ and x_l is non-stationary. In Model 3, x_l is a zero-mean step function with $x_l = -0.15$ for the 85% of the locations closest to $s = 0$ and $x_l = 0.85$ for the remaining locations closest to $s = 1$. This induces a further size distortion in both Kernel and SCPC. In the final DGP, x_l follows a demeaned random walk, which results in rejection frequencies similar to Model 2.

Table 2, which focuses on SCPC, presents a more nuanced view of the rejection frequencies in Models 2-4 by summarizing rejection frequencies conditional on the values of x_l and s_l in each experiment. That is, in these experiments, (x_l, s_l) are sampled as described above, and rejection frequencies are computed over repeated draws of e_l. This process is repeated for many random draws of (x_l, s_l), and the table reports selected quantiles of the resulting distribution of rejection frequencies. The 95th quantile indicates null rejection frequencies larger than 11% in Models 2-4: a researcher unluckily enough to observe these values of

(x_l, s_l) and using a nominal 5% SCPC test, would in fact be using a test with conditional null rejection probability larger than 11%.

Taken together these experiments suggest that SCPC offers substantial improvements on methods that ignore spatial correlation (and improvements over KERNEL), but may exhibit quantitatively important size distortions for some DGPs. This raises two questions. First, how well does the method perform for the range of DGPs typically encountered in empirical work? And second, are there modifications that enhance its performance in situations where it performs poorly? We take up both questions in this paper.

To answer the first, we begin by augmenting the simple regression model to include additional control variables and allow clustering of observations by location. This setup covers panel data models with fixed effects, difference-in-difference designs, and clustered versions of spatial-correlation-robust standard errors. We model empirically relevant spatial designs by considering the spatial density of economic activity over the continental United States and the location of countries over the globe. We generate the variables using parametric models, but also by sampling variables from the World Development Indicators (WDI) data set from the World Bank. This dataset contains hundreds of variables for many countries over several years, and thus provides a wide range of spatial patterns in both cross-section and panel-data regressions.

To answer the second question, we propose a modification of the SCPC method so that it also controls size, conditional on the regressors, for a particular conditionally heteroskedastic Gaussian model for e_l. The details are described below. While not offering perfect size control in all settings, this modification provides a quantitatively important improvement over SCPC in several of the models we consider.

The outline of the paper is as follows. In Section 3 we outline the spatial regression model and inference methods. Notably, the section presents a method for robustifying SCPC inference to control size after conditioning on the values of the regressors and locations. Section 4 presents experiments using the spatial distribution of economic activity in each of the 48 states making up the continental US and hundreds of variables chosen from the World Bank's WDI dataset. These results expand and reinforce the results shown for the simple experiments in Tables 1 and 2. The results indicate that the conditional-SCPC method developed in Section 2.3 offers improved size control over SCPC, which in turn improves upon KERNEL for a wide range of empirically relevant process for both cross-section and panel-data

regressions with clustered standard errors. Several of these results utilize $\mathcal{G}_{\exp}(c)$ processes for the error term, and Section 4 studies the robustness of the conclusions to other spatial processes. Section 5 takes up three issues. The first involves computing the critical values for the spatial-correlation robust t-statistics. The second involves computing the SCPC-variance estimator when n is large: the estimator uses eigenvectors from a model-based $n \times n$ covariance matrix, and when n is large, computing these eigenvectors is difficult. The third involves using SCPC in IV regressions. The final section offers some concluding remarks.

2 Spatial regression and spatial-correlation robust inference

This section has three purposes. First, it presents the spatial regression model used in our analysis. Second, it provides details for t-statistic inference using the SCPC and KERNEL methods. Finally, it proposes a robustification of the SCPC critical value so that the method controls size, conditional on the regressors and locations, for a particular heteroskedastic model that we describe.

2.1 Spatial regression model

The spatial regression model is

$$y_{i,l} = x_{i,l}\beta + z_{i,l}'\gamma + e_{i,l} \tag{2}$$

where $l = 1, ..., n$ indexes *clusters* (or *groups*, *entities*, etc.) and $i = 1, ..., m_l$ identifies the m_l individual observations in cluster l. The total number of observations is $N = \sum_{l=1}^{n} m_l$. The model with $m_l = 1$ for all l is the cross-section spatial regression model, and $m_l > 1$ allows for clustering and panel data. Cluster l is associated with location $s_l \in \mathbb{R}^d$. The variable $x_{i,l}$ is a scalar, and β is the parameter of interest. The $k \times 1$ vector $z_{i,l}$ allows for additional regressors that may include fixed-effect indicators, and $e_{i,l}$ is the regression error. We assume (without loss of generality) that the regressors have been transformed so that $x_{i,l}$ and $z_{i,l}$ are uncorrelated in the sample, that is $\sum_{i,l} x_{i,l} z_{i,l} = 0$.

The following notation will prove useful. The $m_l \times 1$ vector $x_l = (x_{1,l}, x_{2,l}, ..., x_{m_l,l})'$ collects

the x-variables for cluster l, and the $N \times 1$ vector $\mathbf{X} = (x_1', x_2', ..., x_n')'$ collects the x-variables in the sample. The matrices and vectors z_l, e_l, $\mathbf{Z}$, $\mathbf{e}$ as well as the $n \times 1$ vector of locations $\mathbf{s}$ are defined analogously. We denote sample second-moments as $S_{xx} = n^{-1} \sum_{l=1}^{n} x_l' x_l$ and $S_{xy} = n^{-1} \sum_{l=1}^{n} x_l' y_l$, where dividing by n instead of N simplifies the formulae below. Because x and z are uncorrelated in the sample, $\mathbf{X'Z} = 0$. Using this notation, the OLS estimator for β is $\hat{\beta} = S_{xx}^{-1} S_{xy} = \beta + S_{xx}^{-1} n^{-1} \sum_l u_l$, where $u_l = x_l' e_l$. The OLS residuals are $\hat{e}_{i,l}$, and $\hat{u}_l = x_l' \hat{e}_l$. These are collected in the vectors $\mathbf{u}$, $\hat{\mathbf{e}}$ and $\hat{\mathbf{u}}$.

Inference about β is based on the t-statistic

$$\tau = \frac{S_{xx} \sqrt{n} (\hat{\beta} - \beta_0)}{\hat{\sigma}} \tag{3}$$

where $\hat{\sigma}^2$ is an estimator of $\sigma^2 = \text{Var}\left(n^{-1/2} \sum_{l=1}^{n} u_l\right)$ and β_0 is the null value of β.

2.2 Two spatial correlation robust inference methods

The results summarized in Tables 1 and 2 rely on two inference methods. The KERNEL method estimates σ^2 by

$$\hat{\sigma}_K^2 = \frac{1}{n} \sum_{l,\ell} K\left(\frac{||s_l - s_\ell||}{b}\right) \hat{u}_l \hat{u}_\ell \tag{4}$$

where K is the Bartlett weighting function, $K(x) = 1 - |x|$ for $|x| \leq 1$ and $K(x) = 0$ otherwise, and $b > 0$ is a bandwidth parameter. Empirical researchers typically use standard normal critical values for the resulting t-statistic, relying on $b \to 0$ as $n \to \infty$ arguments in Conley (1999). We implement the 'oracle' version of this method that combines the standard normal critical value with the value of b that minimizes the small sample size distortion in each experiment. This choice of b is infeasible in practice, but yields a useful lower bound on size distortions associated with the method.[2]

The SCPC method of Müller and Watson (2021) is based on a principal component estimator of σ^2 based on a pre-specified 'worst-case' exponential covariance function conditional on the observed locations. The method exploits the large sample equivalence between inference

[2]As specified in (4), $\hat{\sigma}_K^2 \geq 0$ is not guaranteed when $d > 1$. In the case of $d = 2$, Conley (1999) suggests using the product of two univariate Bartlett kernels based on distance in each of the two dimensions. Such an approach produces a positive semi-definite estimator of σ^2, but is not invariant to rotation of the axes for the spatial locations. In our calculations shown below, we use the spectral decomposition of the Bartlett weights in (4) with negative eigenvalues set to zero.

about β in (2) and inference about the mean of $y_l^o = \{nS_{xx}^{-1}\hat{u}_l + \hat{\beta}\}_{l=1}^n$, which holds as long as (u_l, x_l) is stationary and weakly dependent. Suppose that $u_l^o = y_l^0 - \beta \sim \mathcal{G}_{\exp}(c)$, and let $\mathbf{\Sigma}(c)$ be the corresponding $n \times n$ covariance matrix of $\mathbf{u}^o$ evaluated at the sample locations $\mathbf{s}$, so that the l, ℓ th element of $\mathbf{\Sigma}(c)$ is $\Sigma_{l,\ell}(c) = \exp(-c||s_l - s_\ell||)$. Let $\bar{\rho}(c)$ denote the resulting average pairwise correlation

$$\bar{\rho}(c) = \frac{1}{n(n-1)} \sum_l \sum_{\ell \neq l} \Sigma_{l,\ell}(c). \tag{5}$$

Suppose a researcher desires a test that controls size for values $\bar{\rho}$ that may be as large as $\bar{\rho} = \bar{\rho}_{\max}$, for a given value of $\bar{\rho}_{\max}$. The results summarized in Tables 1 and 2 are based on $\bar{\rho}_{\max} = 0.03$, for instance. Let $c_{\min}$ satisfy $\bar{\rho}(c_{\min}) = \bar{\rho}_{\max}$, where the notation emphasizes that $\bar{\rho}(c)$ is a decreasing function of c. Note that $\mathbf{\Sigma}(c_{\min})$ is the associated worst case covariance matrix for $\mathbf{u}^o$ in the sense that it induces the largest value of σ^2 among all $\mathbf{\Sigma}(c)$ with $\bar{\rho} \leq \bar{\rho}_{\max}$. Let $\mathbf{1}$ denote a $n \times 1$ vector of 1s and $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$. The matrix $\mathbf{M}_1\mathbf{\Sigma}(c_{\min})\mathbf{M}_1$ denotes the associated covariance matrix for $\hat{\mathbf{u}}^o = \mathbf{u}^o - \bar{u}^o\mathbf{1}$, where $\bar{u}^o = n^{-1}\sum_{l=1}^n u_l^o$. The SCPC estimator of σ^2 uses weighted averages of $\hat{u}_l^o$ with weights constructed from the eigenvectors of $\mathbf{M}_1\mathbf{\Sigma}(c_{\min})\mathbf{M}_1$. These are the principal components of $\hat{\mathbf{u}}^o$ under $\mathcal{G}_{\exp}(c_{\min})$ evaluated at the observed locations $\mathbf{s}$. Let $\mathbf{r}_j$ denote the eigenvector of $\mathbf{M}_1\mathbf{\Sigma}(c_{\min})\mathbf{M}_1$ corresponding to the jth largest eigenvalue, normalized so that $\mathbf{r}_j'\mathbf{r}_j/n = 1$ – see Figure 6 in Section 5.2 below for an example of these eigenvectors. The SCPC estimator of σ^2, based the first q principal components, is

$$\hat{\sigma}^2_{\text{SCPC}}(q) = \frac{1}{q}\sum_{j=1}^{q}(n^{-1/2}\mathbf{r}_j'\hat{\mathbf{u}}^o)^2. \tag{6}$$

The critical value for the SCPC t-statistic is chosen so that the t-test controls size for all values of $c \geq c_{\min}$ (equivalently, $\bar{\rho} \leq \bar{\rho}_{\max}$). Denoting the SCPC t-statistic and critical values by τ_{SCPC} and cv_{SCPC}, the critical value solves

$$\sup_{c \geq c_{\min}} \mathbb{P}^{\text{Normal}}_{\mathbf{\Sigma}(c)}(|\tau_{\text{SCPC}}| > \text{cv}_{\text{SCPC}}) = \alpha \tag{7}$$

where α is the desired level of the test, and the notation emphasizes that the probability is computed under $\mathbf{u}^o \sim \mathcal{N}(0, \mathbf{\Sigma}(c))$. The final ingredient in the method is the choice of q, the number of principal components used to construct $\hat{\sigma}_{\text{SCPC}}$. This is chosen to minimize the

expected length of the associated confidence interval under the i.i.d. model with $\mathbf{\Sigma} = \mathbf{I}_n$.

