Econometric Theory, **40**, 2024, 278–319.
doi:10.1017/S0266466622000445

CONSISTENT SPECIFICATION TESTING UNDER SPATIAL DEPENDENCE

ABHIMANYU GUPTA
University of Essex

XI QU
Shanghai Jiao Tong University

We propose a series-based nonparametric specification test for a regression function when data are spatially dependent, the "space" being of a general economic or social nature. Dependence can be parametric, parametric with increasing dimension, semiparametric or any combination thereof, thus covering a vast variety of settings. These include spatial error models of varying types and levels of complexity. Under a new smooth spatial dependence condition, our test statistic is asymptotically standard normal. To prove the latter property, we establish a central limit theorem for quadratic forms in linear processes in an increasing dimension setting. Finite sample performance is investigated in a simulation study, with a bootstrap method also justified and illustrated. Empirical examples illustrate the test with real-world data.

1. INTRODUCTION

Models for spatial dependence have recently become the subject of vigorous research. This burgeoning interest has roots in the needs of practitioners who frequently have access to datasets featuring interconnected cross-sectional units. Motivated by these practical concerns, we propose a specification test for a regression function in a general setup that covers a vast variety of commonly employed spatial dependence models and permits the complexity of dependence to increase with sample size. Our test is consistent, in the sense that a parametric specification is tested with asymptotically unit power against a nonparametric

We thank the Editor, the Co-Editor, and three referees for insightful comments that improved the paper. We are grateful to Swati Chandna, Miguel Delgado, Emmanuel Guerre, Fernando López Hernandéz, Hon Ho Kwok, Arthur Lewbel, Daisuke Murakami, Ryo Okui, and Amol Sasane for helpful comments, and audiences at YEAP 2018 (Shanghai University of Finance and Economics), NYU Shanghai, Carlos III Madrid, SEW 2018 (Dijon), Aarhus University, SEA 2018 (Vienna), EcoSta 2018 (Hong Kong), Hong Kong University, AFES 2018 (Cotonou), ESEM 2018 (Cologne), CFE 2018 (Pisa), University of York, Penn State, Michigan State, University of Michigan, Texas A&M, 1st Southampton Workshop on Econometrics and Statistics, and MEG 2019 (Columbus). We also thank Xifeng Wen from the Experiment and Data Center of Antai College of Economics and Management (SJTU) for expert computing assistance. The research of the first author was supported by ESRC grant ES/R006032/1. The research of the second author was supported by the National Natural Science Foundation of China (Project Nos. 72222007, 71973097, and 72031006). Address correspondence to Abhimanyu Gupta, Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK; e-mail: a.gupta@essex.ac.uk.

alternative. The "spatial" models that we study are not restricted in any way to be geographic in nature, indeed "space" can be a very general economic or social space. Our empirical examples feature conflict alliances and technology externalities as examples of "spatial dependence," for instance.

Specification testing is an important problem, and this is reflected in a huge literature studying consistent tests. Much of this is based on independent, and often also identically distributed, data. However, data frequently exhibit dependence, and consequently a branch of the literature has also examined specification tests under time series dependence. Our interest centers on dependence across a "space," which differs quite fundamentally from dependence in a time series context. Time series are naturally ordered and locations of the observations can be observed, or at least the process generating these locations may be modeled. It can be imagined that concepts from time series dependence be extended to settings where the data are observed on a geographic space and dependence can be treated as a decreasing function of distance between observations. Indeed, much work has been done to extend notions of time series dependence in this type of setting (see, e.g., Jenish and Prucha, 2009, 2012).

However, in a huge variety of economics and social science applications, agents influence each other in ways that do not conform to such a setting. For example, farmers affect the demand of farmers in the same village but not in different villages, as in Case (1991). Likewise, price competition among firms exhibits spatial features (Pinkse, Slade, and Brett, 2002), input–output relations lead to complementarities between sectors (Conley and Dupor, 2003), co-author connections form among scientists (Oettl, 2012; Mohnen, 2022), R&D spillovers occur through technology and product market spaces (Bloom, Schankerman, and van Reenen, 2013), networks form due to allegiances in conflicts (König et al., 2017), and overlapping bank portfolios lead to correlated lending decisions (Gupta, Kokas, and Michaelides, 2021). Such examples cannot be studied by simply extending results developed for time series and illustrate the growing need for suitable methods.

A very popular model for general spatial dependence is the spatial autoregressive (SAR) class, due to Cliff and Ord (1973). The key feature of SAR models, and various generalizations such as SAR moving average (SARMA) and matrix exponential spatial specification (MESS, due to LeSage and Pace (2007)), is the presence of one or more spatial weight matrices whose elements characterize the links between agents. As noted above, these links may form for a variety of reasons, so the "spatial" terminology represents a very general notion of space, such as social or economic space. Key papers on the estimation of SAR models and their variants include Kelejian and Prucha (1998) and Lee (2004), but research on various aspects of these is active and ongoing (see, e.g., Robinson and Rossi, 2015; Hillier and Martellosio, 2018a, 2018b; Hahn, Kuersteiner, and Mazzocco, 2020; Kuersteiner and Prucha, 2020; Han, Lee, and Xu, 2021).

Unlike work focusing on independent or time series data, a general drawback of spatially oriented research has been the lack of general unified theory. Typically, individual papers have studied specific special cases of various spatial specifications. A strand of the literature has introduced the notion of a cross-sectional linear process to help address this problem, and we follow this approach. This representation can accommodate SAR models in the error term (so-called spatial error models (SEMs)) as a special case, as well as variants like SARMA and MESS, whence its generality is apparent. The linear-process structure shares some similarities with that familiar from the time series literature (see, e.g., Hannan, 1970). Indeed, time series versions may be regarded as very special cases, but, as stressed before, the features of spatial dependence must be taken into account in the general formulation. Such a representation was introduced by Robinson (2011) and further examined in other situations by Robinson and Thawornkaiwong (2012) (partially linear regression), Delgado and Robinson (2015) (nonnested correlation testing), Lee and Robinson (2016) (series estimation of nonparametric regression), and Hidalgo and Schafgans (2017) (cross-sectionally dependent panels).

In this paper, we propose a test statistic similar to that of Hong and White (1995), based on estimating the nonparametric specification via series approximations. Assuming an independent and identically distributed (i.i.d.) sample, their statistic is based on the sample covariance between the residual from the parametric model and the discrepancy between the parametric and nonparametric fitted values. Allowing additionally for spatial dependence through the form of a linear process as discussed above, our statistic is shown to be asymptotically standard normal, consistent and possessing nontrivial power against local alternatives of a certain type. To prove asymptotic normality, we present a new central limit theorem (CLT) for quadratic forms in linear processes in an increasing dimension setting that may be of independent interest. A CLT for quadratic forms under time series dependence in the context of series estimation can be found in Gao and Anh (2000), and our result can be viewed as complementary to this. The setting of Su and Qu (2017) is a very special case of our framework. There has been recent interest in specification testing for spatial models (see, for example, Sun, 2020, for a kernel-based model specification test and Lee, Phillips, and Rossi, 2020, for a consistent omnibus test). We contribute to this literature by studying a linear process-based increasing parameter dimension framework.

Our linear process framework permits spatial dependence to be parametric, parametric with increasing dimension, semiparametric, or any combination thereof, thus covering a vast variety of settings. A class of models of great empirical interest are "higher-order" SAR models in the outcome variables, but with spatial dependence structure also in the errors. We initially present the familiar nonparametric regression to clarify the exposition, and then cover this class as the main model of interest. Our theory covers as special cases SAR, SMA, SARMA, and MESS models for the error term. These specifications may be of any fixed spatial order, but our theory also covers the case where they are of increasing order.

Thus, we permit a more complex model of spatial dependence as more data become available, which encourages a more flexible approach to modeling such dependence as stressed by Gupta and Robinson (2015, 2018) in a higher-order SAR context, Huber (1973), Portnoy (1984, 1985), and Anatolyev (2012) in a regression context, and Koenker and Machado (1999) for the generalized method of moments setting, among others. This literature focuses on a sequence of true models, rather than a sequence of models approximating an infinite true model. Our paper also takes the same approach. On the other hand, in the spatial setting, Gupta (2018a) considers increasing lag models as approximations to an infinite lag model with lattice data and also suggests criteria for choice of lag length.

Our framework is also extended to the situation where spatial dependence occurs through nonparametric functions of raw distances (these may be exogenous economic or social distances, say), as in Pinkse et al. (2002). This allows for greater flexibility in modeling spatial weights as the practitioner only has to choose an exogenous economic distance measure and allow the data to determine the functional form. It also adds a degree of robustness to the theory by avoiding potential parametric misspecification. The case of geographical data is also covered, for example, the important classes of Matérn and Wendland (see, e.g., Gneiting, 2002) covariance functions. Finally, we introduce a new notion of smooth spatial dependence that provides more primitive, and checkable, conditions for certain properties than extant ones in the literature.

To illustrate the performance of the test in finite samples, we present Monte Carlo simulations that exhibit satisfactory small sample properties. The test is demonstrated in three empirical examples, including two based on recently published work on social networks: Bloom et al. (2013) (R&D spillovers in innovation) and König et al. (2017) (conflict alliances during the Congolese civil war). Another example studies cross-country spillovers in economic growth. Our test may or may not reject the null hypothesis of a linear regression in these examples, illustrating its ability to distinguish well between the null and alternative models.

The next section introduces our basic setup using a nonparametric regression with no SAR structure in responses. We treat this abstraction as a base case, and Section 3 discusses estimation and defines the test statistic, whereas Section 4 introduces assumptions and the key asymptotic results of the paper. Section 5 examines the most commonly employed higher-order SAR models, whereas Section 6 deals with nonparametric spatial error structures. Nonparametric specification tests are often criticized for poor finite sample performance when using the asymptotic critical values. In Section 7, we present a bootstrap version of our testing procedure. Sections 8 and 9 contain a study of finite sample performance and the empirical examples, respectively, whereas Section 10 concludes. Proofs are contained in the Appendix as well as in the Supplementary Material, which also contains additional simulation results.

For the convenience of the reader, we collect some frequently used notation here. First, we introduce three notational conventions for any parameter ν for the rest of the paper: $\nu \in \mathbb{R}^{d_\nu}$, ν_0 denotes the true value of ν and for any scalar, vector or matrix valued function $f(\nu)$, we denote $f \equiv f(\nu_0)$. Let $\overline{\varphi}(\cdot)$ (resp. $\underline{\varphi}(\cdot)$) denote the largest (resp. smallest) eigenvalue of a generic square nonnegative definite matrix argument. For a generic matrix A, denote $\|A\| = \left[\overline{\varphi}(A'A)\right]^{1/2}$, i.e., the

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

spectral norm of A which reduces to the euclidean norm if A is a vector. $\|A\|_R$ denotes the maximum absolute row sum norm of a generic matrix A, whereas $\|A\|_F = \left[tr(AA')\right]^{1/2}$, the Frobenius norm. Throughout the paper, $|\cdot|$ is an absolute value when applied to a scalar and determinant when applied to a matrix. Denote by c (C) generic positive constants, independent of any quantities that tend to infinity, and arbitrarily small (big).

2. SETUP

To illustrate our approach, we first consider the nonparametric regression

$$y_i = \theta_0(x_i) + u_i, i = 1, \ldots, n, \tag{2.1}$$

where $\theta_0(\cdot)$ is an unknown function and x_i is a vector of strictly exogenous explanatory variables with support $\mathcal{X} \subset \mathbb{R}^k$. Spatial dependence is explicitly modeled via the error term u_i, which we assume is generated by

$$u_i = \sum_{s=1}^{\infty} b_{is}\varepsilon_s, \tag{2.2}$$

where ε_s are independent random variables, with zero mean and identical variance σ_0^2. Further conditions on the ε_s will be assumed later. The linear process coefficients b_{is} can depend on n, as may the covariates x_i. This is generally the case with spatial models and implies that asymptotic theory ought to be developed for triangular arrays. There are a number of reasons to permit dependence on sample size. The b_{is} can depend on spatial weight matrices, which are usually normalized for both stability and identification purposes.

Such normalizations, e.g., row standardization or division by spectral norm, may be n-dependent. Furthermore, x_i often includes underlying covariates of “neighbors” defined by spatial weight matrices. For instance, for some $n \times 1$ covariate vector z and exogenous spatial weight matrix $W \equiv W_n$, a component of x_i can be $e_i'Wz$, where e_i has unity in the ith position and zeros elsewhere, which depends on n. Thus, subsequently, any spatial weight matrices will also be allowed to depend on n. Finally, treating triangular arrays permits relabeling of quantities that is often required when dealing with spatial data, due to the lack of natural ordering (see, e.g., Robinson, 2011). We suppress explicit reference to this n-dependence of various quantities for brevity, although mention will be made of this at times to remind the reader of this feature.

Now, assume the existence of a $d_\gamma \times 1$ vector γ_0 such that $b_{is} = b_{is}(\gamma_0)$, possibly with $d_\gamma \to \infty$ as $n \to \infty$, for all $i = 1, \ldots, n$ and $s \geq 1$. Let u be the $n \times 1$ vector with typical element u_i, let ε be the infinite-dimensional vector with typical element ε_s, and let B be an infinite-dimensional matrix (Cooke, 1950) with typical element b_{is}. In matrix form,

$$u = B\varepsilon \text{ and } \mathcal{E}\left(uu'\right) = \sigma_0^2 BB' = \sigma_0^2\Sigma \equiv \sigma_0^2\Sigma(\gamma_0). \tag{2.3}$$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

We assume that $\gamma_0 \in \Gamma$, where Γ is a compact subset of $\mathbb{R}^{d_\gamma}$. With d_γ diverging, ensuring Γ has bounded volume requires some care (see Gupta and Robinson, 2018). For a known function $f(\cdot)$, our aim is to test

$$H_0 : P[\theta_0(x_i) = f(x_i, \alpha_0)] = 1, \text{ for some } \alpha_0 \in \mathcal{A} \subset \mathbb{R}^{d_\alpha}, \tag{2.4}$$

against the global alternative $H_1 : P[\theta_0(x_i) \neq f(x_i, \alpha)] > 0$, for all $\alpha \in \mathcal{A}$.

We now nest commonly used models for spatial dependence in (2.3). Introduce a set of $n \times n$ spatial weight (equivalently network adjacency) matrices W_j, $j = 1, \ldots, m_1 + m_2$. Each W_j can be thought of as representing dependence through a particular space. Now, consider models of the form $\Sigma(\gamma) = A^{-1}(\gamma)A'^{-1}(\gamma)$. For example, with ξ denoting a vector of i.i.d. disturbances with variance σ_0^2, the model with SARMA(m_1, m_2) errors is $u = \sum_{j=1}^{m_1} \gamma_j W_j u + \sum_{j=m_1+1}^{m_1+m_2} \gamma_j W_j \xi + \xi$, with $A(\gamma) = \left(I_n + \sum_{j=m_1+1}^{m_1+m_2} \gamma_j W_j\right)^{-1} \left(I_n - \sum_{j=1}^{m_1} \gamma_j W_j\right)$, assuming conditions that guarantee the existence of the inverse. Such conditions can be found in the literature (see, e.g., Lee and Liu, 2010; Gupta and Robinson, 2018). The SEM model is obtained by setting $m_2 = 0$, whereas the model with SMA errors has $m_1 = 0$. The model with MESS(m) errors (LeSage and Pace, 2007; Debarsy, Jin, and Lee, 2015) is $u = \exp\left(\sum_{j=1}^{m} \gamma_j W_j\right) \xi, A(\gamma) = \exp\left(-\sum_{j=1}^{m} \gamma_j W_j\right)$.

In some cases, the space under consideration is geographic, i.e., the data may be observed at irregular points in euclidean space. Making the identification $u_i \equiv U(t_i)$, $t_i \in \mathbb{R}^d$, for some $d > 1$, and assuming covariance stationarity, $U(t)$ is said to follow an isotropic model if, for some function δ on $\mathbb{R}$, the covariance at lag s is $r(s) = \mathcal{E}[U(t)U(t+s)] = \delta(\|s\|)$. An important class of parametric isotropic models is that of Matérn (1986), which can be parameterized in several ways (see, e.g., Stein, 1999). Denoting by Γ_f the Gamma function and by $\mathcal{K}_{\gamma_1}$ the modified Bessel function of the second kind (Gradshteyn and Ryzhik, 1994), take $\delta(\|s\|, \gamma) = \left(2^{\gamma_1 - 1}\Gamma_f(\gamma_1)\right)^{-1} \left(\gamma_2^{-1}\sqrt{2\gamma_1}\,\|s\|\right)^{\gamma_1} \mathcal{K}_{\gamma_1}\left(\gamma_2^{-1}\sqrt{2\gamma_1}\,\|s\|\right)$, with $\gamma_1, \gamma_2 > 0$ and $d_\gamma = 2$. With $d_\gamma = 3$, another model takes $\delta(\|s\|, \gamma) = \gamma_1 \exp\left(-\|s/\gamma_2\|^{\gamma_3}\right)$ (see, e.g., De Oliveira, Kedem, and Short, 1997; Stein, 1999). Fuentes (2007) considers this model with $\gamma_3 = 1$, as well as a specific parameterization of the Matèrn covariance function.

3. TEST STATISTIC

We estimate $\theta_0(\cdot)$ via a series approximation. Certain technical conditions are needed to allow for $\mathcal{X}$ to have unbounded support. To this end, for a function $g(x)$ on $\mathcal{X}$, define a weighted sup-norm (see, e.g., Chen, Hong, and Tamer, 2005; Chen, 2007; Lee and Robinson, 2016) by $\|g\|_w = \sup_{x \in \mathcal{X}} |g(x)| \left(1 + \|x\|^2\right)^{-w/2}$, for some $w > 0$. Assume that there exists a sequence of functions $\psi_i := \psi(x_i) : \mathbb{R}^k \mapsto \mathbb{R}^p$, where $p \to \infty$ as $n \to \infty$, and a $p \times 1$ vector of coefficients β_0 such that

$$\theta_0(x_i) = \psi_i' \beta_0 + e(x_i), \tag{3.1}$$

where $e(\cdot)$ satisfies the following.

