A nonparametric spatial regression model using partitioning estimators

Jose Olmo[a,b,*], Marcos Sanso-Navarro[a]

[a]*Departamento de Análisis Económico, Universidad de Zaragoza, Facultad de Economía y Empresa. Gran Vía 2, Zaragoza, 50005, Spain*

[b]*Division of Economics, University of Southampton, School of Social Sciences, Southampton, SO17 1BJ, United Kingdom*

Abstract

Conventional spatial regression models are extended by modelling the spatial effects of the exogenous regressor model (SLX) as a functional coefficient. This coefficient is estimated by partitioning the domain of the spatial variable into a set of disjoint intervals and approximating the function using local Taylor expansions. The asymptotic properties of the proposed partitioning estimator are derived, and pointwise and uniform tests for the presence of spatial effects are developed. An empirical application of this work is used to study environmental Engel curves and provides strong evidence of neighbouring effects in the relationship between households' income and the amount of pollution embodied in the goods and services they consume.

Keywords: Spatial regression, partitioning estimator, interaction matrix, asymptotic theory, environmental Engel curve

1. Introduction

Spillovers, which are interpreted as exogenous interactions in the explanatory variables, are one of the three types of interactions across units in a cross section of observations. The other two types are endogenous interactions that affect the dependent variable and interaction effects among the error terms. Each of these models is represented in the spatial econometrics literature by a different specification of the spatial effects. The first type, which considers exogenous interactions, is usually specified as an SLX model ($Y = X\beta + WX\gamma + u$) in which the dependent variable (Y) is a linear function of the regressors X. There is a direct effect of the regressors on the dependent variable through the β parameters and an indirect effect through the spatial matrix W that allows for spillovers from the covariate X_j on Y_i for $i \neq j$. The second model specification is the spatial autoregressive (SAR) model,

*Corresponding author.

Email addresses: joseolmo@unizar.es and J.B.Olmo@soton.ac.uk (Jose Olmo), marcossn@unizar.es (Marcos Sanso-Navarro)

which adds a weighted average of nearby values of the dependent variable to the base set of explanatory variables: $Y = WY + X\beta + u$. The third specification given by the spatial error model (SEM) uses a similar structure to directly model spatial relationships among the errors: $Y = X\beta + u$, with the error variable $u = \theta W u + e$, where θ captures the spatial correlation between the error terms. The spatial Durbin model combines spatial features in the dependent variable and exogenous regressors.

Spatial econometric models suffer, in general, from identification problems. Halleck Vega and Elhorst (2015) discuss three types of identification problems. First, different spatial econometric models are generally impossible to distinguish without assuming prior knowledge about the true data-generating process, including the spatial W matrix (see Gibbons and Overman (2012), Corrado and Fingleton (2012) and Partridge et al. (2012)). For this reason, empirical analyses usually report estimation results under different specifications of the dependence structure. Kelejian (2008) and Kelejian and Piras (2011) develop test statistics to select a spatial weights matrix from a set of candidates. Lam and Souza (2020) estimate the best linear combination of candidates and a sparse adjustment matrix using the least absolute shrinkage and selection operator (LASSO). Higgins and Martellosio (2022) develop a similar approach based on a penalised quasi-maximum likelihood estimator that controls for unobserved factors. Bhattacharjee and Jensen-Butler (2013) estimate the interaction matrix from the spatial autocovariance matrix with panel data, showing that identification is only partial. Rose (2017) identifies peer effects in a social network from the fluctuations in the variance and covariance of the outcomes. Second, spatial models are characterised by N(N-1) potential relationships among the observations, but only N data observations are available (see McMillen (2012)). The third identification problem occurs when the unknown parameters of a model cannot be uniquely recovered from their reduced-form specification even if the spatial econometric model and W are correctly specified. Although this problem can arise in models exhibiting spatial endogeneity, such as the SAR model, the spatial econometrics literature has made significant progress in this direction by developing techniques for the consistent estimation of the model parameters given the correct specification of the spatial model and certain assumptions about the weight matrix (see Kelejian and Prucha (1998, 1999), Lee (2004), Bramoullé et al. (2009) and Sun (2016)). See also Anselin (1988) for an excellent monograph on spatial econometrics models.

In this paper, we will focus on the first two issues by proposing a functional specification of an SLX model. Our contribution is to model the spatial effect as a functional coefficient that is approximated nonparametrically using a series of Taylor expansions applied over disjoint intervals covering a partition of the spatial variable. This approach is nonparametric because the Taylor approximation and the partition require a number of regressors that increases with the sample size (see Pinkse et al. (2002), Sun (2016) and Koroglu and Sun (2016) for similar frameworks). The second contribution is to extend the standard SLX model by allowing for spatial effects that are characterised by random variables different from the geographical distance. This extension has nontrivial implications for modelling purposes. While the geographical distance is treated as a non-stochastic variable, the latter variable indexing the functional coefficient is a random variable, adding another layer of complexity to the model.

The identification of $N(N-1)$ interactions in a cross section of N observations is possible due to the specification of the functional coefficient characterising spillovers between observations. Each pairwise interaction is interpreted as a realisation of the functional parameter. This setting takes advantage of smoothing techniques for approximating unknown functional parameters (see Fan and Gijbels (1996), Cai and Li (2008), Cai and Xu (2008) and Cai and Xiao (2012) for local polynomial approximations in different contexts). Sun (2016) applies a similar procedure in a spatial model by only considering endogenous SAR effects, and Koroglu and Sun (2016) apply a similar procedure in a nonparametric spatial Durbin model. In contrast to these authors, we do not use kernel methods to control for the local character of the approximation. Instead, we approximate the functional coefficient using Taylor expansions over an increasing number of disjoint intervals defining a partition of the compact support of the spatial variable. This methodology allows us to estimate the coefficients defining the *local* Taylor expansions by minimising the residual sum of squares over each interval. In this respect, our estimation approach can be framed in the sieve regression literature (see Newey (1997) for a general setting and Pinkse et al. (2002) for an application of series estimators to endogenous spatial models) and, more specifically, it belongs to the class of partitioning estimators (see Cattaneo and Farrell (2013) and Cattaneo et al. (2020)), which we extend to the spatial literature. The accuracy of our approximation also depends on the order of the Taylor expansion of the functional coefficient, which, in contrast to the number of intervals defining the partition, is assumed to be finite.

As an additional contribution, we extend the existing theory on partitioning estimators to derive the pointwise and uniform convergence of the partitioning estimator of the functional coefficient for the SLX model and leave the analysis of the model with endogenous features for future research. In particular, we derive pointwise estimates based on realisations of the functional coefficient at specific locations that are shown to converge at a rate N to a standard normal distribution. This convergence rate is due to the presence of $N(N-1)$ potential neighbours and is similar to the square-root-of-N convergence of the partitioning estimator in a nonparametric setting (see Cattaneo and Farrell (2013); Cattaneo et al. (2020)). We use the asymptotic results derived by these authors to extend the convergence of the partitioning estimator to a centred Gaussian process in the functional space. We also develop pointwise t-tests to evaluate the statistical significance of the spatial effects on specific locations and a uniform test that extends the analysis to the compact support of the spatial variable. The implementation of the uniform test is not straightforward, as it is a composite hypothesis. Under the null hypothesis, we face Davies (1977, 1987)'s problem of the lack of identification of the nuisance parameter. Thus, the asymptotic null distribution of the composite test H_0 is a zero-mean Gaussian process with a covariance kernel that cannot be tabulated. Nevertheless, we follow the theory of Cattaneo et al. (2020) to derive the asymptotic distribution of the test, and the wild bootstrap methods of Hansen (1996) are used to approximate its finite-sample distribution under the null hypothesis.

The finite-sample performance of these tests is evaluated using Monte Carlo methods. We compute the bias and root mean square error (RMSE) of the parameter estimators to demonstrate their consistency. Second, we analyse the size and power of the marginal t-tests and uniform test. The simulations show a very good performance for both tests in terms of power and a reliable empirical size. The proposed methodology is illustrated in an empirical application that involves studying the environmental Engel curves (EECs) discussed in an influential work by Levinson and O'Brien (2019). We extend the analysis carried out by these authors by incorporating neighbouring effects in the relationship between households' income and pollution measures. The variable characterising the spatial structure of the model is the L_1 distance between the pollution content at two different units, such that two neighbouring observations are characterised by similar pollution patterns. Building on recent studies about peer effects in household consumption and energy behaviours (Agarwal et al., 2021; Wolske et al., 2020; De Giorgi et al., 2020), we provide strong empirical evidence of neighbouring effects in the relationship between different forms of environmental pollution and after-tax household income discovered by Levinson and O'Brien (2019). The sign of this relationship is positive, suggesting a reinforcing effect of income on pollution coming from households with similar income levels.

The rest of the paper is structured as follows. Section 2 introduces a functional SLX model. In Section 3, we propose a nonparametric estimator for the functional parameter based on the theory on partitioning estimators and derive the asymptotic theory of the proposed estimators. In particular, we show the consistency and uniform convergence of the functional parameter estimator obtained by our augmented regression model. This section also derives pointwise convergence results of the convergence of the estimators to a normal distribution and weak convergence to a Gaussian process. Section 4 presents different hypothesis tests to statistically assess the presence of spatial effects in the data. This section also discusses model selection and the optimal choice of tuning parameters such as the width of the intervals defining the partition and the order of the Taylor expansion. Section 5 presents a Monte Carlo exercise to evaluate parameter estimation and hypothesis testing in finite samples. Section 6 contains the empirical application, and Section 7 concludes the paper. An appendix contains the mathematical proofs of the main results of the paper.

In what follows, $\|A\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)^{1/2}$ denotes the L^2 norm for A, an $m \times n$ matrix, and $\|a\| = \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}$ denotes the L^2 norm for a vector a of dimension n.

2. A novel spatial regression model

This section introduces the nonparametric SLX model obtained by considering a functional coefficient for the spatial effects. This section also proposes estimators of the spatial parameters based on the theory on partitioning estimators (see Cattaneo and Farrell (2013) and Cattaneo et al. (2020), which are recent seminal contributions).

2.1. The baseline model

The proposed model is

$$y_i = \lambda x_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} w(d_{ij}) x_j + \varepsilon_i, \ \ i = 1, \ldots, N, \tag{1}$$

where y_i is the dependent variable and x_i is the exogenous regressor evaluated at unit i. Both variables are demeaned such that there is no intercept in the model; λ is the slope coefficient measuring the direct relationship between the dependent variable and the regressor and ε_i is a zero-mean error term. The indirect effect between y_i and the regressors evaluated at different units is what we call the spatial effect. This effect is captured by an $N \times N$ interaction matrix with elements $w(d_{ij})$. The diagonal elements of this matrix are zero such that $w(d_{ii}) = 0$ for $d_{ii} = 0$ and the off-diagonal terms are characterised by the distance d_{ij} between the different units. This distance is constructed as $d_{ij} = f(u_i, u_j)$, with $f(\cdot, \cdot) \in \chi \subset \mathbb{R}^+$. It is a metric defined by realisations $\{u_i, u_j\}$ of a spatial variable U evaluated at units i and j; χ is a compact set. To illustrate the model and estimation procedure, we consider one regressor, but the methods below can be extended naturally to the case of a finite number of regressors. Our objective is to estimate the $N(N-1)$ parameters of the interaction matrix from a sample of N observations. To do this, we model $w(d_{ij})$ as a functional coefficient (see Pinkse et al. (2002) and Sun (2016) for similar settings).

The following assumptions impose the exogeneity of the regressors and the independence of the errors, and they introduce regularity conditions on the functional coefficient and the elements of the partition. Before this, we introduce the following notation. Let $[z_k - h_N, z_k + h_N)$ be a generic interval of the partition, with h_N being some bandwidth parameter such that $h_N \to 0$ as $N \to \infty$. This nuisance parameter characterises a partition of the compact set χ given by $\bigcup_{k=1}^{K} [z_k - h_N, z_k + h_N)$. Let $1_{kN}(d)$ be an indicator function with $1_{kN}(d) = 1$ if d belongs to the interval $[z_k - h_N, z_k + h_N)$ and $1_{kN}(d) = 0$ otherwise. Then, $p_{kN} = E[1_{kN}(d)] = P\{d \in [z_k - h_N, z_k + h_N)\}$ such that $\sum_{k=1}^{K} p_{kN} = 1$.

Assumption A.

(A1) $\{(x_i, u_i, \varepsilon_i)\}$ is an independent and identically distributed (*iid*) sequence across index i and y_i is generated from model (1). The regressor $E[x_i^4] < \infty$, for $i = 1, \ldots, N$.

(A2) The functional coefficient $w(d)$ is $(q+1)$-times continuously differentiable on (and an extension of) the compact set $\chi \subset \mathbb{R}^+$, with $q \geq 0$ fixed. The compact set is partitioned into K_N disjoint intervals centred at the knots $\{z_{1N}, \ldots, z_{KN}\}$. The subscript k refers to the specific knot k and the subscript N refers to the dependence of the knots on the sample size through the choice of the bandwidth parameter h_N. In what follows, we remove the subscript N whenever possible from $1_{kN}(\cdot)$, h_N, K_N, z_{kN} and p_{kN}.

(A3) The distance variable $d \in \chi$ is continuously distributed with a Lebesgue density that is bounded, and it is bounded away from zero on χ.

(A4) $E[\varepsilon_i \mid X_i = x, d_i = d] = 0$ for $d_i = \{d_{i1}, \ldots, d_{i,i-1}, d_{i,i+1}, \ldots, d_{iN}\}$ and $d \in \chi^{N-1}$; $\sigma^2(x, d) = E(\varepsilon_i^2 \mid X_i = x, d_i = d)$ is continuous and bounded away from zero, and $E[\varepsilon_i^4 \mid X_i = x, d_i = d] < \infty$, for all i and any $(x, d) \in \mathbb{R} \times \chi^{N-1}$.

(A5) $K/N \to 0$ and $N/K^{q+1} \to 0$ as $K, N \to \infty$.

Under assumption A1, the regressors are *iid* across individuals and their fourth statistical moment is finite. This will be required for proving the consistency of the sample covariance matrices. Assumption A2 is a classical smoothness condition on the functional coefficient that allows us to approximate the unknown function $w(d)$ using different Taylor expansions applied to a disjoint set of shrinking intervals covering the compact set. Assumption A3 guarantees that all the intervals in the partitioning of the compact set are non-empty. Assumption A4 imposes moment conditions on the error term of the regression equation (1) conditional on the vector of exogenous covariates x and d. The assumption also guarantees the smoothness of the conditional variance as a function of the covariates and the existence of the conditional fourth moment of the error term. Importantly, the model accommodates the presence of conditional heteroscedasticity as a function of the covariates. Assumption A5 imposes some regularity conditions on the number of intervals characterising the partition with respect to the order of the Taylor expansion and the sample size. Moreover, the definition of the partition of χ entails the condition $h \asymp K^{-1}$ such that according to the conditions in assumption A5, it follows that $Nh \to \infty$ and $Nh^{q+1} \to 0$ as $N \to \infty$.

Under the above partition and noting that $\sum_{k=1}^{K} 1_k(d) = 1$, the functional coefficient can be expressed as a Taylor expansion of order q, with q fixed, such that

$$w(d) = \sum_{k=1}^{K} \sum_{m=0}^{q} \frac{1}{m!} w^{(m)}(z_k)(d - z_k)^m 1_k(d) + R(d), \tag{2}$$

where $w^{(m)}(z_k)$ is the m^{th} derivative of $w(\cdot)$ evaluated at z_k, $w^{(0)}(z_k) = w(z_k)$ is the functional coefficient evaluated at z_k and $R(d) = \sum_{k=1}^{K} w^{(q+1)}(c_k)(d - z_k)^{q+1} 1_k(d)$ is the remainder of the Taylor expansion, with $c_k \in (z_k - h, z_k + h)$.

Local polynomial approximations of functional coefficients are proposed by Fan and Gijbels (1996), Cai and Li (2008), Cai and Xu (2008) and Cai and Xiao (2012), among others. However, in contrast to these studies, the approximation proposed below is not based on kernel smoothers of the neighbouring observations but on a partitioning of the compact set into disjoint intervals. More formally, let $\widetilde{x}_i^{(km)} = \sum_{\substack{j=1 \\ j \neq i}}^{N} x_j (d_{ij} - z_k)^m 1_k(d_{ij})$ be regression variables indexed by $k = 1, \ldots, K$ and $m = 0, \ldots, q$, and let $\gamma_{km} = \frac{1}{m!} w^{(m)}(z_k)$

be the corresponding spatial regression coefficients. Similarly, let $\overline{R}_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} R(d_{ij})x_j$ be the aggregate remainder term. By plugging in the Taylor expansion in equation (1) and using the above notation, we obtain the following regression model:

$$y_i = \lambda x_i + \sum_{k=1}^{K}\sum_{m=0}^{q}\gamma_{km}\widetilde{x}_i^{(km)} + \overline{R}_i + \varepsilon_i. \tag{3}$$

By applying a Taylor expansion to $w(d_{ij})$ around the different knots in the partition, we reduce the above infinite-dimensional problem with N^2 parameters to a regression model with $K(q+1)$ parameters with $K, N \to \infty$, $K/N \to 0$ and q fixed, under assumption A5.

A more convenient specification for estimation purposes is its matrix form:

$$Y = X\lambda + \mathbb{X}\,\Gamma + \overline{R} + \varepsilon, \tag{4}$$

where $Y = (Y_1, \dots, Y_N)'$ and $\mathbb{X} = [\mathbb{X}_1, \dots, \mathbb{X}_K]$ are matrices of dimensions $N \times K(q+1)$. Each $\mathbb{X}_k$ defines an $N \times (q+1)$ matrix with elements $(\widetilde{x}_i^{(k0)}, \dots, \widetilde{x}_i^{(kq)})$. Similarly, the vector of coefficients satisfies $\Gamma = (\Gamma'_1, \dots, \Gamma'_K)'$, with $\Gamma_k = (\gamma_{k0}, \dots, \gamma_{kq})'$. The vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$ is the error term and $\overline{R}$ is a vector with the remainder terms $\overline{R}_i$ for $i = 1, \dots, N$.

3. Partitioning estimator for the SLX model

3.1. Estimation

Using the partitioned inverse, a suitable estimator of λ is

$$\widehat{\lambda} = \left(\widehat{X}'_u \widehat{X}_u\right)^{-1} \widehat{X}'_u (Y - \widetilde{Y}), \tag{5}$$

with $\widehat{X}_u = M_{\mathbb{X}} X$, where $M_{\mathbb{X}} = I_N - \mathbb{P}_{\mathbb{X}}$ and $\mathbb{P}_{\mathbb{X}} = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$. Similarly, $\widetilde{Y} = \mathbb{P}_{\mathbb{X}} Y$ is the projection of Y on $\mathbb{X} = [\mathbb{X}_1, \dots, \mathbb{X}_K]$. This matrix is partitioned into blocks such that each of the spatial parameters is estimated from the partitioned regression as follows:

$$\widehat{\Gamma}_k = \left(\sum_{i=1}^{N} \mathbb{X}'_{ki}\mathbb{X}_{ki}\right)^{-1} \sum_{i=1}^{N} \mathbb{X}'_{ki}(y_i - \widehat{\lambda} x_i), \tag{6}$$

for each $\widehat{\Gamma}_k = (\widehat{\gamma}_{k0}, \dots, \widehat{\gamma}_{kq})'$ (see Cattaneo and Farrell (2013) for details on partitioned regressors). The estimator of $w(d)$ is obtained from the Taylor expansion (2) as follows:

$$\widehat{w}(d) = \sum_{k=1}^{K}\sum_{m=0}^{q}\widehat{\gamma}_{km}(d - z_k)^m 1_k(d) = \sum_{k=1}^{K}\widehat{\Gamma}'_k v_k(d) 1_k(d), \tag{7}$$

with $v_k(d) = [1, (d - z_k), (d - z_k)^2, \dots, (d - z_k)^q]'$.