Since $\sum_{l=1}^{n} \hat{u}_l = 0$, the numerator of the SCPC t-statistic is $\bar{u}^o = \hat{\beta}$. Furthermore, the terms $\mathbf{r}_j'\hat{\mathbf{u}}^o$ in (6) simplify to $\mathbf{r}_j'\hat{\mathbf{u}}^o = nS_{xx}^{-1}\mathbf{r}_j'\hat{\mathbf{u}}$ because $\mathbf{r}_j'\mathbf{1} = 0$. Thus, when applied to the general regression model, the variance estimator of the SCPC method is a quadratic form in $\hat{u}_l$, just like the KERNEL method.

As discussed in Müller and Watson (2021) the SCPC method has several desirable properties. We list four. First, as already noted, size is controlled in the Gaussian model for $u_l^o \sim \mathcal{G}_{\exp}(c)$ for any value of $c \geq c_{\min}$. Second, in large samples, size control is not limited to Gaussian $\mathcal{G}_{\exp}(c)$ settings, but holds more generally in covariance stationary models with weak dependence. Third, in finite-sample settings, size control extends to arbitrary mixtures of Gaussian models that satisfy certain restrictions. Fourth, SCPC confidence intervals enjoy a near optimality property in terms of their expected length among confidence intervals that control size under stationary models whose spectrum are weakly less steep than that of $\mathcal{G}_{\exp}(c_{\min})$.

These results suggest that SCPC inference will perform well in an important class of spatial regression models, at least under weak dependence. That said, the simulations reported in the introduction suggest non-negligible size distortions in finite-sample settings with stationary, but spatially persistent data (Model 2), and more concerning size distortions when x_l is non-stationary (Model 3). As it turns out, these size distortions can be eliminated or mitigated by using an alternative critical value chosen to control size after conditioning on the regressors. We outline that modification in the next subsection.

2.3 Conditional SCPC inference

The required modification is straightforward: instead of computing the critical value for τ_{SCPC} so that size is controlled under $\mathbf{u} \sim \mathcal{N}(0, \mathbf{\Sigma}(c))$, $c \geq c_{\min}$, we additionally impose that size is also controlled under a set of conditional distributions of $\mathbf{u}$ (and $\hat{\mathbf{u}}$) given $\mathbf{V} = (\mathbf{X}, \mathbf{Z})$ (and in both cases, also conditional on the locations $\mathbf{s}$).[3] This amounts to specifying the distribution of $\mathbf{e}$ given $\mathbf{V}$, which we assume to be Gaussian, $\mathbf{e}|\mathbf{V} \sim \mathcal{N}(0, \mathbf{\Sigma}_{\mathbf{e}|\mathbf{V}})$. To motivate our parameterization of $\mathbf{\Sigma}_{\mathbf{e}|\mathbf{V}}$, note that randomness in $\hat{\beta}$ is driven by the random variable

[3]Similarly, Hansen (2021) suggests the use of critical values for heteroskedasticity robust tests that are small sample valid under i.i.d. Gaussian errors.

$n^{-1/2}\sum_l u_l$, with conditional variance equal to

$$\text{Var}\left(n^{-1/2}\sum_l u_l|\mathbf{V}\right) = n^{-1}\sum_{l,\ell} x_l'\mathbb{E}[e_l e_\ell'|\mathbf{V}]x_\ell. \tag{8}$$

This highlights two distinct issues: first, how to parameterize the covariance of e between clusters ($l \neq \ell$), and second, how to parameterize within-cluster covariation.

To focus on the first issue, consider the cross-section regression, so that x_l and e_l are scalars and between-cluster covariation is the only relevant issue. A straightforward modification of the SCPC formulation assumes that $e_l|\mathbf{V} \sim \mathcal{G}_{\text{exp}}(c)$, so that e_l is independent of $\mathbf{V}$. Examination of (8) shows that, using this formulation, realizations with $x_l x_\ell < 0$ will result in a lower value of $\text{Var}\left(n^{-1/2}\sum_l u_l|\mathbf{V}\right)$ than realizations with x's of the same magnitude with $x_l x_\ell > 0$. This formulation lacks robustness in this respect. A more robust formulation sets $e_l = \text{sign}(x_l)\cdot a_l$, where $a_l|\mathbf{V} \sim \mathcal{G}_{\text{exp}}(c)$; this yields $\text{Var}\left(n^{-1/2}\sum_l u_l|\mathbf{V}\right) = n^{-1}\sum_l |x_l x_\ell| \exp(-c||s_l - s_\ell||)$.

We extend this to the panel-data regression model using

$$e_l = x_l^s a_l, \tag{9}$$

where $x_l^s = x_l/(x_l'x_l)^{1/2}$ and a_l is scalar with $a_l \sim \mathcal{G}_{\text{exp}}(c)$. This yields $\text{Var}\left(n^{-1/2}\sum_l u_l|\mathbf{V}\right) = \sum_{l,\ell}(x_l'x_l)^{1/2}(x_\ell'x_\ell)^{1/2}\exp(-c||s_l - s_\ell||)$. Note that this expression for the variance reduces to that of the cross-section regression when $m_l = 1$ for all l.

This formulation has several desirable features. First, it is a tractable extension of SCPC, leading to computations that are no harder than those of the baseline SCPC method irrespective of the cluster sizes m_l (we discuss computational issues in Section 5.1). Second, as $c \to \infty$, (8) coincides with what is obtained from the i.i.d. model where $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}_N)$. And finally, for panel data, the conditionally singular within-cluster model for e_l in (9) yields the largest variance of u_l among all models with $\text{tr}[\text{Var}(e_l|\mathbf{V})] = 1$, and in this sense maximizes the effect of within-cluster covariation on the variance of $\hat{\beta}$.

We use (9) to modify the critical value for SCPC as follows. First, we compute cv_{SCPC} as in the last subsection. We then also compute a new critical value, say $\text{cv}_{\mathbf{V}}$ that satisfies

$$\sup_{c \geq c_{\min}} \mathbb{P}^{\text{Normal}}_{\mathbf{\Sigma}(c)}(|\tau_{\text{SCPC}}| > \text{cv}_{\mathbf{V}}\,|\mathbf{V}) = \alpha \tag{10}$$

Table 3: Distribution of conditional null rejection frequencies of nominal 5% C-SCPC tests

Model	$c_{\min}$	Avg	Quantile 0.05	0.25	0.50	0.75	0.95
$x_l = 1$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	$c_{0.03}$	0.05	0.05	0.05	0.05	0.05	0.05
$x_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$	$c_{0.03}/2$	0.04	0.03	0.03	0.04	0.04	0.05
$x_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$	$c_{0.03}$	0.06	0.04	0.05	0.06	0.06	0.07
$x_l \sim$ step function, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	$c_{0.03}$	0.05	0.04	0.04	0.05	0.05	0.05
$x_l \sim$ random walk, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$	$c_{0.03}$	0.05	0.05	0.05	0.05	0.05	0.05

Notes: Quantiles of conditional null rejection frequencies of C-SCPC given $\{x_l, s_l\}_{l=1}^n$ for inference on β in regression model $y_l = x_l\beta + e_l$ with $n = 250$, $s_l \sim$ i.i.d. $U(0,1)$. $c_{\min}$ denotes the lower bound on c used in the construction of the C-SCPC critical value.

where the probability is computed under $e_l = x_l^s a_l$ with $a_l|\mathbf{V} \sim \mathcal{G}_{\exp}(c)$. The conditionally-robust critical value for SCPC is then the larger of cv_{SCPC} and $\text{cv}_{\mathbf{V}}$,

$$\text{cv}_{\text{C-SCPC}} = \max(\text{cv}_{\text{SCPC}}, \text{cv}_{\mathbf{V}}).$$

Section 5.1 provides details for computing $\text{cv}_{\mathbf{V}}$ and $\text{cv}_{\text{C-SCPC}}$.

In what follows we will investigate rejection frequencies of τ_{SCPC} using cv_{SCPC} and using $\text{cv}_{\text{C-SCPC}}$. We will refer to these methods as SCPC, and C-SCPC, respectively.

Table 3 shows the performance of C-SCPC for the four models considered in Tables 1 and 2, which reported results for SCPC. Recall that in Model 2, $x_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$ and $e_l \sim \mathcal{G}_{\exp}(c_{0.03}/2)$ so that $u_l = x_l e_l$ had a exponential covariance function with parameter $c_{0.03}$. Table 3 shows the rejection for Model 2 with $\text{cv}_{\text{C-SCPC}}$ computed using $c_{0.03}/2$ (which is the DGP of e_l) and using $c_{0.03}$ (which is the correlation of u_l). In both cases, and in the other models, C-SCPC offers improved size control, across the range of x_l realizations. Note that in the models in Table 3, $e_l \sim \mathcal{G}_{\exp}(c)$, which is different than the model used to compute $\text{cv}_{\text{C-SCPC}}$, namely (9). This provides a hint about the robustness properties of C-SCPC over alternative DGPs, something that we investigate more fully in following sections.

We find the results summarized in Tables 1-3, both enlightening and encouraging. That said, they are based on what are arguably contrived designs. The next section examines the performance of the methods under more empirically relevant designs.

3 Finite sample properties of spatial-correlation robust t-statistics

The various t-statistics discussed above have probability distributions that depend on two distinct features of the population under study: (i) the spatial process that generates $(y(s), x(s), z(s))$ for arbitrary locations $s \in \mathbb{R}^d$, and (ii) the probability distribution that governs which locations are sampled, that is the spatial distribution of s. This section reports results from a variety of experiments in which we choose both features to mimic empirically plausible settings.

In particular, we consider two sets of spatial distributions. The first involves drawing locations from each of the 48 continental US states where the spatial density is proportional to economic activity within the state, and where economic activity is proxied by the intensity of light measured from space. These experiments yield 48 distinct spatial densities with different support (the boundaries for the states) and shapes (the concentration of economic activity within the state), and was used previously in Müller and Watson (2021). The second set of experiments uses data from countries scattered over the globe.