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Assumption R.1. There exists a constant $\mu > 0$ such that $\|e\|_{w_x} = O\left(p^{-\mu}\right)$, as $p \to \infty$, where $w_x \geq 0$ is the largest value such that $\sup_{i=1,\ldots,n} \mathcal{E}\|x_i\|^{w_x} < \infty$, for all n.

By Lemma 1 in Appendix B of Lee and Robinson (2016), this assumption implies that

$$\sup_{i=1,\ldots,n} \mathcal{E}\left(e^2(x_i)\right) = O\left(p^{-2\mu}\right). \tag{3.2}$$

Due to the large number of assumptions in the paper, sometimes with changes reflecting only the various setups we consider, we prefix assumptions with R in this section and the next, to signify "regression" In Section 5, the prefix is SAR, for "spatial autoregression," whereas in Section 6 we use NPN, for "nonparametric network."

Let $y = (y_1, \ldots, y_n)', \theta_0 = (\theta_0(x_1), \ldots, \theta_0(x_n))', \Psi = (\psi_1, \ldots, \psi_n)'$. We will estimate γ_0 using a quasi-maximum likelihood estimator (QMLE) based on a Gaussian likelihood, although Gaussianity is nowhere assumed. For any admissible values β, σ^2, and γ, the (multiplied by $\frac{2}{n}$) negative quasi-log-likelihood function based on using the approximation (3.1) is

$$L(\beta,\sigma^2,\gamma) = \ln\left(2\pi\sigma^2\right) + \frac{1}{n}\ln|\Sigma(\gamma)| + \frac{1}{n\sigma^2}(y-\Psi\beta)'\Sigma(\gamma)^{-1}(y-\Psi\beta), \tag{3.3}$$

which is minimized with respect to β and σ^2 by

$$\bar{\beta}(\gamma) = \left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}y, \tag{3.4}$$

$$\bar{\sigma}^2(\gamma) = n^{-1}y'E(\gamma)'M(\gamma)E(\gamma)y, \tag{3.5}$$

where $M(\gamma) = I_n - E(\gamma)\Psi\left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'E(\gamma)'$ and $E(\gamma)$ is the $n \times n$ symmetric matrix such that $E(\gamma)E(\gamma)' = \Sigma(\gamma)^{-1}$. The use of the approximate likelihood relies on the negligibility of $e(\cdot)$, which in turn permits the replacement of $\theta_0(\cdot)$ by $\psi'\beta_0$ with asymptotically negligible cost. Thus, the concentrated likelihood function is

$$\mathcal{L}(\gamma) = \ln(2\pi) + \ln\bar{\sigma}^2(\gamma) + \frac{1}{n}\ln|\Sigma(\gamma)|. \tag{3.6}$$

We define the QMLE of γ_0 as $\widehat{\gamma} = \arg\min_{\gamma\in\Gamma}\mathcal{L}(\gamma)$ and the QMLEs of β_0 and σ_0^2 as $\widehat{\beta} = \bar{\beta}(\widehat{\gamma})$ and $\widehat{\sigma}^2 = \bar{\sigma}^2(\widehat{\gamma})$. At a given $x_1, \ldots, x_n$, the series estimate of θ_0 is defined as

$$\widehat{\theta} = \left(\hat{\theta}(x_1), \ldots, \hat{\theta}(x_n)\right)' = \left(\psi(x_1)'\widehat{\beta}, \ldots, \psi(x_n)'\widehat{\beta}\right)'. \tag{3.7}$$

Let $\widehat{\alpha}_n \equiv \widehat{\alpha}$ denote an estimator consistent for α_0 under H_0, for example, the (nonlinear) least-squares estimator. Note that $\widehat{\alpha}$ is consistent only under H_0, so we introduce a general probability limit of $\widehat{\alpha}$, as in Hong and White (1995).

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Assumption R.2. There exists a deterministic sequence $\alpha_n^* \equiv \alpha^*$ such that $\widehat{\alpha} - \alpha^* = O_p\left(1/\sqrt{n}\right)$.

Examples of estimators that satisfy this assumption include (nonlinear) least squares, generalized method of moments estimators, or adaptive efficient weighted least squares (Stinchcombe and White, 1998).

Following Hong and White (1995), define the regression error $u_i \equiv y_i - f(x_i, \alpha^*)$ and the specification error $v_i \equiv \theta_0(x_i) - f(x_i, \alpha^*)$. Our test statistic is based on a scaled and centered version of $\widehat{m}_n = \widehat{\sigma}^{-2}\widehat{v}'\Sigma(\widehat{\gamma})^{-1}\widehat{u}/n = \widehat{\sigma}^{-2}\left(\widehat{\theta} - f(x,\widehat{\alpha})\right)'\Sigma(\widehat{\gamma})^{-1}(y - f(x,\widehat{\alpha}))/n$, where $f(x,\alpha) = (f(x_1,\alpha),\ldots,f(x_n,\alpha))'$. Precisely, it is defined as

$$\mathscr{T}_n = \frac{n\widehat{m}_n - p}{\sqrt{2p}}. \tag{3.8}$$

The motivation for such a centering and scaling stems from the fact that, for fixed p, $n\widehat{m}_n$ has an asymptotic χ_p^2 distribution. Such a distribution has mean p and variance $2p$, and it is a well-known fact that $\left(\chi_p^2 - p\right)/\sqrt{2p} \xrightarrow{d} N(0,1)$, as $p \to \infty$. This motivates our use of (3.8) and explains why we aspire to establish a standard normal distribution under the null hypothesis. Intuitively, the test statistic is based on the sample covariance between the residual from the parametric model and the discrepancy between the parametric and nonparametric fitted values, as in Hong and White (1995).

Hong and White (1995) also note that, due to the nonparametric nature of the problem, such a statistic vanishes faster than the parametric ($n^{\frac{1}{2}}$) rate, and thus an $n^{\frac{1}{2}}$-normalization leads to degeneracy of the test. A proper normalization as in (3.8) will yield a nondegenerate limiting distribution. As Hong and White (1995) noted, our test is one-sided. This is because asymptotically negative values of our test statistic can occur only under the null, whereas under the alternative it tends to a positive, increasing number. Thus, we reject the null if our test statistic is on the right tail.

4. ASYMPTOTIC THEORY

4.1. Consistency of $\widehat{\gamma}$

We first provide conditions under which our estimator $\widehat{\gamma}$ of γ_0 is consistent. Such a property is necessary for the results that follow. The following assumption is a rather standard type of asymptotic boundedness and full-rank condition on $\Sigma(\gamma)$.

Assumption R.3.

$$\overline{\lim_{n\to\infty}} \sup_{\gamma\in\Gamma} \bar{\varphi}(\Sigma(\gamma)) < \infty \text{ and } \underline{\lim}_{n\to\infty} \inf_{\gamma\in\Gamma} \underline{\varphi}(\Sigma(\gamma)) > 0.$$

Assumption R.4. The $u_i, i = 1,\ldots,n,$ satisfy the representation (2.2). The ε_s, $s \geq 1$, have zero-mean, finite third and fourth moments μ_3 and μ_4, respectively,

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

and, denoting by $\sigma_{ij}(\gamma)$ the (i,j)th element of $\Sigma(\gamma)$ and defining $b^*_{is} = b_{is}/\sigma_{ii}^{\frac{1}{2}}$, $i = 1,\ldots,n,\ n \geq 1, s \geq 1,$ we have

$$\overline{\lim_{n\to\infty}} \sup_{i=1,\ldots,n} \sum_{s=1}^{\infty} \left|b^*_{is}\right| + \sup_{s\geq 1} \overline{\lim_{n\to\infty}} \sum_{i=1}^{n} \left|b^*_{is}\right| < \infty. \tag{4.1}$$

By Assumption R.3, σ_{ii} is bounded and bounded away from zero, so the normalization of the b_{is} in Assumption R.4 is well defined. The summability conditions in (4.1) are typical conditions on linear process coefficients that are needed to control dependence; for instance, in the case of stationary time series $b^*_{is} = b^*_{i-s}$. The infinite linear process assumed in (2.2) is further discussed by Robinson (2011), who introduced it, and also by Delgado and Robinson (2015). These assumptions imply an increasing-domain asymptotic setup and preclude infill asymptotics.

Because we often need to consider the difference between values of the matrix-valued function $\Sigma(\cdot)$ at distinct points, it is useful to introduce an appropriate concept of "smoothness." This concept has been employed before in economics (see, e.g., Chen, 2007), and is defined below.

Definition 1. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces, let $\mathscr{L}(X,Y)$ be the Banach space of linear continuous maps from X to Y with norm $\|T\|_{\mathscr{L}(X,Y)} = \sup_{\|x\|_X\leq 1} \|T(x)\|_Y$, and let U be an open subset of X. A map $F : U \to Y$ is said to be Fréchet-differentiable at $u \in U$ if there exists $L \in \mathscr{L}(X,Y)$ such that

$$\lim_{\|h\|_X\to 0} \frac{F(u+h) - F(u) - L(h)}{\|h\|_X} = 0. \tag{4.2}$$

L is called the Fréchet-derivative of F at u. The map F is said to be Fréchet-differentiable on U if it is Fréchet-differentiable for all $u \in U$.

The above definition extends the notion of a derivative that is familiar from real analysis to functional spaces and allows us to check high-level assumptions that past literature has imposed. To the best of our knowledge, this is the first use of such a concept in the literature on spatial/network models. Denote by $\mathcal{M}^{n\times n}$ the set of real, symmetric, and positive semidefinite $n \times n$ matrices. Let Γ^o be an open subset of Γ and consider the Banach spaces $\left(\Gamma, \|\cdot\|_g\right)$ and $\left(\mathcal{M}^{n\times n}, \|\cdot\|\right)$, where $\|\cdot\|_g$ is a generic ℓ_p norm, $p \geq 1$. The following assumption ensures that $\Sigma(\cdot)$ is a "smooth" function, in the sense of Fréchet-smoothness.

Assumption R.5. The map $\Sigma : \Gamma^o \to \mathcal{M}^{n\times n}$ is Fréchet-differentiable on Γ^o with Fréchet-derivative denoted by $D\Sigma \in \mathscr{L}\left(\Gamma^o, \mathcal{M}^{n\times n}\right)$. Furthermore, the map $D\Sigma$ satisfies

$$\sup_{\gamma\in\Gamma^o} \|D\Sigma(\gamma)\|_{\mathscr{L}(\Gamma^o,\mathcal{M}^{n\times n})} \leq C. \tag{4.3}$$

Assumption R.5 is a functional smoothness condition on spatial dependence. It has the advantage of being checkable for a variety of commonly employed models. For example, a first-order SEM has $\Sigma(\gamma) = A^{-1}(\gamma)A'^{-1}(\gamma)$ with $A =$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

$I_n - \gamma W$. Corollary CS.1 in the Supplementary Material shows $(D\Sigma(\gamma))\left(\gamma^{\dagger}\right) = \gamma^{\dagger}A^{-1}(\gamma)\left(G'(\gamma)+G(\gamma)\right)A'^{-1}(\gamma)$, at a given point $\gamma \in \Gamma^o$, where $G(\gamma) = WA^{-1}(\gamma)$. Then, taking

$$\|W\| + \sup_{\gamma\in\Gamma}\left\|A^{-1}(\gamma)\right\| < C \quad \textbf{(4.4)}$$

yields Assumption R.5. Condition (4.4) limits the extent of spatial dependence and is very standard in the spatial literature (see, e.g., Lee, 2004, and numerous subsequent papers employing similar conditions).

Fréchet derivatives for higher-order SAR, SMA, SARMA, and MESS error structures are computed in Lemmas LS.5 and LS.6 and Corollaries CS.1 and CS.2 in Appendix S.D. of the Supplementary Material. Strictly speaking, Gâteaux differentiability might suffice for the type of results that we target. We opt for Fréchet differentiability because this derivative map is linear and continuous or, equivalently, a bounded linear operator, a property that makes Assumption R.5 more reasonable.

The following proposition is very useful in "linearizing" perturbations in the $\Sigma(\cdot)$.

PROPOSITION 4.1. *If Assumption R.5 holds, then, for any $\gamma_1, \gamma_2 \in \Gamma^o$,*

$$\|\Sigma(\gamma_1) - \Sigma(\gamma_2)\| \le C\|\gamma_1 - \gamma_2\|. \quad \textbf{(4.5)}$$

To illustrate how the concept of Fréchet differentiability allows us to check high-level assumptions extant in the literature, a consequence of Proposition 4.1 is the following corollary, a version of which appears as an assumption in Delgado and Robinson (2015).

COROLLARY 4.1. *For any $\gamma^* \in \Gamma^o$ and any $\eta > 0$,*

$$\overline{\lim_{n\to\infty}} \sup_{\gamma\in\{\gamma:\|\gamma-\gamma^*\|<\eta\}\cap\Gamma^o} \left\|\Sigma(\gamma) - \Sigma\left(\gamma^*\right)\right\| < C\eta. \quad \textbf{(4.6)}$$

We now introduce regularity conditions needed to establish the consistency of $\hat{\gamma}$. Define

$$\sigma^2(\gamma) = n^{-1}\sigma^2 tr\left(\Sigma(\gamma)^{-1}\Sigma\right) = n^{-1}\sigma^2\left\|E(\gamma)E^{-1}\right\|_F^2,$$

which is nonnegative by definition and bounded by Assumption R.3, red with the matrix $E(\gamma)$ defined after (3.5).

Assumption R.6. $c \le \sigma^2(\gamma) \le C$ for all $\gamma \in \Gamma$.

Assumption R.7. $\gamma_0 \in \Gamma$ and, for any $\eta > 0$,

$$\underline{\lim}_{n\to\infty} \inf_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)} \frac{n^{-1}tr\left(\Sigma(\gamma)^{-1}\Sigma\right)}{\left|\Sigma(\gamma)^{-1}\Sigma\right|^{1/n}} > 1, \quad \textbf{(4.7)}$$

where $\overline{\mathcal{N}}^{\gamma}(\eta) = \Gamma \setminus \mathcal{N}^{\gamma}(\eta)$ and $\mathcal{N}^{\gamma}(\eta) = \{\gamma : \|\gamma - \gamma_0\| < \eta\} \cap \Gamma$.

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Assumption R.8. $\left\{\underline{\varphi}\left(n^{-1}\Psi'\Psi\right)\right\}^{-1}+\overline{\varphi}\left(n^{-1}\Psi'\Psi\right)=O_p(1)$.

Assumption R.6 is a boundedness condition originally considered in Gupta and Robinson (2018), whereas Assumptions R.7 and R.8 are identification conditions. Indeed, Assumption R.7 requires that $\Sigma(\gamma)$ be identifiable in a small neighborhood around γ_0. This is apparent on noticing that the ratio in (4.7) is at least one by the inequality between arithmetic and geometric means, and equals one when $\Sigma(\gamma)=\Sigma$. Similar assumptions arise frequently in related literature (see, e.g., Lee, 2004; Delgado and Robinson, 2015). Assumption R.8 is a typical asymptotic boundedness and non-multicollinearity condition (see, e.g., Newey, 1997, and much other literature on series estimation). Primitive conditions for this assumption to hold require the convergence (in matrix norm) of $n^{-1}\Psi'\Psi$ to its expectation, and this entails restrictions on the extent of spatial dependence in the x_i. A reference is Lee and Robinson (2016), wherein consider Assumption A.4 and the proof of Theorem 1. By Assumption R.3, Assumption R.8 implies $\sup_{\gamma\in\Gamma}\left\{\underline{\varphi}\left(n^{-1}\Psi'\Sigma(\gamma)^{-1}\Psi\right)\right\}^{-1}=O_p(1)$.

THEOREM 4.1. *Under either H_0 or H_1, Assumptions R.1–R.8, and $p^{-1}+\left(d_\gamma+p\right)/n\to 0$ as $n\to\infty$, $\left\|\left(\widehat{\gamma},\hat{\sigma}^2\right)-\left(\gamma_0,\sigma_0^2\right)\right\|\xrightarrow{p}0$.*

4.2. Asymptotic Properties of the Test Statistic

Write $\Sigma_j(\gamma)=\partial\Sigma(\gamma)/\partial\gamma_j$, $j=1,\ldots,d_\gamma$, the matrix differentiated elementwise. While Assumption R.5 guarantees that these partial derivatives exist, the next assumption imposes a uniform bound on their spectral norms.

Assumption R.9. $\overline{\lim}_{n\to\infty}\sup_{j=1,\ldots,d_\gamma}\left\|\Sigma_j(\gamma)\right\|<C$.

We will later consider the sequence of local alternatives

$$H_{\ell n}\equiv H_\ell : f(x_i,\alpha_n^*)=\theta_0(x_i)+(p^{1/4}/n^{1/2})h(x_i), a.s., \tag{4.8}$$

where h is square-integrable on the support $\mathcal{X}$ of the x_i. Under the null H_0, we have $h(x_i)=0$, a.s.

Assumption R.10. For each $n\in\mathbb{N}$ and $i=1,\ldots,n$, the function $f:\mathcal{X}\times\mathcal{A}\to\mathbb{R}$ such that $f(x_i,\alpha)$ is measurable for each $\alpha\in\mathcal{A}$, $f(x_i,\cdot)$ is a.s. continuous on $\mathcal{A}$, with $\sup_{\alpha\in\mathcal{A}}f^2(x_i,\alpha)\le D_n(x_i)$, where $\sup_{n\in\mathbb{N}}D_n(x_i)$ is integrable and $\sup_{\alpha\in\mathcal{A}}\|\partial f(x_i,\alpha)/\partial\alpha\|^2\le D_n(x_i)$, $\sup_{\alpha\in\mathcal{A}}\left\|\partial^2 f(x_i,\alpha)/\partial\alpha\partial\alpha'\right\|\le D_n(x_i)$, all holding a.s.

Define the infinite-dimensional matrix $\mathscr{V}=B'\Sigma^{-1}\Psi\left(\Psi'\Sigma^{-1}\Psi\right)^{-1}\Psi'\Sigma^{-1}B$, which is symmetric, idempotent, and has rank p. We now show that our test statistic is approximated by a quadratic form in ε, weighted by $\mathscr{V}$.