Another important quantity for making statistical inferences about the spatial parameters is the variance of the parameter estimators. Let $\Phi_0 = E[X'M_{\mathbb{X}}X]$ and $\Psi_0 = E[X'M_{\mathbb{X}}X\varepsilon^2]$. The sample counterparts are $\widehat{\Phi} = \frac{1}{N}\sum_{i=1}^{N}\widehat{X}'_{ui}\widehat{X}_{ui}$ and $\widehat{\Psi} = \frac{1}{N}\sum_{i=1}^{N}\widehat{X}'_{ui}\widehat{X}_{ui}e_i^2$, where $e_i = y_i - \widehat{\lambda}x_i - \sum_{k=1}^{K}\mathbb{X}_{ki}\ \widehat{\Gamma}_k$. Then,

$$\widehat{V}\left(\widehat{\lambda}\right) = \frac{1}{N}\widehat{\Phi}^{-1}\widehat{\Psi}\widehat{\Phi}^{-1}. \tag{8}$$

Similarly, let $Q_k = E[\mathbb{X}'_{ki}\mathbb{X}_{ki}]/p_k$ and $A_k = E[\mathbb{X}'_{ki}\mathbb{X}_{ki}\varepsilon_i^2]/p_k$ be the spatial covariance matrices. An appropriate estimator of the variance of the vector of spatial parameter estimators is

$$\widehat{V}\left(\widehat{\Gamma}_k\right) = \frac{1}{\alpha_N p_k}\widetilde{Q}_k^{-1}\widetilde{A}_k\widetilde{Q}_k^{-1}, \tag{9}$$

where $\alpha_N = N(N-1)$ is a standardising constant for the spatial coefficients, $\widetilde{Q}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\mathbb{X}'_{ki}\mathbb{X}_{ki}$, $\widetilde{A}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\mathbb{X}'_{ki}\mathbb{X}_{ki}e_i^2$ and p_k captures the probability of belonging to a given interval of the partition of χ. From these expressions, a natural estimator of the variance of $\widehat{w}(d)$ in equation (7) is

$$\widehat{V}(\widehat{w}(d)) = \sum_{k=1}^{K} v'_k(d)\widehat{V}(\widehat{\Gamma}_k)v_k(d)1_k(d). \tag{10}$$

3.2. Asymptotic convergence of the SLX estimator

This subsection presents pointwise and uniform convergence results for the functional estimator $\widehat{w}(d)$ in equation (7). This subsection also presents results that are necessary for making asymptotic inferences on the pointwise estimates and explores uniform approximations and convergence results when the estimator is considered a process in $d \in \chi$. The following regularity conditions guarantee the existence of the population covariance matrices and suitable conditions for applying the law of large numbers and the central limit theorem in an *iid* setting.

Assumption B. The matrices Φ_0 and Ψ_0 are positive definite, such that $\|\Phi_0\| < \infty$, $\|\Psi_0\| < \infty$, $\|\Phi_0^{-1}\| < \infty$ and $\|\Psi_0^{-1}\| < \infty$. Similarly, we impose $E[\|X'M_{\mathbb{X}}X\|^2] < \infty$. We also assume that $\|Q_k\| < \infty$ and $\|A_k\| < \infty$, for $k = 1, \ldots, K$.

The conditions in assumption B guarantee the existence and positive-definiteness of the population covariance matrices Φ_0, Ψ_0, Q_k and A_k, for $k = 1, \ldots, K$. This assumption is sufficient to show that $\|\widehat{\Phi} - \Phi_0\| = o_p(1)$. Convergence results for the spatial covariance matrices are presented as follows. Let $\widehat{Q}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\overline{X}'_{k,ij}\overline{X}_{k,ij}$ and $\widehat{A}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\overline{X}'_{k,ij}\overline{X}_{k,ij}e_i^2$, where $\overline{X}_{k,ij} = (x_j 1_k(d_{ij}), x_j(d_{ij} - z_k)1_k(d_{ij}), \ldots, x_j(d_{ij} - z_k)^q 1_k(d_{ij}))$ and e_i represents the residuals defined above.

Lemma 1: *Under assumption A1, it follows that* $\|\widetilde{Q}_k - \widehat{Q}_k\| = O_p\left(\frac{1}{\sqrt{N}}\right)$ *for* $k = 1, \ldots, K$. *Similarly,* $\|\widetilde{A}_k - \widehat{A}_k\| = O_p\left(\frac{1}{\sqrt{N}}\right)$ *as* $N \to \infty$.

The result in Lemma 1 allows us to replace the sample covariance matrices $\widetilde{Q}_k$ with $\widehat{Q}_k$ for estimation purposes. The latter matrix is more convenient for analytical and computational purposes, as it does not consider the sample cross-correlations between x_i and x_j for $i, j = 1, \ldots, N$. These correlations converge to a probability of zero under the *iid* assumption in A1. The consistency of the latter sample covariance matrix is derived below.

Lemma 2: *Under assumptions A and B, for* $k = 1, \ldots, K$, *it follows that* $\|\widehat{Q}_k - Q_k\| = O_p\left(\frac{\sqrt{K}}{\sqrt{N}}\right)$ *as* $K, N \to \infty$.

The above result in Lemma 2 is refined in the proof in the appendix and differentiates between the cases $d_{ij} \neq d_{ji}$ and $d_{ij} = d_{ji}$. Nevertheless, in a spatial setting, we are only concerned with the symmetric case.

These results also allow us to derive the consistency of the slope parameter estimator $\widehat{\lambda}$ and the spatial parameter estimators $\widehat{\Gamma}_k$, for each $k = 1, \ldots, K$.

Proposition 1: *Under assumptions A and B, it follows that* $\|\widehat{\lambda} - \lambda\| = O_p(1/\sqrt{N})$ *and* $\|\widehat{\Gamma}_k - \Gamma_k\| = O_p(\sqrt{K}/N)$, *for* $k = 1, \ldots, K$, *as* $K, N \to \infty$.

The above result illustrates the effect of considering a partitioning estimator for the spatial effects. The convergence rate includes an additional effect produced by dividing the compact set into K disjoint intervals.

Proposition 2: *Under assumptions A and B,*

$$\sqrt{N}(\widehat{\lambda} - \lambda) \xrightarrow{d} N\left(0, \Phi_0^{-1}\Psi_0\Phi_0^{-1}\right). \tag{11}$$

These results allow us to derive the uniform convergence of the estimator of the functional coefficient.

Theorem 1: *Under assumptions A and B,*

$$\sup_{d\in\chi} |\widehat{w}(d) - w(d)| = O_p\left(\sqrt{K}/N + K^{-(q+1)}\right). \tag{12}$$

The uniform convergence is determined by a variance term $\sqrt{K}/N$ given by the estimation of the spatial parameters and a bias term $K^{-(q+1)}$ driven by the approximation error due to the remainder terms of the Taylor expansions evaluated at different intervals. The following auxiliary results are useful for obtaining the asymptotic distribution of the estimator of the functional coefficient.

Lemma 3: *Under assumptions A and B and the result in Proposition 1, for every* $k = 1, \ldots, K$, *we have* $\|\widehat{A}_k - A_k\| = O_p\left(\frac{1}{\sqrt{N}}\right)$.

Lemma 4: *Under assumptions A and B, for* $k = 1, \ldots, K$, *the estimator* (9) *satisfies*

$$V(\widehat{\Gamma}_k) = \frac{1}{\alpha_N p_k} Q_k^{-1} A_k Q_k^{-1} + O_p\left(\frac{K}{N^3}\right).$$

This result can be extended to derive the asymptotic convergence of the variance estimator (10). To do this, we introduce further notation. Let $V_K(d) \equiv \alpha_N \sum_{k=1}^{K} v_k'(d) V(\widehat{\Gamma}_k) v_k(d) 1_k(d)$. Similarly, we define $\widehat{V}_K(d) = \sum_{k=1}^{K} v_k'(d) \widehat{Q}_k^{-1} \widehat{A}_k \widehat{Q}_k^{-1} v_k(d) 1_k(d)/p_k$. Note that knowledge of the probability p_k is not required to obtain $\widehat{V}_K(d)$. This is because p_k cancels out the covariance estimators $\widehat{Q}_k^{-1} \widehat{A}_k \widehat{Q}_k^{-1}$. Note also that the definition of $\widehat{V}_K(d)$ considers the estimators $\widehat{Q}_k$ and $\widehat{A}_k$ instead of the sample covariance matrices $\widetilde{Q}_k$ and $\widetilde{A}_k$ used in the definition of $\widehat{V}\left(\widehat{\Gamma}_k\right)$ in equation (9). As shown in Lemmas 1 and 2, this is done for analytical convenience but does not affect the asymptotic results in Proposition 3 due to the differences in convergence rates.

Proposition 3: *Under assumptions A and B, for* $k = 1, \ldots, K$ *and any fixed* $d \in \chi$, *it holds that*

(i) $|\widehat{V}_K(d) - V_K(d)| = O_p(K/N)$,

(ii) $V\left(\sqrt{\alpha_N}(\widehat{w}(d) - w(d))\right) = V_K(d) + O(N/K^{2(q+1)})$.

Therefore, under the regularity conditions in assumption A5, a consistent estimator of $V\left(\sqrt{\alpha_N}(\widehat{w}(d) - w(d))\right)$ is $\widehat{V}_K(d)$. More formally, by applying the triangular inequality, Proposition 3 can be used to show that

$$|V\left(\sqrt{\alpha_N}(\widehat{w}(d) - w(d))\right) - \widehat{V}_K(d)| = O(N/K^{2(q+1)}) + O_p(K/N) = O_p(K/N), \qquad (13)$$

under assumption A5.

The following theorem presents the asymptotic distribution of the estimator of the functional coefficient.

Theorem 2: *Under assumptions A and B, for any fixed $d \in \chi$, it follows that*

$$\sqrt{\alpha_N}\,\frac{\widehat{w}(d) - w(d)}{\widehat{V}_K^{1/2}(d)} \xrightarrow{d} N(0,1)\,. \tag{14}$$

The convergence rate of the estimator reflects the influence of neighbouring effects. This result is the basis of pointwise tests for the presence of spatial effects given by the null hypothesis $H_0 : w(d) = 0$ vs. the alternative hypothesis $H_A : w(d) \neq 0$, for some fixed $d \in \chi$. Importantly, this result can be also extended to the functional space if $\widehat{w}(d)$ is considered a process in $d \in \chi$. Unfortunately, the stochastic process $\widehat{w}(d)$ is not asymptotically tight and, therefore, does not converge weakly in $\mathcal{L}^\infty$, where $\mathcal{L}^\infty$ denotes the set of all uniformly bounded real functions on χ equipped with the uniform norm. Nevertheless, the weak convergence of the above process can be obtained by adapting the strong approximation results derived in Section 6 of Cattaneo et al. (2020). We state the following result, the proof of which is obtained through the application of the asymptotic results found by these authors.

Proposition 4: *Under assumptions A and B, the estimator $\widehat{w}(d)$, for $d \in \chi$, satisfies*

$$\sqrt{\alpha_N}\,\frac{\widehat{w}(d) - w(d)}{\widehat{V}_K^{1/2}(d)} \xrightarrow{w} \mathbb{G}(d), \tag{15}$$

where $\xrightarrow{w}$ denotes weak convergence and $\mathbb{G}(d)$ is a zero-mean Gaussian process defined on $d \in \chi$.

The asymptotic distribution of the supremum functional can be obtained as a byproduct of this result such that

$$\sqrt{\alpha_N}\,\sup_{d\in\chi}\left|\frac{\widehat{w}(d) - w(d))}{\widehat{V}_K^{1/2}(d)}\right| \xrightarrow{d} \sup_{d\in\chi}|\mathbb{G}(d)|\,, \text{ as } N \to \infty. \tag{16}$$

The proof of the above follows from the continuous mapping theorem applied to the supremum. The next section introduces a test for the presence of spatial effects based on the above results and discusses different methods for model selection.

4. Hypothesis testing and model selection

4.1. Hypothesis testing

This subsection exploits the above asymptotic theory to construct different hypothesis tests. Although the focus is on testing for the presence of spatial effects, we also introduce a framework to statistically assess the functional form of $w(d)$. In our context, testing for the presence of spatial effects can be formulated as $H_0 : \sup_{d\in\chi} |w(d)| = 0$, against the alternative $H_A : \sup_{d\in\chi} |w(d)| > 0$. The null hypothesis can be modified to evaluate specific functional forms of $w(d)$. In this case, the hypothesis of interest is $H_{0f} : \sup_{d\in\chi} |w(d) - f(d)| = 0$, where $f(\cdot)$ is some known functional specification of $d \in \chi$, against the alternative $H_{Af} : \sup_{d\in\chi} |w(d) - f(d)| > 0$.

Following Davies (1977, 1987), we propose the test statistic

$$T_N = \sqrt{\alpha_N} \sup_{d\in\chi} \left| \frac{\widehat{w}(d) - f(d)}{\widehat{V}_K^{1/2}(d)} \right|, \tag{17}$$

where the functional form $f(d)$ depends on the null hypothesis under study. Hypothesis tests involving nuisance parameters under the null hypothesis have been widely investigated in the time-series literature and, in particular, in threshold models and structural break testing. The seminal contribution of Andrews and Ploberger (1994) proposes alternative tests based on average weighted and average exponential statistics. Hansen (1996) develops a Wald-type test that is made operational through a p-value transformation.

Theorem 3: *Under assumptions A and B, and the null hypothesis of interest (H_0 or H_{0f}), it holds that*

$$T_N \xrightarrow{d} \sup_{d\in\chi} |\mathbb{G}(d)|, \text{ as } N \to \infty, \tag{18}$$

where $\mathbb{G}(d)$ is the zero-mean Gaussian process defined above.

The proof of this theorem follows as an application of the asymptotic result (16), and it is obtained by replacing $w(d)$ with the null hypothesis of interest.

Obtaining asymptotic critical values for these tests is difficult because the asymptotic distribution is non-standard and cannot be tabulated. Fortunately, simulation and resampling methods can be applied to approximate the critical values in finite samples (see Andrews (1993), Hansen (1996) and, more recently, Cattaneo et al. (2020)). We now discuss a p-value transformation method for testing the null hypothesis of interest. We operate conditionally on a realisation of $\{(x_i, y_i)\}_{i=1}^N$, which is denoted as ω_N. Expression (A.17) in the mathematical appendix shows that

$$\sqrt{\alpha_N}\frac{\widehat{w}(d) - w(d)}{V_k^{1/2}(d)} = \frac{\frac{1}{\sqrt{\alpha_N p_k}}\sum_{i=1}^{N}\sum_{k=1}^{K} v_k(d)'\widehat{Q}_k^{-1}\mathbb{X}'_{ki}e_{0i}1_k(d)}{V_k^{1/2}(d)} + o_p(1),$$

where e_{0i} represents the residuals of the data-generating process obtained under the null hypothesis, e.g. $e_{0i} = y_i - \widehat{\lambda}x_i$, where $\widehat{\lambda}$ is theordinary least square (OLS) estimator of the regression model without spatial effects obtained under the null hypothesis $H_0 : \sup_{d\in\chi} |w(d)| = 0$.

The objective of this approximation is to construct independent replicas of the test statistic T_N for this case. Let $\mathbb{G}_N^\star$ be a conditional zero-mean Gaussian process with the same covariance kernel as $\mathbb{G}(d)$. This process can be simulated by generating a vector of *iid* random variables $\epsilon = (\epsilon_1, \ldots, \epsilon_N)'$ to construct the simulated residuals $e_0^\star = e_0 \otimes \epsilon$, with $\otimes$ denoting element-by-element multiplication. Then,

$$\mathbb{G}_N^\star(d) = \frac{\frac{1}{\sqrt{\alpha_N p_k}} \sum_{i=1}^{N} \sum_{k=1}^{K} v_k(d)' \widehat{Q}_k^{-1} \mathbb{X}_{ki}' e_{0i}^\star 1_k(d)}{V_k^{1/2}(d)}, \tag{19}$$

and $T_N^\star = \sup_{d\in\chi} |\mathbb{G}_N^\star(d)|$.

Using the same arguments used by Cattaneo et al. (2020), we assert that the p-value obtained from the simulated process $\mathbb{G}_N^\star$ converges to the asymptotic p-value of the test under the null hypothesis. More formally,

$$P_{\omega_N} \{T_N^\star > T_N\} \rightarrow P_{H_0} \left\{ T_N > \sup_{d\in\chi} |\mathbb{G}(d)| \right\}, \quad \text{as } N \rightarrow \infty, \tag{20}$$

where P_{ω_N} denotes a probability distribution function conditional on the realisation of the sample ω_N, and P_{H_0} is the probability distribution of $\sup_{d\in\chi} |\mathbb{G}(d)|$. Although the distribution of $T_N^\star$ is not directly observed, it can be approximated to any degree of accuracy by conditionally operating on ω_N. The algorithm used to compute the p-value of the test is described below.

Algorithm:

1. Construct a grid of K points $\bar{\mathbb{Z}} = [z_1, \ldots, z_K]$, with $z_1 = h$ and $z_k = z_{k-1} + 2h$, for $k = 2, \ldots, K$. This grid characterises a partition of the set $\chi = [0, C]$ that spans the support of the network variable d_{ij}, which measures the distance between the regressors x_i and x_j, for $i, j = 1, \ldots, N$, such that $d_{ij} \in \chi$. For a given h and C, we choose the number of intervals K as $K = C/2h$.
2. Compute the test statistic $T_N = \sqrt{\alpha_N} \sup_{d\in\mathbb{D}} \left| \frac{\widehat{w}(d) - f(d)}{\widehat{V}_K^{1/2}(d)} \right|$, where $f(d)$ denotes the null hypothesis; $\mathbb{D}$ denotes a discrete set of equally distant points inside χ. This set of points characterises a finer grid of the interval χ than $\bar{\mathbb{Z}}$.
3. For a given realisation $\omega_N = \{x_i, y_i\}_{i=1}^N$, execute the following steps for $b = 1, \ldots, B$:
    (a) Generate $\{\epsilon_i^{(b)}\}_{i=1}^N$ *iid* $N(0, 1)$ random variables independent of the data to construct the simulated residuals $e_0^{\star(b)} = e_0 \otimes \epsilon^{(b)}$, where e_0 is the vector of residuals of the regression model (1) under the null hypothesis H_0. Then, compute the simulated process (19).

(b) Store the bootstrap test statistic

$$T_N^{\star(b)} = \sup_{d \in \mathbb{D}} |\mathbb{G}_N^{\star(b)}(d)|.$$

This algorithm yields a random sample of B observations from the distribution of $\sup_{d \in \chi} |\mathbb{G}_N^{\star}(d)|$. Using the Glivenko-Cantelli theorem and previous assumptions, the empirical p-value conditional on ω_N defined by

$$\widehat{p}_{N,B}^{\star} = \frac{1}{B} \sum_{b=1}^{B} 1(T_N^{\star(b)} > T_N)$$

converges to a probability $P_{\omega_N} \left\{ T_N^{\star(b)} > T_N \right\}$ as $B \to \infty$.

4.2. Model selection

The estimation of the model parameters depends on the choice of h. This choice determines the number of intervals K covering the compact set χ and, hence, the quality of the approximation of the function $w(d)$. Choosing an optimal bandwidth in finite samples is important for improving the quality of the fit. Suitable methods for partitioning estimators are proposed by Cattaneo and Farrell (2013) and Cattaneo et al. (2020) (seminal examples). These authors explore how to construct plug-in estimators of the bandwidth parameter and, correspondingly, the number of intervals defining the partition, using an integrated mean square error expansion of the partition estimator. These methods are cumbersome in most instances; thus, in this subsection, for simplicity, we adapt off-the-shelf methods for bandwidth selection developed for nonparametric regression models. For illustrative purposes, we consider a compact set given by the interval $[0, C]$.

We first discuss two different methods proposed for series estimation (see Mallows (1973), Li (1987) and Wahba (1985)) that are adapted to our setting. A review of these methods can be found in the monograph by Li and Racine (2007). Thus, Mallows (1973) selects h such that

$$\widehat{h}_M = \arg\min_{\{h\}} \left\{ \widehat{\sigma}_e^2 \left(1 + \frac{C}{Nh} \right) \right\}, \tag{21}$$

where $\widehat{\sigma}_e^2 = \frac{1}{N} \sum_{i=1}^{N} e_i^2$ is obtained under the conditional homoscedasticity of the error term. Craven and Wahba (1978) propose a generalised cross-validation method. Other more sophisticated model selection procedures for series estimators, such as the leave-one-out cross-validation method of Stone (1974), can be found in the literature. These authors select $\widehat{h}$ such that

$$\widehat{h}_{GCV} = \arg\min_{\{h\}} \left\{ \frac{\widehat{\sigma}_e^2}{\left(1 - \frac{C}{Nh}\right)^2} \right\}. \tag{22}$$

Standard information criteria can be also adapted to this framework to optimally choose the order of the Taylor expansion in the approximation of the functional coefficient $w(d)$.