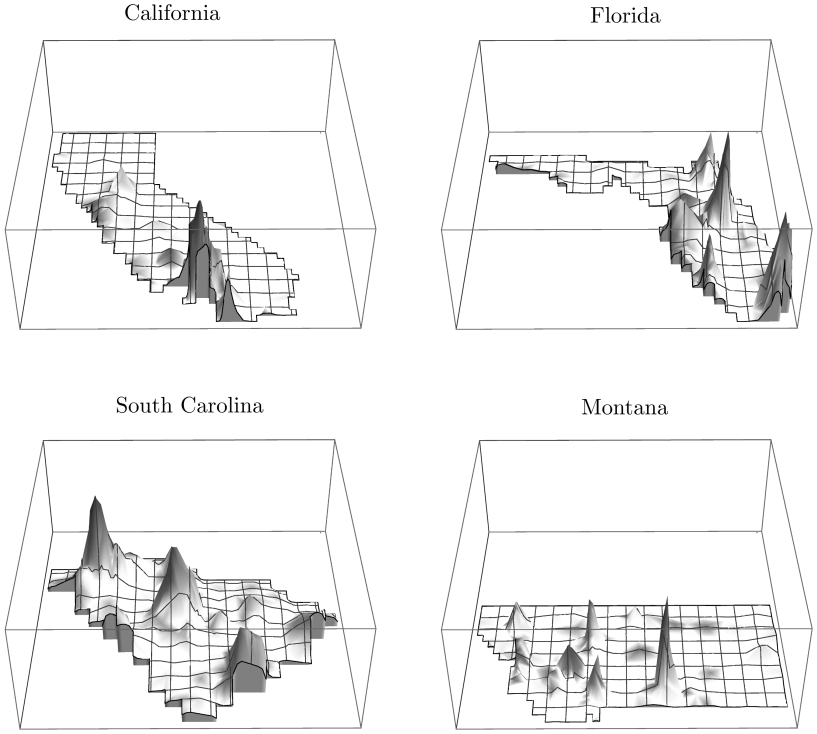
We use two methods for generating realizations of $(y(s), x(s), z(s))$. The first, used in the US states designs, generates data from parametric models like those used in the experiments already discussed. The second, used in the country designs, uses actual data sampled from the World Bank's World Data Indicators (WDI) dataset.

3.1 48 US states

The first set of experiments use data generated from designs that capture the spatial distribution of economic activity in the 48 continental United States. Economic activity is proxied by light intensity which is estimated using the fine grid of light measurements reported in Henderson, Squires, Storeygard, and Weil (2018). States differ by their shape and spatial concentration of economic activity, and Figure 1 shows four examples. Looking across 48 states allows us to study the behavior of the test statistics under a wide range of empirically-relevant spatial distributions.

In the 48-states experiments, realizations of $(y(s), x(s), z(s))$ are generated from parametric models that focus on stylized characteristics of data used in empirical work. We begin by

Figure 1: Spatial densities for four U.S. states



Notes: These are densities of light, as measured from space, estimated from data provided in Henderson, Squires, Storeygard, and Weil (2018).

discussing experiments for cross-section regressions ($m_l = 1$ in equation (2)) and then discuss panel regressions. Table 4 summarizes the experiments for cross-section regressions. The first experiment is a benchmark with $x_l = 1$, $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$ and no additional control variables. In Experiments 2-6, x_l and e_l are generated by independent $\mathcal{G}_{\exp}(c)$ models, where the experiments differ in the values of c and whether additional control variables are included. In Experiments 7-10, x_l is generated by a step function with a shift between southern and northern locations; these regressors are stylized versions of spatially correlated binary treatments. In Experiments 11 and 12, x_l is generated by a $\mathcal{G}_{\exp}(c)$ model with spatial heteroskedasticity, and in 13 and 14, x_l follows a spatial random walk. In all experiments reported in this paper, the x_l regressor is projected off the control regressors to ensure $\sum_{l=1}^{n} x_l' z_l = 0$. We present results for $e_l \sim \mathcal{G}_{\exp}(c)$ and $e_l = \text{sign}(x_l) \cdot a_l$ with $a_l \sim \mathcal{G}_{\exp}(c)$, the latter model is used to compute the critical value of C-SCPC. All simulations use $n = 250$. The rejection frequency of nominal 5% level tests is computed for each draw of $\{x_l, z_l, s_l\}_{l=1}^{n}$, which yields a distribution

Table 4: Spatial regression designs using U.S. states spatial distributions of locations

Design	a_l	x_l	z_l
1	$\mathcal{G}_{\exp}(c_{0.03})$	1	none
2	$\mathcal{G}_{\exp}(c_{0.03})$	$\mathcal{G}_{\exp}(c_{0.03})$	1
3	$\mathcal{G}_{\exp}(c_{0.03})$	$\mathcal{G}_{\exp}(c_{0.03})$	none
4	$\mathcal{G}_{\exp}(c_{0.03})$	$\mathcal{G}_{\exp}(c_{0.03})$	$[1, \mathcal{G}_{\exp}(c_{0.03})^3]$
5	$\mathcal{G}_{\exp}(c_{0.03})$	iid$\mathcal{N}(0,1)$	1
6	$\mathcal{G}_{\exp}(c_{0.03})$	iid$\mathcal{N}(0,1)$	$[1, \mathcal{G}_{\exp}(c_{0.03})^3]$
7	$\mathcal{G}_{\exp}(c_{0.03})$	Step(0.50)	1
8	$\mathcal{G}_{\exp}(c_{0.03})$	Step(0.15)	1
9	$\mathcal{G}_{\exp}(c_{0.03})$	Step(0.50)	$[1, \mathcal{G}_{\exp}(c_{0.03})^3]$
10	$\mathcal{G}_{\exp}(c_{0.03})$	Step(0.15)	$[1, \mathcal{G}_{\exp}(c_{0.03})^3]$
11	$\mathcal{G}_{\exp}(c_{0.03})$	$(1+w_l)v_l$, $w_l \sim$ Step(0.5), $v_l \sim \mathcal{G}_{\exp}(c_{0.03})$	1
12	$\mathcal{G}_{\exp}(c_{0.03})$	$(1+w_l)v_l$, $w_l \sim$ Step(0.15), $v_l \sim \mathcal{G}_{\exp}(c_{0.03})$	1
13	$\mathcal{G}_{\exp}(c_{0.03})$	random walk	1
14	$\mathcal{G}_{\exp}(c_{0.03})$	random walk	$[1, \mathcal{G}_{\exp}(c_{0.03})^3]$

Notes: The regression model is $y_l = x_l\beta + z_l'\gamma + e_l$ with $e_l = a_l$ or $e_l = \text{sign}(x_l) \cdot a_l$, with $s_l \sim$ i.i.d. from a U.S. state-specific density with examples given in Figure 1. All $\mathcal{G}_{\exp}$ processes are mutually independent. In Designs 7-12, Step(λ) is a step function that is equal to 0 for the southern most $\lfloor \lambda n \rfloor$ observations and 1 otherwise. Random walk x_l are approximated by $\mathcal{G}_{\exp}(c)$ for a small c. $\mathcal{G}_{\exp}(c)^3$ in the z_l column denotes 3 independent realizations of a $\mathcal{G}_{\exp}$ process. All models use $n = 250$.

of rejection frequencies for each experiment.

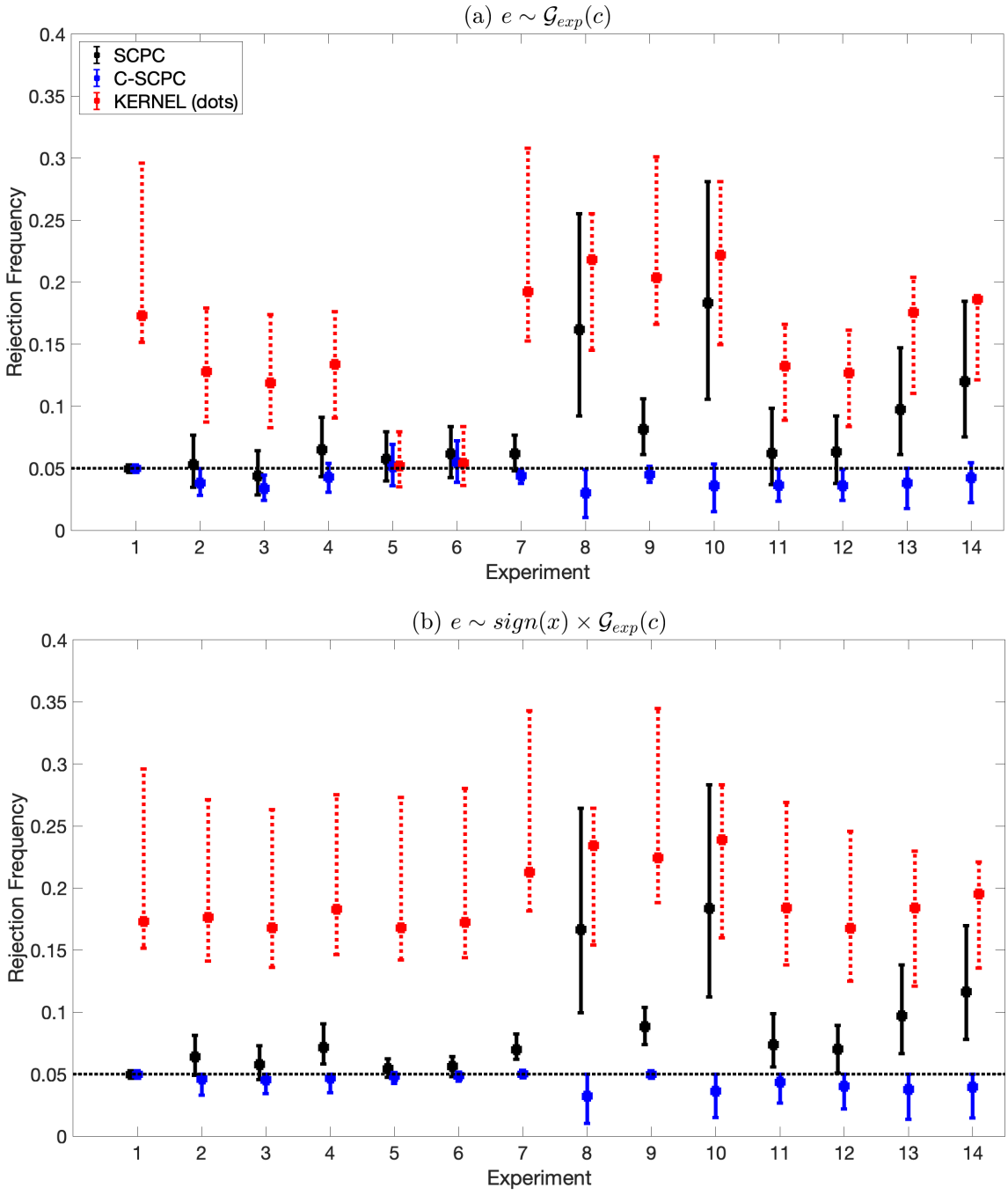
Figure 2 shows the 5th and 95th quantiles, and the mean of these conditional null rejection probabilities. Results are shown for Kernel, SCPC and C-SCPC. The SCPC critical value is computed using $c_{0.03}$ in all experiments.[4] The critical value of C-SCPC is computed using the value of c that generated e_l. Figure 2 does not show results for the usual heteroskedasticity-robust version of the t-statistic that ignores spatial correlation; suffice it to say the resulting test yields null rejection frequencies that greatly exceed its nominal 5% level.

The results for this rich set of spatial designs is similar to what we found earlier in the simple $U(0,1)$ spatial design. We highlight four results.

First, despite the optimal choice of bandwidth, the Kernel method suffers from substantial size distortions. This is consistent with a large body of research that finds similar

[4]In Experiments 2-6, x and e follow independent $\mathcal{G}_{\exp}(c)$ processes so that $u_l = x_l e_l$ has the $\mathcal{G}_{\exp}(2c)$ covariance function. Thus, the c-value used to compute cv is smaller than the c-value used to generate for u in these experiments. A smaller critical value (and larger rejection frequency) results when $2c$ is used to compute cv.