THEOREM 4.2. *Under Assumptions R.1–R.10, $p^{-1}+p\left(p+d_\gamma^2\right)/n+\sqrt{n}/p^{\mu+1/4}\to 0$, as $n\to\infty$, and H_0, $\mathscr{T}_n-\left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon-p\right)/\sqrt{2p}=o_p(1)$.*

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Assumption R.11. $\overline{\lim}_{n\to\infty} \left\|\Sigma^{-1}\right\|_R < \infty$.

Because $\left\|\Sigma^{-1}\right\| \leq \left\|\Sigma^{-1}\right\|_R$, this restriction on spatial dependence is somewhat stronger than a restriction on spectral norm but is typically imposed for CLTs in this type of setting (cf. Lee, 2004; Delgado and Robinson, 2015; Gupta and Robinson, 2018). The next assumption is needed in our proofs to check a Lyapunov condition. A typical approach would be to assume moments of order $4+\epsilon$, for some $\epsilon > 0$. Due to the linear process structure under consideration, taking $\epsilon = 4$ makes the proof tractable (see, for example, Delgado and Robinson, 2015).

Assumption R.12. The ε_s, $s \geq 1$, have finite eighth moment.

The next assumption is strong if the basis functions $\psi_{ij}(\cdot)$ are polynomials, requiring all moments to exist in that case.

Assumption R.13. $\mathcal{E}\left|\psi_{ij}(x)\right| < C$, $i = 1,\ldots,n$ and $j = 1,\ldots,p$.

The next theorem establishes the asymptotic normality of the approximating quadratic form introduced above.

THEOREM 4.3. *Under Assumptions R.3, R.4, R.8, and R.11–R.13 and* $p^{-1} + p^3/n \to 0$, *as* $n \to \infty$, $\left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon - p\right)/\sqrt{2p} \xrightarrow{d} N(0,1)$.

This is a new type of CLT, integrating both a linear process framework and an increasing dimension element. A linear-quadratic form in i.i.d. disturbances is treated by Kelejian and Prucha (2001), whereas a quadratic form in a linear process framework is treated by Delgado and Robinson (2015). However, both results are established in a parametric framework, entailing no increasing dimension aspect of the type we face with $p \to \infty$.

Next, we summarize the properties of our test statistic in a theorem that records its asymptotic normality under the null, consistency, and ability to detect local alternatives at $p^{1/4}/n^{1/2}$ rate. This rate has been found also by De Jong and Bierens (1994) and Gupta (2018b). Introduce the quantity $\varkappa = \left(\sqrt{2}\sigma_0^2\right)^{-1} \operatorname{plim}_{n\to\infty} n^{-1}h'\Sigma^{-1}h$, where $h = (h(x_1),\ldots,h(x_n))'$ and $h(x_i)$ is from (4.8).

THEOREM 4.4. *Under the conditions of Theorems 4.2 and 4.3, (1)* $\mathscr{T}_n \xrightarrow{d} N(0,1)$ *under* H_0, *(2)* $\mathscr{T}_n$ *is a consistent test statistic, and (3)* $\mathscr{T}_n \xrightarrow{d} N(\varkappa,1)$ *under local alternatives* H_ℓ.

5. MODELS WITH SAR STRUCTURE IN RESPONSES

We now introduce the SAR model

$$y_i = \sum_{j=1}^{d_\lambda} \lambda_{0j} w_{i,j}' y + \theta_0(x_i) + u_i, i = 1,\ldots,n, \tag{5.1}$$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

where $W_j, j = 1, \ldots, d_\lambda$, are known spatial weight matrices with ith rows denoted by $w'_{i,j}$, as discussed earlier, and λ_{0j} are unknown parameters measuring the strength of spatial dependence. We take d_λ to be fixed for convenience of exposition. The error structure remains the same as in (2.2). Here, spatial dependence arises not only in errors but also in responses. For example, this corresponds to a situation where agents in a network influence each other both in their observed and unobserved actions. Note that the error term u_i can be generated by the same W_j, or different ones.

While the model in (5.1) is new in the literature, some related ones are discussed here. Models such as (5.1) but without dependence in the error structure are considered by Su and Jin (2010) and Gupta and Robinson (2015, 2018), but the former consider only $d_\lambda = 1$ and the latter only parametric $\theta_0(\cdot)$. Linear $\theta_0(\cdot)$ and $d_\lambda > 1$ are permitted by Lee and Liu (2010), but the dependence structure in errors differs from what we allow in (5.1). Using the same setup as Su and Jin (2010) and independent disturbances, a specification test for the linearity of $\theta_0(\cdot)$ is proposed by Su and Qu (2017). In comparison, our model is much more general and our test can handle more general parametric null hypotheses. We thank a referee for pointing out that (5.1) is a particular case of Sun (2016) when u_i are i.i.d. and of Malikov and Sun (2017) when $d_\lambda = 1$.

Denoting $S(\lambda) = I_n - \sum_{j=1}^{d_\lambda} \lambda_j W_j$, the quasi-likelihood function based on Gaussianity and conditional on covariates is

$$L(\beta, \sigma^2, \phi) = \log(2\pi\sigma^2) - \frac{2}{n}\log|S(\lambda)| + \frac{1}{n}\log|\Sigma(\gamma)| + \frac{1}{\sigma^2 n}(S(\lambda)y - \Psi\beta)'\Sigma(\gamma)^{-1}(S(\lambda)y - \Psi\beta), \tag{5.2}$$

at any admissible point $(\beta', \phi', \sigma^2)'$ with $\phi = (\lambda', \gamma')'$, for nonsingular $S(\lambda)$ and $\Sigma(\gamma)$. For given $\phi = (\lambda', \gamma')'$, (5.2) is minimized with respect to β and σ^2 by

$$\bar{\beta}(\phi) = \left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}S(\lambda)y, \tag{5.3}$$

$$\bar{\sigma}^2(\phi) = n^{-1}y'S'(\lambda)E(\gamma)'M(\gamma)E(\gamma)S(\lambda)y. \tag{5.4}$$

The QMLE of ϕ_0 is $\widehat{\phi} = \arg\min_{\phi\in\Phi}\mathcal{L}(\phi)$, where

$$\mathcal{L}(\phi) = \log\bar{\sigma}^2(\phi) + n^{-1}\log\left|S'^{-1}(\lambda)\Sigma(\gamma)S^{-1}(\lambda)\right|, \tag{5.5}$$

and $\Phi = \Lambda \times \Gamma$ is taken to be a compact subset of $\mathbb{R}^{d_\lambda + d_\gamma}$. The QMLEs of β_0 and σ_0^2 are defined as $\bar{\beta}(\widehat{\phi}) \equiv \widehat{\beta}$ and $\bar{\sigma}^2(\widehat{\phi}) \equiv \widehat{\sigma}^2$, respectively. The following assumption controls spatial dependence and is discussed below equation (4.4).

Assumption SAR.1. $\max_{j=1,\ldots,d_\lambda}\|W_j\| + \|S^{-1}\| < C$.

Writing $T(\lambda) = S(\lambda)S^{-1}$ and $\phi = (\lambda', \gamma')'$, define the quantity

$$\sigma^2(\phi) = n^{-1}\sigma_0^2 tr\left(T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right) = n^{-1}\sigma_0^2\left\|E(\gamma)T(\lambda)E^{-1}\right\|_F^2,$$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

which is nonnegative by definition and bounded by Assumptions R.3 and SAR.1. The assumptions below directly extend Assumptions R.6 and R.7 to the present setup.

Assumption SAR.2. $c \le \sigma^2(\phi) \le C$, for all $\phi \in \Phi$.

Assumption SAR.3. $\phi_0 \in \Phi$ and, for any $\eta > 0$,

$$\varliminf_{n\to\infty} \inf_{\phi \in \overline{\mathcal{N}}^{\phi}(\eta)} \frac{n^{-1} tr\left(T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right)}{\left|T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right|^{1/n}} > 1, \tag{5.6}$$

where $\overline{\mathcal{N}}^{\phi}(\eta) = \Phi \setminus \mathcal{N}^{\phi}(\eta)$ and $\mathcal{N}^{\phi}(\eta) = \{\phi : \|\phi - \phi_0\| < \eta\} \cap \Phi$.

We now introduce an identification condition that is required in the setup of this section.

Assumption SAR.4. $\beta_0 \neq 0$ and, for any $\eta > 0$,

$$P\left(\varliminf_{n\to\infty} \inf_{(\lambda',\gamma')' \in \Lambda \times \overline{\mathcal{N}}^{\gamma}(\eta)} n^{-1}\beta_0'\Psi' T'(\lambda)E(\gamma)'M(\gamma)E(\gamma)T(\lambda)\Psi\beta_0/\|\beta_0\|^2 > 0\right) = 1. \tag{5.7}$$

Upon performing minimization with respect to β, the event inside the probability in (5.7) is equivalent to the event

$$\varliminf_{n\to\infty} \min_{\beta\in\mathbb{R}^p} \inf_{(\lambda',\gamma')' \in \Lambda \times \overline{\mathcal{N}}^{\gamma}(\eta)} n^{-1}(\Psi\beta - T(\lambda)\Psi\beta_0)'\Sigma(\gamma)^{-1}(\Psi\beta - T(\lambda)\Psi\beta_0)/\|\beta_0\|^2 > 0,$$

which is analogous to the identification condition for the nonlinear regression model with a parametric linear factor in Robinson (1972), weighted by the inverse of the error covariance matrix. This reduces the condition to a scalar form of a rank condition, making the identifying nature of the assumption transparent. A similar identifying assumption is used by Gupta and Robinson (2018).

THEOREM 5.1. *Under either H_0 or H_1, Assumptions R.1–R.5, R.8, and SAR.1–SAR.4 and*

$$p^{-1} + \left(d_\gamma + p\right)/n \to 0, \text{ as } n \to \infty,$$

$$\left\|\left(\widehat{\phi}, \widehat{\sigma}^2\right) - \left(\phi_0, \sigma_0^2\right)\right\| \xrightarrow{p} 0 \text{ as } n \to \infty.$$

The test statistic $\mathscr{T}_n$ can be constructed as before but with the null residuals redefined to incorporate the spatially lagged terms, i.e., $\hat{u} = S(\hat{\lambda})y - f(x,\hat{\alpha})$. Then we have the following theorem.

THEOREM 5.2. *Under Assumptions R.1–R.5, R.8–R.10, and SAR.1–SAR.4,*

$$p^{-1} + p\left(p + d_\gamma^2\right)/n + \sqrt{n}/p^{\mu+1/4} + d_\gamma^2/p \to 0, \text{ as } n \to \infty,$$

and H_0, $\mathscr{T}_n - \left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon - p\right)/\sqrt{2p} = o_p(1)$.

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

THEOREM 5.3. *Under the conditions of Theorems 4.3, 5.1, and 5.2, (1)* $\mathscr{T}_n \xrightarrow{d} N(0,1)$ *under* H_0, *(2)* $\mathscr{T}_n$ *is a consistent test statistic, and (3)* $\mathscr{T}_n \xrightarrow{d} N(\varkappa,1)$ *under local alternatives* H_ℓ.

6. NONPARAMETRIC SPATIAL WEIGHTS

In this section, we are motivated by settings where spatial dependence occurs through nonparametric functions of raw distances (this may be geographic, social, economic, or any other type of distance), as is the case in Pinkse et al. (2002), for example. In their kind of setup, d_{ij} is a raw distance between units i and j and the corresponding element of the spatial weight matrix is given by $w_{ij} = \zeta_0\left(d_{ij}\right)$, where $\zeta_0(\cdot)$ is an unknown nonparametric function. Pinkse et al. (2002) use such a setup in a SAR model like (5.1), but with a linear regression function. In contrast, in keeping with the focus of this paper, we instead model dependence in the errors in this manner. Our formulation is rather general, covering, for example, a specification like $w_{ij} = f\left(\gamma_0, \zeta_0\left(d_{ij}\right)\right)$, with $f(\cdot)$ a *known* function, γ_0 an *unknown* parameter of possibly increasing dimension, and $\zeta_0(\cdot)$ an *unknown* nonparametric function. For the sake of simplicity, we do not permit the x_i in this section to be generated by such nonparametric weight matrices although they can be generated from other, known weight matrices.

Let Ξ be a compact space of functions, on which we will specify more conditions later. For notational simplicity, we abstract away from the SAR dependence in the responses. Thus, we consider (2.1), but with

$$u_i = \sum_{s=1}^{\infty} b_{is}\left(\gamma_0, \zeta_0\left(z_i\right)\right)\varepsilon_s, \tag{6.1}$$

where $\zeta_0(\cdot) = \left(\zeta_{01}(\cdot), \ldots, \zeta_{0d_\zeta}(\cdot)\right)'$ is a fixed-dimensional vector of real-valued nonparametric functions with $\zeta_{0\ell} \in \Xi$ for each $\ell = 1, \ldots, d_\zeta$, and z_i a fixed-dimensional vector of data, independent of the ε_s, $s \geq 1$, with support $\mathcal{Z}$. One can also take z_i to be a fixed distance measure. We base our estimation on approximating each $\zeta_{0\ell}(z_i)$, $\ell = 1, \ldots, d_\zeta$, with the series representation $\delta_{0\ell}'\varphi_\ell(z_i)$, where $\varphi_\ell\left(z_i\right) \equiv \varphi_\ell$ is an $r_\ell \times 1$ ($r_\ell \to \infty$ as $n \to \infty$) vector of basis functions with typical function $\varphi_{\ell k}$, $k = 1, \ldots, r_\ell$. The set of linear combinations $\delta_\ell'\varphi_\ell(z_i)$ forms the sequence of sieve spaces $\Phi_{r_\ell} \subset \Xi$ as $r_\ell \to \infty$, for any $\ell = 1, \ldots, d_\zeta$, and

$$\zeta_{0\ell}\left(z\right) = \delta_{0\ell}'\varphi_\ell + \nu_\ell, \tag{6.2}$$

with the following restriction on the function space Ξ.

Assumption NPN.1. For some scalars $\kappa_\ell > 0$, $\|\nu_\ell\|_{w_z} = O\left(r_\ell^{-\kappa_\ell}\right)$, as $r_\ell \to \infty$, $\ell = 1, \ldots, d_\zeta$, where $w_z \geq 0$ is the largest value such that $\sup_{z\in\mathcal{Z}} \mathcal{E}\|z\|^{w_z} < \infty$.

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Just as Assumption R.1 implied (3.2), by Lemma 1 of Lee and Robinson (2016), we obtain

$$\sup_{z\in\mathcal{Z}} \mathcal{E}\left(\nu_\ell^2\right) = O\left(r_\ell^{-2\kappa_\ell}\right), \ell = 1,\ldots,d_\zeta. \tag{6.3}$$

Thus, we now have an infinite-dimensional nuisance parameter $\zeta_0(\cdot)$ and increasing-dimensional nuisance parameter γ. Writing $\sum_{\ell=1}^{d_\zeta} r_\ell = r$ and $\tau = (\gamma', \delta_1', \ldots, \delta_{d_\zeta}')'$, which has increasing dimension $d_\tau = d_\gamma + r$, define $\varsigma(r) = \sup_{z\in\mathcal{Z};\,\ell=1,\ldots,d_\zeta} \|\varphi_\ell\|$. Write $\Sigma(\tau)$ for the covariance matrix of the $n\times 1$ vector of u_i in (6.1), with $\delta_\ell'\varphi_\ell$ replacing each admissible function $\zeta_\ell(\cdot)$. This is analogous to the definition of $\Sigma(\gamma)$ in earlier sections, and indeed after conditioning on z, it can be treated in a similar way because $d_\gamma \to \infty$ was already permitted. For example, suppose that $u = (I_n - W)^{-1}\varepsilon$, where $\|W\| < 1$ and the elements satisfy $w_{ij} = \zeta_0\left(d_{ij}\right)$, $i,j = 1,\ldots,n$, for some fixed distances d_{ij} and unknown function $\zeta_0(\cdot)$ (see, e.g., Pinkse, 1999). Approximating $\zeta_0(z) = \tau_0'\varphi(z) + \nu$, for some $r\times 1$ basis function vector $\varphi(z)$ and approximation error ν, we define $W(\tau)$ as the $n\times n$ matrix with elements $w_{ij}(\tau) = \tau_0'\varphi\left(d_{ij}\right)$, and set $\Sigma(\tau) = \mathrm{var}\left((I_n - W(\tau))^{-1}\varepsilon\right) = \sigma_0^2(I_n - W(\tau))^{-1}(I_n - W'(\tau))^{-1}$.

For any admissible values β, σ^2, and τ, the redefined (multiplied by $\frac{2}{n}$) negative quasi-log-likelihood function based on using the approximations (3.1) and (6.2) is

$$L(\beta,\sigma^2,\tau) = \ln\left(2\pi\sigma^2\right) + \frac{1}{n}\ln|\Sigma(\tau)| + \frac{1}{n\sigma^2}(y-\Psi\beta)'\Sigma(\tau)^{-1}(y-\Psi\beta), \tag{6.4}$$

which is minimized with respect to β and σ^2 by

$$\bar{\beta}(\tau) = \left(\Psi'\Sigma(\tau)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\tau)^{-1}y, \tag{6.5}$$

$$\bar{\sigma}^2(\tau) = n^{-1}y'E(\tau)'M(\tau)E(\tau)y, \tag{6.6}$$

where $M(\tau) = I_n - E(\tau)\Psi\left(\Psi'\Sigma(\tau)^{-1}\Psi\right)^{-1}\Psi'E(\tau)'$ and $E(\tau)$ is the $n\times n$ symmetric matrix such that $E(\tau)E(\tau)' = \Sigma(\tau)^{-1}$. Thus, the concentrated likelihood function is

$$\mathcal{L}(\tau) = \ln(2\pi) + \ln\bar{\sigma}^2(\tau) + \frac{1}{n}\ln|\Sigma(\tau)|. \tag{6.7}$$

Again, for compact Γ and sieve coefficient space Δ, the QMLE of τ_0 is $\widehat{\tau} = \arg\min_{\tau\in\Gamma\times\Delta}\mathcal{L}(\tau)$ and the QMLEs of β and σ^2 are $\widehat{\beta} = \bar{\beta}(\widehat{\tau})$ and $\widehat{\sigma}^2 = \bar{\sigma}^2(\widehat{\tau})$, respectively. The series estimate of θ_0 is defined as in (3.7). Define also the product Banach space $\mathcal{T} = \Gamma\times\Xi^{d_\zeta}$ with norm $\left\|\left(\gamma',\zeta'\right)'\right\|_{\mathcal{T}_w} = \|\gamma\| + \sum_{\ell=1}^{d_\zeta}\|\zeta_\ell\|_w$, and consider the map $\Sigma: \mathcal{T}^o \to \mathcal{M}^{n\times n}$, where $\mathcal{T}^o$ is an open subset of $\mathcal{T}$.