The choice of q has a non-negligible effect on the accuracy of the approximation because it directly affects the number of regressors in equation (3). In particular, we adapt the Akaike information criterion (AIC) and Bayesian information criterion (BIC) such that

$$\{\widehat{h}_{AIC}, q_{AIC}\} = \underset{\{h,q\}}{\arg\min} \left\{ \ln \widehat{\sigma}_e^2 + 2\frac{(q+1)\lceil C/2h\rceil + 1}{N} \right\} \tag{23}$$

and

$$\{\widehat{h}_{BIC}, q_{BIC}\} = \underset{\{h,q\}}{\arg\min} \left\{ \ln \widehat{\sigma}_e^2 + \frac{((q+1)\lceil C/2h\rceil + 1)ln\ N}{N} \right\}. \tag{24}$$

5. Monte Carlo simulations

This section explores the finite-sample approximation of the asymptotic results using Monte-Carlo simulations. We present three different exercises that illustrate i) the consistency of the parameter estimates, ii) the rejection rates associated with the marginal t-tests using the asymptotic distribution in Theorem 2 and iii) the empirical size and power of the uniform tests H_0 and H_{0f} obtained from Theorem 3.

The data-generating process (DGP) considered for the simulation exercise is

$$y_i = \lambda x_i + \sum_{\substack{j=1 \\ j\neq i}}^{N} w(d_{ij})x_j + \varepsilon_i, \text{ for } i = 1, \ldots, N, \tag{25}$$

where x_i represents realisations of a single covariate X with a $N(0,1)$ distribution. For simplicity, the regressor also acts as the spatial variable Z, which establishes the proximity between units in the cross section such that $d_{ij} = |x_i - x_j|$. The error term ε_i is modelled as a $N(0, \log^2|1 + x_i|)$ random variable that is uncorrelated with X but exhibits conditional heteroscedasticity.

Although our estimation procedure does not require knowledge of the parametric form of the functional parameter $w(\cdot)$, we do need to impose a specification to fully characterise the DGP in the simulation exercise. With this aim, two alternatives have been considered. The first specification corresponds to the exponential function $w(d_{ij}) = \beta \exp(-\theta d_{ij})$, while the second one is the Gaussian kernel $w(d_{ij}) = \beta \exp\left(-\frac{1}{2}(\theta d_{ij})^2\right)$, with $\beta, \theta > 0$. Both formulations are standard in the spatial econometrics literature for describing neighbouring effects (see Fischer and Wang (2011)). The first specification represents exponentially decaying spillover effects of x_j on y_i as d_{ij} increases. The second formulation corresponds to the standard Gaussian kernel used in the nonparametric econometrics literature (Li and Racine, 2007), as well as in locally weighted and geographically weighted regressions (see Cleveland and Devlin (1988) and Wheeler and Páez (2010), respectively).

Throughout the Monte Carlo exercise, we implement $B = 500$ simulations, and the compact set is $\chi = [0, 1]$, such that $K = 1/2h$. By doing this, we explore the spatial

Table 1: Bias of estimators of λ and $w(d)$ in equations (5) and (7).

				Model 1: Exponential function					Model 2: Gaussian kernel function				
λ	h	θ	N	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$
1	0.050	5	100	0.144	-0.006	0.011	0.018	0.017	0.004	0.002	0.001	0.003	0.007
			250	0.095	0.010	0.014	0.015	0.014	0.046	0.005	0.003	0.003	0.005
			500	0.064	0.007	0.010	0.011	0.012	0.078	0.005	0.004	0.005	0.006
		9	100	0.011	0.000	-0.001	-0.005	-0.005	0.010	0.010	-0.002	-0.001	0.005
			250	0.030	-0.005	0.004	0.003	0.002	-0.002	0.009	-0.002	-0.001	0.003
			500	0.033	-0.003	0.003	0.001	-0.001	-0.002	0.005	-0.003	-0.001	0.003
	0.075	5	100	0.055	-0.001	-0.001	-0.002	0.000	0.006	0.013	0.008	0.006	0.003
			250	0.115	-0.005	0.000	-0.001	-0.002	0.001	0.003	-0.002	-0.001	0.001
			500	0.016	-0.006	-0.002	-0.002	-0.003	0.004	0.003	-0.002	-0.001	0.002
		9	100	0.008	-0.016	0.004	0.002	0.000	-0.019	0.024	-0.010	0.000	0.008
			250	-0.001	-0.017	0.005	0.001	-0.005	-0.007	0.030	-0.009	-0.002	0.009
			500	0.009	-0.012	0.005	-0.001	-0.005	0.000	0.023	-0.011	0.000	0.012
	0.100	5	100	0.015	-0.010	0.003	0.002	-0.003	0.003	0.012	-0.004	-0.001	0.003
			250	0.032	-0.007	0.001	-0.001	-0.004	0.008	0.009	-0.004	-0.001	0.004
			500	0.058	-0.008	0.000	-0.002	-0.003	-0.001	0.009	-0.004	0.000	0.005
		9	100	-0.020	-0.020	0.013	0.001	-0.008	0.000	0.058	-0.024	0.003	0.026
			250	0.002	-0.021	0.011	-0.001	-0.010	-0.003	0.055	-0.026	0.001	0.024
			500	0.002	-0.022	0.010	-0.001	-0.009	-0.008	0.054	-0.026	0.002	0.025
0.25	0.050	5	100	0.030	0.002	0.002	0.002	0.004	0.006	-0.005	0.001	0.004	0.003
			250	0.075	0.003	0.003	0.003	0.004	0.013	0.004	0.001	-0.001	0.000
			500	0.044	0.001	0.002	0.003	0.002	0.020	0.001	0.001	0.001	0.001
		9	100	-0.006	0.000	0.002	0.001	0.002	-0.011	0.003	-0.001	-0.001	0.003
			250	0.007	-0.003	0.000	0.001	0.001	0.000	0.003	0.001	0.000	0.000
			500	0.009	0.002	0.001	-0.001	-0.001	-0.001	0.002	-0.001	-0.001	0.000
	0.075	5	100	0.004	0.007	0.002	-0.001	0.000	0.006	0.006	0.002	0.000	-0.002
			250	0.032	0.002	0.002	0.001	0.000	0.000	-0.002	-0.001	0.000	0.001
			500	0.058	0.000	-0.001	-0.001	-0.001	-0.005	0.000	0.000	0.001	0.000
		9	100	-0.001	-0.005	0.001	-0.001	-0.004	-0.002	-0.002	-0.003	0.004	0.006
			250	0.002	-0.005	0.000	-0.001	-0.002	-0.002	0.007	-0.003	-0.001	0.001
			500	0.002	-0.002	0.001	-0.001	-0.002	0.001	0.006	-0.003	-0.001	0.003
	0.100	5	100	-0.013	-0.008	0.002	0.004	0.003	-0.003	0.002	0.000	0.001	-0.001
			250	0.015	-0.006	0.000	0.001	0.001	-0.005	0.006	-0.001	-0.002	-0.001
			500	0.013	0.000	0.000	-0.001	-0.001	-0.002	0.003	0.000	0.001	0.001
		9	100	0.011	0.003	0.004	-0.003	-0.006	0.005	0.007	-0.007	0.004	0.012
			250	0.001	-0.005	0.002	-0.001	-0.002	0.000	0.013	-0.006	0.002	0.006
			500	0.005	-0.004	0.003	-0.001	-0.002	-0.008	0.015	-0.007	0.000	0.006

Note: This table reports the estimation bias under two specifications for the functional parameter for $d \in [0, C]$, with $C = 1$. The number of intervals is $K = 1/2h$. The number of simulations is 500.

effects given by close observations and discard values of d greater than one. Furthermore, we consider the values $h = 0.05, 0.075, 0.1$ to assess the sensitivity of the estimates to the choice of the bandwidth parameter. The sample size is equal to $N = 100, 250, 500$. The parameters characterising the functional form of $w(d)$ are $\beta = 0.1$ and $\theta = 5, 9$. This choice of parameters results in small spatial effects; however, as shown below, the test statistic (14) has considerable power to reject the null hypothesis under the presence of such effects. We also consider $\theta = 7$ in the study of the asymptotic coverage rate α corresponding to the

Table 2: Root mean square error of estimators of λ and $w(d)$ in equations (5) and (7).

				Model 1: Exponential function					Model 2: Gaussian kernel function				
λ	h	θ	N	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$
1	0.050	5	100	0.251	0.242	0.109	0.117	0.102	0.205	0.239	0.100	0.108	0.100
			250	0.128	0.091	0.046	0.048	0.043	0.125	0.086	0.037	0.041	0.036
			500	0.070	0.052	0.030	0.030	0.028	0.116	0.043	0.019	0.021	0.019
		9	100	0.201	0.242	0.101	0.108	0.095	0.210	0.232	0.101	0.107	0.097
			250	0.122	0.086	0.037	0.039	0.036	0.120	0.086	0.037	0.041	0.036
			500	0.093	0.039	0.017	0.019	0.018	0.084	0.042	0.018	0.020	0.018
	0.075	5	100	0.234	0.205	0.084	0.091	0.083	0.220	0.201	0.087	0.090	0.080
			250	0.176	0.072	0.030	0.033	0.029	0.135	0.071	0.030	0.033	0.029
			500	0.142	0.037	0.016	0.017	0.016	0.099	0.034	0.014	0.016	0.015
		9	100	0.220	0.200	0.083	0.092	0.081	0.224	0.206	0.085	0.091	0.080
			250	0.134	0.069	0.029	0.032	0.030	0.137	0.077	0.033	0.034	0.032
			500	0.099	0.036	0.016	0.016	0.015	0.101	0.042	0.019	0.016	0.018
	0.100	5	100	0.244	0.179	0.072	0.079	0.069	0.253	0.183	0.074	0.078	0.071
			250	0.164	0.064	0.025	0.030	0.027	0.162	0.064	0.026	0.029	0.026
			500	0.143	0.033	0.013	0.014	0.013	0.128	0.032	0.013	0.014	0.014
		9	100	0.259	0.180	0.077	0.083	0.071	0.248	0.189	0.078	0.079	0.074
			250	0.157	0.067	0.028	0.029	0.027	0.162	0.086	0.037	0.029	0.035
			500	0.124	0.039	0.017	0.014	0.016	0.128	0.062	0.029	0.014	0.028
0.25	0.050	5	100	0.219	0.244	0.099	0.109	0.099	0.204	0.242	0.101	0.110	0.096
			250	0.136	0.085	0.037	0.041	0.037	0.121	0.086	0.037	0.041	0.035
			500	0.066	0.041	0.019	0.020	0.017	0.086	0.041	0.018	0.020	0.017
		9	100	0.199	0.247	0.101	0.108	0.098	0.215	0.250	0.105	0.115	0.100
			250	0.125	0.086	0.037	0.039	0.035	0.122	0.085	0.038	0.041	0.037
			500	0.084	0.041	0.017	0.019	0.017	0.088	0.041	0.018	0.019	0.017
	0.075	5	100	0.207	0.199	0.084	0.091	0.082	0.219	0.202	0.081	0.087	0.078
			250	0.133	0.072	0.029	0.033	0.030	0.141	0.072	0.031	0.034	0.031
			500	0.113	0.035	0.014	0.016	0.014	0.103	0.034	0.014	0.016	0.014
		9	100	0.216	0.203	0.082	0.089	0.080	0.215	0.208	0.085	0.094	0.082
			250	0.135	0.072	0.031	0.033	0.030	0.133	0.070	0.031	0.032	0.029
			500	0.097	0.035	0.014	0.016	0.014	0.100	0.034	0.015	0.016	0.014
	0.100	5	100	0.262	0.179	0.075	0.078	0.069	0.252	0.171	0.075	0.077	0.069
			250	0.161	0.065	0.026	0.028	0.026	0.167	0.064	0.027	0.028	0.025
			500	0.126	0.031	0.013	0.014	0.013	0.125	0.031	0.013	0.014	0.012
		9	100	0.247	0.183	0.072	0.078	0.070	0.248	0.180	0.075	0.082	0.072
			250	0.170	0.068	0.025	0.030	0.027	0.171	0.067	0.027	0.029	0.027
			500	0.128	0.031	0.013	0.014	0.012	0.127	0.034	0.015	0.014	0.014

Note: This table reports the root mean square error under two specifications for the functional parameter for $d \in [0, C]$, with $C = 1$. The number of intervals is $K = 1/2h$. The number of simulations is 500.

$(1-\alpha)$-confidence interval of $w(d)$, which is constructed as follows:

$$\left[\widehat{w}(d) - z_{1-\alpha/2}\widehat{V}_K^{1/2}(d)/\sqrt{\alpha_N}, \widehat{w}(d) + z_{1-\alpha/2}\widehat{V}_K^{1/2}(d)/\sqrt{\alpha_N}\right], \tag{26}$$

where $\widehat{V}_K(d) = \sum_{k=1}^{K} v_k'(d)\widehat{Q}_k^{-1}\widehat{A}_k\widehat{Q}_k^{-1}v_k(d)1_k(d)/p_k$ is defined in Proposition 3, and z_α is the critical value of a standard normal distribution function at a significance level of α.

Table 3: Empirical coverage rates for the confidence interval (26) at a significance level (coverage rate) of $\alpha = 0.05$.

			Model 1: Exponential function					Model 2: Gaussian kernel function				
h	θ	N	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$
0.050	5	100	0.178	0.096	0.118	0.126	0.118	0.112	0.092	0.106	0.120	0.106
		250	0.232	0.080	0.092	0.090	0.068	0.066	0.058	0.084	0.084	0.084
		500	0.324	0.050	0.064	0.056	0.082	0.084	0.078	0.046	0.052	0.058
	7	100	0.120	0.100	0.090	0.116	0.124	0.112	0.098	0.126	0.112	0.098
		250	0.082	0.070	0.078	0.086	0.070	0.088	0.060	0.078	0.092	0.060
		500	0.066	0.044	0.058	0.060	0.048	0.042	0.058	0.064	0.074	0.074
	9	100	0.116	0.114	0.132	0.122	0.118	0.092	0.110	0.108	0.112	0.120
		250	0.082	0.080	0.082	0.060	0.054	0.074	0.068	0.058	0.066	0.076
		500	0.052	0.050	0.036	0.044	0.048	0.080	0.050	0.064	0.064	0.060
0.075	5	100	0.102	0.114	0.094	0.102	0.110	0.104	0.100	0.100	0.104	0.112
		250	0.088	0.060	0.064	0.066	0.070	0.074	0.068	0.062	0.072	0.068
		500	0.058	0.082	0.080	0.054	0.064	0.064	0.048	0.050	0.044	0.068
	7	100	0.106	0.086	0.098	0.108	0.114	0.092	0.124	0.102	0.116	0.094
		250	0.072	0.056	0.066	0.080	0.068	0.076	0.072	0.072	0.060	0.072
		500	0.068	0.052	0.054	0.046	0.056	0.072	0.056	0.086	0.068	0.054
	9	100	0.098	0.108	0.096	0.102	0.100	0.110	0.122	0.088	0.102	0.090
		250	0.078	0.064	0.060	0.064	0.052	0.044	0.066	0.076	0.072	0.068
		500	0.068	0.056	0.050	0.044	0.082	0.066	0.066	0.062	0.074	0.080
0.100	5	100	0.118	0.092	0.106	0.072	0.080	0.138	0.098	0.116	0.106	0.092
		250	0.096	0.058	0.060	0.076	0.060	0.102	0.092	0.094	0.072	0.082
		500	0.056	0.054	0.050	0.052	0.074	0.084	0.046	0.048	0.062	0.054
	7	100	0.138	0.102	0.122	0.102	0.106	0.112	0.110	0.092	0.088	0.138
		250	0.074	0.062	0.072	0.056	0.062	0.102	0.062	0.076	0.072	0.100
		500	0.056	0.054	0.044	0.054	0.046	0.074	0.044	0.048	0.048	0.066
	9	100	0.080	0.094	0.104	0.096	0.090	0.132	0.110	0.158	0.092	0.130
		250	0.080	0.054	0.068	0.076	0.050	0.078	0.094	0.094	0.062	0.056
		500	0.066	0.052	0.070	0.058	0.064	0.064	0.082	0.084	0.066	0.078

Note: This table reports the coverage probability of the confidence interval for the functional parameter $w(d)$ under two specifications for $d \in [0, C]$, with $C = 1$. Hence, $K = 1/2h$. The DGP is given by $\lambda = 1$, $\beta = 0.1$, and the number of simulations is 500.

5.1. Consistency of the parameter estimates

The consistency of the parameter estimators (5) and (7) is assessed through the analysis of the bias and root mean square error (RMSE). Table 1 reports the biases of the parameter estimators $\widehat{\lambda}$ and $\widehat{w}(d)$ for two regression models given by $\lambda = 1, 0.25$, respectively. The left panel corresponds to the specification of $w(d)$ given by an exponential function, while the right panel considers $w(d)$ given by a Gaussian kernel. Table 2 reports the corresponding RMSE for the different DGPs.

For the sake of presentation, we restrict our simulation exercise to show the influence of the closest neighbours, which are given by $d = \{h/2, h, 3h/2, 2h\}$. The figures displayed in Table 1 do not show evidence of bias in any direction. This bias decreases in most instances as the sample size increases. The results in Table 2 are more conclusive; the RMSE decreases monotonically to zero as the sample size increases, providing strong empirical evidence of

Table 4: Empirical coverage rates for the confidence interval (26) at a significance level (coverage rate) of $\alpha = 0.05$.

			Model 1: Exponential function					Model 2: Gaussian kernel function				
h	θ	N	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$
0.050	5	100	0.186	0.100	0.090	0.096	0.110	0.094	0.098	0.112	0.112	0.128
		250	0.260	0.078	0.062	0.068	0.094	0.090	0.068	0.088	0.072	0.062
		500	0.328	0.054	0.076	0.070	0.056	0.052	0.068	0.052	0.052	0.070
	7	100	0.104	0.108	0.094	0.098	0.112	0.104	0.122	0.118	0.120	0.114
		250	0.062	0.056	0.060	0.054	0.066	0.058	0.050	0.062	0.064	0.062
		500	0.068	0.060	0.052	0.046	0.060	0.056	0.056	0.062	0.042	0.052
	9	100	0.116	0.128	0.084	0.130	0.130	0.096	0.128	0.114	0.100	0.108
		250	0.068	0.060	0.066	0.078	0.086	0.056	0.066	0.074	0.062	0.064
		500	0.062	0.058	0.070	0.056	0.064	0.052	0.048	0.054	0.048	0.060
0.075	5	100	0.106	0.092	0.082	0.112	0.090	0.122	0.106	0.104	0.108	0.144
		250	0.082	0.078	0.072	0.060	0.062	0.080	0.070	0.054	0.048	0.060
		500	0.092	0.058	0.064	0.066	0.054	0.048	0.060	0.084	0.072	0.062
	7	100	0.088	0.100	0.118	0.110	0.118	0.122	0.104	0.098	0.088	0.118
		250	0.070	0.068	0.072	0.064	0.056	0.072	0.074	0.068	0.068	0.072
		500	0.076	0.054	0.060	0.056	0.066	0.068	0.068	0.072	0.056	0.064
	9	100	0.112	0.108	0.096	0.114	0.082	0.122	0.106	0.124	0.148	0.120
		250	0.072	0.068	0.072	0.074	0.064	0.084	0.064	0.066	0.066	0.066
		500	0.056	0.060	0.062	0.084	0.070	0.042	0.072	0.064	0.060	0.038
0.100	5	100	0.124	0.092	0.090	0.086	0.112	0.110	0.096	0.110	0.100	0.088
		250	0.076	0.078	0.072	0.080	0.070	0.084	0.068	0.072	0.072	0.068
		500	0.074	0.046	0.052	0.040	0.050	0.098	0.066	0.068	0.072	0.070
	7	100	0.104	0.100	0.114	0.154	0.112	0.102	0.104	0.112	0.102	0.100
		250	0.080	0.068	0.068	0.066	0.058	0.078	0.054	0.060	0.070	0.080
		500	0.082	0.048	0.068	0.060	0.080	0.046	0.064	0.072	0.052	0.056
	9	100	0.112	0.104	0.114	0.108	0.110	0.096	0.102	0.128	0.062	0.092
		250	0.118	0.090	0.066	0.070	0.066	0.078	0.080	0.074	0.082	0.082
		500	0.068	0.054	0.062	0.064	0.066	0.052	0.074	0.086	0.054	0.074

Note: This table reports the coverage probability of the confidence interval for the functional parameter $w(d)$ under two specifications for $d \in [0, C]$, with $C = 1$. Hence, $K = 1/2h$. The DGP is given by $\lambda = 0.25$, $\beta = 0.1$, and the number of simulations is 500.

the consistency of the parameter estimators of $w(d)$ for different values of d in the interval $[0, 1]$.