Figure 2: Rejection frequencies for 5% nominal tests: spatial regressions, 48-states experiments



Notes: The bars show the 5th through 95th quantiles of the distribution of rejection frequencies conditional on the regressors for nominal 5% tests. Squares denote the mean of the distribution, which is the unconditional rejection frequency. See Table 4 and the text for a description of the experiments.

size distortions in time series regressions using HAC standard errors together with standard normal critical values.[5]

Second, SCPC does a reasonably good job controlling size in several designs, but has uncomfortably large size distortions in some cases. For example, the mean null rejection frequency of SCPC is over 0.15 when x_l is generated by a step function in Experiments 8 and 10.

The third result is that C-SCPC has much improved size control, both for models for which the critical value was designed (that is when $e_l = \text{sign}(x_l)a_l$ with $a_l|\mathbf{V} \sim \mathcal{G}_{\text{exp}}(c)$ as in panel (b)) but also when $e_l|\mathbf{V} \sim \mathcal{G}_{\text{exp}}(c)$ (panel (a)); this is consistent with the results from the $s_l \sim$ i.i.d. $U(0,1)$ designs previously shown in Table 3. One might wonder about the rejection frequencies below 5% evident in panel (b) for some realizations of $\{(x_l, z_l, s_l)\}_{l=1}^n$. These arise when the size constraint in (10) is binding for a value of c other than $c_{\min}$, reflecting the fact that size is also controlled for processes that are *less* spatially correlated than the DGP used in the experiment.

Finally, the fourth result is that the rejection frequencies for Kernel and SCPC are somewhat larger in panel (b) than in panel (a). This reflects the increase in spatial correlation that motived the construction of the C-SCPC critical value, that is replacing $x_l x_\ell$ (which is sometimes negative) with $|x_l x_\ell|$.

Figure 3 summarizes results from a panel version of these designs with $m_l = 4$ observations in each of $n = 250$ clusters. The panel data models differ from their non-panel counterparts in three ways. First, these designs rely on a cluster version of the $\mathcal{G}_{\text{exp}}(c)$ model, which we now describe. Let w_l denote an $m \times 1$ vector of variables associated with location s_l; in our applications w represents e, x, or one column of z. Within-cluster covariance is generated by an AR(1) structure with AR parameter ϕ, and between-cluster covariation generated by the exponential covariance function used above. This results in the covariance function $\text{cov}(w_{i,l}w_{j,\ell}) = \phi^{|i-j|}\exp(-c||s_l - s_\ell||)$. The results shown in panel (a) of Figure 3 use data generated as described in Table 4 with $\phi = 0.9$ for within-cluster covariation and the values of c shown in the table for between-cluster covariation (unreported results show little sensitivity to the choice of ϕ). Results in panel (b) use the same regressors but with e_l generated by the model (9) used to compute $\text{cv}_{\text{C-SCPC}}$, that is $e_l = x_l^s a_l$ where $a_l \sim \mathcal{G}_{\text{exp}}(c)$ is a scalar process.

[5]These size distortions were documented in an early contribution by den Haan and Levin (1997). Lazarus, Lewis, Stock, and Watson (2018) provide a recent set of simulations and references.

The second difference is that cluster-specific fixed effects are included as controls in all models; this eliminates Experiment 1 in which $x_l = 1$. The third difference concerns the step-function regressor designs in Experiments 8-12. In the panel data models these step functions are used to generate the regressors for the final two observations in each cluster ($i = 3$ and 4); observations for $i = 1$ and 2 are set to zero. This produces a differences-in-differences design where treatment occurs in the north during the final two time periods and the observations in the south are untreated throughout.

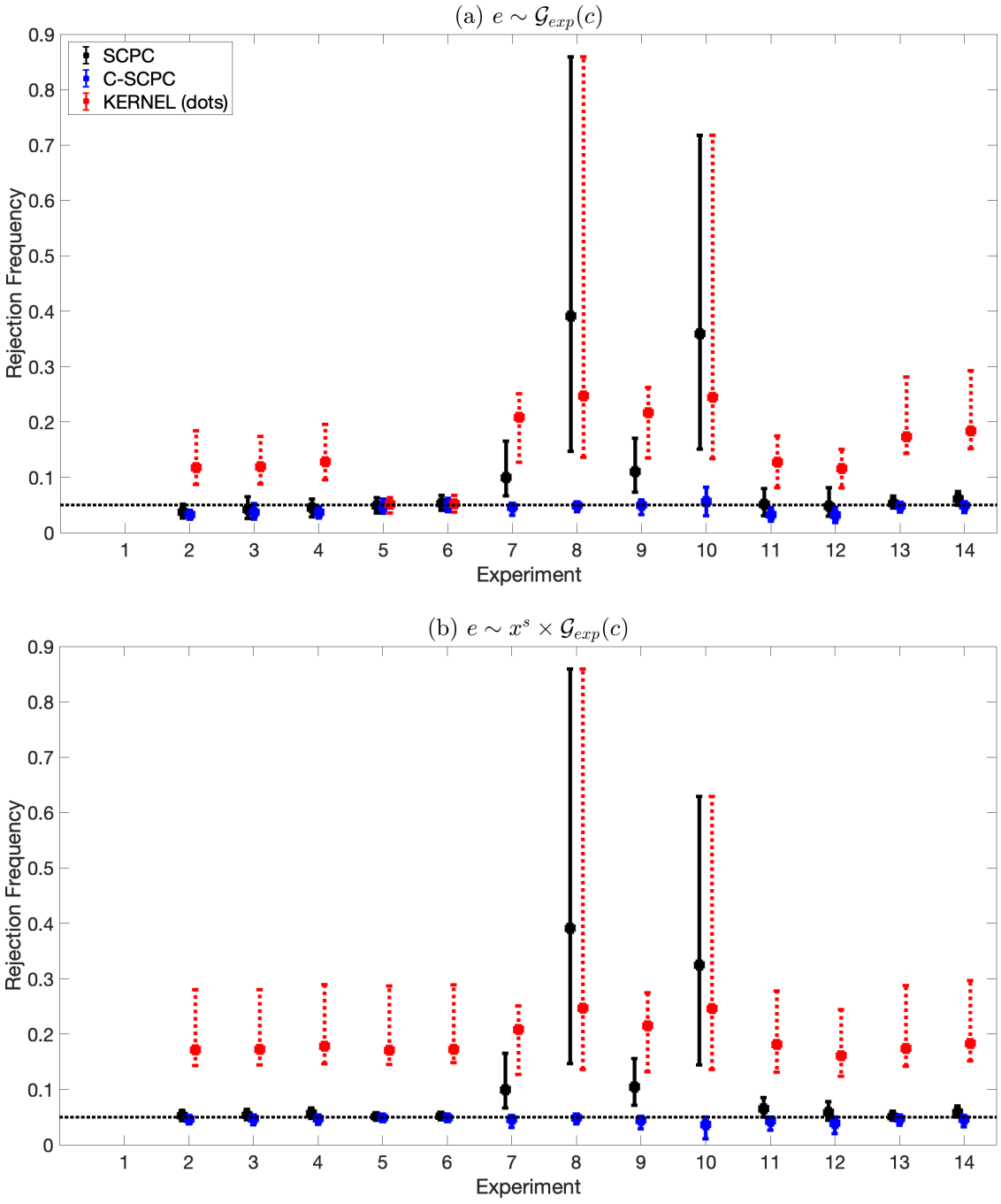
With these differences, the panel data results in Figure 3 look a lot like its non-panel counterpart in Figure 2. The most obvious difference are the more pronounced size distortions for KERNEL and SCPC in Experiments 8 and 10. These arise because the differences-in-differences design results in a much smaller effective sample size because only 15% of the clusters are treated.

One might wonder how the power of the different tests compare. Since the SCPC and C-SCPC methods are based on the same t-statistic τ_{SCPC}, and only differ in their critical values, the size-adjusted power (or, equivalently, average length of confidence intervals) of SCPC and C-SCPC are identical. The critical value of the C-SCPC method not only depends on the observed locations s_l, but also on the observed regressors (x_l, z_l), and is "conditional" in this sense. The cost of insisting on such conditional size control is small: An unreported comparison of the C-SCPC method in the design (9) used to compute $\text{cv}_{\text{C-SCPC}}$ with an unconditionally (over (s_l, x_l, z_l)) size-adjusted version of SCPC reveals small differences in average confidence interval lengths. Finally, comparing SCPC (or C-SCPC) with the size-adjusted KERNEL method reveals the KERNEL method to be somewhat more efficient. This is because $\hat{\sigma}^2_K$ is relatively less variable than $\hat{\sigma}^2_{\text{SCPC}}$, so after the size-adjustment that adjusts for the larger bias in $\hat{\sigma}^2_K$, the KERNEL method is akin to the oracle t-test that uses the true value of σ^2 in the denominator. In practice, of course, this size adjustment is not feasible, so we do not interpret this as a reason to prefer the KERNEL method.

3.2 World Development Indicators

In this section we report results from experiments using data that are drawn from the World Bank's World Development Indicators (WDI) database. Using these series as regressors and/or error terms allows us to investigate test performance in settings with data that closely

Figure 3: Rejection frequencies for 5% nominal tests: spatial panel regressions, 48-states experiments



Notes: See notes to Figure 2. The inclusion of fixed effects eliminates Experiment 1 from the spatial panel regressions.

matches empirical applications. We begin with a short description of the dataset, and then describe the experiments.

The WDI is a panel data set containing over 1400 socioeconomic variables for 266 countries. The dataset includes typical economic measurements like GDP, balance of payments and national debt, but also measurements of education, health and infrastructure. For our purposes, the data present three challenges: they are measured in a variety of units, some variables have missing values for some years and countries, and many series contain large outliers. To address these challenges, we use logarithms for several of the variables, discard series that cover fewer than 100 countries, and treat large outliers as missing data. Appendix A describes these steps in detail. We construct two datasets from the WDI. The first contains decadal differences for each variable, say $w_{2015,l} - w_{2005,l}$, where $w_{t,l}$ denotes the value of variable w in year t for country l. We used these 2015-2005 differences in cross section regressions. The second dataset is a balanced panel containing 10 years of data, from 2006 through 2015; these data are used for the panel data regressions. There are 749 variables in the cross section dataset and 644 variables in the panel dataset. We compute distances between countries via the great-circle formula.

3.2.1 Mixed empirical and parametric models

The first set of experiments use the WDI variables as regressors with error terms generated by the parametric models used in the 48-states experiments. Six experiments are run. The first two use cross-section datasets, using each the dataset's 749 variables, one at a time, as the x_l regressor. The first experiment includes only a constant as a control, and the second adds three additional series chosen at random from the WDI. The other four experiments use the panel dataset, where each of the 644 series is used as x_l. The panel regressions include country fixed effects, and the four experiments differ in their inclusion of additional series selected at random from the WDI and time fixed effects. As in the 48-states experiments, the error terms are generated as $e|\mathbf{V} \sim \mathcal{G}_{\exp}(c_{0.03})$ with within-cluster covariance parameterized by the AR(1) model with $\phi = 0.9$ for the panel data experiments, or as $e_l = x_l^s a_l$ where $a|\mathbf{V} \sim \mathcal{G}_{\exp}(c_{0.03})$. Table 5 provides a summary.