Assumption NPN.2. The map $\Sigma: \mathcal{T}^o \to \mathcal{M}^{n\times n}$ is Fréchet-differentiable on $\mathcal{T}^o$ with Fréchet-derivative denoted by $D\Sigma \in \mathscr{L}\left(\mathcal{T}^o, \mathcal{M}^{n\times n}\right)$. Furthermore,

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

conditional on z, the map $D\Sigma$ satisfies

$$\sup_{t\in\mathcal{T}^o} \|D\Sigma(t)\|_{\mathscr{L}(\mathcal{T}^o,\mathcal{M}^{n\times n})} \le C, \tag{6.8}$$

on its domain $\mathcal{T}^o$.

This assumption can be checked in a similar way to how we checked Assumption R.5, where a diverging dimension for the argument was already permitted.

PROPOSITION 6.1. *If Assumption NPN.2 holds, then, for any $t_1, t_2 \in \mathcal{T}^o$, conditional on z,*

$$\|\Sigma(t_1) - \Sigma(t_2)\| \le C\varsigma(r)\|t_1 - t_2\|. \tag{6.9}$$

COROLLARY 6.1. *For any $t^* \in \mathcal{T}^o$ and any $\eta > 0$, conditional on z,*

$$\overline{\lim_{n\to\infty}} \sup_{t\in\{t:\|t-t^*\|<\eta\}\cap\mathcal{T}^o} \left\|\Sigma(t) - \Sigma\left(t^*\right)\right\| < C\varsigma(r)\eta. \tag{6.10}$$

Assumption NPN.3. $c \le \sigma^2(\tau) \le C$ for $\tau \in \Gamma \times \Delta$, conditional on z.

Denote $\Sigma(\tau_0) = \Sigma_0$. Note that this is not the true covariance matrix, which is $\Sigma \equiv \Sigma(\gamma_0, \zeta_0)$.

Assumption NPN.4. $\tau_0 \in \Gamma \times \Delta$ and, for any $\eta > 0$, conditional on z,

$$\underline{\lim}_{n\to\infty} \inf_{\tau\in\overline{\mathcal{N}}^\tau(\eta)} \frac{n^{-1}tr\left(\Sigma(\tau)^{-1}\Sigma_0\right)}{\left|\Sigma(\tau)^{-1}\Sigma_0\right|^{1/n}} > 1, \tag{6.11}$$

where $\overline{\mathcal{N}}^\tau(\eta) = (\Gamma\times\Delta)\setminus\mathcal{N}^\tau(\eta)$ and $\mathcal{N}^\tau(\eta) = \{\tau : \|\tau - \tau_0\| < \eta\} \cap (\Gamma\times\Delta)$.

Remark 1. Expressing the identification condition in Assumption NPN.4 in terms of τ implies that identification is guaranteed via the sieve spaces Φ_{r_ℓ}, $\ell = 1,\dots,d_\zeta$. This approach is common in the sieve estimation literature (see, e.g., Chen, 2007, Condition 3.1, p. 5589).

THEOREM 6.1. *Under either H_0 or H_1, Assumptions R.1–R.4 (with Assumptions R.3 and R.4 holding for $t \in \mathcal{T}$ rather than $\gamma \in \Gamma$), R.8, and NPN.1–NPN.4, and $p^{-1} + \left(\min_{\ell=1,\dots,d_\zeta} r_\ell\right)^{-1} + \left(d_\gamma + p + \max_{\ell=1,\dots,d_\zeta} r_\ell\right)/n \to 0$ as $n\to\infty$, $\left\|\left(\widehat{\tau},\hat{\sigma}^2\right) - \left(\tau_0,\sigma_0^2\right)\right\| \xrightarrow{p} 0$.*

THEOREM 6.2. *Under the conditions of Theorems 4.2 and 6.1, but with τ and $\mathcal{T}$ replacing γ and Γ in assumptions prefixed with R and $p\to\infty$,*

$$\left(\min_{\ell=1,\dots,d_\zeta} r_\ell\right)^{-1} + \frac{p^2}{n} + \frac{\sqrt{n}}{p^{\mu+1/4}} + p^{1/2}\varsigma(r)\left(\frac{d_\gamma + \max_{\ell=1,\dots,d_\zeta} r_\ell}{\sqrt{n}} + \sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right) \to 0,$$

as $n\to\infty$, and H_0, $\mathscr{T}_n - \left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon - p\right)/\sqrt{2p} = o_p(1)$.

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

THEOREM 6.3. *Let the conditions of Theorems 4.3 and 6.2 hold, but with τ and $\mathcal{T}$ replacing γ and Γ in assumptions prefixed with R. Then, (1) $\mathscr{T}_n \xrightarrow{d} N(0,1)$ under H_0, (2) $\mathscr{T}_n$ is a consistent test statistic, and (3) $\mathscr{T}_n \xrightarrow{d} N(\varkappa,1)$ under local alternatives H_ℓ.*

7. FIXED-REGRESSOR RESIDUAL-BASED BOOTSTRAP TEST

The performance of nonparametric tests based on asymptotic distributions often leaves something to be desired in finite samples. An alternative approach is to use the bootstrap approximation. In this section, we propose a bootstrap version of our test, focusing on the setting of Section 5. In our simulations and empirical studies, we consider test statistics based on both $\widehat{m}_n = \widehat{\sigma}^{-2}\widehat{v}'\Sigma(\widehat{\gamma})^{-1}\widehat{u}/n$ and $\widetilde{m}_n = \widehat{\sigma}^{-2}(\widehat{u}'\Sigma(\widehat{\gamma})^{-1}\widehat{u} - \widehat{\eta}'\Sigma(\widehat{\gamma})^{-1}\widehat{\eta})/n$, where $\widehat{\eta} = S(\hat{\lambda})y - \widehat{\theta}$, i.e., the residual from nonparametric estimation, $\hat{u} = S(\hat{\lambda})y - f(x,\hat{\alpha})$, and $\hat{v} = \hat{\theta} - f(x,\hat{\alpha})$. Analogous to the definition of $\mathscr{T}_n$, define the statistic $\mathscr{T}_n^a = (n\widetilde{m}_n - p)/\sqrt{2p}$. In the case of no SAR term, and under the power series, $\mathscr{T}_n^a$ and $\mathscr{T}_n$ are numerically identical, as was observed by Hong and White (1995). However, in the SAR with spatial errors (SARSE) setting, a difference arises due to the spatial structure in the response y. We show that $\mathscr{T}_n^a - \mathscr{T}_n = o_p(1)$ under the null or local alternatives in Theorem TS.1 in the Supplementary Material.

The bootstrap versions of the test statistics $\mathscr{T}_n$ and $\mathscr{T}_n^a$ are

$$\mathscr{T}_n^* = \frac{n\widehat{m}_n^* - p}{\sqrt{2p}} = \frac{\widehat{\sigma}^{*-2}\widehat{v}^{*\prime}\Sigma(\widehat{\gamma}^*)^{-1}\widehat{u}^* - p}{\sqrt{2p}},$$
$$\mathscr{T}_n^{a*} = \frac{n\widetilde{m}_n^* - p}{\sqrt{2p}} = \frac{\widehat{\sigma}^{*-2}(\widehat{u}^{*\prime}\Sigma(\widehat{\gamma}^*)^{-1}\widehat{u}^* - \widehat{\eta}^{*\prime}\Sigma(\widehat{\gamma}^*)^{-1}\widehat{\eta}^*) - p}{\sqrt{2p}},$$

respectively, where $\widehat{u}^*$ is the bootstrap residual vector under the null, $\widehat{\eta}^*$ is the bootstrap residual vector under the alternative, $\widehat{v}^* = \widehat{\theta}^*(x) - f(x,\widehat{\alpha}^*)$, and $\left(\widehat{\gamma}^*, \lambda^*, \widehat{\sigma}^{*2}, \widehat{\theta}^*, \widehat{\alpha}^*\right)$ is the estimator using the bootstrap sample. We elaborate on the bootstrap statistics using the SARARMA(m_1,m_2,m_3) model as an example:

$$y = \sum_{k=1}^{m_1}\lambda_k W_{1k}y + \theta(x) + u,\ u = \sum_{l=1}^{m_2}\gamma_{2l}W_{2l}u + \sum_{l=1}^{m_3}\gamma_{3l}W_{3l}\xi + \xi.$$

Following Jin and Lee (2015), we first deduct the empirical mean of the residual vector from

$$\widehat{\xi} = \left(\sum_{l=1}^{m_3}\widehat{\gamma}_{3l}W_{3l} + I_n\right)^{-1}\left(I_n - \sum_{l=1}^{m_2}\widehat{\gamma}_{2l}W_{2l}\right)\left(y - \sum_{k=1}^{m_1}\widehat{\lambda}_k W_{1k}y - \widehat{\theta}_n\right)$$

to obtain $\widetilde{\xi} = (I_n - \frac{1}{n}l_n l_n')\widehat{\xi}$. Next, we sample randomly with replacement n times from elements of $\widetilde{\xi}$ to obtain a vector of ξ^*. After this, we generate the bootstrap

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

sample y^* by treating $\widehat{f} = f(x, \widehat{\alpha})$, $\hat{\lambda}$ and $\widehat{\gamma}$ as the true parameter:

$$y^* = \left(I_n - \sum_{k=1}^{m_1} \widehat{\lambda}_k W_{1k}\right)^{-1} \left(\widehat{f} + \left(I_n - \sum_{l=1}^{m_2} \widehat{\gamma}_{2l} W_{2l}\right)^{-1} \left(\sum_{l=1}^{m_3} \widehat{\gamma}_{3l} W_{3l} + I_n\right) \xi^*\right).$$

We estimate the model based on the bootstrap sample y^* using QMLE to obtain the estimator $\widehat{\theta}^* = \psi' \widehat{\beta}^*, \widehat{\lambda}^*$, and $\widehat{\gamma}^*$ under the alternative hypothesis and $\widehat{\alpha}^*$ under the null hypothesis of $\theta(x) = f(x, \alpha_0)$. Then, $\widehat{\eta}^* = y^* - \sum_{k=1}^{m_1} \widehat{\lambda}_k^* W_{1k} y^* - \widehat{\theta}^*$, $\widehat{u}^* = y^* - \sum_{k=1}^{m_1} \widehat{\lambda}_k^* W_{1k} y^* - f(x, \widehat{\alpha}^*)$.

This procedure is repeated B times to obtain the sequence $\left\{\mathscr{T}_{nj}^*\right\}_{j=1}^{B}$. We reject the null when $p^* = B^{-1} \sum_{j=1}^{B} \mathbf{1}(\mathscr{T}_n < \mathscr{T}_{nj}^*)$ is smaller than the given level of significance. An identical procedure holds for the test based on $\mathscr{T}_n^{a*}$. The asymptotic validity of the bootstrap method can be shown as in Theorem 4 of Su and Qu (2017) and Lemma 2 in Jin and Lee (2015), and a detailed analysis can be found in the Supplementary Material (see the proof of Theorem TS.1).

8. FINITE SAMPLE PERFORMANCE

8.1. Parametric Error Spatial Structure

Taking $n = 60, 100, 200$, we choose two specifications to generate y from the SARARMA(m_1,m_2, m_3) models:

SARARMA(0,1,0): $\quad y = \theta(x) + u, \ u = \gamma_2 W_2 u + \xi,$

SARARMA(1,0,1): $\quad y = \lambda_1 W_1 y + \theta(x) + u, \ u = \gamma_3 W_3 \xi + \xi,$

where ξ is $N(0, I_n)$. The data generating process (DGP) of $\theta(x)$ is

$$\theta(x_i) = x_i'\alpha + c p^{1/4} n^{-1/2} \sin(x_i'\alpha),$$

where $x_i'\alpha = 1 + x_{1i} + x_{2i}$, with $x_{1i} = (z_i + z_{1i})/2$ and $x_{2i} = (z_i + z_{2i})/2$. We choose two settings: compactly supported regressors where z_i, z_{1i}, and z_{2i} are i.i.d., $U[0, 2\pi]$, and unboundedly supported regressors where z_i, z_{1i}, and z_{2i} are i.i.d., $N(0, 1)$. We report the compact support setting in the main text, whereas the results for unbounded support are reported in the Supplementary Material.

We use three series bases for our experiments: power (polynomial) series of the third and fourth order ($p = 10$ and $p = 15$), trigonometric series $trig_1 = (1, \sin(x_1), \sin(x_1/2), \sin(x_2), \sin(x_2/2), \cos(x_1), \cos(x_1/2), \cos(x_2), \cos(x_2/2))'$ and $trig_2 = \left(trig_1', \sin\left(x_1^2\right), \cos\left(x_1^2\right), \sin\left(x_2^2\right), \cos\left(x_2^2\right)\right)'$, and the B-spline bases of the fourth and seventh order ($p = 9$ and $p = 14$), We also set $\gamma_2 = 0.3$, $\lambda_1 = 0.3$, and $\gamma_3 = 0.4$; the value $c = 0, 3, 6$ indicates the null hypothesis and the local alternatives. The spatial weight matrices are generated using LeSage's code make_neighborsw from http://www.spatial-econometrics.com/, where the row-normalized sparse matrices are generated by choosing a specific number of the closest locations from randomly generated coordinates and we set the number of

TABLE 1. Rejection probabilities of SARARMA(0,1,0) using bootstrap test $\mathscr{T}_n^*$ at 1%, 5%, and 10% levels, power series (**PS**), trigonometric (**Trig**), and B-spline (**B-s**) bases.

$\mathscr{T}_n^*$	SARARMA(0,1,0)								
	PS			**Trig**			**B-s**		
$n = 60$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.1
$c = 0$	0.008	0.032	0.084	0.004	0.04	0.092	0.006	0.048	0.104
	0.004	0.038	0.096	0.004	0.038	0.094	0.006	0.034	0.098
$c = 3$	0.036	0.154	0.296	0.092	0.276	0.39	0.098	0.292	0.470
	0.154	0.414	0.62	0.056	0.22	0.374	0.036	0.150	0.292
$c = 6$	0.22	0.544	0.748	0.454	0.794	0.908	0.432	0.814	0.938
	0.844	0.992	1	0.314	0.714	0.872	0.174	0.542	0.732
$n = 100$									
$c = 0$	0.006	0.044	0.098	0.002	0.04	0.09	0.008	0.038	0.110
	0.012	0.046	0.096	0.006	0.036	0.102	0.01	0.056	0.108
$c = 3$	0.294	0.578	0.72	0.214	0.508	0.626	0.272	0.572	0.712
	0.37	0.662	0.824	0.194	0.45	0.632	0.188	0.46	0.63
$c = 6$	0.95	0.99	0.996	0.902	0.99	0.998	0.922	0.994	1
	0.992	0.998	1	0.856	0.988	1	0.852	0.98	0.998
$n = 200$									
$c = 0$	0.006	0.038	0.104	0.008	0.042	0.112	0.024	0.074	0.132
	0.006	0.048	0.088	0.016	0.038	0.082	0.022	0.074	0.144
$c = 3$	0.178	0.402	0.55	0.162	0.374	0.532	0.314	0.516	0.654
	0.282	0.56	0.694	0.136	0.346	0.468	0.19	0.37	0.542
$c = 6$	0.846	0.968	0.984	0.796	0.95	0.98	0.89	0.976	0.986
	0.982	0.998	1	0.776	0.934	0.974	0.852	0.946	0.982

neighbors to be $\frac{n}{20}$. We employ 100 bootstrap replications in each of 500 Monte Carlo replications except for the SARARMA(1,0,1) design with $n = 200$, where we set 50 bootstrap replications in view of the computation time. We report the rejection frequencies of tests based on bootstrap critical values in the main text, whereas tests based on asymptotic critical values are reported in the Supplementary Material.

Tables 1–4 report the empirical rejection frequencies using the bootstrap test statistics $\mathscr{T}_n^*$ (Tables 1 and 3) and $\mathscr{T}_n^{a*}$ (Tables 2 and 4), when nominal levels are given by 1%, 5%, and 10%. To see how the choice of p and the basis functions affect small sample outcomes, we report two sets of results for each basis function family: the first row for each value of c is from the smaller p ($p = 9$ or 10), whereas the second row is from the larger p ($p = 14$ or 15). We summarize some important findings. First, we see that for most DGPs, our bootstrap test is closer to the

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Table 2. Rejection probabilities of SARARMA(0,1,0) using bootstrap test $\mathscr{T}_n^{a*}$ at 1%, 5%, and 10% levels, power series (**PS**), trigonometric (**Trig**), and B-spline (**B-s**) bases.