5.2. Empirical coverage rate and rejection probabilities

This exercise studies the finite-sample coverage probability of the asymptotic confidence intervals for λ and $w(d)$ for a discrete grid of values $d = \{h/2, h, 3h/2, 2h\}$ under the heteroscedasticity of the error term. To do this, we compute the empirical fraction of times that the true parameters λ and $w(d)$ are outside the above $(1-\alpha)$-confidence intervals for $\alpha = 0.05$. Tables 3 and 4 report, respectively, the empirical coverage rates $\widehat{\alpha}$ for the regression models characterised by $\lambda = 1, 0.25$. In line with the previous subsection, the left panel corresponds to the exponential function and the right panel considers the functional specification of $w(d)$ given by the Gaussian kernel. The simulated results show empirical

Table 5: Empirical power of marginal t-test for $H_0 : w(d) = 0$.

			Model 1: Exponential function					Model 2: Gaussian kernel function				
h	θ	N	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$
0.050	5	100	0.990	0.128	0.234	0.180	0.200	0.988	0.152	0.286	0.210	0.214
		250	1.000	0.250	0.624	0.444	0.442	1.000	0.216	0.682	0.588	0.628
		500	1.000	0.554	0.958	0.882	0.886	1.000	0.576	0.982	0.968	0.982
	7	100	0.992	0.146	0.216	0.180	0.170	0.998	0.138	0.238	0.196	0.232
		250	1.000	0.196	0.566	0.372	0.332	1.000	0.218	0.676	0.562	0.614
		500	1.000	0.558	0.952	0.856	0.808	1.000	0.552	0.988	0.950	0.956
	9	100	0.994	0.154	0.210	0.144	0.138	0.994	0.108	0.252	0.210	0.196
		250	1.000	0.206	0.560	0.372	0.310	1.000	0.234	0.646	0.560	0.536
		500	1.000	0.554	0.942	0.794	0.690	1.000	0.542	0.976	0.930	0.940
0.075	5	100	0.988	0.138	0.288	0.186	0.182	0.980	0.134	0.314	0.286	0.260
		250	1.000	0.298	0.716	0.474	0.470	1.000	0.302	0.846	0.712	0.692
		500	1.000	0.682	0.986	0.946	0.902	1.000	0.700	0.992	0.988	0.990
	7	100	0.982	0.132	0.222	0.170	0.148	0.988	0.120	0.320	0.250	0.266
		250	0.998	0.304	0.676	0.382	0.288	1.000	0.252	0.788	0.668	0.640
		500	1.000	0.648	0.968	0.864	0.816	1.000	0.742	0.996	0.982	0.970
	9	100	0.994	0.134	0.218	0.160	0.152	0.986	0.136	0.318	0.244	0.216
		250	1.000	0.278	0.626	0.352	0.272	1.000	0.302	0.764	0.590	0.512
		500	1.000	0.684	0.962	0.768	0.628	1.000	0.706	0.986	0.956	0.912
0.100	5	100	0.966	0.154	0.278	0.186	0.168	0.960	0.134	0.402	0.308	0.274
		250	0.996	0.350	0.752	0.490	0.432	0.998	0.372	0.878	0.724	0.720
		500	1.000	0.762	0.984	0.922	0.876	1.000	0.788	0.998	0.998	0.992
	7	100	0.972	0.150	0.248	0.150	0.144	0.948	0.148	0.336	0.282	0.230
		250	1.000	0.320	0.722	0.392	0.296	0.998	0.318	0.852	0.680	0.586
		500	1.000	0.766	0.982	0.874	0.686	1.000	0.804	0.998	0.990	0.950
	9	100	0.972	0.188	0.240	0.118	0.132	0.948	0.144	0.320	0.182	0.164
		250	0.998	0.328	0.690	0.344	0.204	0.996	0.378	0.818	0.556	0.368
		500	1.000	0.764	0.966	0.722	0.474	1.000	0.816	0.994	0.936	0.822

Note: This table reports the rejection rates of marginal t-tests for the null hypothesis $H_0 : w(d) = 0$ vs. the alternative $H_A : w(d) \neq 0$. The data were generated under the alternative hypothesis considering two specifications for the functional parameter, which were exponential and Gaussian functions, for $d \in [0, C]$, with $C = 1$. The number of intervals characterising the partition is $K = 1/2h$. The DGP is given by $\lambda = 1$ and the number of simulations is 500.

rates close to 0.05 that, in most cases, are slightly above the nominal coverage rate. To study the relationship between the empirical coverage probability, the sample size and the functional form of $w(d)$, we have also considered $\theta = 7$ as an additional DGP. The empirical coverage rates provide very satisfactory results across the two functional specifications of $w(d)$, for different values of θ and for different sample sizes. Furthermore, the coverage rates converge to the nominal coverage rates at $\alpha = 0.05$ as the sample size increases.

Tables 5 and 6 present the power of the marginal t-tests obtained from the asymptotic convergence result (14) for the pointwise null hypothesis $H_0 : \lambda = 0$ vs. the alternative $H_A : \lambda \neq 0$, for $\lambda = 1, 0.25$, and $H_0 : w(d) = 0$ vs. $H_A : w(d) \neq 0$, for $d = \{h/2, h, 3h/2, 2h\}$. We should note that the DGP is generated under the alternative hypothesis given by $w(d)$ following an exponential function (left panel) or a Gaussian kernel function (right panel). To

Table 6: Empirical power of marginal t-test for $H_0 : w(d) = 0$.

			Model 1: Exponential function					Model 2: Gaussian kernel function				
h	θ	N	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$	λ	$w(h/2)$	$w(h)$	$w(3h/2)$	$w(2h)$
0.050	5	100	0.420	0.110	0.236	0.202	0.188	0.402	0.108	0.306	0.276	0.226
		250	0.704	0.224	0.614	0.476	0.430	0.636	0.190	0.710	0.588	0.602
		500	0.960	0.578	0.960	0.856	0.866	0.902	0.552	0.984	0.964	0.984
	7	100	0.378	0.116	0.206	0.176	0.182	0.390	0.152	0.302	0.214	0.214
		250	0.646	0.208	0.576	0.390	0.370	0.652	0.214	0.696	0.578	0.590
		500	0.918	0.578	0.950	0.828	0.788	0.882	0.526	0.978	0.958	0.956
	9	100	0.382	0.138	0.196	0.134	0.132	0.404	0.144	0.278	0.240	0.196
		250	0.630	0.208	0.580	0.368	0.282	0.652	0.262	0.656	0.532	0.538
		500	0.902	0.578	0.944	0.768	0.660	0.850	0.586	0.974	0.944	0.928
0.075	5	100	0.372	0.154	0.260	0.192	0.180	0.354	0.140	0.312	0.274	0.284
		250	0.548	0.264	0.720	0.524	0.470	0.510	0.262	0.812	0.706	0.734
		500	0.788	0.664	0.980	0.904	0.874	0.792	0.752	0.990	0.982	0.984
	7	100	0.342	0.130	0.238	0.160	0.144	0.372	0.162	0.358	0.252	0.226
		250	0.520	0.304	0.658	0.412	0.338	0.536	0.274	0.800	0.658	0.632
		500	0.712	0.692	0.982	0.854	0.782	0.708	0.704	0.996	0.978	0.966
	9	100	0.326	0.166	0.230	0.164	0.174	0.372	0.174	0.310	0.226	0.218
		250	0.566	0.276	0.612	0.348	0.226	0.554	0.312	0.760	0.596	0.526
		500	0.726	0.666	0.964	0.738	0.572	0.728	0.722	0.988	0.972	0.924
0.100	5	100	0.308	0.152	0.276	0.186	0.190	0.354	0.174	0.372	0.268	0.290
		250	0.452	0.372	0.782	0.492	0.426	0.450	0.366	0.878	0.750	0.752
		500	0.634	0.768	0.986	0.934	0.862	0.568	0.754	0.998	0.992	0.986
	7	100	0.306	0.134	0.250	0.122	0.126	0.324	0.144	0.318	0.218	0.210
		250	0.442	0.356	0.720	0.410	0.294	0.446	0.374	0.842	0.674	0.586
		500	0.562	0.732	0.978	0.856	0.716	0.564	0.768	0.994	0.978	0.944
	9	100	0.312	0.138	0.228	0.132	0.096	0.316	0.174	0.376	0.226	0.152
		250	0.416	0.338	0.672	0.310	0.186	0.456	0.360	0.808	0.548	0.378
		500	0.566	0.786	0.978	0.756	0.436	0.606	0.834	0.992	0.958	0.808

Note: This table reports the rejection rates of marginal t-tests for the null hypothesis $H_0 : w(d) = 0$ vs. the alternative $H_A : w(d) \neq 0$. The data were generated under the alternative hypothesis considering two specifications for the functional parameter, which were exponential and Gaussian functions, for $d \in [0, C]$, with $C = 1$. The number of intervals characterising the partition is $K = 1/2h$. The DGP is given by $\lambda = 0.25$ and the number of simulations is 500.

be consistent with the study of the empirical coverage probability at $\alpha = 0.05$, we consider $\beta = 0.1$ and $\theta = 5, 7, 9$ as data-generating processes for both the exponential and Gaussian kernel functions. The results of this simulation exercise show the strong performance of the t-tests when it comes to rejecting the null hypothesis across values of d in the grid and for different DGPs. The empirical power of the test is large in most instances and achieves values above 0.80 for $N = 500$ in most scenarios.

5.3. Size and power of the uniform test

The study of the power of the marginal t-tests confirms empirically their suitability for detecting spatial effects for specific values of d given by $d = \{h/2, h, 3h/2, 2h\}$. This subsection extends this analysis by evaluating the finite-sample size and power of the uniform test presented in equation (17) and Theorem 3. We consider two different null hypotheses given by *i*) the absence of spatial effects and *ii*) a specific functional form for $w(d)$ given by the exponential function.

Table 7: Empirical size and power of the uniform test (17).

		H_0: No network structure					
		Size			Power		
θ	N	h=0.050	0.075	0.100	h=0.050	0.075	0.100
5	100	0.071	0.056	0.065	0.998	0.998	1.000
	250	0.049	0.052	0.036	1.000	1.000	1.000
	500	0.073	0.032	0.040	1.000	1.000	1.000
9	100	0.065	0.062	0.071	0.985	0.997	0.998
	250	0.039	0.042	0.054	1.000	1.000	1.000
	500	0.035	0.036	0.052	1.000	1.000	1.000
		H_{0f}: Exponential function					
		Size			Power		
θ	N	h=0.050	0.075	0.100	h=0.050	0.075	0.100
5	100	0.056	0.058	0.046	0.994	1.000	0.999
	250	0.024	0.042	0.041	1.000	1.000	1.000
	500	0.030	0.030	0.034	1.000	1.000	1.000
9	100	0.055	0.057	0.071	0.980	0.994	0.998
	250	0.037	0.042	0.028	1.000	1.000	1.000
	500	0.038	0.033	0.034	1.000	1.000	1.000

Note: This table reports the rejection rates of the uniform test for two different DGPs under the null hypothesis. The nominal size is $\alpha = 0.05$, and the number of simulations is 500.

To assess the presence of spatial effects, data are generated under the null hypothesis $w(d) = 0$ for every $d \in \chi$. This implies that the DGP is a standard cross-sectional regression model. For the simulation exercise, we consider $\lambda = 1$ and $\beta = 0.1$. The top panel of Table 7 reports the empirical size and power for the null hypothesis for a nominal size $\alpha = 0.05$. Due to space constraints, we only consider $\theta = 5$ and 9. The figures show reliable empirical size estimates for the uniform test T_N for different values of h across sample sizes. The same procedure has been implemented to evaluate the specification of the functional coefficient $w(d)$. In this case, the null hypothesis of interest is $H_{0f} : w(d) = \exp(-\theta d)$, for $d \in [0, 1]$. The empirical size and power of the test are reported in the bottom panel of Table 7. We observe results similar to those obtained for the case of spatial effects, i.e. the empirical power is extremely high even when the test is slightly undersized.

6. Empirical application

This section extends the analysis carried out by Levinson and O'Brien (2019) by incorporating neighbouring effects in the relationship between households' income and the pollution generated to produce the goods and services they consume. These authors construct a rich dataset obtained from a random survey of households to study environmental Engel curves (EECs) for the US for each year from 1984 to 2012. The number of individuals available for their study varies from 3000 to 3500 per year.

Levinson and O'Brien (2019) have two main objectives. The first one is to find the shape of the relationship between income and pollution, and the magnitude of the slope, and to study its curvature. The second aim is to analyse shifts in EECs in terms of income increases (movements along the curve) or in terms of regulation-induced price increases (movements of the curve). By conducting the analysis separately for each year, Levinson and O'Brien (2019) are able to control for prices, available products and regulations. These authors find that EECs are upward-sloping, indicating that richer households cause more pollution and that the rate at which pollution increases with income is less than one-to-one. In addition, pollution increases at a decreasing rate with income over time, i.e. EECs are concave. The latter result shows that households have consumed a basket of goods that cause less pollution, both directly and indirectly, in recent years.

Table 8: Empirical application: Descriptive statistics of the variables for selected years (cross section).

	Year 1984 (N=3,184)				Year 1998 (N=2,706)				Year 2012 (N=3,538)			
	Mean	St. Dev.	Minimum	Maximum	Mean	St. Dev.	Minimum	Maximum	Mean	St. Dev.	Minimum	Maximum
PM10	11.874	7.027	0.726	87.029	12.178	6.532	0.840	59.402	11.626	6.780	0.399	89.510
VOCs	19.990	14.522	1.180	309.383	18.948	13.101	1.139	179.079	16.352	12.331	0.234	227.510
NOx	73.270	39.559	3.559	362.799	77.408	37.872	3.763	311.762	71.658	38.244	2.634	539.093
SO_2	117.869	66.667	5.649	553.229	126.882	63.329	6.203	497.223	121.411	65.327	4.928	1,044.552
CO	45.723	34.375	2.346	374.047	43.277	31.623	2.318	289.077	39.195	30.928	0.815	357.851
Income	3.913	2.819	-0.083	16.866	4.306	3.255	-0.043	20.312	4.706	3.954	-0.060	25.192
Income sq.	23.258	33.400	0.000	284.448	29.134	44.722	0.000	412.589	37.772	68.971	0.000	634.628

Notes: Pollutants are measured in pounds and have been calculated according to the technologies and emission intensities of the year 2002. After-tax income is expressed in units of 10,000 2002 US dollars.

Levinson and O'Brien (2019) construct two types of EECs: one that uses only income as a covariate, and a multivariate model that incorporates households' characteristics (up to 18 regressors). Table 2 in Levinson and O'Brien (2019) presents a detailed description of these variables. As pointed out by these authors, *'adding those common demographic variables has little effect on the conclusions about the shapes of EECs or how they have changed over time'* (p. 122). Furthermore, endogeneity between household income and household pollution is not an issue in this context, as discussed by these authors on page 124. In the present application, we focus on their first model, which considers pollution as a function of after-tax income and its square. These authors estimate separate curves for five major air pollutants – particulates smaller than 10 microns (PM10), volatile organic compounds (VOC_s), nitrogen oxides (NO_x), sulfur dioxide (SO_2) and carbon monoxide (CO) – because they are not measured in the same units and have different environmental consequences; see Table 8 for cross-sectional descriptive statistics of selected years of the sample. By adopting this approach, and taking into account the literature on peer effects in household consumption and energy behaviour (Agarwal et al., 2021; De Giorgi et al., 2020;

Wolske et al., 2020), we estimate the following specification:

$$p_{it} = \lambda_{1t} y_{it} + \sum_{\substack{j=1 \\ j \neq i}}^{N} w(d_{ij,t}) y_{jt} + \lambda_{2t} y_{it}^2 + \varepsilon_{it}, \tag{27}$$

where p_{it} and y_{it} are the pollution and after-tax income, respectively; ε_{it} is the error term that satisfies $E[\varepsilon_{it} \mid Y_t] = 0$, with $Y_t = (y_{1t}, \ldots, y_{Nt})$. The coefficients are indexed by t because we run separate regressions for each year.

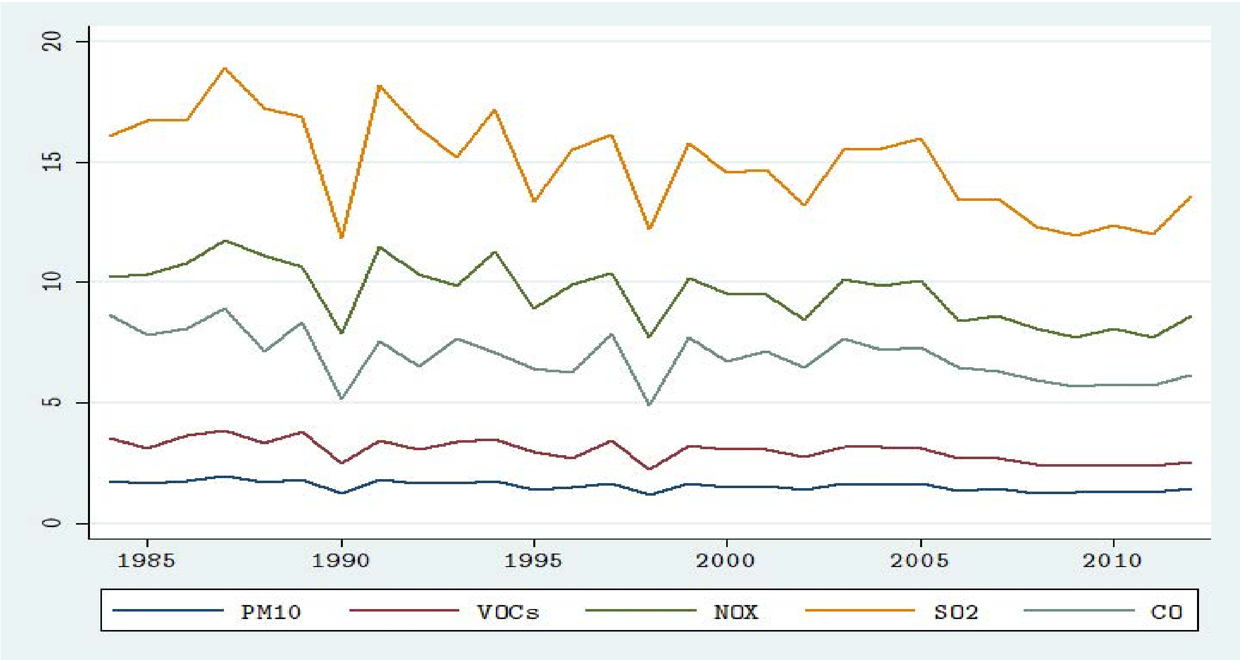
The above specification incorporates indirect effects on the relationship between pollution and income from households with similar levels of after-tax income. The model aims to capture these effects using the functional coefficient $w(d_{ij,t})$, with $d_{ij,t} = |y_{it} - y_{jt}|$, for $i, j = 1, \ldots, N_t$, where N_t is the number of households included in the sample for a given year. For simplicity, we restrict the spatial effects to a linear relationship between pollution and income. The estimation equation can be extended, at the expense of a larger regression model, by also assuming that there are spatial effects on the quadratic component. Using the specification presented in expression (3), the above model can be approximated by

$$p_{it} = \lambda_{1t} y_{it} + \sum_{k=1}^{K} \sum_{m=0}^{q} \gamma_{km,t} y_{it}^{(km)} + \lambda_{2t} y_{it}^2 + \varepsilon_{it}, \tag{28}$$

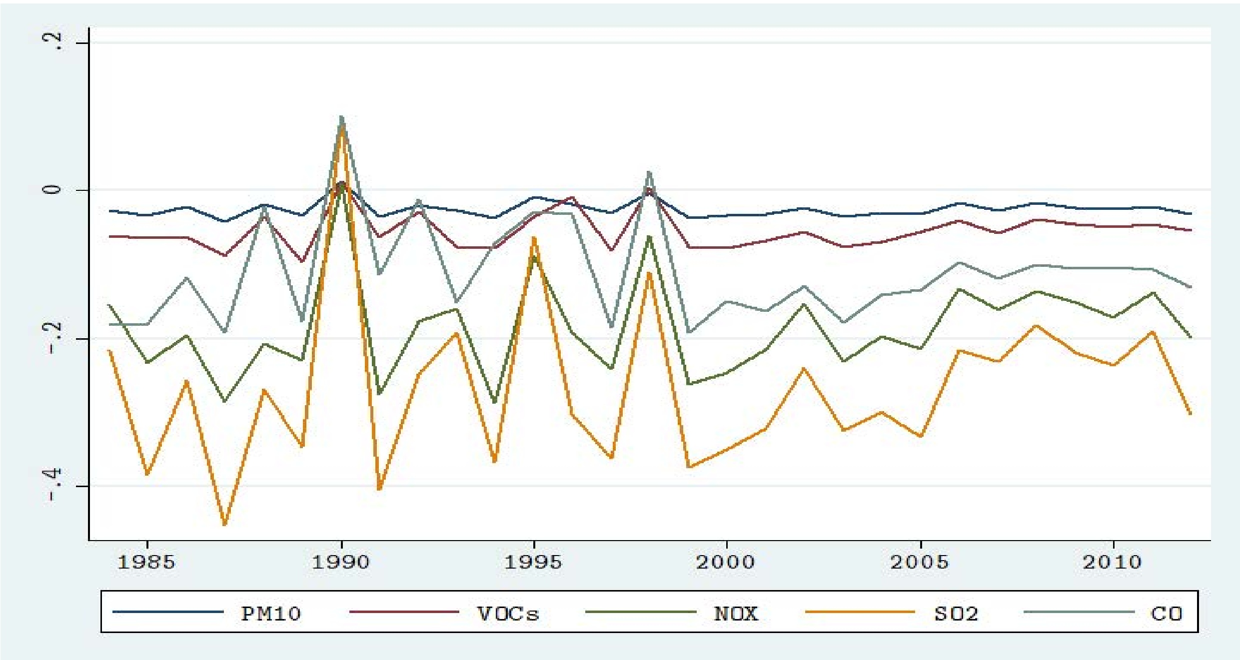
where $y_{it}^{(km)} = \sum_{\substack{j=1 \\ j \neq i}}^{N} y_{jt} (d_{ij,t} - z_k)^m 1_k(d_{ij,t})$, and $1_k(d_{ij,t}) = 1(|d_{ij,t} - z_k| \leq h)$. The regression coefficients are $\gamma_{km,t} = \frac{1}{m!} w^{(m)}(z_k)$, and they correspond to the Taylor expansion for $m = 0, 1, \ldots, q$, with $q = 2$ in this application.