Results are depicted in Figure 4 and suggest two conclusions. First, size distortions are evident for Kernel and SCPC, although they are not as severe as in some of the 48-states designs. Evidently, the regressors in the WDI exhibit less spatial correlation than those

Table 5: Spatial Regression Designs Using the WDI dataset

Design	Panel	z
1	N	1
2	N	[1, 3 WDI variables]
3	Y	Country Fixed Effects
4	Y	Country and Year Fixed Effects
5	Y	[Country Fixed Effects, 3 WDI variables]
6	Y	[Country and Year Fixed Effects, 3 WDI variables]

Notes: Experiments 1 and 2 use each of the 749 variables in the 2015-2005 dataset as x_l. Experiments 3-6 use each of the 644 variables in the 2006-2015 panel dataset as x. $e_l \sim \mathcal{G}_{\exp}(c_{0.03})$ (with within cluster AR(1) covariances with $\phi = 0.9$ in the panel data models) or $e_l = x_l^s a_l$ with $a_l \sim \mathcal{G}_{\exp}(c_{0.03})$.

generated for some of the 48-states experiments. Indeed, for the WDI experiments, the mean rejection frequency for SCPC is less than 0.10 in all of the experiments and close to its nominal 0.05 value in the panel regressions. That said, null rejection frequencies exceed 0.10 for more than 5% of the regressors chosen from the WDI datasets. The second conclusion is that C-SCPC has null rejection frequencies close to 0.05 in all experiments and regressors. This is consistent with the earlier experiments and provides reassuring evidence about size control C-SCPC.

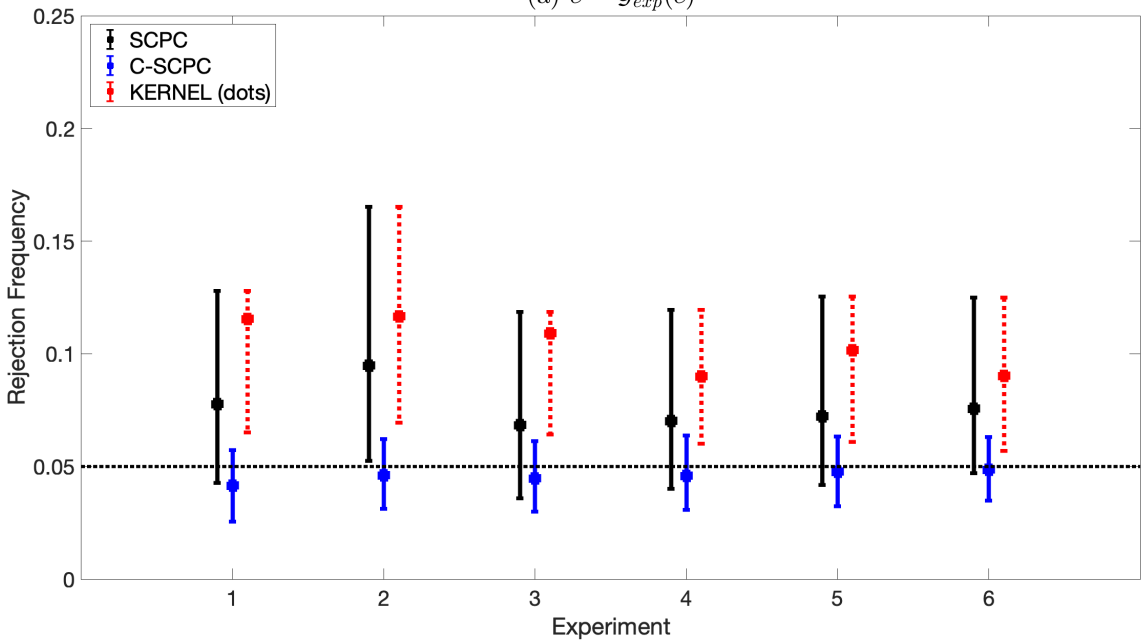
3.2.2 Empirical models for both regressor and dependent variable

In a final exercise we generate both regressors and dependent variables from the WDI dataset. This requires some care to ensure that (i) we have a well-defined population with known regression coefficient, and (ii) we control the degree of spatial correlation in the exercise. The idea is to generate artificial data by a discrete Markov chain that more likely selects countries that are geographically close to the previously selected country. With an appropriate restriction of the transition matrix, the stationary distribution is uniform across countries, so that the population regression simply becomes the regression using all countries in the WDI dataset. Furthermore, by varying the degree of preference for conditionally closer countries, we can control the average pairwise correlation across samples.

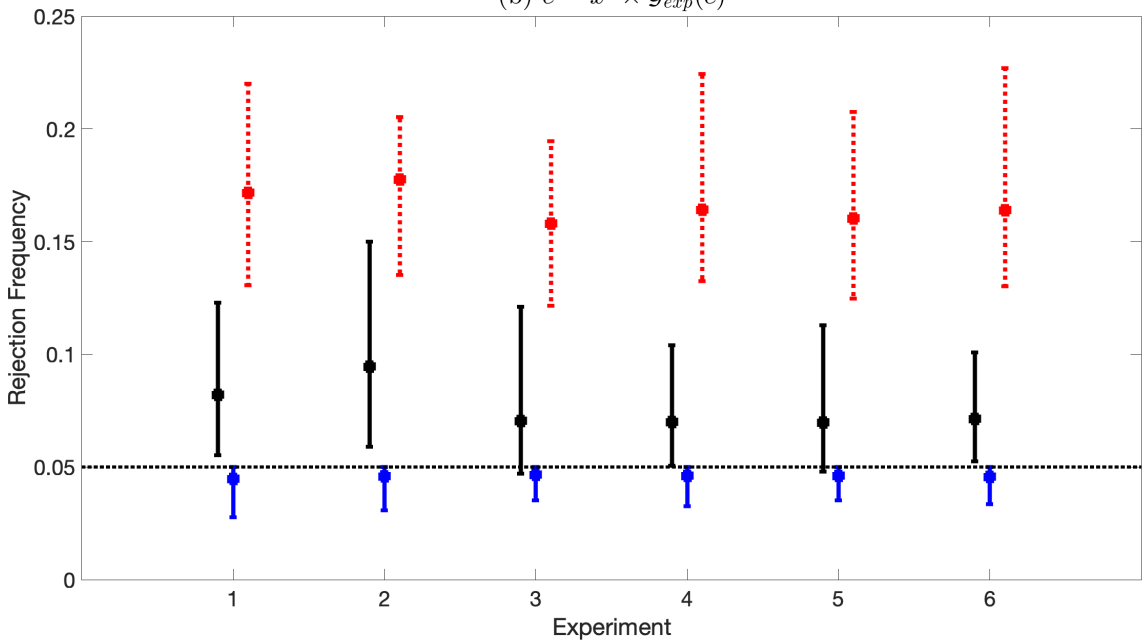
The details are as follows. Suppose we select series labelled y, x, and (if relevant) z from the WDI dataset, and suppose that the resulting (x_l, y_l, z_l) provide data for $l = 1, ..., M$ countries. Think of these countries as defining a population, where every country is assigned a

Figure 4: Rejection frequencies for 5% nominal tests: spatial regressions, WDI experiments

(a) $e \sim \mathcal{G}_{exp}(c)$



(b) $e \sim x^s \times \mathcal{G}_{exp}(c)$



Notes: See Table 5 for a description of the experiments and the notes to Figure 2.

probability of $1/M$. The population regression is then the OLS regression using data for these M countries; let $u_{(l)} = x'_{(l)}e_{(l)}$ with $e_{(l)}$ the residual from this population regression. We will compute the various HR, KERNEL, SCPC, and C-SCPC tests using random samples drawn this population. The challenge is to devise a sampling scheme that (i) puts uniform weight $(1/M)$ on each of the locations and (ii) induces a pre-specified average spatial correlation. We do this using a Markov chain, which we now describe.

Let Π be an $M \times M$ Markov transition matrix. As is well known, if Π is doubly stochastic (so that columns and rows sum to 1), then the stationary distribution is uniform. For a given value of c, let $\tilde{\Pi}(c)$ be the $M \times M$ transition matrix such that in row l, the transition probability is proportional to $\exp(-c||s_l - s_\ell||)$, $\ell \neq l$, except that the diagonal of $\tilde{\Pi}(c)$ is zero; thus locations closer to s_l are assigned a higher probability than more distant locations. Let $\Pi(c)$ be the resulting doubly stochastic matrix after appropriate diagonal scaling, that is, $\Pi(c) = \Lambda\tilde{\Pi}(c)\Lambda$ for a numerically determined diagonal matrix Λ (which is unique up to scale). We generate a sample of size n of country indices $L_i \in \{1, 2, \ldots, M\}$, $i = 1, \ldots, n$ from the stationary Markov chain with transition matrix $\Pi(c)$, with the first index L_1 drawn uniformly.

The average pairwise correlation of $u_{(L_i)}$ and $u_{(L_{i+k})}$ depends on the parameter c and is easily calculated. For all $i \geq 1$ and $k \geq 0$, we have

$$\mathbb{E}[u_{(L_i)}u_{(L_{i+k})}] = M^{-1}\sum_{l=1}^{M}\sum_{\ell=1}^{M}[\Pi(c)^k]_{l\ell}u_{(l)}u_{(\ell)}$$

where $[\Pi(c)^k]_{l\ell}$ is the l,ℓ th element of $\Pi(c)^k$. Thus, the average pairwise correlation $u_{(L_i)}$ and $u_{(L_{i+k})}$ across samples is given by

$$\begin{aligned}\bar{\rho} &= \frac{2}{n(n-1)}\sum_{k=1}^{n-1}(n-k)\frac{\mathbb{E}[u_{(L_1)}u_{(L_{1+k})}]}{\mathbb{E}[u^2_{(L_1)}]} \\ &= \frac{2}{n(n-1)}\sum_{k=1}^{n-1}(n-k)\frac{M^{-1}\sum_{l=1}^{M}\sum_{\ell=1}^{M}[\Pi(c)^k]_{l\ell}u_{(l)}u_{(\ell)}}{M^{-1}\sum_{l=1}^{M}u^2_{(l)}}.\end{aligned}$$

Thus, it is straightforward to numerically to obtain the value of c that induces a given value of $\bar{\rho}$.

With $\Pi(c)$ determined in this fashion, we generate 5,000 random samples of countries

Table 6: Rejection frequency of nominal 5% level tests using WDI-Markov-Chain design

Model	Panel	5th, 50th, and 95th quantiles			
		HR	KERNEL	SCPC	C-SCPC
1	N	0.20, 0.26, 0.33	0.10, 0.15, 0.24	0.09, 0.14, 0.21	0.04, 0.08, 0.14
2	N	0.21, 0.27, 0.36	0.11, 0.17, 0.27	0.10, 0.17, 0.27	0.03, 0.08, 0.18
3	Y	0.23, 0.27, 0.33	0.08, 0.14, 0.21	0.06, 0.10, 0.18	0.04, 0.08, 0.12
4	Y	0.23, 0.27, 0.32	0.08, 0.15, 0.22	0.07, 0.11, 0.19	0.04, 0.08, 0.14

Notes: See text for description of the design. Models 1 and 2 include constants as control variables and Models 3 and 4 include country fixed effects; Models 2 and 4 include 3 additional regressors selected at random from the WDI datasets.

and corresponding sample regressors and dependent variable using $n = 50$ and $\bar{\rho} = 0.03$. Each set of (y, x, z) from the WDI corresponds to a different population, with associated rejection frequencies for the HR, KERNEL, SCPC, and C-SCPC tests.[6] We construct two experiments using spatial regressions from the 2015-2005 dataset; the first includes (y, x), a constant and no additional controls, and the second adds three randomly selected additional series as controls. Similarly we construct two experiments using panel regressions from the 2006-2015 dataset, where fixed country effects are included the first, and the second adds three additional randomly selected series. Table 6 shows the 5th, 50th, and 95th quantiles of rejection frequencies across these 5,000 populations.