$\mathscr{T}_n^{a*}$	SARARMA(0,1,0)								
	PS			**Trig**			**B-s**		
$n=60$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.1
$c=0$	0.008	0.032	0.084	0.004	0.04	0.092	0.01	0.07	0.132
	0.004	0.038	0.096	0.004	0.038	0.094	0.004	0.038	0.096
$c=3$	0.036	0.154	0.296	0.09	0.274	0.384	0.164	0.376	0.558
	0.154	0.414	0.62	0.056	0.22	0.376	0.036	0.152	0.288
$c=6$	0.22	0.544	0.748	0.444	0.794	0.906	0.56	0.892	0.956
	0.844	0.992	1	0.312	0.714	0.87	0.174	0.532	0.732
$n=100$									
$c=0$	0.006	0.044	0.098	0.004	0.038	0.092	0.012	0.048	0.112
	0.012	0.046	0.096	0.006	0.036	0.106	0.01	0.056	0.106
$c=3$	0.294	0.578	0.72	0.214	0.504	0.63	0.28	0.564	0.72
	0.37	0.662	0.824	0.194	0.45	0.632	0.196	0.466	0.64
$c=6$	0.95	0.99	0.996	0.900	0.99	0.998	0.932	0.992	1
	0.992	0.998	1	0.856	0.988	1	0.86	0.984	0.998
$n=200$									
$c=0$	0.006	0.038	0.104	0.012	0.046	0.114	0.014	0.048	0.132
	0.006	0.048	0.088	0.016	0.042	0.08	0.022	0.07	0.14
$c=3$	0.178	0.402	0.55	0.162	0.38	0.524	0.282	0.476	0.608
	0.282	0.56	0.694	0.134	0.35	0.466	0.198	0.37	0.514
$c=6$	0.846	0.968	0.984	0.802	0.952	0.978	0.848	0.95	0.982
	0.982	0.998	1	0.774	0.934	0.972	0.84	0.932	0.97

nominal level than the asymptotic test (reported in the Supplementary Material), although the sizes of both types of tests improve generally as the sample size increases. Second, both bootstrap and asymptotic tests are powerful in detecting any deviations from linearity in the local alternatives. The patterns are similar across all cases: the bootstrap generally affords better size control, albeit not always.

All three types of bases give qualitatively similar results, but we note that $\mathscr{T}_n^* = \mathscr{T}_n^{*a}$ when using polynomial series under the SARARMA(0,1,0) model, as observed in Hong and White (1995). When using trigonometric and B-spline series, tests based on these two statistics give slightly different rejection rates. However, under the SARARMA(1,0,1) model, all series give quantitatively different results, as illustrated in Tables 3 and 4. When using B-spline bases, $p = 14$

TABLE 3. Rejection probabilities of SARARMA(1,0,1) using bootstrap test $\mathscr{T}_n^*$ at 1%, 5%, and 10% levels, power series (**PS**), trigonometric (**Trig**), and B-spline (**B-s**) bases.

$\mathscr{T}_n^*$	SARARMA(1,0,1)								
	PS			**Trig**			**B-s**		
$n = 60$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.1
$c = 0$	0.006	0.054	0.08	0.012	0.062	0.106	0.016	0.044	0.086
	0.016	0.062	0.118	0.026	0.09	0.138	0.016	0.048	0.088
$c = 3$	0.08	0.264	0.402	0.082	0.256	0.406	0.08	0.288	0.475
	0.132	0.41	0.578	0.096	0.222	0.354	0.048	0.192	0.282
$c = 6$	0.266	0.588	0.748	0.266	0.616	0.782	0.218	0.604	0.772
	0.444	0.804	0.894	0.204	0.474	0.658	0.198	0.496	0.612
$n = 100$									
$c = 0$	0.006	0.054	0.116	0.012	0.046	0.114	0.014	0.042	0.09
	0.02	0.056	0.112	0.012	0.044	0.088	0.034	0.058	0.118
$c = 3$	0.134	0.366	0.496	0.132	0.346	0.514	0.162	0.46	0.59
	0.222	0.556	0.732	0.242	0.542	0.698	0.08	0.234	0.372
$c = 6$	0.566	0.832	0.916	0.59	0.888	0.96	0.548	0.898	0.952
	0.732	0.964	0.986	0.476	0.846	0.918	0.432	0.796	0.874
$n = 200$									
$c = 0$	0.04	0.086	0.11	0.026	0.076	0.108	0.02	0.06	0.09
	0.03	0.078	0.114	0.032	0.074	0.118	0.038	0.086	0.112
$c = 3$	0.186	0.4	0.524	0.242	0.432	0.526	0.29	0.516	0.626
	0.402	0.636	0.754	0.244	0.42	0.542	0.184	0.36	0.458
$c = 6$	0.718	0.904	0.962	0.78	0.942	0.982	0.73	0.948	0.978
	0.872	0.98	0.998	0.794	0.948	0.98	0.772	0.914	0.94

does not perform well compared with $p = 9$. In the other cases, both choices of p work well.

8.2. Nonparametric Error Spatial Structure

Now, we examine finite sample performance in the setting of Section 6. The three DGPs of $\theta(x)$ are the same as the parametric setting, but we generate the $n \times n$ matrix W^* as $w_{ij}^* = \Phi(-d_{ij})I(c_{ij} < 0.05)$ if $i \neq j$, and $w_{ii}^* = 0$, where $\Phi(\cdot)$ is the standard normal cdf, $d_{ij} \sim$iid $U[-3,3]$, and $c_{ij} \sim$iid $U[0,1]$. From this construction, we ensure that W^* is sparse with no more than 5% of elements being nonzero. Then, y is generated from $y = \theta(x) + u$, $u = Wu + \xi$, where $\xi \sim N(0, I_n)$ and $W = W^*/1.2\overline{\varphi}(W^*)$, ensuring the existence of $(I - W)^{-1}$. In estimation, we know the distance d_{ij} and the indicator $I(c_{ij} < 0.05)$, but we do not know the functional form

TABLE 4. Rejection probabilities of SARARMA(1,0,1) using bootstrap test $\mathscr{T}_n^{a*}$ at 1%, 5%, and 10% levels, power series (**PS**), trigonometric (**Trig**), and B-spline (**B-s**) bases.

$\mathscr{T}_n^{a*}$	SARARMA(1,0,1)								
	PS			**Trig**			**B-s**		
$n=60$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.1
$c=0$	0.006	0.052	0.084	0.014	0.064	0.096	0.012	0.044	0.104
	0.012	0.068	0.114	0.024	0.088	0.13	0.018	0.038	0.068
$c=3$	0.092	0.27	0.396	0.08	0.25	0.406	0.118	0.382	0.56
	0.164	0.408	0.596	0.102	0.242	0.37	0.046	0.15	0.23
$c=6$	0.268	0.596	0.752	0.248	0.61	0.792	0.23	0.56	0.808
	0.518	0.824	0.9	0.206	0.484	0.658	0.176	0.43	0.56
$n=100$									
$c=0$	0.008	0.058	0.122	0.01	0.046	0.116	0.004	0.04	0.82
	0.024	0.062	0.118	0.014	0.044	0.096	0.028	0.056	0.074
$c=3$	0.14	0.36	0.494	0.122	0.354	0.52	0.186	0.4	0.524
	0.252	0.566	0.73	0.272	0.568	0.696	0.04	0.148	0.214
$c=6$	0.536	0.818	0.914	0.554	0.884	0.948	0.58	0.914	0.95
	0.786	0.958	0.974	0.478	0.834	0.916	0.328	0.586	0.678
$n=200$									
$c=0$	0.04	0.08	0.116	0.03	0.076	0.102	0.016	0.036	0.072
	0.026	0.064	0.108	0.028	0.06	0.122	0.008	0.014	0.02
$c=3$	0.176	0.382	0.516	0.22	0.438	0.526	0.262	0.45	0.55
	0.41	0.632	0.738	0.256	0.428	0.538	0.06	0.124	0.164
$c=6$	0.704	0.894	0.948	0.746	0.934	0.976	0.69	0.916	0.974
	0.914	0.986	0.996	0.776	0.93	0.976	0.482	0.612	0.66

of w_{ij}, so we approximate elements in W by $\widehat{w}_{ij} = \sum_{l=0}^{r} a_l d_{ij}^l I(c_{ij} < 0.05)$ if $i \neq j$; $\widehat{w}_{ii} = 0$.

Table 5 reports the rejection rates using 500 Monte Carlo simulations at the 5% asymptotic level 1.645 using polynomial bases with $r = 2,3,4,5$ and $p = 10,15,20$. We take $n = 150,300,500,600,700$ larger sample sizes than earlier because two nonparametric functions must be estimated in this spatial setting. The two largest bandwidths ($r = 5$ and $p = 20$) are only employed for the largest sample size $n = 700$. We observe a clear pattern of rejection rates approaching the theoretical level as sample size increases. Power improves as c increases for all designs and is nontrivial in all cases even for $c = 3$. Sizes are acceptable for $n = 500$, particularly when $p = 15$. Size performance improves further as $n = 600$, indicating asymptotic stability. Note that with two diverging bandwidths (p and r), we expect sizes to improve in a diagonal pattern going from top-left corner to bottom-right corner

TABLE 5. Rejection probabilities of $\mathscr{T}_n$ at the 5% asymptotic level, nonparametric spatial error structure.

	$r=2$		$r=3$		$r=4$		$r=5$		
$n=150$	$p=10$	$p=15$	$p=10$	$p=15$	$p=10$	$p=15$	$p=10$	$p=15$	$p=20$
$c=0$	0.0860	0.2020	0.1180	0.2060	0.1420	0.2240			
$c=3$	0.3320	0.6340	0.3700	0.6380	0.3760	0.6700			
$c=6$	0.9060	0.9920	0.9180	0.9940	0.9220	0.9960			
$n=300$									
$c=0$	0.0820	0.0960	0.0880	0.1080	0.1060	0.1100			
$c=3$	0.2680	0.5980	0.2600	0.6120	0.2780	0.6220			
$c=6$	0.8140	0.9980	0.8160	0.9980	0.8220	0.9980			
$n=500$									
$c=0$	0.0280	0.0420	0.0260	0.0400	0.0360	0.0480			
$c=3$	0.2320	0.6660	0.2400	0.6620	0.2460	0.6680			
$c=6$	0.8920	1	0.9040	1	0.9000	1			
$n=600$									
$c=0$	0.0320	0.0500	0.0340	0.0540	0.0360	0.0540			
$c=3$	0.3140	0.6480	0.3080	0.6280	0.3120	0.6460			
$c=6$	0.9220	1	0.9180	1	0.9180	1			
$n=700$									
$c=0$	0.0260	0.0300	0.0280	0.0380	0.0280	0.0380	0.0280	0.0420	0.0580
$c=3$	0.2420	0.5540	0.2400	0.5480	0.2520	0.5500	0.2420	0.5600	0.6920
$c=6$	0.9580	0.9980	0.9560	0.9980	0.9600	0.9980	0.9500	0.9980	1

in Table 5. This is indeed the case. For $n = 700$, we observe that the pairs $(r,p) = (5,15),(5,20)$ deliver acceptable sizes.

9. EMPIRICAL APPLICATIONS

In this section, we illustrate the specification test presented in previous sections using several empirical examples.

9.1. Conflict Alliances

This example is based on a study of how a network of military alliances and enmities affects the intensity of a conflict, conducted by König et al. (2017). They stress that understanding the role of informal networks of military alliances and enmities is important not only for predicting outcomes, but also for designing and implementing policies to contain or put an end to violence. König et al. (2017) obtain a closed-form characterization of the Nash equilibrium and perform an

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

empirical analysis using data on the Second Congo War, a conflict that involves many groups in a complex network of informal alliances and rivalries.

To study the fighting effort of each group, the authors use a panel data model with individual fixed effects, where key regressors include total fighting effort of allies and enemies. They further correct the potential spatial correlation in the error term by using a spatial heteroskedasticity and autocorrelation robust standard error. We use their data and the main structure of the specification and build a cross-sectional SAR(2) model with two weight matrices, W^A ($W^A_{ij} = 1$ if groups i and j are allies, and $W^A_{ij} = 0$ otherwise) and W^E ($W^E_{ij} = 1$ if groups i and j are enemies, and $W^E_{ij} = 0$ otherwise):

$$y = \lambda_1 W^A y + \lambda_2 W^E y + \mathbf{1}_n \beta_0 + X\beta + u,$$

where y is a vector of fighting efforts of each group and X includes the current rainfall, rainfall from the last year, and their squares.[1] To consider the spatial correlation in the error term, we consider both the Error SARMA(1,0) and Error SARMA(0,1) structures. For these, we employ a spatial weight matrix W^d, based on the inverse distance between group locations and set to be 0 after 150 km, following König et al. (2017). The idea is that geographical spatial correlation dies out as groups become further apart. We also report results using a nonparametric estimator of the spatial weights, as described in Section 6 and studied in simulations in Section 8. For the nonparametric estimator, we take $r = 2$.

In the original dataset, there are 80 groups, but groups 62 and 63 have the same variables and the same locations, so we drop one group and end up with a sample of 79 groups. We use data from 1998 as an example and further use the pooled data from all years as a robustness check. H_0 stands for restricted model where the linear functional form of the regression is imposed, whereas H_1 stands for the unrestricted model where we use basis functions comprising of power series with $p = 10$. In all our specifications, the test statistics are negative, so we cannot reject the null hypothesis that the model is correctly specified. As Table 6 indicates, this failure to reject the null persists when we use pooled data from 13 years, yielding 1,027 observations. Thus, we conclude that a linear specification is not inappropriate for this setting. One possible reason is that the original regression, though linear, has already included the squared terms of the rainfall as regressors. This finding is robust to using the bootstrap tests of Section 7, which generally yield smaller p-values but unchanged conclusions.

9.2. Innovation Spillovers

This example is based on the study of the impact of R&D on growth from Bloom et al. (2013). They develop a general framework incorporating two types of spillovers: a positive effect from technology (knowledge) spillovers and a negative

[1] We follow the analysis in the original paper and do not row normalize. This is because the economic content of the weight matrices is defined by total fights of allies or enemies.

Table 6. The estimates and test statistics for the conflict data.

	1998				Pooled			
	H_0	p-value	H_1	p-value	H_0	p-value	H_1	
				SARARMA(2,1,0)				
$W^A y$	−0.005	<0.001	−0.003	<0.001	0.013	<0.001	0.013	<0.001
$W^E y$	0.130	<0.001	0.129	<0.001	0.121	<0.001	0.121	<0.001
W^d	−0.159	0.281	−0.225	<0.001	−0.086	0.033	−0.086	0.033
$\mathscr{T}_n$			−1.921	0.973			−2.531	0.994
$\mathscr{T}_n^*$				0.840				0.940
$\mathscr{T}_n^a$			−1.918	0.972			−2.547	0.995
$\mathscr{T}_n^{a*}$				0.870				0.730
				SARARMA(2,0,1)				
$W^A y$	0.001	<0.01	0.011	<0.01	0.013	<0.01	0.013	<0.01
$W^E y$	0.127	<0.01	0.122	<0.01	0.121	<0.01	0.121	<0.01
W^d	−0.153	<0.01	−0.050	<0.01	−0.086	<0.01	−0.086	0.025
$\mathscr{T}_n$			−1.763	0.961			−2.421	0.992
$\mathscr{T}_n^*$				0.900				0.990
$\mathscr{T}_n^a$			−2.349	0.991			−2.423	0.992
$\mathscr{T}_n^{a*}$				0.850				0.790
				Nonparametric				
$W^A y$	−0.052	<0.001	−0.011	<0.001	0.033	<0.001	0.033	<0.001
$W^E y$	0.149	<0.001	0.133	<0.001	0.110	<0.001	0.109	<0.001
W^d								
$\mathscr{T}_n$			−1.294	0.902			−2.314	0.990
$\mathscr{T}_n^*$				0.830				0.850
$\mathscr{T}_n^a$			−1.898	0.971			−2.530	0.994
$\mathscr{T}_n^{a*}$				0.660				0.910

Note: * Denotes the bootstrap p-value.

"business stealing" effect from product market rivals. They implement this model using panel data on U.S. firms.

We consider the Productivity Equation in Bloom et al. (2013):

$$\ln y = \varphi_1 \ln(R\&D) + \varphi_2 \ln(Sptec) + \varphi_3 \ln(Spsic) + \varphi_4 X + error, \qquad \textbf{(9.1)}$$

where y is a vector of sales, $R\&D$ is a vector of R&D stocks, and regressors in X include the log of capital (*Capital*), log of labor (*Labor*), $R\&D$, a dummy for missing values in $R\&D$, a price index, and two spillover terms constructed as the log of $W_{SIC}R\&D$ (*Spsic*) and the log of $W_{TEC}R\&D$ (*Sptec*), where W_{SIC} measures the product market proximity and W_{TEC} measures the technological proximity.

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Table 7. The estimates and test statistics for the R&D data, SARARMA(0,1,0).

Variables	FE		SARARMA(0,1,0), W_{TEC}			
		p-value	H_0	p-value	H_1	p-value
ln(*Spsic*)	−0.005	0.649	0.007	0.574	0.015	0.166
ln(*Sptec*)	0.191	<0.001	0.006	0.850	−0.001	0.998
ln(*Lab.*)	0.636	<0.001	0.572	<0.001		
ln(*Cap.*)	0.154	<0.001	0.336	<0.001		
ln(*R&D*)	0.043	<0.001	0.081	<0.001		
W_{TEC}			0.835	<0.001	0.829	<0.001
$\mathscr{T}_n$					15.528	<0.001
$\mathscr{T}_n^*$						0.050

Variables	SARARMA(0,1,0), W_{SIC}			
	H_0	p-value	H_1	p-value
ln(*Spsic*)	0.008	0.620	0.017	0.193
ln(*Sptec*)	0.039	0.157	0.020	0.336
ln(*Lab.*)	0.571	<0.001		
ln(*Cap.*)	0.318	<0.001		
ln(*R&D*)	0.082	<0.001		
W_{SIC}	0.722	<0.001	0.724	<0.001
$\mathscr{T}_n$			10.451	<0.001
$\mathscr{T}_n^*$				<0.001

Notes: * Denotes the bootstrap p-value. The price index and a dummy variable for missing value in R&D are included, but we only report the coefficients reported in Bloom et al. (2013).

Specifically, they define

$$W_{SIC,ij} = S_i S_j' / (S_i S_i')^{1/2} (S_j S_j')^{1/2}, W_{TEC,ij} = T_i T_j' / (T_i T_i')^{1/2} (T_j T_j')^{1/2},$$

where $S_i = (S_{i1}, S_{i2}, \ldots, S_{i597})'$, with S_{ik} being the share of patents of firm i in the four digit industry k and $T_i = (T_{i1}, T_{i2}, \ldots, T_{i426})'$, with $T_{i\tau}$ being the share of patents of firm i in technology class τ. Focusing on a cross-sectional analysis, we use observations from the year 2000 and obtain a sample size of 577. Both weight matrices are row normalized.