The dynamics of the parameters associated with the relationship between the different pollutants and households' after-tax income (λ_{1t}) and its square (λ_{2t}) are plotted in Figure 1. Although these parameters display different magnitudes across pollutants, they suggest a positive relationship between pollution and income that, in line with Levinson and O'Brien (2019), tends to decrease over time. In fact, the magnitude of the estimated coefficients is very similar to that obtained by these authors for the quadratic model (see the first column of their Table 2). The estimated EEC for PM10 using household data from 1984 is concave, with a linear coefficient associated with after-tax income of 1.95 and a negative coefficient associated with income squared of -0.03; both are statistically significant. Our estimates are of a similar magnitude and significant at the 1% level. For completeness, we also report the results for the other pollutants being studied. In all cases, they also indicate a concave-shaped, and statistically significant at 1%, relationship with household income.

More importantly, Figure 2 reports the estimates of the functional coefficients $w(d)$ in equation (28) that capture spatial effects for selected values of the distance variable d. We consider $d = \{h/2, h, 3h/2, 2h\}$, as in the Monte Carlo exercise. However, the choice of the bandwidth parameter h is determined by the optimisation of the Bayesian information criterion and slightly varies between $h_{opt} = 0.225$ and $h_{opt} = 0.25$. Similar results are obtained when the Akaike information criterion is used. In this application, the spatial effects

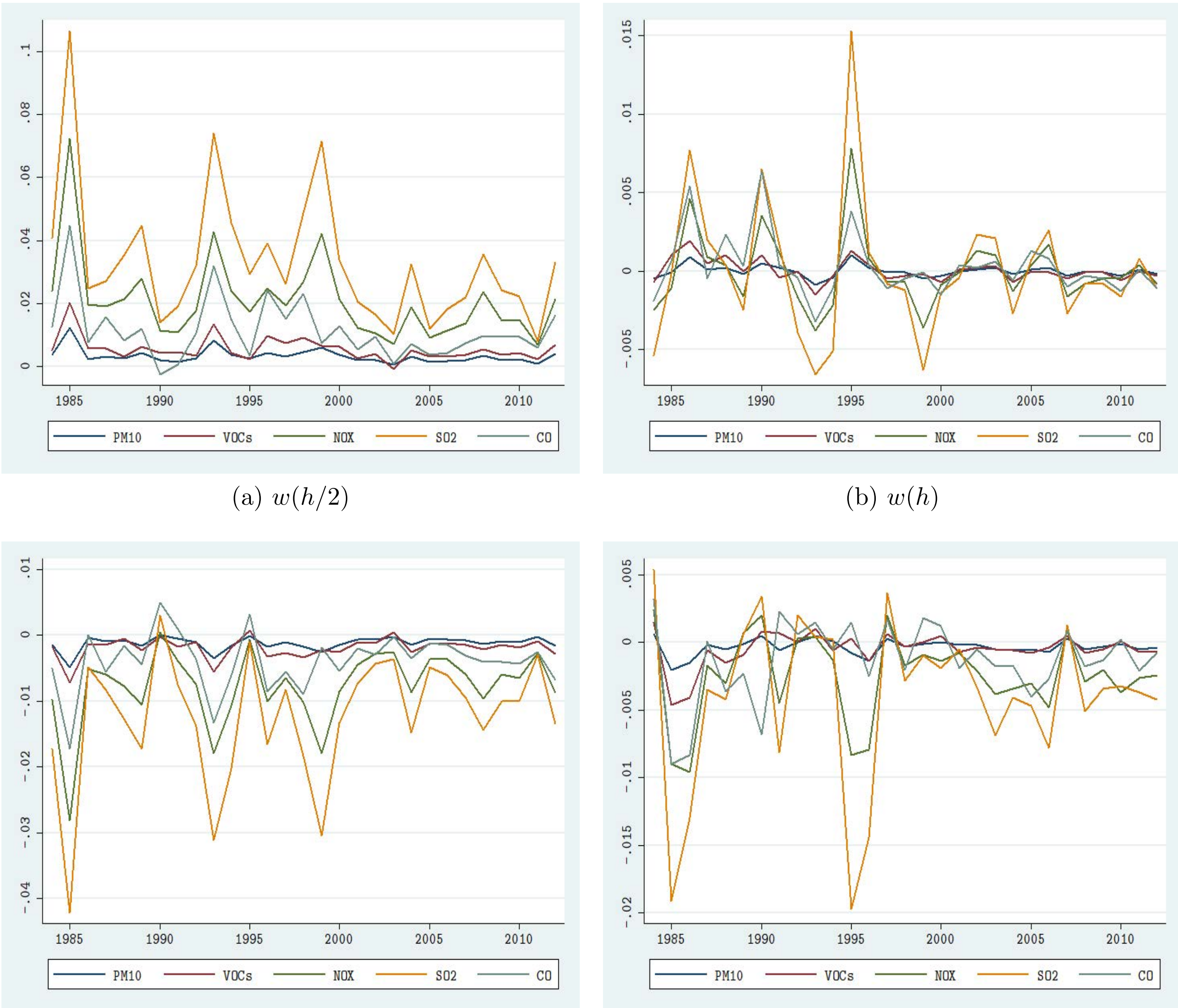


(a) Coefficient for after-tax household income ($\gamma_{1t,net}$).



(b) Coefficient for squared after-tax household income ($\gamma_{2t,net}$).

Figure 1: Spatial regression estimation of environmental Engel curves (EECs) in the US from 1984–2012.



(a) $w(h/2)$ (b) $w(h)$

(c) $w(3h/2)$ (d) $w(2h)$

Figure 2: Spatial regression estimation of EECs in the US from 1984–2012. Functional coefficient.

describe the explanatory power of households with similar after-tax incomes on the pollution generated to produce the goods and services that these households consume. These neighbouring effects are, in most periods, of small magnitude but statistically significant at the 5% level, according to the p-values of the marginal t-tests reported for each coefficient. The analysis of the NO_x and SO_2 pollutants yields parameter estimates of a larger magnitude. Interestingly, the sign of the functional coefficients varies across h_{opt}: although we find positive neighbouring effects for $w(h_{opt}/2)$, they are negative for $w(3h_{opt}/2)$ and $w(2h_{opt})$. The statistical significance of these neighbouring effects is illustrated in Figure 3. This chart reports the p-values of the uniform test T_N in expression (17) over the course of the evaluation period. Despite their fluctuation, the results provide ample support for the

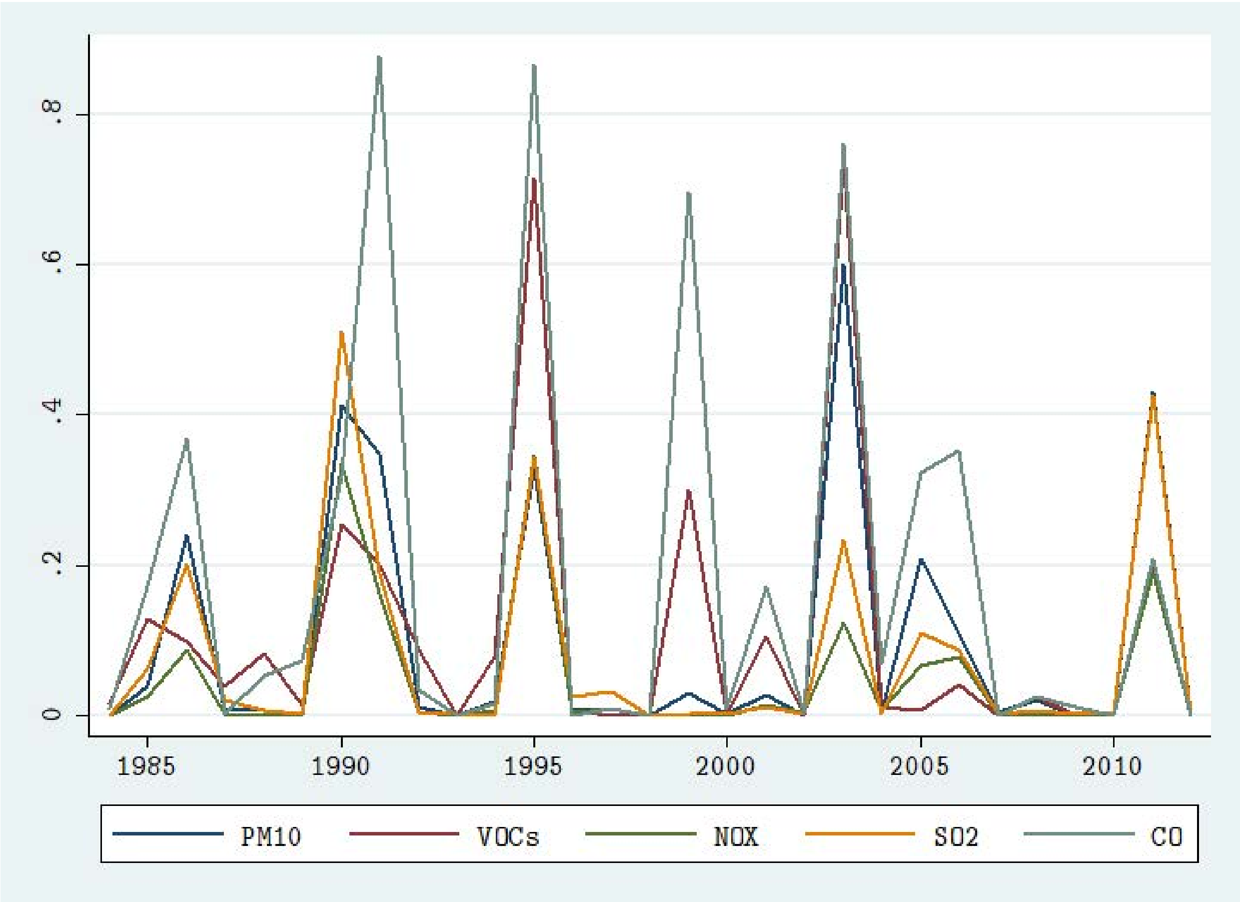
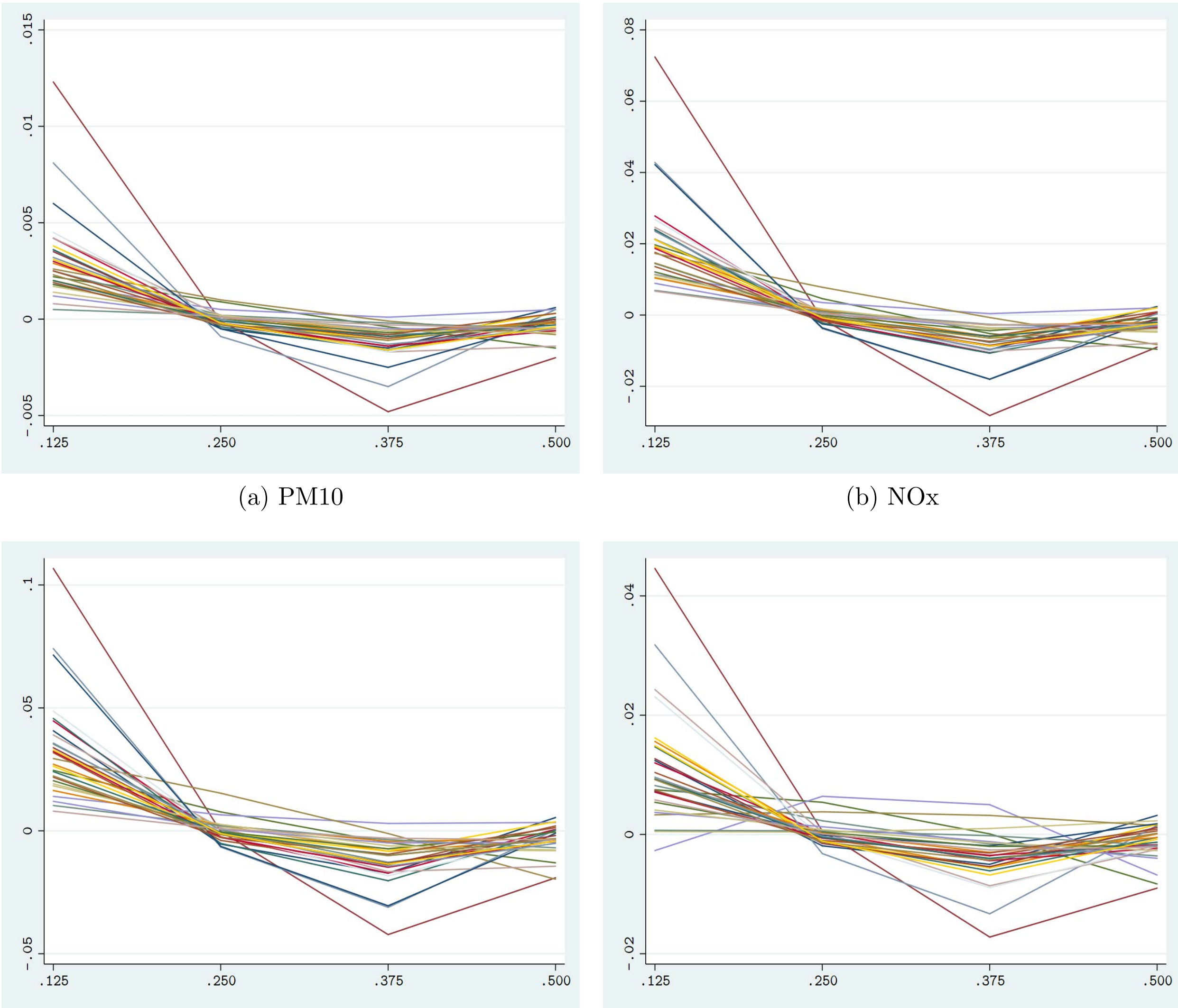


Figure 3: Spatial regression estimation of the functional coefficient. Uniform test statistic, p-values.

significance of neighbouring effects in this context, adding further evidence to that obtained from the marginal t-tests for the different realisations of the function $w(d)$ discussed above.

Figure 4 presents a similar analysis but reports the estimator $\widehat{w}(d)$ as a function of the distance $d = \{h/2, h, 3h/2, 2h\}$. For ease of presentation, we only report the results for the pollutants PM10, NO_x, SO_2 and CO. The results for VOC_s are similar and omitted from the figure due to space considerations. The estimates show a declining pattern in the spatial effect and some small negative values of the function as the distance d increases. The most important effect occurs for the neighbouring observations at $d = h/2$, with $h = 0.25$. Although the magnitudes of the spatial effects are quite small, we should note that they are similar to more conventional spatial models characterised by a fixed interaction matrix. In these models, it is common practice to impose spatial weights given by $1/N$ for contiguous observations, implying spatial effects of similar or even smaller magnitudes than in our empirical exercise.

The analysis is completed by studying the adjusted coefficient of determination (R^2) of the spatial regression model (28). The presence of heterogeneity in the explanatory power across models and over time is shown at the top of Figure 5. More specifically, the adjusted



(a) PM10

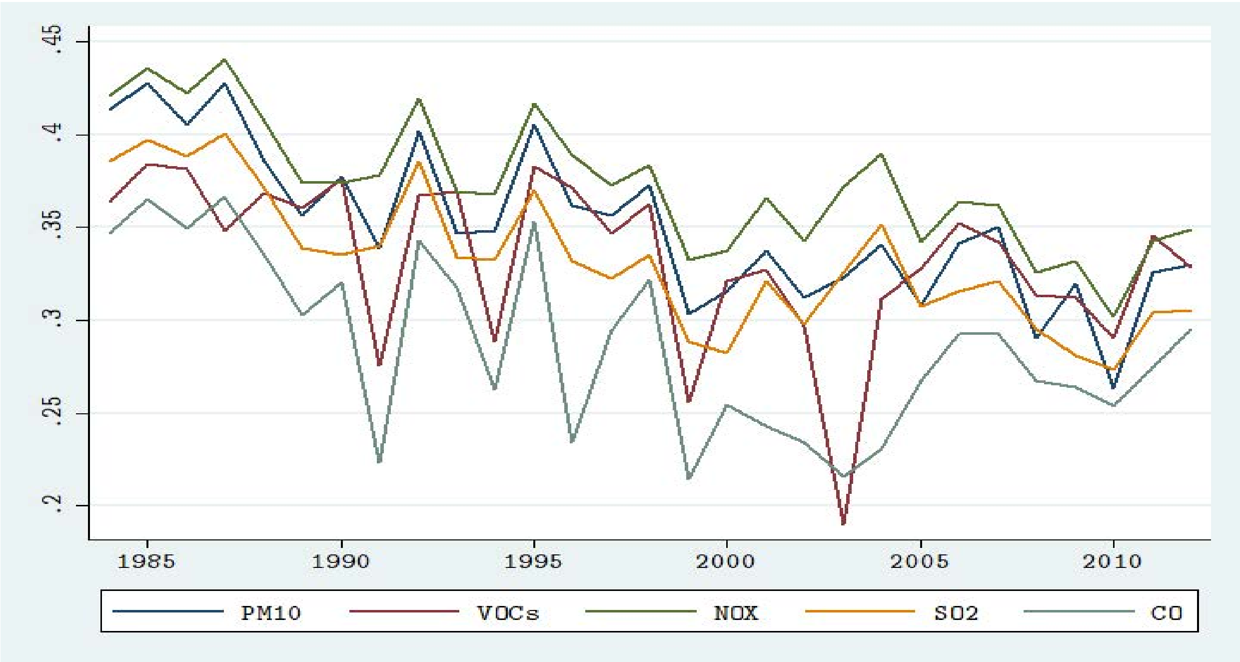
(b) NOx

(c) SO_2

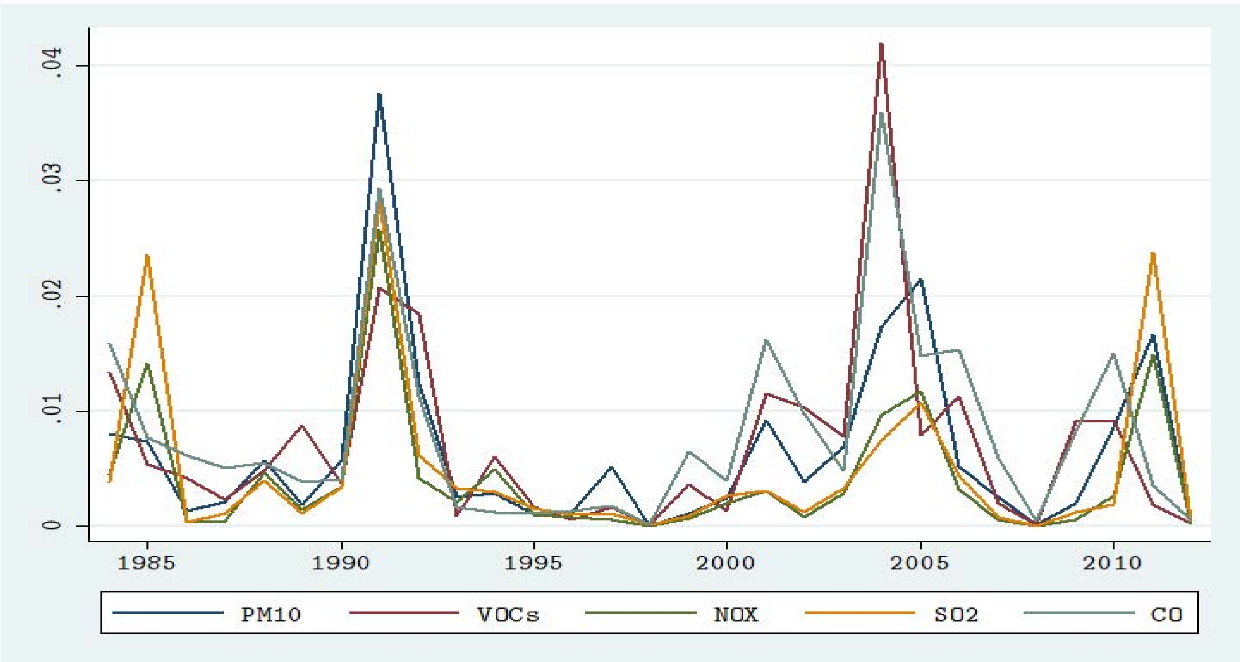
(d) CO

Figure 4: $\widehat{w}(d)$ for $d \in [0.125, 0.5]$ for the spatial regression estimation of EECs in the US from 1984–2012.