These experiments differ from those reported earlier in two distinct ways. First, as emphasized above, the values of both e and x correspond to real-world data; so, for example, neither is generated by a Gaussian model as in the earlier experiments. Second, earlier experiments used $n = 250$, while these use $n = 50$, where the smaller sample size reflects the fact that many of the series in the WDI are available for as few as $M = 100$ countries. The rejection frequencies for these experiments share some of the features of the earlier experiments, but there are notable differences. For example, the smaller sample size means that HR rejects less frequently; the rejection frequencies fall from around 50% in Table 1 to half that value in Table 6. KERNEL and SCPC improve on HR, but (as in some of the previous experiments) exhibit quantitatively important size distortions. C-SCPC does better, with a median re-

[6]To insure that the Markov chain can attain $\overline{\rho} = 0.03$ without re-sampling the same locations frequently, we restrict our attention to (y, x, z) variables in the WDI that exhibit spatial correlation. Specifically we use data in which the kernel-estimated variance of $u_{(l)}$ with a bandwidth of 0.3 of the largest distance in data set is at least three times larger than the HR-estimated variance.

jection frequency of 8% – close the nominal 5% value – although for 5% of the DGPs, the rejection frequency exceeds 0.15 for the spatial (non-panel) regressions.

4 Conditional size control in more general model

The analysis thus far has investigated the size properties of tests under two parametric spatial processes for the errors, $\mathcal{G}_{\exp}(c)$, namely the Gaussian model with exponential covariance function, and the product of x^s and a variable following the same process. A natural question asks how the tests perform under alternate error processes. For the model with $x_l = 1$, Müller and Watson (2021) shows that SCPC controls size in large samples for all weakly dependent covariance stationary processes: SCPC-inference is robust to the form of the covariance function. In this section we present some results for the spatial regression model after conditioning on $\{x_l, z_l, s_l\}$, by investigating the robustness of C-SCPC in a finite sample Gaussian setting.

We carry out a straightforward but instructive exercise. Specifically, we carry out C-SCPC inference as in the other experiments — that is using $\bar{\rho}_{\max} = 0.03$ and using the critical value cv chosen so that conditional size is controlled for $e_l = x_l^s a_l$ where $a_l \sim \mathcal{G}_{\exp}(c)$ for $c \geq c_{\bar{\rho}_{\max}}$ (equivalently for $\bar{\rho} \leq \bar{\rho}_{\max}$) — however, we now use non-$\mathcal{G}_{\exp}$ stochastic processes to generate the errors e_l or a_l. In particular, we generate the errors from five different Matérn processes and study the robustness for all values of $\bar{\rho} \leq \bar{\rho}_{\max}$ for each of these five processes.

The Matérn class of stochastic processes is indexed by the parameter $\theta = (\nu, c)$, where ν and c are positive constants. If a follows a Matérn process, its covariance function $\sigma_a(r-s)$ depends on the locations only through $\Delta = ||r-s||$, with $\sigma_a(\Delta) \propto (c\Delta)^\nu K_\nu(c\Delta)$, where K_ν is modified Bessel function of the second kind. When $\nu \in \{1/2, 3/2, 5/2, \infty\}$, the expression for the covariance simplifies:

- $\nu = 1/2$: $\sigma_a(\Delta) \propto \exp[-c\Delta]$
- $\nu = 3/2$: $\sigma_a(\Delta) \propto (1 + \Delta c)\exp[-c\Delta]$
- $\nu = 5/2$: $\sigma_a(\Delta) \propto (3 + 3\Delta c + \Delta^2 c^2)\exp[-c\Delta]$
- $\nu = \infty$: $\sigma_a(\Delta) \propto \exp[-c^2\Delta^2/2]$.

So $\nu = 1/2$ is the exponential model used throughout the paper but the other processes have different decay. We consider these four processes together with the process with $\nu = 1/4$.

Gneiting (2013) shows that for $\nu > 1/2$, the Matérn covariance function is not positive definite when distances are measured by the great-circle formula. So in this section, we instead map each country's location on the surface of the earth into a point in $\mathbb{R}^3$, and then use the Euclidian norm to compute distances. This effectively also changes the covariance matrices for $\nu = 1/2$, providing an additional degree of separation from the baseline specification.

We re-run the 48-states and WDI experiments from Figures 2-4 using the same regressors but with errors generated by the Matérn processes with $\nu \in \{1/4, 1/2, 3/2, 5/2, \infty\}$. For each process we find a range of values of c to trace out values of $\bar{\rho} \in (0.001, \bar{\rho}_{\max})$ with $\bar{\rho}_{\max} = 0.03$, and generate errors for each value of c. We record the largest rejection frequency over each of the processes and c-values. As in the previous experiments, this generates a distribution of rejection frequencies over realizations of $\{x_l, z_l, s_l\}$ but now the rejection frequencies represent the largest rejection frequency over this Matérn class with $\bar{\rho} \leq \bar{\rho}_{\max}$. Figure 5 summarizes the results, showing for each experiment the 5th, 50th and 95th quantiles.

The results are reassuring: the 95th percentile of the rejection frequency across all values of $\{x_l, z_l, s_l\}$ is less that 0.09 across all five Matérn models, for all values of $\bar{\rho} \in (0.001, \bar{\rho}_{\max})$ and for all of the experiments shown in the figure. In most experiments the 95th percentile rejection frequency is less than 0.07. We conclude that C-SCPC inference is robust for this class of DGPs.

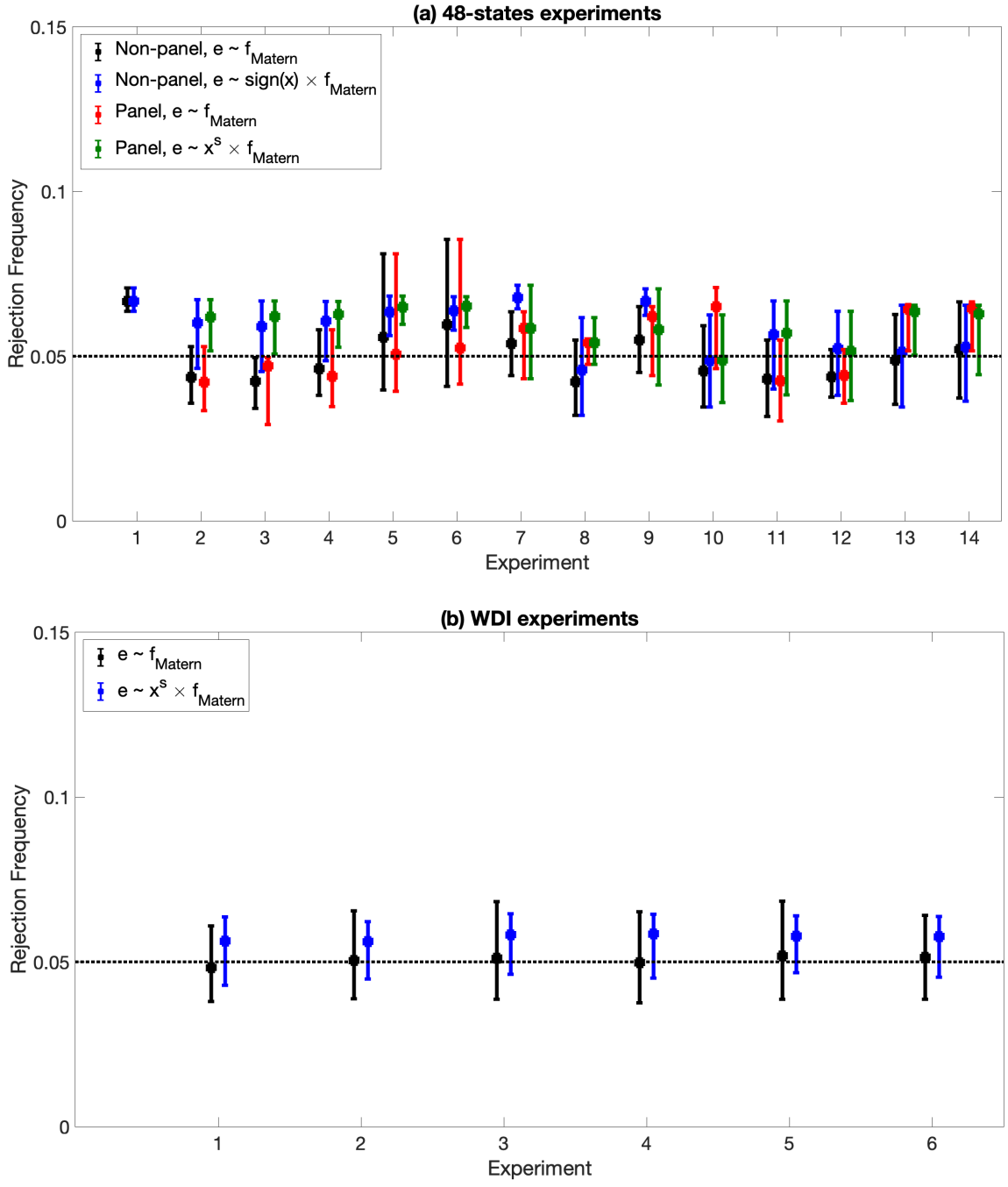
5 Critical values, IV regression and large-n numerical methods

5.1 Computing $\text{cv}_{\mathbf{SCPC}}$ and $\text{cv}_{\mathbf{V}}$ critical values

Let $\mathbf{\Sigma}(c)$ denote an $n \times n$ covariance matrix using an exponential covariance function with parameter c evaluated at the sample locations $\mathbf{s}$. Let $\mathbf{R}$ denote an $n \times q$ matrix whose columns are the eigenvectors of $\mathbf{M_1 \Sigma}(c_{\min})\mathbf{M_1}$ corresponding the largest q eigenvectors. These eigenvectors are used to compute $\hat{\sigma}^2_{\text{SCPC}}$ (see (6)). The rejection region for the τ_{SCPC} test using a critical value cv corresponds to values of $\mathbf{u}$ that satisfy $|\tau_{\text{SCPC}}| > \text{cv}$, or equivalently as

$$\mathbf{u'11'u} - \text{cv}\, q^{-1} \hat{\mathbf{u}}' \mathbf{RR}' \hat{\mathbf{u}} > 0. \tag{11}$$

Figure 5: Largest rejection frequency across a set of Matérn processes



Notes: See the text for a description of the experiments and the notes to Figure 2.

The cv_{SCPC} and $\text{cv}_{\mathbf{V}}$ use this expression to compute rejection frequencies under different probability laws for $\mathbf{u}$. We discuss these in turn.

Computing cv_{SCPC}:

The cv_{SCPC} critical value uses $u_l \sim \mathcal{G}_{\text{exp}}(c)$, so that $\text{Var}(\mathbf{u}|\mathbf{s}) = \mathbf{\Sigma}(c)$. It uses the regression model with $x_l = 1$ so that $\hat{u}_l = u_l - \bar{u}$. Because the columns of $\mathbf{R}$ are eigenvectors of $\mathbf{M_1}\mathbf{\Sigma}(c_{\min})\mathbf{M_1}$, $\mathbf{R}'\mathbf{1} = 0$, so $\mathbf{R}'\hat{\mathbf{u}} = \mathbf{R}'\mathbf{u}$; this implies that the values of $\hat{\mathbf{u}}$ in (11) can be replaced with $\mathbf{u}$.