The column FE of Table 7 is from Table 5 of Bloom et al. (2013) based on their panel fixed effects estimation and we use it as a baseline for comparison. This table reports results for SARARMA(0,1,0) models using W_{SIC} and W_{TEC} separately. We use both W_{SIC} and W_{TEC} simultaneously in SARARMA(0,2,0), SARARMA(0,2,0), and Error MESS(2) models, reported in Table 8. In all of these specifications, the test statistics are larger than 1.645, so we reject the null hypothesis of the linear specification. This rejection also persists with the bootstrap tests, albeit the p-values go up compared to the asymptotic ones. However,

TABLE 8. The estimates and test statistics for the R&D data, SARARMA(0,2,0) and Error MESS(2).

Variables	SARARMA(0,2,0)			
	H_0	p-value	H_1	p-value
ln (*Spsic*)	0.009	0.587	0.018	0.170
ln (*Sptec*)	0.044	0.112	0.026	0.236
ln (*Lab.*)	0.573	<0.001		
ln (*Cap.*)	0.315	<0.001		
ln (*R&D*)	0.082	<0.001		
W_{SIC}	0.696	<0.001	0.693	<0.001
W_{TEC}	0.157	0.092	0.164	0.079
$\mathscr{T}_n$			10.485	<0.001
$\mathscr{T}_n^*$				0.060
Variables	SARARMA(0,0,2)			
	H_0	p-value	H_1	p-value
ln (*Spsic*)	−0.0002	0.991	0.013	0.266
ln (*Sptec*)	0.033	0.200	0.017	0.434
ln (*Lab.*)	0.565	<0.01		
ln (*Cap.*)	0.334	<0.01		
ln (*R&D*)	0.076	<0.01		
W_{SIC}	0.624	<0.01	0.728	<0.001
W_{TEC}	0.312	0.123	0.321	0.112
$\mathscr{T}_n$			15.144	<0.001
$\mathscr{T}_n^*$				0.020
Variables	Error MESS(2)			
	H_0	p-value	H_1	p-value
ln (*Spsic*)	0.002	0.788	0.014	0.040
ln (*Sptec*)	0.045	0.025	0.027	0.088
ln (*Lab.*)	0.569	<0.001		
ln (*Cap.*)	0.323	<0.001		
ln (*R&D*)	0.077	<0.001		
W_{SIC}	0.775	<0.001	0.836	<0.001
W_{TEC}	0.338	0.010	0.380	0.004
$\mathscr{T}_n$			12.776	<0.001
$\mathscr{T}_n^*$				0.050

Notes: * Denotes the bootstrap p-value. The price index and a dummy variable for missing value in R&D are included, but we only report the coefficients reported in Bloom et al. (2013).

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

we can say even more as our estimation also sheds light on spatial effects in the disturbances in (9.1). As before, H_0 imposes linear functional form of the regressors, whereas H_1 uses the nonparametric series estimate employing power series with $p = 10$. Regardless of the specification of the regression function, the disturbances suggest a strong spatial effect as the coefficients on W_{TEC} and W_{SIC} are large in magnitude.

9.3. Economic Growth

The final example is based on the study of economic growth rate in Ertur and Koch (2007). Knowledge accumulated in one area might depend on knowledge accumulated in other areas, especially in its neighborhoods, implying the possible existence of spatial spillover effects. These questions are of interest to both economists and regional scientists. For example, Autant-Bernard and LeSage (2011) examine spatial spillovers associated with research expenditures for French regions, whereas Ho, Wang, and Yu (2013) examine the international spillover of economic growth through bilateral trade among OECD countries, Crespo Cuaresma and Feldkircher (2013) study spatially correlated growth spillovers in the income convergence process of Europe, and Evans and Kim (2014) study the spatial dynamics of growth and convergence in Korean regional incomes.

In this section, we want to test the linear SAR model specification in Ertur and Koch (2007). Their dataset covers a sample of 91 countries over the period of 1960–1995, originally from Heston, Summers, and Aten (2002), obtained from the Penn World Tables (PWT version 6.1). The variables in use include per worker income in 1960 ($y60$) and 1995 ($y95$), average rate of growth between 1960 and 1995 (gy), average investment rate of this period (s), and average rate of growth of working-age population (n_p).

Ertur and Koch (2007) consider the model

$$y = \lambda Wy + X\beta + WX\theta + \varepsilon, \tag{9.2}$$

where the dependent variable is log real income per worker $\ln(y95)$, elements of the explanatory variable $X = (x_1', x_2')$ include log investment rate $\ln(s) = x_1$ and log physical capital effective rate of depreciation $\ln(n_p + 0.05) = x_2$, with corresponding subscripted coefficients $\beta_1, \beta_2, \theta_1, \theta_2$. A restricted regression based on the joint constraints $\beta_1 = -\beta_2$ and $\theta_1 = -\theta_2$ (these constraints are implied by economic theory) is also considered in Ertur and Koch (2007). The model (9.2) has regressors (X, WX) and i.i.d. errors, so the test derived in Section 5 can be directly applied here. Denoting by d_{ij} the great-circle distance between the capital cities of countries i and j, one construction of W takes $w_{ij} = d_{ij}^{-2}$, whereas the other takes $w_{ij} = e^{-2d_{ij}}$, following Ertur and Koch (2007).

Table 9 presents the estimation and testing results based on using linear and quadratic power series basis functions with $p = 10$ and a sample size of $n = 91$. We impose additive structure in our estimation to at least alleviate the curse of dimensionality, always a concern in nonparametric estimation. We also use only

Table 9. The estimates and test statistics of the linear SAR model for the growth data.

Variable	$w_{ij}^* = d_{ij}^{-2}$ for $i \neq j$		$w_{ij}^* = e^{-2d_{ij}}$ for $i \neq j$	
	Estimate	p-value	Estimate	p-value
Constant	1.0711	0.608	0.5989	0.798
$\ln(s)$	0.8256	< 0.001	0.7938	< 0.001
$\ln(n_p + 0.05)$	−1.4984	0.008	−1.4512	0.009
$W\ln(s)$	−0.3159	0.075	−0.3595	0.020
$W\ln(n_p + 0.05)$	0.5633	0.498	0.1283	0.856
Wy	0.7360	< 0.001	0.6510	< 0.001
$\mathscr{T}_n$	−1.88	0.970	−2.08	0.981
$\mathscr{T}_n^*$		0.850		0.900
$\mathscr{T}_n^a$	−1.90	0.971	−2.05	0.980
$\mathscr{T}_n^{a*}$		0.820		0.810
Restricted regression				
Constant	2.1411	< 0.001	2.9890	< 0.001
$\ln(s) - \ln(n + 0.05)$	0.8426	< 0.001	0.8195	< 0.001
$W[\ln(s) - \ln(n_p + 0.05)]$	−0.2675	0.122	−0.2589	0.098
$W\ln(y)$	0.7320	< 0.001	0.6380	< 0.001
$\mathscr{T}_n$	0.30	0.382	4.04	< 0.001
$\mathscr{T}_n^*$		0.500		< 0.001
$\mathscr{T}_n^a$	0.10	0.460	4.50	< 0.001
$\mathscr{T}_n^{a*}$		0.560		0.040

Note: * Denotes the bootstrap p-value.

linear and quadratic basis functions to reduce the number of terms for series estimation.

We cannot reject linearity of the regression function for the unrestricted model. On the other hand, linearity is rejected for the restricted model, which is the preferred specification of Ertur and Koch (2007), with $w_{ij} = e^{-2d_{ij}}$. Thus, not only can we conclude that the specification of the model is under suspicion, but we can also infer this is due to constraints from economic theory. The findings are supported by the bootstrap tests of Section 7.

10. CONCLUSION

This paper justifies a specification test for the regression function in a model where data are spatially dependent. The test is based on a nonparametric series approximation and is consistent. The paper also allows for some robustness in error spatial dependence by permitting this to be a nonparametric function of an underlying economic distance. On the other hand, our Section 5 imposes correct

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

specification of the spatial weight matrices W_j in the SAR context, whereas Sun (2020) allows these to be nonparametric functions as well. Thus, our work acts as a complement to existing results in the literature and future work might combine both aspects.

A. APPENDIX: Proofs of Theorems and Propositions

Proof of Proposition 4.1. Because the map $\Sigma : \Gamma^o \to \mathcal{M}^{n\times n}$ is Fréchet-differentiable on Γ^o, it is also Gâteaux-differentiable and the two derivative maps coincide. Thus, by Theorem 1.8 of Ambrosetti and Prodi (1995), $\|\Sigma(\gamma_1) - \Sigma(\gamma_2)\| \le \sup_{\gamma\in\ell[\gamma_1,\gamma_2]} \|D\Sigma(\gamma)\| \|\gamma_1 - \gamma_2\|$, where $\ell[\gamma_1,\gamma_2] = \{t\gamma_1 + (1-t)\gamma_2 : t \in [0,1]\}$. The claim now follows by (4.3). □

Proof of Theorem 4.1. This a particular case of the proof of Theorem 5.1 with $\lambda = 0$, and so $S(\lambda) = I_n$. □

Proof of Theorem 4.2. In the Supplementary Material. □

Proof of Theorem 4.3. We would like to establish the asymptotic unit normality of

$$\frac{\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon - p}{\sqrt{2p}}. \tag{A.1}$$

Writing $q = \sqrt{2p}$, the ratio in (A.1) has zero mean and variance equal to one, and may be written as $\sum_{s=1}^{\infty} w_s$, where $w_s = \sigma_0^{-2}q^{-1}v_{ss}\left(\varepsilon_s^2 - \sigma_0^2\right) + 2\sigma_0^{-2}q^{-1}\mathbf{1}(s\ge 2)\varepsilon_s\sum_{t<s} v_{st}\varepsilon_t$, with v_{st} the typical element of $\mathscr{V}$, with $s,t = 1,2,\ldots$. We first show that

$$w_* \xrightarrow{p} 0, \tag{A.2}$$

where $w_* = w - w_S$, $w_S = \sum_{s=1}^{S} w_s$, and $S = S_n$ is a positive integer sequence that is increasing in n. All expectations in the sequel are taken conditional on X. By Chebyshev's inequality proving

$$\mathcal{E}w_*^2 \xrightarrow{p} 0 \tag{A.3}$$

is sufficient to establish (A.2). Notice that $\mathcal{E}w_s^2 \le Cq^{-2}v_{ss}^2 + Cq^{-2}\mathbf{1}(s\ge 2)\sum_{t<s} v_{st}^2 \le Cq^{-2}\sum_{t\le s} v_{st}^2$, so that, writing $\mathscr{M} = \Sigma^{-1}\Psi[\Psi'\Sigma^{-1}\Psi]^{-1}\Psi'\Sigma^{-1}$,

$$\begin{aligned}
\sum_{s=S+1}^{\infty}\mathcal{E}w_s^2 &\le Cq^{-2}\sum_{s=S+1}^{\infty}\sum_{t\le s} v_{st}^2 \le Cq^{-2}\sum_{s=S+1}^{\infty} b_s' M \sum_{t\le s} b_t b_t' \mathscr{M} b_s \\
&\le Cq^{-2}\|\Sigma\|\sum_{s=S+1}^{\infty} b_s'\mathscr{M}^2 b_s \le Cq^{-2}\sum_{s=S+1}^{\infty}\sum_{i,j,k=1}^{n} b_{is}b_{kt}m_{ij}m_{kj} \\
&\le Cq^{-2}\sum_{s=S+1}^{\infty}\sum_{i,k=1}^{n}|b_{is}^*||b_{ks}^*|\sum_{j=1}^{n}\left(m_{kj}^2 + m_{ij}^2\right),
\end{aligned} \tag{A.4}$$

where m_{ij} is the (i,j)th element of $\mathscr{M}$ and we have used the inequality $|ab| \le \left(a^2+b^2\right)/2$ in the last step. Now, denote by h_i' the ith row of the $n\times p$ matrix $\Sigma^{-1}\Psi$. Denoting the

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

elements of Σ^{-1} by Σ_{ij}^{-1} and $\psi_{jl} = \psi(x_{jl})$, h_i has entries $h_{il} = \sum_{j=1}^{n} \Sigma_{ij}^{-1} \psi_{jl}$, $l = 1, \ldots, p$. We have $|h_{il}| = O_p\left(\sum_{j=1}^{n} \left|\Sigma_{ij}^{-1}\right|\right) = O_p\left(\left\|\Sigma^{-1}\right\|_R\right) = O_p(1)$, uniformly, by Assumptions R.11 and R.13. Thus, we have $\|h_i\| = O_p(\sqrt{p})$, uniformly in i. As a result,

$$|m_{ij}| = n^{-1}\left|h_i'\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1} h_j\right| = O_p\left(n^{-1}\|h_i\|\|h_j\|\right) = O_p\left(pn^{-1}\right), \tag{A.5}$$

uniformly in i,j, by Assumption R.11. Similarly, note that

$$\begin{aligned}\sum_{j=1}^{n} m_{ij}^2 &= n^{-1} h_i'\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1}\left(n^{-1}\Psi'\Sigma^{-2}\Psi\right)\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1} h_i \\ &\leq n^{-1}\|h_i\|^2\left\|\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|^2\left\|n^{-1}\Psi'\Sigma^{-2}\Psi\right\| \\ &= O_p\left(pn^{-2}\|\Psi\|^2\left\|\Sigma^{-1}\right\|^2\right) = O_p\left(pn^{-1}\right),\end{aligned} \tag{A.6}$$

uniformly in i. Thus, (A.4) is

$$O_p\left(q^{-2}pn^{-1}\sum_{i=1}^{n}\sum_{s=S+1}^{\infty}|b_{is}^*|\sum_{t=1}^{n}|b_{ks}^*|\right) = O_p\left(q^{-2}p\sup_{i=1,\ldots,n}\sum_{s=S+1}^{\infty}|b_{is}^*|\right), \tag{A.7}$$

by Assumption R.4. By the same assumption, there exists S_{in} such that $\sum_{s=S_{in}+1}^{\infty}|b_{is}^*| \leq \epsilon_n$ for any decreasing sequence $\epsilon_n \to 0$ as $n \to \infty$. Choosing $S = \max_{i=1,\ldots,n} S_{in}$ in w_S, we deduce that (A.7) is $O_p\left(q^{-2}p\epsilon_n\right) = O_p(\epsilon_n) = o_p(1)$, proving (A.3). Thus, we need to only focus on w_S, and seek to establish that

$$w_S \longrightarrow_d N(0,1), \text{ as } n \to \infty. \tag{A.8}$$

From Scott (1973), (A.8) follows if

$$\sum_{s=1}^{S}\mathcal{E}w_s^4 \xrightarrow{p} 0, \text{ as } n \to \infty, \tag{A.9}$$

and

$$\sum_{s=1}^{S}\left[\mathcal{E}\left(w_s^2 \mid \varepsilon_t, t<s\right) - \mathcal{E}\left(w_s^2\right)\right] \xrightarrow{p} 0, \text{ as } n \to \infty. \tag{A.10}$$

We show (A.9) first. Evaluating the expectation and using (A.6) yields

$$\begin{aligned}\mathcal{E}w_s^4 &\leq Cq^{-4}v_{ss}^4 + Cq^{-4}\sum_{t<s}v_{st}^4 \leq Cq^{-4}\left(\sum_{t\leq s}v_{st}^2\right)^2 \leq Cq^{-4}\left(b_s'\mathscr{M}\sum_{t\leq s}b_t b_t'\mathscr{M}b_s\right)^2 \\ &\leq Cq^{-4}\left(b_s'\mathscr{M}^2 b_s\right)^2 = Cq^{-4}\sum_{i,j,k=1}^{n} b_{is}b_{ks}m_{ij}m_{kj} \leq Cq^{-4}\sum_{i,k=1}^{n}|b_{is}^*||b_{ks}^*|\sum_{j=1}^{n}\left(m_{ij}^2 + m_{kj}^2\right) \\ &= O_p\left(q^{-4}pn^{-1}\left(\sum_{i=1}^{n}|b_{is}^*|\right)^2\right),\end{aligned}$$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

whence

$$\sum_{s=1}^{S}\mathcal{E}w_s^4 = O_p\left(q^{-4}pn^{-1}\sum_{s=1}^{S}\left(\sum_{i=1}^{n}|b_{is}^*|\right)^2\right) = O_p\left(q^{-4}pn^{-1}\sum_{s=1}^{S}\left(\sum_{i=1}^{n}|b_{is}^*|\right)\right) = O_p\left(q^{-4}p\right),$$

by Assumption R.4. Thus, (A.9) is established. Notice that $\mathcal{E}\left(w_s^2\middle|\epsilon_t, t<s\right)$ equals

$$4q^{-2}\sigma_0^{-4}\left\{\left(\mu_4-\sigma_0^4\right)v_{ss}^2+2\mu_3\mathbf{1}(s\geq 2)\sum_{t<s}v_{st}v_{ss}\varepsilon_t\right\}+4q^{-2}\sigma_0^{-2}\mathbf{1}(s\geq 2)\left(\sum_{t<s}v_{st}\varepsilon_t\right)^2,$$

and $\mathcal{E}w_s^2 = 4q^{-2}\sigma_0^{-4}\left(\mu_4-\sigma_0^4\right)v_{ss}^2+4q^{-2}\mathbf{1}(s\geq 2)\sum_{t<s}v_{st}^2$, so that (A.10) is bounded by a constant times

$$q^{-2}\sum_{s=2}^{S}\sum_{t<s}v_{st}v_{ss}\varepsilon_t+\left\{\sum_{s=2}^{S}\left(\sum_{t<s}v_{st}\varepsilon_t\right)^2-\sigma_0^2\sum_{t<s}v_{st}^2\right\}. \tag{A.11}$$