R^2 value fluctuates between 0.25 and 0.45. We should note that these figures are particularly high given that the number of regressors of model (28) is $(q+1)K_{opt}+2$, with $K_{opt} = \frac{C}{2h_{opt}}$. This number varies with the choice of the optimal bandwidth parameter h obtained for each period and model but is between 10 and 15 regressors. To attach a statistical figure to these values, we compute the F-test using the difference between the unadjusted R^2 value obtained from the spatial regression model and its cross-sectional counterpart, which is given by the quadratic regression model estimated by Levinson and O'Brien (2019). The p-values of the F-test, plotted at the bottom of Figure 5, show overwhelming evidence that the statistical significance of the augmented model given by considering the spatial variables is greater than that of the benchmark model given by the cross-sectional quadratic regression model.



(a) Adjusted coefficients of determination (R^2).



(b) F-test for adjusted R^2 values, p-values.

Figure 5: Spatial regression estimation of EECs in the US from 1984–2012.

7. Conclusions

This paper proposes a spatial regression model that extends conventional spatial regression models in several dimensions. Importantly, the spatial effects are modelled as a functional coefficient indexed by a spatial variable. Our model is approximated by local piecewise polynomials estimated over disjoint intervals of a partition of the domain of the spatial variable. By doing this, we avoid model misspecification issues produced by imposing a certain parametric structure on the spatial dependence. The second innovation is to extend the SLX model by considering a broader definition of spatial dependence in which neighbouring observations are determined by a spatial variable that is potentially different from the geographical distance. The technical implications of this extension are not trivial because the spatial variable becomes stochastic. These results are formalised by studying the asymptotic properties of the nonparametric estimators and proposing statistical tests for assessing the presence of spillover effects. This is done using pointwise and uniform tests in which the presence of spatial effects is tested over the domain of the spatial variable.

The proposed methodology is illustrated in an empirical application that involves studying the environmental Engel curves discussed in a recent influential work by Levinson and O'Brien (2019). We find strong empirical evidence of neighbouring effects on the relationship between different forms of environmental pollution and after-tax household income.

Acknowledgements

The authors acknowledge financial support from Gobierno de Aragón (grants LMP21-21 and S39-20R ADETRE Research Group) and Ministerio de Ciencia e Innovacion/Agencia Estatal de Investigación (grants PID2019-104326GB-I00 and PID2020-112773GB-I00). Jose Olmo also acknowledges financial support from Fundación Agencia Aragonesa para la Investigación y el Desarrollo.

A. Mathematical Proofs

Proof of Lemma 1. Let $\widetilde{Q}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\mathbb{X}'_{ki}\mathbb{X}_{ki}$ and $\widetilde{A}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\mathbb{X}'_{ki}\mathbb{X}_{ki}e_i^2$, with $e_i = y_i - \widehat{\lambda}x_i - \sum_{k=1}^{K}\mathbb{X}_{ki}\,\widehat{\Gamma}_k$ the residual term, $\alpha_N = N(N-1)$ a standardizing constant for the spatial coefficients, and p_k the probability of belonging to a given interval of the partition of χ. Similarly, let $\widehat{Q}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}\overline{X}_{k,ij}$ and $\widehat{A}_k = \frac{1}{\alpha_N p_k}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}\overline{X}_{k,ij}e_i^2$, where $\overline{X}_{k,ij} = (x_j 1_k(d_{ij}), x_j(d_{ij}-z_k)1_k(d_{ij}), \ldots, x_j(d_{ij}-z_k)^q 1_k(d_{ij}))$.

To prove the asymptotic convergence in probability between $\widetilde{Q}_k$ and $\widehat{Q}_k$, it is sufficient to show that $E[||\widetilde{Q}_k - \widehat{Q}_k||^2] = o(1)$ as $N \to \infty$. Note that

$$E[\|\widetilde{Q}_k - \widehat{Q}_k\|^2] = \frac{1}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}E\left[\sum_{i=1}^{N}\left(\widetilde{x}_i^{(kr)}\widetilde{x}_i^{(ks)} - \sum_{\substack{j=1\\j\neq i}}^{N}x_j^2(d_{ij}-z_k)^{r+s}1_k(d_{ij})\right)^2\right],$$

with q the order of the Taylor polynomial such that the matrices are of dimension $q+1$. Simple algebra applied to each element of the matrix shows that

$$\widetilde{x}_i^{(kr)}\widetilde{x}_i^{(ks)} - \sum_{\substack{j=1 \\ j\neq i}}^{N} x_j^2 (d_{ij}-z_k)^{r+s} 1_k(d_{ij}) = \sum_{\substack{j=1 \\ j\neq i}}^{N}\sum_{\substack{l=1 \\ l\neq i,j}}^{N} x_j x_l (d_{ij}-z_k)^r 1_k(d_{ij})(d_{il}-z_k)^s 1_k(d_{il}).$$

Let $1_{k,ij}$ denote the event $1_k(d_{ij}) = 1$. Then, the above expression can be expressed as $E[\|\widetilde{Q}_k - \widehat{Q}_k\|^2] =$

$$\frac{1}{N(N-1)^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q} E\left[\left(\sum_{\substack{j=1 \\ j\neq i}}^{N}\sum_{\substack{l=1 \\ l\neq i,j}}^{N} x_j x_l (d_{ij}-z_k)^r 1_{k,ij}(d_{il}-z_k)^s 1_{k,il}\right)^2\right]$$

$$= \frac{1}{N(N-1)^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{\substack{j=1 \\ j\neq i}}^{N}\sum_{\substack{l=1 \\ l\neq i,j}}^{N} E\left[E\left[x_j^2 x_l^2 \mid 1_{k,ij}, 1_{k,il}\right](d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s} 1_{k,ij}1_{k,il}\right]\right),$$

applying the law of iterated expectations and noting that $E[x_j x_l x_r x_s \mid 1_{k,ij}, 1_{k,il}, 1_{k,ir}, 1_{k,is}] = E[x_j \mid 1_{k,ij}]E[x_l \mid 1_{k,il}]E[x_r \mid 1_{k,ir}]E[x_s \mid 1_{k,is}] = 0$ for $j \neq l \neq r \neq s$, under assumption A.1. Similarly, $E[x_j^2 x_l x_r \mid 1_{k,ij}, 1_{k,il}, 1_{k,ir}] = 0$ for $j \neq l \neq r$. Furthermore, the *iid* assumption in A1 also implies that

$$\frac{1}{N(N-1)^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{\substack{j=1 \\ j\neq i}}^{N}\sum_{\substack{l=1 \\ l\neq i,j}}^{N} E\left[E\left[x_j^2 x_l^2 \mid 1_{k,ij}, 1_{k,il}\right](d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s} 1_{k,ij}1_{k,il}\right]\right)$$

$$\leq \frac{C}{N}\sum_{r=0}^{q}\sum_{s=0}^{q} h_k^{2(r+s)},$$

given that $E\left[x_j^2 x_l^2 \mid 1_{k,ij}1_{k,il}\right] < C$ under assumption A.1, for C some positive constant, and $E[(d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s}1_{k,ij}1_{k,il}] = E[(d_{ij}-z_k)^{2r} \mid 1_{k,ij}]E[(d_{il}-z_k)^{2s} \mid 1_{k,il}]p_k^2 \leq h^{2(r+s)}p_k^2$. Finally,

$$E[\|\widetilde{Q}_k - \widehat{Q}_k\|^2] \leq \frac{C}{N}\sum_{r=0}^{q}\sum_{s=0}^{q} h_k^{2(r+s)} = \frac{C}{N}\left(\frac{1-h_k^{2(q+1)}}{1-h_k^2}\right)^2,$$

which converges to zero given that $N \to \infty$, and $\left(\frac{1-h_k^{2(q+1)}}{1-h_k^2}\right)^2 < \infty$ for $0 < h_k < 1$. Therefore, $E[\|\widetilde{Q}_k - \widehat{Q}_k\|^2] = O\left(\frac{1}{N}\right)$ that implies $||\widetilde{Q}_k - \widehat{Q}_k|| = O_p\left(\frac{1}{\sqrt{N}}\right)$ as $N \to \infty$, applying standard probability theory convergence results.

The proof for the asymptotic convergence between $\widetilde{A}_k$ and $\widehat{A}_k$ follows analogously. □

Proof of Lemma 2. To prove the asymptotic convergence of $\widehat{Q}_k$ to Q_k, it is sufficient to show that $E[||\widehat{Q}_k - Q_k||^2] = o(1)$. Let $\widehat{Q}_k = \frac{1}{\alpha_N}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}\overline{X}_{k,ij}/p_k$, with $\overline{X}_{k,ij} = (x_j 1_k(d_{ij}), x_j(d_{ij} - z_k)1_k(d_{ij}), \ldots, x_j(d_{ij} - z_k)^q 1_k(d_{ij}))$, such that

$$E[\|\widehat{Q}_k - Q_k\|^2] =$$

$$\sum_{r=0}^{q}\sum_{s=0}^{q} E\left[\frac{1}{\alpha_N^2 p_k^2}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\left(x_j^2(d_{ij}-z_k)^r(d_{ij}-z_k)^s 1_k(d_{ij}) - E\left[x_j^2(d_{ij}-z_k)^r(d_{ij}-z_k)^s 1_k(d_{ij})\right]\right)\right)^2\right].$$

Following similar algebra to the proof of Lemma 1, we use $1_{k,ij}$ to denote the event $1_k(d_{ij}) = 1$, and apply the law of iterated expectations in several instances. Thus, we obtain $E[||\widehat{Q}_k - Q_k||^2] =$

$$\frac{1}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N} E\left[E[x_j^4 \mid 1_{k,ij}](d_{ij}-z_k)^{2(r+s)}1_{k,ij}\right]\right)$$

$$+\frac{N-2}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N} E\left[E[x_j^4 \mid 1_{k,ij}, 1_{k,ji}](d_{ij}-z_k)^{2r}(d_{ji}-z_k)^{2s}1_{k,ij}1_{k,ji}\right]\right)$$

$$+\frac{N-1}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i,j}}^{N} E\left[E[x_j^2 x_l^2 \mid 1_{k,ij}, 1_{k,il}](d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s}1_{k,ij}1_{k,il}\right]\right)$$

$$-\frac{N-1}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i}}^{N} E\left[E[x_j^2 \mid 1_{k,ij}](d_{ij}-z_k)^{2r}1_{k,ij}\right] E\left[E[x_l^2 \mid 1_{k,il}](d_{il}-z_k)^{2s}1_{k,il}\right]\right).$$

The *iid* cross-sectional assumption in A1 entails the condition $E[x_j^2 x_l^2 \mid 1_{k,ij}1_{k,il}] = E[x_j^2 \mid 1_{k,ij}]E[x_l^2 \mid 1_{k,il}]$ for $j \neq l$, such that the only remaining elements are those defining the fourth central moments. More specifically,

$$E[||\widehat{Q}_k - Q_k||^2] = \frac{1}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N} E\left[\left(E[x_j^4 \mid 1_{k,ij}] - E^2[x_j^2 \mid 1_{k,ij}]\right)(d_{ij}-z_k)^{2(r+s)}1_{k,ij}\right]\right)$$

$$+\frac{N-2}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N} E\left[\left(E[x_j^4 \mid 1_{k,ij}1_{k,ji}] - E^2[x_j^2 \mid 1_{k,ij}1_{k,ji}]\right)(d_{ij}-z_k)^{2r}(d_{ji}-z_k)^{2s}1_k(d_{ij})1_k(d_{ji})\right]\right)$$

Under A.1, the sequence $\{x_i\}_{i=1}^N$ is *iid* implying that

$$E[||\widehat{Q}_k - Q_k||^2] = \frac{1}{\alpha_N p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q} E\left[\left(E[x_j^4 \mid 1_{k,ij}] - E^2[x_j^2 \mid 1_{k,ij}]\right)(d_{ij} - z_k)^{2(r+s)} 1_{k,ij}\right] \quad \text{(A.1)}$$

$$+ \frac{N-2}{\alpha_N p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q} E\left[\left(E[x_j^4 \mid 1_{k,ij}1_{k,ji}] - E^2[x_j^2 \mid 1_{k,ij}1_{k,ji}]\right)(d_{ij} - z_k)^{2r}(d_{ji} - z_k)^{2s} 1_k(d_{ij})1_k(d_{ji})\right].$$

Furthermore, $E[x_j^4 \mid 1_{k,ij}1_{k,ji}] \leq C_0 < \infty$ for $i, j = 1, \dots, N$, such that the above expression can be bounded as

$$\begin{aligned}
E[||\widehat{Q}_k - Q_k||^2] &\leq \frac{C_0}{\alpha_N}\sum_{r=0}^{q}\sum_{s=0}^{q} E[(d_{ij} - z_k)^{2(r+s)} \mid 1_{k,ij}]/p_k \\
&+ \frac{C_0}{N}\sum_{r=0}^{q}\sum_{s=0}^{q} E[(d_{ij} - z_k)^{2r}(d_{ji} - z_k)^{2s} \mid 1_k(d_{ij})1_k(d_{ji})] \\
&\leq \frac{C_0}{\alpha_N}\sum_{r=0}^{q}\sum_{s=0}^{q} h^{2(r+s)}/p_k + \frac{C_0}{N}\sum_{r=0}^{q}\sum_{s=0}^{q} h^{2(r+s)} = C_0\left(\frac{K}{\alpha_N} + \frac{1}{N}\right)\left(\frac{1 - h^{2(q+1)}}{1 - h^2}\right)^2 \\
&= O\left(\frac{K}{\alpha_N} + \frac{1}{N}\right),
\end{aligned}$$

by noting that $p_k \asymp K^{-1}$ given that $p_k \asymp h$ as imposed in assumption A.5. This result is also obtained by noting (see also the proof of Lemma 1) that $\sum_{r=0}^{q}\sum_{s=0}^{q} h^{2(r+s)} = \sum_{r=0}^{q} h^{2r}\sum_{s=0}^{q} h^{2s} = \left(\frac{1-h^{2(q+1)}}{1-h^2}\right)^2$. Finally, noting from assumption A.5 that $K/N \to 0$, we obtain $E[||\widehat{Q}_k - Q_k||^2] = O\left(\frac{1}{N}\right)$ such that $||\widehat{Q}_k - Q_k|| = O_p\left(\frac{1}{\sqrt{N}}\right)$.

Expression (A.1) simplifies under symmetry of the distance variable ($d_{ij} = d_{ji}$) for all $i, j = 1, \dots, N$. In this case, the sum of the two expressions in (A.1) collapses into one single term such that

$$\begin{aligned}
E[||\widehat{Q}_k - Q_k||^2] &= \frac{N-1}{\alpha_N^2 p_k^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} E\left[\left(E[x_j^4 \mid 1_{k,ij}] - E^2[x_j^2 \mid 1_{k,ij}]\right)(d_{ij} - z_k)^{2(r+s)} 1_{k,ij}\right]\right) \\
&= \frac{1}{N}\sum_{r=0}^{q}\sum_{s=0}^{q} E[(d_{ij} - z_k)^{2(r+s)} \mid 1_{k,ij}]/p_k \\
&\leq \frac{C_0}{N p_k}\left(\frac{1 - h^{2(q+1)}}{1 - h^2}\right)^2 = O\left(\frac{K}{N}\right),
\end{aligned}$$

under assumption A.5. Then, $||\widehat{Q}_k - Q_k|| = O_p\left(\frac{\sqrt{K}}{\sqrt{N}}\right)$. □

Proof of Proposition 1. The orthogonality condition $\widehat{X}_u'\mathbb{X} = X'M_{\mathbb{X}}\mathbb{X} = \mathbf{0}$, with $\mathbf{0}$ a $(p+1)\times K(q+1)$ matrix of zeros, implies that

$$\widehat{\lambda} - \lambda = \widehat{\Phi}^{-1}\frac{1}{N}X'M_{\mathbb{X}}\varepsilon + \widehat{\Phi}^{-1}\frac{1}{N}X'M_{\mathbb{X}}\overline{R}. \tag{A.2}$$

The consistency of the vector of parameter estimators is obtained by showing (i) $\|\widehat{\Phi}-\Phi_0\| = o_p(1)$ with $\|\Phi_0\| < \infty$, (ii) $\|\frac{1}{N}X'M_{\mathbb{X}}\varepsilon\| = o_P(1)$ and (iii) $\|\frac{1}{N}X'M_{\mathbb{X}}\overline{R}\| = o_p(1)$ as $N \to \infty$. The proof of condition (i) follows from the law of large numbers for *iid* sequences under assumption A1. For condition (ii), under assumption A1, it is sufficient to show that $\frac{1}{N^2}E[\|X'M_{\mathbb{X}}\varepsilon\|^2] = o(1)$. This is, however, naturally satisfied under assumptions A1 and A4 that entail the existence of finite second moments of x_i and ε_i. More formally, $\frac{1}{N^2}E\left[\left(\sum_{i=1}^{N}X_i'M_{\mathbb{X}i}\varepsilon_i\right)^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[X_i'M_{\mathbb{X}i}\varepsilon_iX_j'M_{\mathbb{X}j}\varepsilon_j\right] = \frac{1}{N}E\left[X_i'M_{\mathbb{X}i}X_i\varepsilon_i^2\right]$, with $M_{\mathbb{X}i}$ and $M_{\mathbb{X}j}$ columns of matrix $M_{\mathbb{X}}$, under the mutual independence between the error terms in assumption A4. Now, applying the Cauchy-Schwarz inequality:

$$\frac{1}{N}E\left[X_i'M_{\mathbb{X}i}X_i\varepsilon_i^2\right] \leq \frac{1}{N}E\left[(X_i'M_{\mathbb{X}i}X_i)^2\right]^{1/2}E\left[\varepsilon_i^4\right]^{1/2} = O(1/N),$$

under assumptions A4 and B.