Let $h = \mathbf{W}'\mathbf{u}$ with $\mathbf{W} = [\mathbf{1}, \mathbf{R}]$. Let $\mathbf{\Omega}(c)$ the $(q+1)\times(q+1)$ covariance matrix of h, where $\mathbf{\Omega}(c) = \mathbf{W}'\mathbf{\Sigma}(c)\mathbf{W}$. The rejection region can then be written as

$$\mathbf{h}'\mathbf{D}(\text{cv})\mathbf{h} > 0$$

where

$$\mathbf{D}(\text{cv}) = \begin{bmatrix} 1 & 0 \\ 0 & -\text{cv}\, q^{-1}\mathbf{I}_q \end{bmatrix}.$$

The rejection frequency under normality is

$$\mathbb{P}(\mathbf{h}'\mathbf{D}(\text{cv})\mathbf{h} > 0) \text{ with } h \sim \mathcal{N}(0, \mathbf{\Omega}(c))$$

and where the probability can be efficiently computed using formula in Bakirov and Székely (2005). (See Lemma 1 in Müller and Watson (2021).) As in (7), cv_{SCPC} is chosen to satisfy

$$\mathbb{P}^{\text{Normal}}_{c \geq c_{\bar{\rho}_{\max}}}(\mathbf{h}'\mathbf{D}(\text{cv}_{\text{SCPC}})\mathbf{h} > 0) = \alpha$$

where α is the desired size of the test.

Computing $\text{cv}_{\mathbf{V}}$:

The $\text{cv}_{\mathbf{V}}$ critical value uses $u_l = x_l' e_l$ with $e_l = x_l^s a_l$ where $a_l|(\mathbf{X}, \mathbf{Z}, \mathbf{s}) \sim \mathcal{G}_{\text{exp}}(c)$ and computes the probability of (11) conditional on $(\mathbf{X}, \mathbf{Z}, \mathbf{s})$.

Some new notation helps explain the details of the calculation. Let $\mathbf{\Xi}$ denote an $N \times n$

block diagonal matrix with x_l on the diagonal

$$\mathbf{\Xi} = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & x_n \end{bmatrix}$$

and let $\mathbf{\Xi}_s$ denote the analogous matrix using x_l^s in place of x_l. Then $\mathbf{e} = \mathbf{\Xi}_s\mathbf{a}$ and $\mathbf{u} = \mathbf{\Xi}'\mathbf{e} = \mathbf{\Xi}'\mathbf{\Xi}_s\mathbf{a}$. Let $\mathbf{M_V} = \mathbf{I}_N - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}$ with $\mathbf{V} =(\mathbf{X}\,\mathbf{Z})$. Then $\hat{\mathbf{e}} = \mathbf{M_V}\mathbf{e}$ and $\hat{\mathbf{u}} = \mathbf{\Xi}'\hat{\mathbf{e}} = \mathbf{\Xi}'\mathbf{M_V}\mathbf{e} = \mathbf{\Xi}'\mathbf{M_V}\mathbf{\Xi}_s\mathbf{a}$.

Paralleling the discussion of cv_{SCPC}, let $\tilde{h}= \tilde{\mathbf{W}}'\mathbf{a}$ with

$$\tilde{\mathbf{W}} = [\mathbf{\Xi}_s'\mathbf{\Xi}\mathbf{1}, \mathbf{\Xi}_s'\mathbf{M_V}\mathbf{\Xi}\mathbf{R}]. \tag{12}$$

Under $\mathbf{a}|(\mathbf{X}, \mathbf{Z}, \mathbf{s}) \sim \mathcal{N}(0, \mathbf{\Sigma}(c))$ then $\tilde{h}|(\mathbf{X}, \mathbf{Z}, \mathbf{s}) \sim \mathcal{N}(0, \tilde{\mathbf{\Omega}}(c))$ with $\tilde{\mathbf{\Omega}}(c) = \tilde{\mathbf{W}}'\mathbf{\Sigma}(c)\tilde{\mathbf{W}}$. The remaining calculations are identical to those of cv_{SCPC} with $\tilde{\mathbf{\Omega}}$ replacing $\mathbf{\Omega}$.

5.2 Computing SCPC eigenvectors for large n

The SCPC estimator for σ^2 given in (6) relies on the eigenvectors, $\mathbf{r}_i$, of the $n \times n$ matrix $\mathbf{M}_1\mathbf{\Sigma}(c_{\min})\mathbf{M}_1$. When n is large (say, n is much larger than 3500) computing these eigenvectors is a challenging computational task. However, it is possible to accurately approximate $\mathbf{r}_i$ by first computing eigenvectors using only $\tilde{n} < n$ locations $s_l \in \mathcal{S}$ and then extending these to encompass all n locations. This is a version of the so-called Nyström method (see, for instance, Rasmussen and Williams (2005) for discussion and references).

The idea of the extension is most easily explained if we initially ignore the demeaning by $\mathbf{M}_1$. So for now, suppose we seek the eigenvectors $\mathbf{v}_i$ of the $n \times n$ matrix $\mathbf{K} =\mathbf{\Sigma}(c_{\min})$ with (l, ℓ) element $k(s_l, s_\ell) = \exp(-c_{\min}||s_l - s_\ell||)$ corresponding to the q largest eigenvalues λ_i. Since $\mathbf{K}\mathbf{v}_i = \lambda_i\mathbf{v}_i$, we trivially have $\mathbf{v}_i = \lambda_i^{-1}\mathbf{K}\mathbf{v}_i$ for $\lambda_i > 0$. If we put the original locations in random order, then the first $\tilde{n}$ locations $\{\tilde{s}_l\}_{l=1}^{\tilde{n}}$ form an i.i.d. random subset of size $\tilde{n} < n$ of all n observed locations. Let $\tilde{\mathbf{K}}$ be the corresponding $\tilde{n} \times \tilde{n}$ matrix computed from the first $\tilde{n}$ locations, with ith eigenvalue-eigenvector pair $(\tilde{\lambda}_i,\tilde{\mathbf{v}}_i)$, where $\tilde{n}^{-1}\tilde{\mathbf{v}}_i'\tilde{\mathbf{v}}_i = 1$. One natural way

to extend the $\tilde{n} \times 1$ vector $\tilde{\mathbf{v}}_i$ to the $n \times 1$ vector $\mathbf{v}_i$ is via

$$\mathbf{v}_i \approx \hat{\mathbf{v}}_i = \tilde{\lambda}_i^{-1} \tilde{\mathbf{K}}_0 \tilde{\mathbf{v}}_i$$

where the $n \times \tilde{n}$ matrix $\tilde{\mathbf{K}}_0$ has (l, ℓ) element $k(s_l, s_\ell)$, $l = 1, \ldots, n$, $\ell = 1, \ldots, \tilde{n}$. Note that the first $\tilde{n}$ elements of $\hat{\mathbf{v}}_i$ are identical to $\tilde{\mathbf{v}}_i$, and for $l > \tilde{n}$,

$$\hat{\mathbf{v}}_{i,l} = \tilde{\lambda}_i^{-1} \sum_{\ell=1}^{\tilde{n}} \tilde{\mathbf{v}}_{i,l} k(s_l, s_\ell), \tag{13}$$

so this method approximates the value of $\mathbf{v}_l$ at the additional locations by a weighted average of the kernel k, with weights proportional to the eigenvector computed from the first $\tilde{n}$ locations.

This analysis can be made rigorous: Suppose the $\tilde{n}$ locations $s_l \in \mathcal{S}$ are sampled i.i.d. with density g. Let $\mathcal{L}_G^2$ the Hilbert space of functions $\mathcal{S} \mapsto \mathbb{R}$ with inner product $\langle f_1, f_2 \rangle = \int f_1(s) f_2(s) g(s) ds$. Then by Mercer's Theorem, k has the representation $k(s, r) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(s) \varphi_i(r)$, where $(\lambda_i, \varphi_i) \in \mathbb{R} \times \mathcal{L}_G^2$ are eigenvalues and eigenfunctions of k, with eigenvalues ordered from largest to smallest, normalized so that $\int \varphi_i(s) \varphi_j(s) g(s) ds = \mathbf{1}[i = j]$ and $\varphi_i(\cdot) = \lambda_i^{-1} \int k(\cdot, r) \varphi_i(r) g(r) dr$ for $\lambda_i > 0$. Now Rosasco, Belkin, and Vito (2010) show that if the eigenvalue λ_i is unique, then $\sup_{s \in \mathcal{S}} ||\tilde{\varphi}_i(s) - \varphi_i(s)|| = O(\tilde{n}^{-1/2})$, where

$$\tilde{\varphi}_i(\cdot) = \tilde{\lambda}_i^{-1} \sum_{\ell=1}^{\tilde{n}} k(\cdot, s_\ell) v_{i,\ell}.$$

That paper also contains analogous results if there are eigenvalues of multiplicity larger than one.

These ideas and results extend to the problem of consideration here, where we seek to approximate the eigenvectors $\mathbf{r}_i$ of a demeaned version of $\mathbf{K}$, namely $\mathbf{M}_1 \mathbf{K} \mathbf{M}_1$. Define

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

so that the (l, ℓ) element of $\mathbf{M}_1 \mathbf{K} \mathbf{M}_1$ is equal to $\hat{k}_n(s_l, s_\ell)$. Let $(\tilde{\omega}_i, \tilde{\mathbf{r}}_i)$ with $\tilde{n}_i^{-1} \tilde{\mathbf{r}}_i' \tilde{\mathbf{r}}_i = 1$ be the eigenvalues and eigenvectors of the corresponding matrix $\tilde{\mathbf{M}}_1 \tilde{\mathbf{K}} \tilde{\mathbf{M}}_1$ computed from the first $\tilde{n}$

locations. Then in analogy to (13), we extend the $\tilde{n} \times 1$ vector $\tilde{\mathbf{r}}_i$ to $\hat{\mathbf{r}}_i \approx \mathbf{r}_i$ via

$$\begin{aligned} \mathbf{r}_{i,l} &\approx \hat{\mathbf{r}}_{i,l} = \tilde{\omega}_i^{-1} \sum_{\ell=1}^{\tilde{n}} \hat{k}_{\tilde{n}}(s_l, s_\ell) \tilde{r}_{i,\ell} \\ &= \tilde{\omega}_i^{-1} \sum_{\ell=1}^{\tilde{n}} \left(\exp[-c_{\min} ||s_l - \tilde{s}_\ell||] - \tilde{n}^{-1} \sum_{j=1}^{\tilde{n}} \exp[-c_{\min} ||\tilde{s}_\ell - \tilde{s}_j||] \right) \tilde{r}_{i,\ell} \end{aligned} \tag{14}$$

where the second equality exploits $\sum_{l=1}^{\tilde{n}} \tilde{r}_{i,l} = 0$.