By transforming the range of summation, the square of the first term in (A.11) has expectation bounded by

$$Cq^{-4}\mathcal{E}\left(\sum_{t=1}^{S-1}\sum_{s=t+1}^{S}v_{st}v_{ss}\varepsilon_t\right)^2 \leq Cq^{-4}\sum_{t=1}^{S-1}\left(\sum_{s=t+1}^{S}v_{st}v_{ss}\right)^2, \tag{A.12}$$

where the factor in parentheses on the RHS of (A.12) is

$$\sum_{s,r=t+1}^{S}b_s'\mathscr{M}b_sb_s'\mathscr{M}b_tb_r'\mathscr{M}b_rb_r'\mathscr{M}b_t \leq \sum_{s,r=t+1}^{S}\left|b_s'\mathscr{M}b_sb_r'\mathscr{M}b_r\right|\left|b_s'\mathscr{M}b_t\right|\left|b_r'\mathscr{M}b_t\right|$$

$$\leq C\sum_{s,r=t+1}^{S}\sum_{i,j,k,l=1}^{n}|b_{is}|\left|m_{ij}\right|\left|b_{jr}\right||b_{ks}||m_{lk}||b_{kr}|\left|b_s'\mathscr{M}b_t\right|\left|b_r'\mathscr{M}b_t\right|$$

$$\leq C\left(\sup_{i,j}|m_{ij}|\right)^2\left(\sup_{s\geq 1}\sum_{i=1}^{n}|b_{is}^*|\right)^4\sum_{s,r=t+1}^{S}\left|b_s'\mathscr{M}b_t\right|\left|b_r'\mathscr{M}b_t\right|$$

$$= O_p\left(p^2n^{-2}\left(\sum_{s=t+1}^{S}\left|b_t'\mathscr{M}b_s\right|\right)^2\right) = O_p\left(p^2n^{-2}\left(\sum_{s=t+1}^{S}\sum_{i,j=1}^{n}|b_{it}^*|\left|m_{ij}\right|\left|b_{js}^*\right|\right)^2\right),$$

where we used Assumption R.4 and (A.5). Now, Assumptions R.4 and R.11 and (A.5) imply that

$$\sum_{s=t+1}^{S}\sum_{i,j=1}^{n}|b_{it}^*|\left|m_{ij}\right|\left|b_{js}^*\right| = O_p\left(\sup_{i,j}|m_{ij}|\sup_t\sum_{i=1}^{n}|b_{it}^*|\sum_{j=1}^{n}\sum_{s=t+1}^{S}\left|b_{js}^*\right|\right) = O_p\left(p\sup_t\sum_{i=1}^{n}|b_{it}^*|\right),$$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

so (A.12) is $O_p\left(q^{-4}p^4n^{-2}\sup_t\left(\sum_{i=1}^n|b_{it}^*|\right)\left(\sum_{i=1}^n\left(\sum_{t=1}^{S-1}|b_{it}^*|\right)\right)\right)$. By Assumption R.4, the latter is $O_p\left(q^{-4}p^4n^{-1}\right)$ and therefore the first term in (A.11) is $O_p\left(p^2n^{-1}\right)$, which is negligible.

Once again transforming the summation range and using the inequality $|a+b|^2 \le C\left(a^2+b^2\right)$, we can bound the square of the second term in (A.11) by a constant times

$$\left(\sum_{t=1}^{S-1}\sum_{s=t+1}^{S} v_{st}^2\left(\varepsilon_t^2-\sigma_0^2\right)\right)^2+\left(2\sum_{t=1}^{S-1}\sum_{r=1}^{t-1}\sum_{s=t+1}^{S} v_{st}v_{sr}\varepsilon_t\varepsilon_r\right)^2. \tag{A.13}$$

Using Assumption R.4, the expectations of the two terms in (A.13) are bounded by a constant times α_1 and a constant times α_2, respectively, where $\alpha_1=\sum_{t=1}^{S-1}\left(\sum_{s=t+1}^{S} v_{st}^2\right)^2$, $\alpha_2=\sum_{t=1}^{S-1}\sum_{r=1}^{t-1}\left(\sum_{s=t+1}^{S} v_{st}v_{sr}\right)^2$. Thus, (A.13) is $O_p(\alpha_1+\alpha_2)$. Now, by (A.5), Assumptions R.4 and R.11, and elementary inequalities, α_2 is bounded by

$$\sum_{t=1}^{S-1}\sum_{r=1}^{t-1}\sum_{s=t+1}^{S}\sum_{u=t+1}^{S} b_s'\mathscr{M}b_tb_s'\mathscr{M}b_rb_u'\mathscr{M}b_tb_u'\mathscr{M}b_r$$

$$=O_p\left(q^{-4}\sum_{s,r,t,u=1}^{S}\sum_{i,j=1}^{n}|b_{ir}^*||m_{ij}|\left|b_{js}^*\right|\sum_{i,j=1}^{n}|b_{ir}^*||m_{ij}|\left|b_{ju}^*\right|\sum_{i,j=1}^{n}|b_{it}^*||m_{ij}|\left|b_{js}^*\right|\sum_{i,j=1}^{n}|b_{it}^*||m_{ij}|\left|b_{ju}^*\right|\right)$$

$$=O_p\left(q^{-4}pn^{-1}\sum_{s,r,t=1}^{S}\left(\sum_{i,j=1}^{n}|b_{ir}^*||m_{ij}|\left|b_{js}^*\right|\right)\left(\sum_{i,j=1}^{n}|b_{ir}^*||m_{ij}|\sum_{u=1}^{S}\left|b_{ju}^*\right|\right)\right.$$

$$\left.\times\sum_{i,j=1}^{n}|b_{it}^*||m_{ij}|\left|b_{js}^*\right|\sum_{i=1}^{n}|b_{it}^*|\sup_u\sum_{j=1}^{n}\left|b_{ju}^*\right|\right)$$

$$=O_p\left(q^{-4}p^2n^{-2}\sum_{s,r=1}^{S}\left(\sum_{i,j=1}^{n}|b_{ir}^*||m_{ij}|\left|b_{js}^*\right|\right)\sum_{i=1}^{n}|b_{ir}^*|\sum_{j=1}^{n}\left(\sum_{u=1}^{S}\left|b_{ju}^*\right|\right)\left(\sum_{i,j=1}^{n}\sum_{t=1}^{S}|b_{it}^*||m_{ij}|\left|b_{js}^*\right|\right)\right)$$

$$=O_p\left(q^{-4}p^2n^{-1}\sum_{i,j=1}^{n}\left(\sum_{r=1}^{S}|b_{ir}^*|\right)|m_{ij}|\left(\sum_{s=1}^{S}\left|b_{js}^*\right|\right)\left(\sup_j\sum_{i=1}^{n}|m_{ij}|\right)\sum_{j=1}^{n}\left|b_{js}^*\right|\right)$$

$$=O_p\left(q^{-4}p^2n^{-1}\sup_k\sum_{i,j=1}^{n}|m_{ij}|\sum_{i=1}^{n}|m_{ik}|\right)=O_p\left(q^{-4}p^2n^{-1}\sup_k\sum_{i,j,\ell=1}^{n}|m_{ij}||m_{\ell k}|\right)$$

$$=O_p\left(q^{-4}p^2n^{-1}\sup_k\sum_{i,j,\ell=1}^{n}\left(m_{ij}^2+m_{\ell k}^2\right)\right)=O_p\left(q^{-4}p^2n^{-1}\sum_{i,j,\ell=1}^{n}\left(m_{ij}^2+m_{\ell j}^2\right)\right)$$

$$=O_p\left(q^{-4}p^2n^{-1}\sum_{i,j=1}^{n}m_{ij}^2\right)=O_p\left(q^{-4}p^2\sup_j\sum_{i=1}^{n}m_{ij}^2\right)=O_p\left(pn^{-1}\right),$$

where we used (A.6) in the last step. A similar use of the conditions of the theorem and (A.5) implies that α_1 is

$$O_p\left(q^{-4}\sum_{t=1}^{S-1}\left\{\sum_{s=t+1}^{S}\left(\sum_{i,j=1}^{n}|m_{ij}|\left|b_{jt}^*\right|\left|b_{is}^*\right|\right)^2\right\}^2\right)$$

$$=O_p\left(q^{-4}\left(\sup_{i,j}|m_{ij}|\right)^4\sum_{t=1}^{S-1}\left\{\sum_{s=t+1}^{S}\left(\sum_{i=1}^{n}\left|b_{is}^*\right|\sum_{j=1}^{n}\left|b_{jt}^*\right|\right)^2\right\}^2\right)$$

$$=O_p\left(q^{-4}p^4n^{-4}\sum_{t=1}^{S-1}\left\{\sum_{s=t+1}^{S}\left(\sum_{i=1}^{n}\left|b_{is}^*\right|\right)^2\left(\sum_{j=1}^{n}\left|b_{jt}^*\right|\right)^2\right\}^2\right)$$

$$=O_p\left(q^{-4}p^4n^{-4}\sum_{t=1}^{S-1}\left(\sum_{s=t+1}^{S}\left(\sum_{i=1}^{n}\left|b_{is}^*\right|\right)^2\right)^2\left(\sum_{j=1}^{n}\left|b_{jt}^*\right|\right)^4\right)$$

$$=O_p\left(q^{-4}p^4n^{-4}\left(\sum_{t=1}^{S-1}\sum_{j=1}^{n}\left|b_{jt}^*\right|\right)\left(\sum_{s=t+1}^{S}\sum_{i=1}^{n}\left|b_{is}^*\right|\right)^2\sup_s\left(\sum_{i=1}^{n}\left|b_{is}^*\right|\right)^2\sup_t\left(\sum_{j=1}^{n}\left|b_{jt}^*\right|\right)^3\right),$$

$$=O_p\left(q^{-4}p^4n^{-1}\right)=O_p\left(p^2n^{-1}\right)$$

proving (A.10), as $p^2/n\to 0$ by the conditions of the theorem. □

Proof of Theorem 4.4. In the Supplementary Material. □

Proof of Theorem 5.1. Due to the similarity with proofs in Delgado and Robinson (2015) and Gupta and Robinson (2018), the details are in the Supplementary Material. □

Proof of Theorem 5.2. Denote θ^* as the solution of $\min_\theta \mathcal{E}\left(y_i-\sum_{j=1}^{d_\lambda}\lambda_j w_{i,j}'y-\theta(x_i)\right)^2$. Put $\theta_i^*=\theta^*(x_i)$, $\theta_{0i}=\theta_0(x_i)$, $\widehat{\theta}_i=\psi_i'\widehat{\beta}$, $\widehat{f}_i=f(x_i,\widehat{\alpha})$, and $f_i^*=f(x_i,\alpha^*)$. Then $\widehat{u}_i=y_i-\sum_{j=1}^{d_\lambda}\widehat{\lambda}_j w_{i,j}'y-f(x_i,\widehat{\alpha})=u_i+\theta_{0i}+\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)w_{i,j}'y-\widehat{f}_i$. Proceeding as in the proof of Theorem 4.2, we obtain $n\widehat{m}_n=\widehat{\sigma}^{-2}u'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}u+\widehat{\sigma}^{-2}\sum_{j=1}^{7}A_j$. Thus, compared to the test statistic with no spatial lag (cf. the proof of Theorem 4.2), we have the additional terms

$$A_5=\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)y'W_j'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)W_jy,$$

$$A_6=\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)y'W_j'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(u+\theta_0-\widehat{f}),$$

$$A_7=\left(\Psi\left(\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(u+e)-e+\theta_0-\widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)W_jy.$$

We now show that $A_\ell = o_p(\sqrt{p}), \ell > 4$, so the leading term in $n\widehat{m}_n$ is the same as before. First, $\|y\| = O_p(\sqrt{n})$ from $y = (I_n - \sum_{j=1}^{d_\lambda} \lambda_{j0} W_j)^{-1}(\theta_0 + u)$. Then, with $\|\lambda_0 - \widehat{\lambda}\| = O_p\left(\sqrt{d_\gamma/n}\right)$ by Lemma LS.2 in the Supplementary Material, we have

$$|A_5| \leq \left\|\lambda_0 - \widehat{\lambda}\right\|^2 \sum_{j=1}^{d_\lambda} \left\|W_j\right\|^2 \sup_{\gamma,j} \left\|\Sigma(\gamma)^{-1} \frac{1}{n}\Psi\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}\right\| \|y\|^2$$
$$= O_p\left(d_\gamma/n\right) O_p(1) O_p(n) = O_p\left(d_\gamma\right) = o_p(\sqrt{p}).$$

Uniformly in γ and j,

$$\mathcal{E}\left(u'S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi[\Psi'\Sigma(\gamma)^{-1}\Psi]^{-1}\Psi'\Sigma(\gamma)^{-1}u\right)$$
$$= \mathcal{E}tr\left(\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Sigma S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi\right) = O_p(p)$$

and

$$\mathcal{E}\left(\theta_0'S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi[\Psi'\Sigma(\gamma)^{-1}\Psi]^{-1}\Psi'\Sigma(\gamma)^{-1}u\right)^2$$
$$= O_p\left(\left\|S^{-1}\right\|^2 \sup_\gamma \left\|\Sigma(\gamma)^{-1}\right\|^4 \left\|\frac{1}{n}\Psi\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\right\|^2 \sup_j \left\|W_j\right\|^2 \|\Sigma\| \|\theta_0\|^2\right)$$
$$= O_p(n).$$

Similarly, $\theta_0'S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi[\Psi'\Sigma(\gamma)^{-1}\Psi]^{-1}\Psi'\Sigma(\gamma)^{-1}W_j\theta_0 = O_p(n)$, uniformly. Therefore,

$$\left|\sum_{j=1}^{d_\lambda}(\lambda_{j0} - \widehat{\lambda}_j)y'W_j'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}u\right|$$
$$= \left|\sum_{j=1}^{d_\lambda}(\lambda_{j0} - \widehat{\lambda}_j)(\theta_0 + u)'S^{-1\prime}W_j'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}u\right|$$
$$\leq d_\lambda \left\|\lambda_0 - \widehat{\lambda}\right\| \sup_{\gamma,j}\left|\theta_0'S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi[\Psi'\Sigma(\gamma)^{-1}\Psi]^{-1}\Psi'\Sigma(\gamma)^{-1}u\right|$$
$$+ d_\lambda \left\|\lambda_0 - \widehat{\lambda}\right\| \sup_{\gamma,j}\left|u'S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi[\Psi'\Sigma(\gamma)^{-1}\Psi]^{-1}\Psi'\Sigma(\gamma)^{-1}u\right|$$
$$= O_p\left(\sqrt{d_\gamma/n}\right)O_p(\sqrt{n}) + O_p\left(\sqrt{d_\gamma/n}\right)O_p(p) = O_p\left(\sqrt{d_\gamma}\right) = o_p\left(\sqrt{p}\right),$$

and

$$\left|\sum_{j=1}^{d_\lambda}(\lambda_{j0} - \widehat{\lambda}_j)y'W_j'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(\theta_0 - \widehat{f})\right|$$
$$\leq d_\lambda \left\|\lambda_0 - \widehat{\lambda}\right\| \|y\| \sup_j \left\|W_j\right\| \sup_\gamma \left\|\frac{1}{n}\Psi\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi\right\| \sup_\gamma \left\|\Sigma(\gamma)^{-1}\right\|^2 \left\|\theta_0 - \widehat{f}\right\|$$
$$= O_p\left(\sqrt{d_\gamma/n}\right)O_p\left(\sqrt{n}\right)O_p\left(p^{1/4}\right) = O_p\left(\sqrt{d_\gamma p^{1/4}}\right) = o_p(\sqrt{p}),$$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

so that $A_6 = o_p(\sqrt{p})$. Finally,

$$\left|\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)y'W_j'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}e\right|$$
$$\le d_\lambda \left\|\lambda_0-\widehat{\lambda}\right\|\|y\|\sup_j\left\|W_j\right\|\sup_\gamma\left\|\frac{1}{n}\Psi\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi\right\|\sup_\gamma\left\|\Sigma(\gamma)^{-1}\right\|^2\|e\|$$
$$= O_p\left(\sqrt{d_\gamma/n}\right)O_p\left(\sqrt{n}\right)O_p\left(p^{-\mu}\sqrt{n}\right) = O_p\left(\sqrt{d_\gamma}p^{-\mu}\sqrt{n}\right) = o_p(\sqrt{p}),$$

and

$$\left|(e+\theta_0-\widehat{f})'\Sigma(\widehat{\gamma})^{-1}\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)W_jy\right|$$
$$\le d_\lambda\left\|\lambda_0-\widehat{\lambda}\right\|\left(\|e\|+\left\|\theta_0-\widehat{f}\right\|\right)\sup_\gamma\left\|\Sigma(\gamma)^{-1}\right\|\sup_j\left\|W_j\right\|\|y\|$$
$$= O_p\left(\sqrt{d_\gamma/n}\right)O_p\left(p^{-\mu}\sqrt{n}+p^{1/4}\right)O_p(\sqrt{n}) = O_p\left(\sqrt{d_\gamma}p^{-\mu}\sqrt{n}+\sqrt{d_\gamma}p^{1/4}\right) = o_p(\sqrt{p}),$$

implying that $A_7 = o_p(\sqrt{p})$. □

Proof of Theorem 5.3. Omitted as it is similar to the proof of Theorem 4.4. □

Proof of Proposition 6.1. Because the map $\Sigma : \mathcal{T}^o \to \mathcal{M}^{n\times n}$ is Fréchet-differentiable on $\mathcal{T}^o$, it is also Gâteaux-differentiable and the two derivative maps coincide. Thus, by Theorem 1.8 of Ambrosetti and Prodi (1995),

$$\|\Sigma(t_1)-\Sigma(t_1)\| \le \sup_{t\in\mathcal{T}^o}\|D\Sigma(t)\|_{\mathscr{L}(\mathcal{T}^o,\mathcal{M}^{n\times n})}\left(\|\gamma_1-\gamma_2\|+\sum_{\ell=1}^{d_\zeta}\left\|(\delta_{\ell 1}-\delta_{\ell 2})'\varphi_\ell\right\|_w\right), \tag{A.14}$$

where

$$\sum_{\ell=1}^{d_\zeta}\left\|(\delta_{\ell 1}-\delta_{\ell 2})'\varphi_\ell\right\|_w = \sum_{\ell=1}^{d_\zeta}\sup_{z\in\mathcal{Z}}\left|(\delta_{\ell 1}-\delta_{\ell 2})'\varphi_\ell\right|\left(1+\|z\|^2\right)^{-w/2}$$
$$\le \sum_{\ell=1}^{d_\zeta}\|\delta_{\ell 1}-\delta_{\ell 2}\|\sup_{z\in\mathcal{Z}}\|\varphi_\ell\|\left(1+\|z\|^2\right)^{-w/2}$$
$$\le C\varsigma(r)\sum_{\ell=1}^{d_\zeta}\|\delta_{\ell 1}-\delta_{\ell 2}\| \le C\varsigma(r)\|t_1-t_2\|.$$