Similarly, for condition (iii), the *iid* assumption in A1 implies that it is sufficient to show that $E[\|\frac{1}{N}X'M_{\mathbb{X}}\overline{R}\|^2] = o(1)$ as $N \to \infty$. To show this, we write the expression as $\frac{1}{N^2}E\left[\left(\sum_{i=1}^{N}X_i'M_{\mathbb{X}i}\overline{R}_i\right)^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[X_i'M_{\mathbb{X}i}\overline{R}_iX_j'M_{\mathbb{X}j}\overline{R}_j\right]$. In contrast to the preceding case, there is cross-sectional dependence between the observations such that, applying the Cauchy-Schwarz inequality,

$$E\left[X_i'M_{\mathbb{X}i}\overline{R}_iX_j'M_{\mathbb{X}j}\overline{R}_j\right] \leq E\left[(X_i'M_{\mathbb{X}i}M_{\mathbb{X}j}'X_j)^2\right]^{1/2}E\left[\overline{R}_i^2\overline{R}_j^2\right]^{1/2}. \tag{A.3}$$

Under assumptions A1 and B, the first term is $O(1)$. To study the convergence of the second term, we have

$$E[\overline{R}_i^2\overline{R}_j^2] = E\left[\left(\sum_{\substack{j=1\\j\neq i}}^{N}R(d_{ij})x_j\right)^2\left(\sum_{\substack{l=1\\l\neq j}}^{N}R(d_{jl})x_l\right)^2\right] \tag{A.4}$$

$$= \sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i}}^{N}\sum_{\substack{j'=1\\j'\neq j}}^{N}\sum_{\substack{l'=1\\l'\neq j}}^{N}E[x_jx_lx_{j'}x_{l'}R(d_{ij})R(d_{il})R(d_{jj'})R(d_{jl'})]. \tag{A.5}$$

Under assumption A1 imposing the independence between the different units, and using similar algebra to the proof of Lemma 2, the leading term is $(N-1)^2E[x_j^2]E[x_l^2]E[R(d_{ij})^2]E[R(d_{il})^2]$, with $E[R(d_{ij})^2] = \sum_{k=1}^{K} \left(\beta^{(q+1)}(c_k)\right)^2 E[(d_{ij}-z_k)^{2(q+1)} \mid 1_k(d_{ij})=1]p_k$, such that

$$
\begin{aligned}
E[\overline{R}_i^2\overline{R}_j^2] &= (N-1)^2E[x_j^2]^2E[R(d_{ij})^2]^2 + o((N-1)^2) \leq C_0(N-1)^2\sum_{k=1}^{K}h^{4(q+1)}p_k^2 + o((N-1)^2) \\
&= C_0(N-1)^2K^{-4q-5} + o((N-1)^2),
\end{aligned}
$$

with $C_0 > 0$ an upper bound of $\max_{k=1,\ldots,K}\{(\beta^{(q+1)}(c_k))^2\}$, and $p_k \asymp K^{-1}$ and $h \asymp K^{-1}$. Therefore, $E\left[X_i'M_{\mathbb{X}i}\overline{R}_iX_j'M_{\mathbb{X}j}\overline{R}_j\right] = O\left(N/K^{2q+5/2}\right)$ such that $\|\frac{1}{N}X'M_{\mathbb{X}}\overline{R}\| = O_p\left(\sqrt{N}/K^{q+5/4}\right)$. Then,

$$
|\widehat{\lambda}-\lambda| = O_p\left(\frac{1}{\sqrt{N}} + \frac{\sqrt{N}}{K^{q+5/4}}\right) = O_p\left(\frac{1}{\sqrt{N}}\right) = o_p(1), \tag{A.6}
$$

under assumption A5.

For the second part of the proof, expression (6) implies that

$$
\widehat{\Gamma}_k - \Gamma_k = \widetilde{Q}_k^{-1}\frac{1}{\alpha_N}\mathbb{X}_k'X(\lambda-\widehat{\lambda})/p_k + \widetilde{Q}_k^{-1}\frac{1}{\alpha_N}\mathbb{X}_k'\overline{R}/p_k + \widetilde{Q}_k^{-1}\frac{1}{\alpha_N}\mathbb{X}_k'\varepsilon/p_k. \tag{A.7}
$$

We study each right hand side term in (A.7) separately. First, using Lemma 2, $\|\widehat{Q}_k^{-1}\| = O_p(1)$. Also,

$$
\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}'x_i(\lambda-\widehat{\lambda})/p_k\| = |\widehat{\lambda}-\lambda|\,\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}'x_i/p_k\|. \tag{A.8}
$$

Note from the analysis above that $|\widehat{\lambda}-\lambda| = O_p(1/\sqrt{N})$. To analyze the asymptotic convergence of the above expression we study $\frac{1}{\alpha_N^2}E\left[\left(\sum_{i=1}^{N}\mathbb{X}_{ki}'x_i/p_k\right)^2\right]$. Under assumption A1, $E[x_jx_k] = 0$ and $E[x_jx_kx_i^2] = 0$, for j,k,i different values, such that the previous expression is equal to $\frac{1}{\alpha_N}B_k/p_k$, with $B_k = E\left[\frac{1}{\alpha_N}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}_{k,ij}'\overline{X}_{k,ij}x_i^2/p_k\right]$. Note that $B_k = E[\overline{X}_{k,ij}'\overline{X}_{k,ij} \mid 1_k(d_{ij})]E[x_i^2]$, that is finite, under assumptions A1 and B1. Then, $\frac{1}{\alpha_N}B_k/p_k = O\left(\frac{K}{N^2}\right)$ such that $\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}'x_i/p_k\| = O_p\left(\frac{\sqrt{K}}{N}\right)$ and $\|\ \widetilde{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}'X_i(\lambda-\widehat{\lambda})/p_k\| = O_p\left(\frac{\sqrt{K}}{N^{3/2}}\right)$.

For the second expression on the right hand side, we use the Cauchy-Schwarz inequality to obtain

$$\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}'_{ki}\overline{R}_i/p_k\|^2 = \|\frac{1}{\alpha_N}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\overline{X}'_{k,ij}\overline{R}_i/p_k\|^2 \leq \left[\|\widetilde{Q}_k - Q_k\| + \|Q_k\|\right]\frac{1}{\alpha_N}\sum_{i=1}^{N}\overline{R}_i^2/p_k.$$

Using Lemmas 1 and 2, $\|\widetilde{Q}_k - Q_k\| = O_p(1/\sqrt{N})$ and, by assumption B, $Q_k = O(1)$. We now study $\frac{1}{\alpha_N}\sum_{i=1}^{N}\overline{R}_i^2/p_k$ to obtain the consistency of the network parameter estimators. A sufficient condition to show this is $\frac{1}{\alpha_N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\overline{R}_i^2\overline{R}_j^2/p_k^2\right]$. This condition is, however, shown in expression (A.4) such that $\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{(N-1)^2}E\left[\overline{R}_i^2\overline{R}_j^2/p_k^2\right] = O(K^{-4q-3})$. Therefore, $\frac{1}{\alpha_N}\sum_{i=1}^{N}\overline{R}_i^2/p_k = O_p\left(K^{-2q-3/2}\right)$ such that $\left(\frac{1}{\alpha_N}\sum_{i=1}^{N}\overline{R}_i^2/p_k\right)^{1/2} = O_p\left(K^{-q-3/4}\right)$. Thus,

$$\|\widetilde{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}'_{ki}\overline{R}_i/p_k\| = O_p(1)\left(O_p(1/\sqrt{N}) + O(1)\right)^{1/2}O_p\left(K^{-q-3/4}\right) = O_p\left(K^{-q-3/4}\right).$$

Finally, we prove that $\|\widetilde{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}'_{ki}\varepsilon_i/p_k\| = O_p\left(\sqrt{K}/N\right)$. To do this, it is sufficient to show that $\frac{1}{\alpha_N^2}E\left[\left(\sum_{i=1}^{N}\mathbb{X}'_{ki}\varepsilon_i/p_k\right)^2\right] = O\left(K/N^2\right)$. Under assumption A4, $E[x_j\varepsilon_i] = 0$ and $E[x_jx_k\varepsilon_i^2] = 0$, for $j \neq k$, such that the previous expression is $\frac{1}{\alpha_N p_k}A_k$, with $A_k = E\left[\frac{1}{\alpha_N}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\overline{X}'_{k,ij}\overline{X}_{k,ij}\varepsilon_i^2/p_k\right]$. Note that $A_k = E[\overline{X}'_{k,ij}\overline{X}_{k,ij}\varepsilon_i^2]/p_k$, that is finite, under assumptions A4 and B. Then, $\frac{1}{\alpha_N p_k}A_k = O\left(\frac{K}{N^2}\right)$ such that $\|\widetilde{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}'_{ki}\varepsilon_i/p_k\| = O_p\left(\frac{\sqrt{K}}{N}\right)$. Putting together the different expressions, and for $q \geq 1$ fixed, we obtain

$$\|\widehat{\Gamma}_k - \Gamma_k\| = O_p\left(\frac{\sqrt{K}}{N^{3/2}}\right) + O_p\left(K^{-q-3/4}\right) + O_p\left(\frac{\sqrt{K}}{N}\right) = O_p\left(\frac{\sqrt{K}}{N}\right), \quad \text{(A.9)}$$

under the conditions in Assumption A5. More specifically, the above convergence rate holds if $K^{-q-3/4}/(K^{1/2}/\sqrt{\alpha_N}) \to 0$ as $K, N \to \infty$. This condition is guaranteed by assumption A5. □

Proof of Proposition 2. First, we note that

$$\widehat{\lambda} - \lambda = \widehat{\Phi}^{-1}\frac{1}{N}\widehat{X}_u'\varepsilon + \widehat{\Phi}^{-1}\frac{1}{N}\widehat{X}_u'\overline{R}, \tag{A.10}$$

using the property $\widehat{X}_u'\mathbb{X} = X'M_{\mathbb{X}}\mathbb{X} = \mathbf{0}$ with $\mathbf{0}$ a $(p+1)\times K(q+1)$ matrix of zeros. To derive the asymptotic normality of the standardized parameter estimator, note from (A.10) that

$$\sqrt{N}\left(\widehat{\lambda} - \lambda\right) = \widehat{\Phi}^{-1}\frac{1}{\sqrt{N}}X'M_{\mathbb{X}}\varepsilon + \widehat{\Phi}^{-1}\frac{1}{\sqrt{N}}X'M_{\mathbb{X}}\overline{R}.$$

Therefore, we need to show that $\|\frac{1}{\sqrt{N}}X'M_{\mathbb{X}}\overline{R}\| = o_p(1)$ as $N\to\infty$. For this, it is sufficient to note from condition (iii) of Proposition 1 that $\|\frac{1}{N}X'M_{\mathbb{X}}\overline{R}\| = O\left(\sqrt{N}/K^{q+5/4}\right)$, for q fixed. Then, $\|\frac{1}{\sqrt{N}}X'M_{\mathbb{X}}\overline{R}\| = O\left(N/K^{q+5/4}\right) = o_p(1)$, under assumption A5.

Now, applying the central limit theorem to the above expression, we obtain

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i'M_{\mathbb{X}}\varepsilon_i \to N(0,\Psi_0), \tag{A.11}$$

with $\Psi_0 = E\left[(X_i'M_{\mathbb{X}}\varepsilon_i)^2\right]$. Furthermore, under assumptions A1 and B, and applying the law of large numbers, $\widehat{\Phi} = \frac{1}{N}\sum_{i=1}^{N}X_i'M_{\mathbb{X}}X_i$ is a consistent estimator of $\Phi_0 = E[X'M_{\mathbb{X}}X]$ such that $\|\widehat{\Phi} - \Phi_0\| = o_P(1)$, with $\|\Phi_0\| < \infty$. With these results, we obtain the asymptotic convergence in distribution:

$$\sqrt{N}\left(\widehat{\lambda} - \lambda\right) \to N(0, \Phi_0^{-1}\Psi_0\Phi_0^{-1}). \tag{A.12}$$

□

Proof of Theorem 1. To prove this result we combine expressions (2) and (7), and apply the triangular inequality, such that

$$\sup_{d\in\chi}|\widehat{w}(d) - w(d)| \leq \sup_{d\in\chi}\left|\sum_{k=1}^{K}\left(\widehat{\Gamma}_k - \Gamma_k\right)' v_k(d)1_k(d)\right| + \sup_{d\in\chi}|R(d)|.$$

For the first term, we note that

$$\sup_{d\in\chi}\left|\sum_{k=1}^{K}\left(\widehat{\Gamma}_k - \Gamma_k\right)' v_k(d)1_k(d)\right| \leq \max_{\{k=1,\ldots,K\}}\left\{\sup_{d\in\chi}\left|(\widehat{\Gamma}_k - \Gamma_k)'v_k(d)1_k(d)\right|\right\},$$

given that $1_k(d) = 1$ for one interval of the partition only.

Furthermore, applying the triangular inequality and the partition of the compact space into K disjoint intervals, the quantity on the right hand side of the above expression is bounded by $\max_{\{k=1,\ldots,K\}} \|\widehat{\Gamma}_k - \Gamma_k\| \max_{\{k=1,\ldots,K\}} \left\{ \sup_{d\in\chi} |v_k(d) 1_k(d)| \right\}$. To prove that $\max_{\{k=1,\ldots,K\}} \|\widehat{\Gamma}_k - \Gamma_k\|$ is bounded in probability, we need to show that $\max_{\{k=1,\ldots,K\}} \|\widehat{Q}_k - Q_k\|$ satisfies this condition. Lemma A.4 in Cattaneo and Farrell (2013) shows this condition for $K = (N/logN)^{\xi}$, with $\xi \in (0,1)$. More generally, using the result in Proposition 1, we obtain that $P\left\{ \max_{\{k=1,\ldots,K\}} \|\widehat{\Gamma}_k - \Gamma_k\| > \left(\frac{\sqrt{K}}{N}\right)^{1/2} \epsilon \right\}$ is arbitrarily small for ϵ large enough. This result is obtained applying Boole's inequality, Bernstein's inequality, and $p_k \asymp K^{-1}$, see proof of Theorem 2 in Cattaneo and Farrell (2013) for additional details. Therefore,

$$\max_{\{k=1,\ldots,K\}} \|\widehat{\Gamma}_k - \Gamma_k\| \leq O_p\left(\frac{\sqrt{K}}{N}\right).$$

Additionally,

$$\max_{\{k=1,\ldots,K\}} \sup_{d\in\chi} |v_k(d) 1_k(d)| \leq \max_{\{k=1,\ldots,K\}} \sum_{m=0}^{q} h_k^m = \max_{\{k=1,\ldots,K\}} \left\{ \frac{1-h_k^{q+1}}{1-h_k} \right\} = \frac{1-h^{q+1}}{1-h},$$

for $h = h_1, \ldots, h_K$. This result is bounded for $0 < h < 1$.

For the second term, $R(d) = \sum_{k=1}^{K} w^{(q+1)}(c_k)(d-z_k)^{q+1} 1_k(d)$. Then, $\sup_{d\in\chi} |R(d)| \leq C_0 \max_{\{k=1,\ldots,K\}} \{h_k^{q+1}\}$, with C_0 a positive constant satisfying that $\max_{\{k=1,\ldots,K\}} |w^{(q+1)}(c_k)| \leq C_0$. Therefore, $\sup_{d\in\chi} |R(d)| = O_p(K^{-(q+1)})$. Then,

$$\sup_{d\in\chi} |\widehat{w}(d) - w(d)| = O_p\left(\frac{\sqrt{K}}{N} + K^{-(q+1)}\right).$$

□

Proof of Lemma 3. We proceed to show the asymptotic convergence between the estimator $\widehat{A}_k = \frac{1}{\alpha_N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} \overline{X}'_{ij} \overline{X}_{ij} e_i^2 / p_k$ and $A_k = E\left[\overline{X}'_{k,ij} \overline{X}_{k,ij} \varepsilon_i^2 / p_k\right]$. To obtain the asymptotic convergence it is sufficient to show that $E[\|\widehat{A}_k - A_k\|^2] = o_p(1)$. Note that

$$E[\|\widehat{A}_k - A_k\|^2] = E\left[\| \frac{1}{\alpha_N} \left(\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} \overline{X}'_{k,ij} \overline{X}_{k,ij} e_i^2 - E\left[\overline{X}'_{k,ij} \overline{X}_{k,ij} \varepsilon_i^2\right] \right) / p_k \|^2 \right].$$

Using the triangular inequality and further algebra, this expression is bounded by

$$E\left[\|\frac{1}{\alpha_N}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}\overline{X}_{k,ij}(e_i^2-\varepsilon_i^2)\right)/p_k\|^2\right]+E\left[\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\left(\overline{X}'_{k,ij}\overline{X}_{k,ij}\varepsilon_i^2-E\left[\overline{X}'_{k,ij}\overline{X}_{k,ij}\varepsilon_i^2\right]\right)/p_k\|^2\right].$$

For the first expression, we note that $e_i=\varepsilon_i+x_i(\lambda-\widehat{\lambda})+\sum_{k=1}^{K}\mathbb{X}_{ki}(\Gamma_k-\widehat{\Gamma}_k)+\overline{R}_i$, such that applying the Cauchy-Schwarz inequality, the asymptotic convergence of $\widehat{Q}_k$, and the convergence of the parameter estimators in Proposition 1, the expression converges to zero in probability as $K,N\to\infty$. For the second expression, using the same steps as in Lemma 2, the conditional zero-mean error term in A4 implies that $E\left[\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\left(\overline{X}'_{k,ij}\overline{X}_{k,ij}\varepsilon_i^2-E\left[\overline{X}'_{k,ij}\overline{X}_{k,ij}\varepsilon_i^2\right]\right)/p_k\|^2\right]=$

$$\sum_{r=0}^{q}\sum_{s=0}^{q}E\left[\frac{1}{\alpha_N^2}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\left(x_j^2\varepsilon_i^2(d_{ij}-z_k)^{r+s}1_k(d_{ij})-E\left[x_j^2\varepsilon_i^2(d_{ij}-z_k)^{r+s}1_k(d_{ij})\right]\right)/p_k\right)^2\right].$$

After tedious algebra, the preceding expression can be written as

$$\frac{1}{N^2(N-1)^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i}}^{N}E[x_j^2x_l^2\varepsilon_i^4]E[(d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s}1_k(d_{ij})1_k(d_{il})]/p_k^2\right)$$
$$+\frac{1}{N^2(N-1)^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i}}^{N}\sum_{\substack{l'=1\\l'\neq j}}^{N}E[x_l^2x_{l'}^2\varepsilon_i^2\varepsilon_j^2]E[(d_{il}-z_k)^{2r}(d_{il'}-z_k)^{2s}1_k(d_{il})1_k(d_{il'})]/p_k^2\right)$$
$$-\frac{1}{N^2(N-1)^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i}}^{N}\sum_{\substack{l'=1\\l'\neq j}}^{N}E[x_l^2\varepsilon_i^2]E[x_{l'}^2\varepsilon_j^2]E[(d_{il}-z_k)^{2r}(d_{il'}-z_k)^{2s}1_k(d_{il})1_k(d_{il'})]/p_k^2\right).$$

Under assumptions A1 and A4, $E[x_l^2x_{l'}^2\varepsilon_i^2\varepsilon_j^2]=E[x_l^2\varepsilon_i^2]E[x_{l'}^2\varepsilon_j^2]$. Then, the above expression is equal to

$$\frac{1}{N^2(N-1)^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}E[x_j^4\varepsilon_i^4]E[(d_{ij}-z_k)^{2(r+s)}1_k(d_{ij})]/p_k^2\right)$$

$$+\frac{1}{N^2(N-1)^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i,j}}^{N}E[x_j^2x_l^2\varepsilon_i^4]E[(d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s}1_k(d_{ij})1_k(d_{il})]/p_k^2\right)$$

$$=\frac{1}{N(N-1)}\sum_{r=0}^{q}\sum_{s=0}^{q}E[x_j^4\varepsilon_i^4]E[(d_{ij}-z_k)^{2(r+s)}\mid 1_k(d_{ij})1_k(d_{il})]/p_k$$

$$+\frac{1}{N}\sum_{r=0}^{q}\sum_{s=0}^{q}E[x_j^2x_l^2\varepsilon_i^4]E[(d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s}\mid 1_k(d_{ij})1_k(d_{il})]$$

$$\leq O\left(\frac{K}{N^2}\right)+\frac{C_0}{N}\sum_{r=0}^{q}\sum_{s=0}^{q}h^{2(r+s)}=O\left(\frac{K}{N^2}\right)+\frac{C_0}{N}\left(\frac{1-h^{2(q+1)}}{1-h^2}\right)^2=O\left(\frac{1}{N}\right),$$

with C_0 some positive constant that reflects the finite character of the first four moments of x_i and ε_i imposed in A1 and A4. Note also that $p_k \asymp K^{-1}$, therefore, we obtain $E[||\widehat{A}_k-A_k||^2]=O\left(\frac{1}{N}\right)$ such that $||\widehat{A}_k-A_k||=O_p\left(\frac{1}{\sqrt{N}}\right)$.