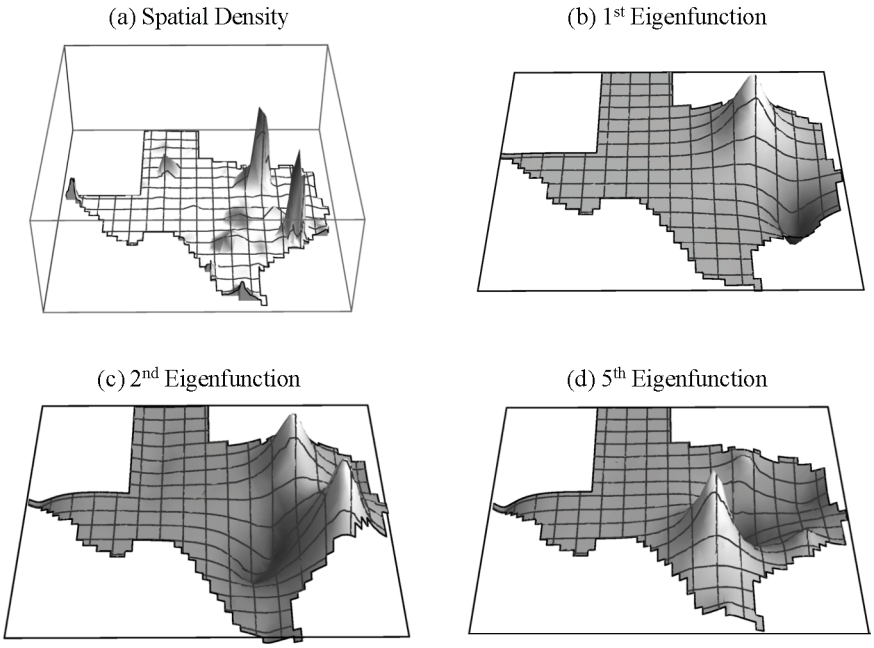
The results of Rosasco, Belkin, and Vito (2010) are not directly applicable, since the demeaned kernel $\hat{k}_n$ is also a function of the observed locations $\{s_l\}_{l=1}^n$, and hence sample dependent. However, one would expect that in large samples, $\hat{k}_n$ is well approximated by the population demeaned kernel

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

with eigenvalue-eigenfunction representation $\bar{k}(s,r) = \sum_{i=1}^{\infty} \omega_i \phi_i(s) \phi_i(r)$. Lemma 6 in Müller and Watson (2021) shows that this is indeed the case, and it implies that if ω_i is unique, then $\sup_{s \in \mathcal{S}} ||\tilde{\phi}_i(s) - \phi_i(s)|| = O(\tilde{n}^{-1/2})$, where $\tilde{\phi}_i(\cdot) = \tilde{\omega}_i^{-1} \sum_{\ell=1}^{\tilde{n}} \hat{k}_{\tilde{n}}(\cdot, s_\ell) r_{i,\ell}$. Figure 6 depicts three of the resulting estimated eigenfunctions $\tilde{\phi}_i$ for the U.S. state of Texas.

We hence conclude that the numerical approximation $\mathbf{r}_{i,l} \approx \hat{\mathbf{r}}_{i,l}$ in (14) is formally justified under $\tilde{n} \to \infty$ asymptotics, and numerical experiments suggest that results become fairly stable with $\tilde{n} = 1000$. In practice, the approximation (14) 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. This further improves the accuracy of the approximation, and reduces the artificial randomness induced by the selection of $\tilde{n}$ locations. The resulting algorithm has $O(n)$ running time, which compares favorably to the $O(n^2)$ running time of a basic implementation of a kernel variance estimator. We provide corresponding STATA and Matlab code in the replication files.

Figure 6: Spatial density and eigenfunctions for Texas



Notes: Panel (a) shows the spatial density of light in Texas. Panels (b)-(d) show three of the estimated eigenfunctions of the kernel $\overline{k}$ using $c_{0.03}$.

5.3 SCPC and C-SCPC for IV regression

Consider a version of (2) in which ζ is the coefficient on a scalar endogenous regressor, say $p_{i,l}$:

$$y_{i,l} = p_{i,l}\zeta + z_{i,l}'\gamma + e_{i,l}$$

where $p_{i,l}$ is potentially correlated with $e_{i,l}$. Suppose $x_{i,l}$ is a scalar instrument for $p_{i,l}$, $z_{i,l}$ is a vector of exogenous regressors and $v_{i,l} = (x_{i,l}, z_{i,l})$ satisfies the assumptions discussed in the previous sections. Minor modifications of the methods discussed above provide SCPC and C-SCPC inference for ζ.

Weak-instrument robust inference about ζ utilizes the Anderson and Rubin (1949) regression

$$y_{i,l} - p_{i,l}\zeta_0 = x_{i,l}\beta + z_{i,l}'\gamma + e_{i,l}^0 \tag{15}$$

where the null hypothesis $\zeta = \zeta_0$ corresponds to $\beta = 0$ in the regression (15). Spatial correlation robust SCPC/C-SCPC tests of the $\beta = 0$ null proceed as described in Sections 2.2

and 2.3. Note that C-SCPC inference for $\beta = 0$ in (15) (equivalently $\zeta = \zeta_0$) produces valid inference in Section 2.3's heteroskedastic Gaussian model conditional on the instrument and exogenous regressors $(\mathbf{X}, \mathbf{Z})$.

When $x_{i,l}$ is known to be a strong instrument, spatial correlation robust inference can be carried out using the IV t-statistic constructed with the SCPC estimator for σ. The IV estimator is $\hat{\zeta}^{IV} = S_{xp}^{-1} S_{xy}$ (recall that x and z are uncorrelated in the sample, so $\mathbf{X}'\mathbf{Z} = 0$) so $\hat{\zeta}^{IV} = \zeta + S_{xp}^{-1} n^{-1} \sum_{l=1}^{n} x_l' e_l$. The resulting IV t-statistic has the same form as (3) with $\hat{\zeta}^{IV}$ replacing the OLS estimator $\hat{\beta}$. SCPC inference then follows as in Section 2.2 with $y_l^o = \{n S_{xp}^{-1} \hat{u}_l^{IV} + \hat{\zeta}^{IV}\}_{l=1}^{n}$, where $\hat{u}_l^{IV} = x_l' \hat{e}_l^{IV}$ with $\hat{e}_{i,l}^{IV} = y_{i,l} - p_{i,l}\hat{\zeta}^{IV} - z_{i,l}'\hat{\gamma}^{IV}$.

The case is more complicated for conducting C-SCPC inference using the τ_{SCPC} IV t-statistic. The complication arises because the IV residuals are given by $\hat{\mathbf{e}}^{IV} = \mathbf{M}\mathbf{e}$ with $\mathbf{M} = \mathbf{I}_N - [\mathbf{P}\,\mathbf{Z}](\mathbf{V}'[\mathbf{P}\,\mathbf{Z}])^{-1}\mathbf{V}'$, so that $\mathbf{M}$ depends on the (potentially endogenous) regressors $\mathbf{P}$. Under the C-SCPC heteroskestastic model, $e_l = x_l^s a_l$, so $\mathbf{e} = \mathbf{\Xi}_s \mathbf{a}$ and $\hat{\mathbf{e}}^{IV} = \mathbf{M}\mathbf{\Xi}_s\mathbf{a}$. Thus, if $a_l|(\mathbf{P}, \mathbf{X}, \mathbf{Z}) \sim \mathcal{G}_{\exp}(c)$ for $c \geq c_{min}$, then C-SCPC has guaranteed size control conditional on $(\mathbf{P}, \mathbf{X}, \mathbf{Z})$. That said, if $\mathbb{E}(e_{i,l}|p_{i,l}, x_{i,l}, z_{i,l}) \neq 0$, that is, if p is endogenous, then a_l is not independent of $(\mathbf{P}, \mathbf{X}, \mathbf{Z})$ invalidating the small-sample conditional size control of C-SCPC.

This discussion has assumed that $x_{i,l}$ is a scalar. When there are multiple instruments, say $\widetilde{x}_{i,l}$, then SCPC inference based on $\hat{\beta}^{IV}$ can be conducted using the scalar instrument $x_{i,l} = \hat{w}'\tilde{x}_{i,l}$, where $\hat{w}$ is a vector of weights computed, for example, by 2SLS.

6 Concluding remarks

This paper has studied the properties of spatial-correlation robust t-statistics in linear regression and panel models using data generating processes designed to mimic spatial correlation patterns relevant for empirical work in economics. We find significant size distortions (i.e., over-rejections of the null hypothesis) using standard kernel-based spatial correlation robust standard errors and standard normal critical values. Size is improved using a projection-based standard error (called SCPC) and critical value proposed in Müller and Watson (2021). That said, the experiments find uncomfortably large size distortions for SCPC in some settings, particularly with non-stationary regressors. The paper proposes a modification of the SCPC method (called C-SCPC), which is designed to induce validity also conditional on the sample values of the regressors. The results indicate that C-SCPC has good size control in a wide

variety of empirically relevant settings.

The analysis has focused on inference for a single regression coefficient. Of course, kernel-based methods are readily extended to test vector-value regression coefficient coefficients. But, as in the scalar case, these kernel-based tests will exhibit significant size distortions when used with large-sample χ^2 critical values and applied to spatially correlated data of the sort studied here. The multivariate extension of SCPC is conceptually straightforward: principal components under a 'worst-case' benchmark spatial model can be used to estimate the relevant covariance matrix, and the critical value of the resulting test statistic can be constructed to guarantee coverage with spatial correlation less than or equal to the worst-case level. In practice, implementing such a test requires the non-trivial tasks of specifying the multivariate worst-case benchmark model and numerically determining the critical value. The extension to C-SCPC is less clear-cut as it requires specifying an appropriate multivariate extension of the heteroskedastic model for e_l given in (9). We leave the multivariate extensions SCPC and C-SCPC to future work.

A Appendix: WDI Data

This appendix summarizes the pre-processing of variables from the World Development Indicators database prior to their use in Section 3.2.

- For the 2015-2005 decadal difference dataset:
  - Transformations: The level of the series was used if (a) its description included the word *index* or included the symbol % or (b) the series included non positive values. Otherwise the logarithm of the series was used. Let $y = x_{2015} - x_{2005}$ denote the decadal difference of the transformed series
  - Outlier adjustments: If y contained 4 or fewer unique values, no outlier adjustment was performed. Otherwise, observations with $|y_l - \text{median}(y)| > 5 \times \text{IQR}$ were replaced with missing values, where IQR denotes the inter-quartile range.
  - Series were discarded if: (1) they contained fewer than 100 countries, (2) had a sample kurtosis greater than 20, (3) had fewer than 10 countries with different values.
  - This resulted in a dataset with 749 series.
- For the 2006-2015 panel dataset:
  - Transformations: Same as above. Let $x_{t,l}$ denote the transformed series, $\overline{x}_l$ denote the mean for country l, $\overline{x}$ denote the sample mean over all t and l, and $y_{t,l} = x_{t,l} - \overline{x}_l$ denote the country-demeaned data
  - Outlier adjustment: Observations with $|y_{t,l} - \text{median}(y)| > 5 \times \text{IQR}$ were replaced with missing values.
  - $y_{t,l}$ was replaced with a missing value if $y_{\tau,l}$ was missing for any value of $\tau \in [2006, 2015]$. Thus, the panel is balanced.
  - The series was discarded if it contained non-missing values for fewer than 100 countries or had a kurtosis greater than 20.
  - This resulted in a dataset with 644 series.

References

Anderson, T. W., and H. Rubin (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *The Annals of Mathematical Statistics*, 21, 570–582.

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

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*.

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

den Haan, W. J., and A. T. Levin (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics 15*, ed. by G. S. Maddala, and C. R. Rao, pp. 299–342. Elsevier, Amsterdam.

Eicker, F. (1967): "Limit Theorems for Regression with Unequal and Dependent Errors," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 59–82.

Gneiting, T. (2013): "Strictly and Non-Strictly Positive Definite Functions on Spheres," *Bernouli*, 4(19), 1327–1349.

Hansen, B. E. (2021): "The Exact Distribution of the White t-Ratio," *Manuscript, University of Wisconsin*.

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.

Huber, P. (1967): "The Behavior of the Maximum Likelihood Estimates under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 221–233, Berkeley. University of California Press.

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

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*.

——— (2021): "Persistence, Randomization, and Spatial Noise," *Manuscript, University College Dublin*.

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

Kim, M. S., and Y. Sun (2011): "Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix," *Journal of Econometrics*, 2(160), 349–371.

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.

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

Müller, U. K., and M. W. Watson (2021): "Spatial Correlation Robust Inference," *Manuscript, Princeton University*.

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. (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.

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

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.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–830.