The claim now follows by (6.8) in Assumption NPN.2, because $\|\gamma_1-\gamma_2\| \le C\varsigma(r)\|t_1-t_2\|$ for some suitably chosen C. □

Proof of Theorem 6.1. The proof is omitted as it is entirely analogous to that of Theorem 5.1, with the exception of one difference when proving equicontinuity. In the setting of Section 6, we obtain via Proposition 6.1 that $\left\|\Sigma(\tau)-\Sigma\left(\tau^*\right)\right\| = O_p(\varepsilon)$, the

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

$\varsigma(r)$ factor being omitted because only finitely many neighborhoods contribute due to compactness of $\mathcal{T}$. □

Proof of Theorem 6.2. Writing $\delta(z) = \left(\widehat{\delta}_1'\varphi_1(z),\ldots,\widehat{\delta}_{d_\zeta}'\varphi_{d_\zeta}(z)\right)'$ and taking $t_1 = \left(\widehat{\gamma}',\hat{\delta}(z)'\right)'$ and $t_2 = \left(\gamma_0',\zeta_0(z)'\right)'$ in Proposition 6.1 implies (we suppress the argument z)

$$\|\Sigma(\widehat{\tau})-\Sigma\| = O_p\left(\varsigma(r)\left(\|\widehat{\gamma}-\gamma_0\|+\left\|\widehat{\delta}-\zeta_0\right\|\right)\right) = O_p\left(\varsigma(r)\left(\|\widehat{\tau}-\tau_0\|+\|\nu\|\right)\right)$$
$$= O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n},\sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right\}\right),$$

uniformly on $\mathcal{Z}$. Thus, we have

$$\left\|\Sigma(\widehat{\tau})^{-1}-\Sigma^{-1}\right\| \le \left\|\Sigma(\widehat{\tau})^{-1}\right\|\|\Sigma(\widehat{\tau})-\Sigma\|\left\|\Sigma^{-1}\right\| = O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n},\sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right\}\right).$$

And similarly,

$$\left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\tau})^{-1}\Psi\right)^{-1}-\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|$$
$$\le \left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\tau})^{-1}\Psi\right)^{-1}\right\|\left\|\frac{1}{n}\Psi'\left(\Sigma(\widehat{\tau})^{-1}-\Sigma^{-1}\right)\Psi\right\|\left\|\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|$$
$$= O_p\left(\left\|\Sigma(\widehat{\tau})^{-1}-\Sigma^{-1}\right\|\right) = O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n},\sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right\}\right).$$

As in the proof of Theorem 4.2, $n\widehat{m}_n = \widehat{\sigma}^{-2}u'\Sigma(\widehat{\tau})^{-1}\Psi[\Psi'\Sigma(\widehat{\tau})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\tau})^{-1}u + \widehat{\sigma}^{-2}\sum_{k=1}^{4}A_k$, where γ in the parametric setting is changed to τ in this nonparametric setting. Then, by the mean value theorem (MVT)

$$\left|u'\left(\Sigma(\widehat{\tau})^{-1}\Psi[\Psi'\Sigma(\widehat{\tau})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\tau})^{-1}-\Sigma^{-1}\Psi[\Psi'\Sigma^{-1}\Psi]^{-1}\Psi'\Sigma^{-1}\right)u\right|$$
$$\le 2\left(\sup_t\left\|\frac{1}{\sqrt{n}}u'\Sigma(t)^{-1}\Psi\right\|\left\|\left(\frac{1}{n}\Psi'\Sigma(t)^{-1}\Psi\right)^{-1}\right\|\right)\sum_{j=1}^{d_\tau}\left\|\frac{1}{\sqrt{n}}\Psi'\left(\Sigma(\widetilde{\tau})^{-1}\Sigma_j(\widetilde{\tau})\Sigma(\widetilde{\tau})^{-1}\right)u\right\|$$
$$\times\left|\widetilde{\tau}_j-\tau_{j0}\right| + 2\sup_t\left\|\frac{1}{\sqrt{n}}u'\Sigma(t)^{-1}\Psi\right\|\left\|\left(\frac{1}{n}\Psi'\Sigma(t)^{-1}\Psi\right)^{-1}\right\|\left\|\frac{1}{\sqrt{n}}\Psi'(\Sigma_0-\Sigma)u\right\|$$
$$+\left\|\frac{1}{\sqrt{n}}u'\Sigma^{-1}\Psi\right\|^2\left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\tau})^{-1}\Psi\right)^{-1}-\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|$$
$$= O_p(\sqrt{p})O_p(d_\tau\sqrt{p}\varsigma(r)/\sqrt{n}) + O_p(\sqrt{p})O_p\left(\sqrt{p}\varsigma(r)\sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right)$$

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

$$+ O_p(p)O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n}, \sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right\}\right)$$
$$= O_p\left(p\varsigma(r)\max\left\{d_\tau/\sqrt{n}, \sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right\}\right) = o_p(\sqrt{p}),$$

where the last equality holds under the conditions of the theorem. Next, it remains to show that $A_k = o_p(p^{1/2}), k = 1, \ldots, 4$. The order of A_k, $k \le 3$, is the same as the parametric case:

$$|A_1| = \left|u'\Sigma(\widehat{\tau})^{-1}\left(\theta_0 - \widehat{f}\right)\right| \le \sup_{\alpha,t}\left\|u'\Sigma(t)^{-1}\frac{\partial f(x,\alpha)}{\partial\alpha_j}\right\|\left|\alpha_j^* - \widetilde{\alpha}_j\right| + \frac{p^{1/4}}{n^{1/2}}\sup_t\left\|u'\Sigma(t)^{-1}h\right\|$$
$$= O_p(\sqrt{n})O_p(\frac{1}{\sqrt{n}}) + O\left(\frac{p^{1/4}}{n^{1/2}}\right)O_p(\sqrt{n}) = O_p(p^{1/4}) = o_p(p^{1/2}),$$
$$|A_2| = \left|(u+\theta_0 - \widehat{f})'\left(\Sigma(\widehat{\tau})^{-1} - \Sigma(\widehat{\tau})^{-1}\Psi[\Psi'\Sigma(\widehat{\tau})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\tau})^{-1}\right)e\right|$$
$$\le \sup_t|u'\Sigma(t)^{-1}e| + \sup_t\left|u'\Sigma(t)^{-1}\Psi[\Psi'\Sigma(t)^{-1}\Psi]^{-1}\Psi'\Sigma(t)^{-1}e\right|$$
$$+ \|\theta_0 - \widehat{f}\|\sup_t\left(\left\|\Sigma(t)^{-1}\right\| + \left\|\Sigma(t)^{-1}\Psi[\Psi'\Sigma(t)^{-1}\Psi]^{-1}\Psi'\Sigma(t)^{-1}\right\|\right)\|e\|$$
$$= O_p(p^{-\mu}n^{1/2}) + O_p(p^{-\mu+1/4}n^{1/2}) = O_p(p^{-\mu+1/4}n^{1/2}) = o_p(\sqrt{p}),$$
$$|A_3| = \left|u'\Sigma(\widehat{\tau})^{-1}\Psi\left(\Psi'\Sigma(\widehat{\tau})^{-1}\Psi\right)^{-1}\Psi'\Sigma(\widehat{\tau})^{-1}(\theta_0 - \widehat{f})\right|$$
$$\le \sup_{\alpha,t}\sum_{j=1}^{d_\alpha}\left\|u'\Sigma(t)^{-1}\Psi\left(\Psi'\Sigma(t)^{-1}\Psi\right)^{-1}\Psi'\Sigma(t)^{-1}\frac{\partial f(x,\alpha)}{\partial\alpha_j}\right\|\left|\alpha_j^* - \widetilde{\alpha}_j\right|$$
$$+ \frac{p^{1/4}}{n^{1/2}}\sup_t\left\|u'\Sigma(t)^{-1}\Psi\left(\Psi'\Sigma(t)^{-1}\Psi\right)^{-1}\Psi'\Sigma(t)^{-1}h\right\|$$
$$= O_p(1) + O_p(p^{1/4}) = O_p(p^{1/4}) = o_p(p^{1/2}).$$

However, A_4 has a different order. Under H_ℓ,

$$A_4 = \left(\theta_0 - \widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\left(\theta_0 - \widehat{f}\right)$$
$$= \left(\theta_0 - \widehat{f}\right)'\Sigma_0^{-1}\left(\theta_0 - \widehat{f}\right) + \left(\theta_0 - \widehat{f}\right)'\left(\Sigma(\widehat{\tau})^{-1} - \Sigma^{-1}\right)\left(\theta_0 - \widehat{f}\right)$$
$$= \frac{p^{1/2}}{n}h'\Sigma_0^{-1}h + o_p(1) + O_p\left(p^{1/2}\right)O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n}, \sqrt{\sum_{\ell=1}^{d_\zeta} r_\ell^{-2\kappa_\ell}}\right\}\right)$$
$$= \frac{p^{1/2}}{n}h'\Sigma_0^{-1}h + o_p(\sqrt{p}),$$

where the last equality holds under the conditions of the theorem. Combining these together, we have $n\widehat{m}_n = \widehat{\sigma}^{-2}\widehat{v}'\Sigma(\widehat{\tau})^{-1}\widehat{u} = \sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon + \left(p^{1/2}/n\right)h'\Sigma_0^{-1}h + o_p(\sqrt{p})$, under H_ℓ, and the same expression holds with $h = 0$ under H_0. □

https://doi.org/10.1017/S0266466622000445 Published online by Cambridge University Press

Proof of Theorem 6.3. Omitted as it is similar to the proof of Theorem 4.4. □

SUPPLEMENTARY MATERIAL

Gupta, A. and Qu, X. (2022): Supplement to "Consistent specification testing under spatial dependence," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466622000445.

REFERENCES

Ambrosetti, A. & G. Prodi (1995) *A Primer of Nonlinear Analysis*. Cambridge University Press.

Anatolyev, S. (2012) Inference in regression models with many regressors. *Journal of Econometrics* 170, 368–382.

Autant-Bernard, C. & J.P. LeSage (2011) Quantifying knowledge spillovers using spatial autoregressive models. *Journal of Regional Science* 51, 471–496.

Bloom, N., M. Schankerman, & J. van Reenen (2013) Identifying technology pillovers and product market rivalry. *Econometrica* 81, 1347–1393.

Case, A.C. (1991) Spatial patterns in household demand. *Econometrica* 59, 953–965.

Chen, X. (2007) *Chapter 76: Large sample sieve estimation of semi-nonparametric models*. In J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, pp. 5549–5632. North-Holland.

Chen, X., H. Hong, & E. Tamer (2005) Measurement error models with auxiliary data. *Review of Economic Studies* 72, 343–366.

Cliff, A.D. & J.K. Ord (1973) *Spatial Autocorrelation*. Pion.

Conley, T.G. & B. Dupor (2003) A spatial analysis of sectoral complementarity. *Journal of Political Economy* 111, 311–352.

Cooke, R.G. (1950) *Infinite Matrices & Sequence Spaces*. Macmillan and Company.

Crespo Cuaresma, J. & M. Feldkircher (2013) Spatial filtering, model uncertainty and the speed of income convergence in Europe. *Journal of Applied Econometrics* 28, 720–741.

De Jong, R.M. & H.J. Bierens (1994) On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory* 10, 70–90.

De Oliveira, V., B. Kedem, & D.A. Short (1997) Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association* 92, 1422–1433.

Debarsy, N., F. Jin, & L.F. Lee (2015) Large sample properties of the matrix exponential spatial specification with an application to FDI. *Journal of Econometrics* 188, 1–21.

Delgado, M. & P.M. Robinson (2015) Non-nested testing of spatial correlation. *Journal of Econometrics* 187, 385–401.

Ertur, C. & W. Koch (2007) Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics* 22, 1033–1062.

Evans, P. & J.U. Kim (2014) The spatial dynamics of growth and convergence in Korean regional incomes. *Applied Economics Letters* 21, 1139–1143.

Fuentes, M. (2007) Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association* 102, 321–331.

Gao, J. & V. Anh (2000) A central limit theorem for a random quadratic form of strictly stationary processes. *Statistics and Probability Letters* 49, 69–79.

Gneiting, T. (2002) Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* 97, 590–600.

Gradshteyn, I.S. & I.M. Ryzhik (1994) *Table of Integrals, Series and Products*, *5th Edition.* Academic Press.

Gupta, A. (2018a) Autoregressive spatial spectral estimates. *Journal of Econometrics* 203, 80–95.

Gupta, A. (2018b) Nonparametric specification testing via the trinity of tests. *Journal of Econometrics* 203, 169–185.

Gupta, A., S. Kokas, & A. Michaelides (2021) Credit market spillovers in a financial network. Working paper. doi:10.2139/ssrn.3063886.

Gupta, A. & P.M. Robinson (2015) Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics* 186, 19–31.

Gupta, A. & P.M. Robinson (2018) Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. *Journal of Econometrics* 202, 92–107.

Hahn, J., G. Kuersteiner, & M. Mazzocco (2020) Joint time-series and cross-section limit theory under mixingale assumptions. *Econometric Theory*, Published online 11 August 2020. doi:10.1017/S0266466620000316.

Han, X., L.-F. Lee, & X. Xu (2021) Large sample properties of Bayesian estimation of spatial econometric models. *Econometric Theory* 37, 708–746.

Hannan, E.J. (1970) *Multiple Time Series*. Wiley.

Heston, A., R. Summers, & B. Aten (2002) *Penn World Tables Verison 6.1. Downloadable Dataset*. Center for International Comparisons at the University of Pennsylvania.

Hidalgo, J. & M. Schafgans (2017) Inference and testing breaks in large dynamic panels with strong cross sectional dependence. *Journal of Econometrics* 196, 259–274.

Hillier, G. & F. Martellosio (2018a) Exact and higher-order properties of the MLE in spatial autoregressive models, with applications to inference. *Journal of Econometrics* 205, 402–422.

Hillier, G. & F. Martellosio (2018b) Exact likelihood inference in group interaction network models. *Econometric Theory* 34, 383–415.

Ho, C.-Y., W. Wang, & J. Yu (2013) Growth spillover through trade: A spatial dynamic panel data approach. *Economics Letters* 120, 450–453.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.

Huber, P.J. (1973) Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics* 1, 799–821.

Jenish, N. & I.R. Prucha (2009) Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150, 86–98.

Jenish, N. & I.R. Prucha (2012) On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170, 178–190.

Jin, F. & L.F. Lee (2015) On the bootstrap for Moran's I test for spatial dependence. *Journal of Econometrics* 184, 295–314.

Kelejian, H.H. & I.R. Prucha (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17, 99–121.

Kelejian, H.H. & I.R. Prucha (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219–257.

Koenker, R. & J.A.F. Machado (1999) GMM inference when the number of moment conditions is large. *Journal of Econometrics* 93, 327–344.

König, M.D., D. Rohner, M. Thoenig, & F. Zilibotti (2017) Networks in conflict: Theory and evidence from the Great War of Africa. *Econometrica* 85, 1093–1132.

Kuersteiner, G.M. & I.R. Prucha (2020) Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica* 88, 2109–2146.

Lee, J., P.C.B. Phillips, & F. Rossi (2020) Consistent misspecification testing in spatial autoregressive models. Cowles Foundation Discussion paper 2256.

Lee, J. & P.M. Robinson (2016) Series estimation under cross-sectional dependence. *Journal of Econometrics* 190, 1–17.

Lee, L.F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899–1925.

Lee, L.F. & X. Liu (2010) Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26, 187–230.

LeSage, J.P. & R. Pace (2007) A matrix exponential spatial specification. *Journal of Econometrics* 140, 190–214.
Malikov, E. & Y. Sun (2017) Semiparametric estimation and testing of smooth coefficient spatial autoregressive models. *Journal of Econometrics* 199, 12–34.
Matérn, B. (1986) *Spatial Variation*. Almaenna Foerlaget.
Mohnen, M. (2022) Stars and brokers: Peer effects among medical scientists. *Management Science* 68, 2377–3174.
Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.
Oettl, A. (2012) Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science* 58, 1122–1140.
Pinkse, J. (1999). Asymptotic Properties of Moran and Related Tests and Testing for Spatial Correlation in Probit Models. Mimeo, Department of Economics, University of British Columbia and University College London.
Pinkse, J., M.E. Slade, & C. Brett (2002) Spatial price competition: A semiparametric approach. *Econometrica* 70, 1111–1153.
Portnoy, S. (1984) Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Annals of Statistics* 12, 1298–1309.
Portnoy, S. (1985) Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Annals of Statistics* 13, 1403–1417.
Robinson, P.M. (1972) Non-linear regression for multiple time-series. *Journal of Applied Probability* 9, 758–768.
Robinson, P.M. (2011) Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics* 165, 5–19.
Robinson, P.M. & F. Rossi (2015) Refined tests for spatial correlation. *Econometric Theory* 31, 1249–1280.
Robinson, P.M. & S. Thawornkaiwong (2012) Statistical inference on regression with spatial dependence. *Journal of Econometrics* 167, 521–542.
Scott, D.J. (1973) Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability* 5, 119–137.
Stein, M. (1999) *Interpolation of Spatial Data*. Springer.
Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.
Su, L. & S. Jin (2010) Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics* 157, 18–33.
Su, L. & X. Qu (2017) Specification test for spatial autoregressive models. *Journal of Business & Economic Statistics* 35, 572–584.
Sun, Y. (2016) Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics* 195, 134–153.
Sun, Y. (2020) The LLN and CLT for U-statistics under cross-sectional dependence. *Journal of Nonparametric Statistics* 32, 201–224.