□

Proof of Lemma 4. Expression (A.7) implies that

$$\widehat{\Gamma}_k-\Gamma_k \quad =\widehat{Q}_k^{-1}\tfrac{1}{\alpha_N p_k}\sum_{i=1}^{N}\mathbb{X}'_{ki}\varepsilon_i-\widehat{Q}_k^{-1}\tfrac{1}{\alpha_N p_k}\sum_{i=1}^{N}\mathbb{X}'_{ki}x_i(\widehat{\lambda}-\lambda)+O_p\left(K^{-q-3/4}\right).$$

Then,

$$E[(\widehat{\Gamma}_k-\Gamma_k)^2]=\frac{1}{p_k}\widehat{Q}_k^{-1}\left(\frac{1}{\alpha_N^2}\sum_{i=1}^{N}E\left[\mathbb{X}'_{ki}\mathbb{X}_{ki}\varepsilon_i^2\right]/p_k\right)\widehat{Q}_k^{-1} \tag{A.13}$$

$$+\frac{1}{p_k}\widehat{Q}_k^{-1}\left(\frac{1}{\alpha_N^2}\sum_{i=1}^{N}E\left[\mathbb{X}'_{ki}x_i(\widehat{\lambda}-\lambda)^2x'_i\mathbb{X}_{ki}\right]/p_k\right)\widehat{Q}_k^{-1} \tag{A.14}$$

$$-\frac{1}{p_k}\widehat{Q}_k^{-1}\left(\frac{1}{\alpha_N^2}\sum_{i=1}^{N}E\left[\mathbb{X}'_{ki}\varepsilon_i(\widehat{\lambda}-\lambda)x'_i\mathbb{X}_{ki}\right]/p_k\right)\widehat{Q}_k^{-1}+O_p\left(K^{-2q-3/2}\right). \tag{A.15}$$

Using Lemma 2 and the definition of A_k, expression (A.13) is equal to $\frac{1}{\alpha_N p_k}Q_k^{-1}A_kQ_k^{-1}+O_p(1/N)$. Similarly, we use the result in Proposition 2 such that expression (A.14) is

$\frac{1}{N\alpha_N^2 p_k}\widehat{Q}_k^{-1}\left(\sum_{i=1}^{N} E\left[\left(\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}x_i\right)\widehat{\Phi}^{-1}\Psi_0\widehat{\Phi}^{-1}\left(\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}x_i\right)'\right]/p_k\right)\widehat{Q}_k^{-1}$. Furthermore, using the consistency of $\widehat{\Phi}$ to Φ_0 and the definition of B_k in the proof of Proposition 1, we note that $\frac{1}{\alpha_N}\sum_{i=1}^{N} E\left[\left(\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}x_i\right)\widehat{\Phi}^{-1}\Psi_0\widehat{\Phi}^{-1}\left(\sum_{\substack{j=1\\j\neq i}}^{N}\overline{X}'_{k,ij}x_i\right)'\right]/p_k = \Phi_0^{-1}\Psi_0\Phi_0^{-1}B_k + o_p(1)$, with $\|B_k\| < \infty$ as shown above. Then,

$$\|\widehat{Q}_k^{-1}\left(\frac{1}{\alpha_N^2 p_k}\sum_{i=1}^{N} E\left[\mathbb{X}'_{ki}X_i(\widehat{\lambda}-\lambda)^2 X_i'\mathbb{X}_{ki}\right]/p_k\right)\widehat{Q}_k^{-1}\| = O_p\left(\frac{K}{N^{3/2}}\right).$$

The asymptotic convergence of expression (A.15) is studied in a similar fashion. More specifically, replacing expression (A.10):

$$\begin{aligned}\frac{1}{\alpha_N^2 p_k^2}\sum_{i=1}^{N} E\left[\mathbb{X}'_{ki}\varepsilon_i(\widehat{\lambda}-\lambda)x_i'\mathbb{X}_{ki}\right] &= \frac{1}{\alpha_N^2 p_k}\left(E\left[\mathbb{X}'_{ki}\varepsilon_i\widehat{\Phi}^{-1}X'M_{\mathbb{X}}\varepsilon x_i'\mathbb{X}_{ki}\right]/p_k\right)\\ &+ \frac{1}{\alpha_N^2 p_k^2}\left(E\left[\mathbb{X}'_{ki}\varepsilon_i\widehat{\Phi}^{-1}X'M_{\mathbb{X}}\overline{R}X_i'\mathbb{X}_{ki}\right]/p_k\right),\end{aligned}$$

such that $\|\frac{1}{\alpha_N^2 p_k^2}\sum_{i=1}^{N} E\left[\mathbb{X}'_{ki}\varepsilon_i(\widehat{\lambda}-\lambda)x_i'\mathbb{X}_{ki}\right]\| = 0$, with $E\left[\mathbb{X}'_{ki}\varepsilon_i\widehat{\Phi}^{-1}X'M_{\mathbb{X}}\varepsilon x_i'\mathbb{X}_{ki}\right]/p_k = 0$, under the assumption $E[x_i] = 0$ in A1, and $E\left[\mathbb{X}'_{ki}\varepsilon_i\widehat{\Phi}^{-1}X'M_{\mathbb{X}}\overline{R}X_i'\mathbb{X}_{ki}\right]/p_k = 0$, by assumption $E[\varepsilon_i] = 0$ in A4.

Thus, putting together the above expressions, the variance of the spatial parameter estimator is

$$E[(\widehat{\Gamma}_k - \Gamma_k)^2] = \frac{1}{\alpha_N p_k} Q_k^{-1}A_kQ_k^{-1} + O_p\left(\frac{K}{N^{3/2}}\right).$$

□

Proof of Proposition 3. Let $V_K(d) \equiv \alpha_N\sum_{k=1}^{K} v_k'(d)V(\widehat{\Gamma}_k)v_k(d)1_k(d)$. Lemmas 2 and 3 imply that $V_K(d) = \sum_{k=1}^{K} v_k'(d)Q_k^{-1}A_kQ_k^{-1}v_k(d)1_k(d)/p_k + O_p\left(\frac{K}{N}\right)$. Similarly, using expression (10), we obtain $\widehat{V}_K(d) = \sum_{k=1}^{K} v_k'(d)\widehat{Q}_k^{-1}\widehat{A}_k\widehat{Q}_k^{-1}v_k(d)1_k(d)/p_k$. Using the convergence results in Lemmas 2 and 3, it follows that $|\widehat{V}_K(d) - V_K(d)| = O_p\left(\frac{K}{N}\right)$, as $N \to \infty$, that proves result (i).

To show result (ii), we put together expressions (2) and (7), and obtain

$$\sqrt{\alpha_N}(\widehat{w}(d) - w(d)) = \sum_{k=1}^{K}\sqrt{\alpha_N}\left(\widehat{\Gamma}_k - \Gamma_k\right)' v_k(d)1_k(d) - \sqrt{\alpha_N}R(d). \tag{A.16}$$

Using the result in Lemma 4, we have

$$V\left(\sqrt{\alpha_N}(\widehat{w}(d)-w(d))\right) = V_K(d) + \alpha_N\ V(R(d)) \\ - \alpha_N \sum_{k=1}^{K} v_k'(d) Cov((\widehat{\Gamma}_k - \Gamma_k)1_k(d), R(d)) v_k(d),$$

with $R(d) = \sum_{k=1}^{K} w^{(q+1)}(c_k)(d-z_k)^{q+1}1_k(d)$. Thus, the variance of the remainder term satisfies

$$\begin{aligned} V(R(d)) &= \sum_{k=1}^{K} (w^{(q+1)}(c_k))^2 \left(E[(d-z_k)^{2(q+1)} \mid 1_k(d)=1]p_k - E[(d-z_k)^{q+1} \mid 1_k(d)=1]^2 p_k^2\right) \\ &= \sum_{k=1}^{K} (w^{(q+1)}(c_k))^2 \left[V\left((d-z_k)^{q+1} \mid 1_k(d)=1\right) p_k + E[(d-z_k)^{q+1} \mid 1_k(d)=1]^2 p_k(1-p_k)\right] \\ &\le 2C_0 \sum_{k=1}^{K} h^{2(q+1)} p_k - C_0 \sum_{k=1}^{K} h^{2(q+1)} p_k^2 = O(K^{-2(q+1)}) + O(K^{-2q-3}), \end{aligned}$$

given that $p_k \asymp 1/K$ and $\max_{k=1,\ldots,K} (\beta^{(q+1)}(c_k))^2 \le C_0$.

Similar tedious calculations for the covariance term yield the same convergence $O(K^{-2(q+1)})$ as above, and we obtain

$$V\left(\sqrt{\alpha_N}(\widehat{w}(d)-w(d))\right) = V_K(d) + O(N^2/K^{2(q+1)}),$$

with $\alpha_N/K^{2(q+1)} \to 0$ under assumption A5. Finally, we note that $V_K(d) = O(K)$, by construction.

□

Proof of Theorem 2. To show the asymptotic distribution in this theorem, note from expressions (A.7) and (A.9) that $\widehat{\Gamma}_k - \Gamma_k = \widehat{Q}_k^{-1} \frac{1}{\alpha_N} \sum_{i=1}^{N} \mathbb{X}_{ki}' \varepsilon_i / p_k + O_p\left(\frac{\sqrt{K}}{N^{3/2}}\right)$. Therefore, using expression (A.16), we obtain

$$\sqrt{\alpha_N}(\widehat{w}(d)-w(d)) = \frac{1}{\sqrt{\alpha_N}} \sum_{i=1}^{N} \sum_{k=1}^{K} v_k(d)' \widehat{Q}_k^{-1} \mathbb{X}_{ki}' \varepsilon_i 1_k(d)/p_k + O_p\left(\frac{\sqrt{K}}{\sqrt{N}}\right) - \sqrt{\alpha_N} R(d). \tag{A.17}$$

Let $z_{iN}(d) = \frac{\sum_{k=1}^{K} v_k(d)' Q_k^{-1} \mathbb{X}_{ki}' \varepsilon_i 1_k(d)/p_k}{\alpha_N^{1/2} V_k^{1/2}}$, with $V_K = \sum_{k=1}^{K} v_k(d)' Q_k^{-1} A_k Q_k^{-1} v_k(d) 1_k(d)/p_k$. The process $\{z_{iN}(d)\}_{i=1}^{N}$ inherits the properties of the error term ε_i, by assumption A4, such that $E[z_{iN}(d) \mid X, D] = 0$ and $E[z_{iN}^2(d) \mid X, D] = 1$. Thus,

$$\frac{\sqrt{\alpha_N}(\widehat{w}(d)-w(d))}{V_k^{1/2}} = \sum_{i=1}^{N} z_{iN}(d) - \frac{\sqrt{\alpha_N}R(d)}{V_k^{1/2}} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

with $V_K^{1/2} = O(\sqrt{K})$, by Proposition 3. Furthermore, the proof of Theorem 1 shows that $|R(d)| = O_p(K^{-(q+1)})$ such that $\frac{\sqrt{\alpha_N}|R(d)|}{V_k^{1/2}} = O_p\left(N/K^{q+3/2}\right)$. Therefore, by assumption A5, this quantity converges to zero in probability, such that

$$\frac{\sqrt{\alpha_N}(\widehat{w}(d)-w(d))}{V_k^{1/2}} = \sum_{i=1}^{N} z_{iN}(d) + O_p\left(\frac{N}{K^{q+3/2}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

The quantity $\sum_{i=1}^{N} z_{iN}(d)$ is of order $O_p(1)$. To show this, note from the proof of Proposition 1 that $\|\widehat{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}'\varepsilon_i/p_k\| = O_p\left(\frac{\sqrt{K}}{N}\right)$. Then, $\sum_{i=1}^{N} z_{iN}(d) = O_p(1)$, given that $\|\widehat{Q}_k^{-1}\frac{1}{\sqrt{\alpha_N}}\sum_{i=1}^{N}\mathbb{X}_{ki}'\varepsilon_i/p_k^2\| = O_p\left(\sqrt{K}\right)$ and $V_K^{1/2} = O(\sqrt{K})$.

It remains to see the asymptotic distribution of the standardized estimator. To do this we note that z_{iN} is a triangular array, and apply a Lindeberg-Levy central limit theorem to $\sum_{i=1}^{N} z_{iN}(d)$. More formally, we need to verify the Lindeberg condition

$$\sum_{i=1}^{N} E[z_{iN}^2(d)1(|z_{iN}(d)|>\delta)\mid X,D] \xrightarrow{p} 0,$$

for any $\delta > 0$. This condition can be represented as

$$\sum_{i=1}^{N} E\left[\left(\frac{\sum_{k=1}^{K} v_k(d)'Q_k^{-1}\mathbb{X}_{ki}'\varepsilon_i 1_k(d)/p_k}{\alpha_N^{1/2}V_k^{1/2}}\right)^2 1\left(\left|\frac{\sum_{k=1}^{K} v_k(d)'Q_k^{-1}\mathbb{X}_{ki}'\varepsilon_i 1_k(d)/p_k}{\alpha_N^{1/2}V_k^{1/2}}\right| > \delta\right) \mid X,D\right] \xrightarrow{p} 0,$$

for any $\delta > 0$. Applying Hölder's inequality, the above expression is bounded by

$$\sum_{i=1}^{N} E\left[\left(\frac{\sum_{k=1}^{K} v_k(d)'Q_k^{-1}\mathbb{X}_{ki}'\varepsilon_i 1_k(d)/p_k}{\alpha_N^{1/2}V_K^{1/2}}\right)^{2+\eta} \mid X,D\right]^{\frac{2}{2+\eta}} \left[P\left(\left|\frac{\sum_{k=1}^{K} v_k(d)'Q_k^{-1}\mathbb{X}_{ki}'\varepsilon_i 1_k(d)/p_k}{\alpha_N^{1/2}V_K^{1/2}}\right| > \delta \mid X,D\right)\right]^{\frac{\eta}{2+\eta}}.$$

Now, using Markov's inequality, we have the following upper bound:

$$N\left[\frac{E\left[\left|\sum_{k=1}^{K} v_k(d)'Q_k^{-1}\mathbb{X}_{ki}'\varepsilon_i 1_k(d)/p_k\right|^{2+\eta} \mid X,D\right]}{\alpha_N^{1+\frac{\eta}{2}}V_K^{1+\frac{\eta}{2}}}\right]^{\frac{2}{2+\eta}} \left[\frac{E\left[\left|\sum_{k=1}^{K} v_k(d)'Q_k^{-1}\mathbb{X}_{ki}'\varepsilon_i 1_k(d)/p_k\right|^{2+\eta} \mid X,D\right]}{\delta^{2+\eta}\alpha_N^{1+\frac{\eta}{2}}V_K^{1+\frac{\eta}{2}}}\right]^{\frac{\eta}{2+\eta}}$$

$$\leq \frac{E\left[\left|\sum_{k=1}^{K} v_k(d)' Q_k^{-1} \mathbb{X}_{ki}' \varepsilon_i 1_k(d)/p_k\right|^{2+\eta} \mid X, D\right]}{\delta^\eta N^{\eta/2} (N-1)^{\frac{2+\eta}{2}} V_k^{1+\eta/2}} = \frac{\sum_{k=1}^{K} \|v_k'(d) Q_k^{-1}\|^{2+\eta} A_{k\eta} 1_k(d)}{\delta^\eta N^{\eta/2} V_k^{1+\eta/2}},$$

with $A_{k\eta} = E\left[\left\|\frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N} \overline{X}_{k,ij}' \overline{X}_{k,ij} \varepsilon_i^2 / p_k\right\|^{\frac{2+\eta}{2}} |X, D\right]$. Note that $A_{k\eta}^{\frac{2}{2+\eta}} \asymp V_k$, and therefore, such that $A_{k\eta}/V_k^{1+\frac{\eta}{2}} = O(1)$. Therefore, under assumption A5, the above expression satisfies

$$\frac{\sum_{k=1}^{K} \|v_k'(d) Q_k^{-1})\|^{2+\eta} A_{k\eta} 1_k(d)}{\delta^\eta N^{\eta/2} V_k^{1+\eta/2}} \asymp \frac{1}{N^{\eta/2}} \to 0, \text{ as } N \to \infty, \text{ for } \eta > 0.$$

Therefore, the central limit theorem applies such that $\frac{\sqrt{\alpha_N}(\widehat{w}(d) - w(d))}{V_k^{1/2}} \xrightarrow{d} N(0,1)$, for $d \in \chi$ fixed. Furthermore, the result $|\widehat{V}_k(d) - V_k(d)| = o_p(1)$ in Proposition 3 implies that $\frac{\widehat{V}_k(d)}{V_k(d)} \xrightarrow{p} 1$, for all $d \in \chi$ such that for $V_k(d) \neq 0$, we obtain

$$\frac{\sqrt{\alpha_N}\,(\widehat{w}(d) - w(d))}{\widehat{V}_k^{1/2}} \xrightarrow{d} N(0,1), \text{ for } \ d \in \chi, \ \text{ as } \ N \to \infty.$$

□

References

Agarwal, S., Qian, W., Zou, X., 2021. Thy neighbor's misfortune: Peer effect on consumption. American Economic Journal: Economic Policy 13, 1–25. doi:10.1257/pol.20170634.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856. doi:10.2307/2951764.

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414. doi:10.2307/2951753.

Anselin, L., 1988. Spatial econometrics: Methods and models. Studies in Operational Regional Science, Springer Verlag, Heidelberg, Germany.

Bhattacharjee, A., Jensen-Butler, C., 2013. Estimation of the spatial weights matrix under structural constraints. Regional Science and Urban Economics 43, 617–634. doi:10.1016/j.regsciurbeco.2013.03.005.

Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. Journal of Econometrics 150, 41–55. doi:10.1016/j.jeconom.2008.12.021.

Cai, Z., Li, Q., 2008. Nonparametric estimation of varying coefficient dynamic panel data models. Econometric Theory 24, 1321–1342. doi:10.2307/20142543.

Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. Journal of Econometrics 167, 413–425. doi:10.1016/j.jeconom.2011.09.025.

Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. Journal of the American Statistical Association 103, 1595–1608. doi:10.1198/016214508000000977.

Cattaneo, M.D., Farrell, M.H., 2013. Optimal convergence rates, bahadur representation, and asymptotic normality of partitioning estimators. Journal of Econometrics 174, 127–143. doi:10.1016/j.jeconom.2013.02.002.

Cattaneo, M.D., Farrell, M.H., Feng, Y., 2020. Large sample properties of partitioning-based series estimators. The Annals of Statistics 48, 1718–1741. doi:10.1214/19-AOS1865.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association 83, 596–610. doi:10.1080/01621459.1988.10478639.

Corrado, L., Fingleton, B., 2012. Where is the economics in spatial econometrics? Journal of Regional Science 52, 210–239. doi:10.1111/j.1467-9787.2011.00726.x.

Craven, P., Wahba, G., 1978. Smoothing noisy data with spline functions. estimating the correct degree of smoothing by the method of generalized cross-validation. Numerische Mathematik 31, 377–404. doi:10.1007/BF01404567.

Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247–254. doi:10.1093/biomet/64.2.247.

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternatives. Biometrika 74, 33–43. doi:10.2307/2336019.

De Giorgi, G., Frederiksen, A., Pistaferri, L., 2020. Consumption network effects. The Review of Economic Studies 87, 130–163. doi:10.1093/restud/rdz026.

Fan, J., Gijbels, I., 1996. Local polynomial modelling and its applications. Monographs on Statistics and Applied Probability, Chapman & Hall/CRC Press, London, United Kingdom.

Fischer, M.M., Wang, J., 2011. Spatial data analysis: Models, methods and techniques. Springer Briefs in Regional Science, Springer-Verlag, Berlin, Germany. doi:10.1007/978-3-642-21720-3.

Gibbons, S., Overman, H.G., 2012. Mostly pointless econometrics? Journal of Regional Science 52, 172–191. doi:10.1111/j.1467-9787.2012.00760.x.

Halleck Vega, S., Elhorst, J.P., 2015. The slx model. Journal of Regional Science 55, 339–363. doi:10.1111/jors.12188.

Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430. doi:10.2307/2171789.

Higgins, A., Martellosio, F., 2022. Shrinkage estimation of network spillovers with factor structured errors. Journal of Econometrics (In press). doi:10.1016/j.jeconom.2021.11.017.

Kelejian, H.H., 2008. A spatial j-test for model specification against a single or a set of non-nested alternatives. Letters in Spatial and Resource Sciences 1, 3–11. doi:10.1007/s12076-008-0001-9.

Kelejian, H.H., Piras, G., 2011. An extension of kelejian's j-test for non-nested spatial models. Regional Science and Urban Economics 41, 281–292. doi:10.1016/j.regsciurbeco.2011.01.002.

Kelejian, H.H., Prucha, I.R., 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. doi:10.1023/A:1007707430416.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533. doi:10.1111/1468-2354.00027.

Koroglu, M., Sun, Y., 2016. Functional-coefficient spatial durbin models with nonparametric spatial weights: An application to economic growth. Econometrics 4, 6. doi:10.3390/econometrics4010006.

Lam, C., Souza, P.C.L., 2020. Estimation and selection of spatial weight matrix in a spatial lag model. Journal of Business & Economic Statistics 38, 693–710. doi:10.1080/07350015.2019.1569526.

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

Levinson, A., O'Brien, J., 2019. Environmental engel curves: Indirect emissions of common air pollutants. The Review of Economics and Statistics 101, 121–133. doi:10.1162/rest_a_00736.

Li, K.C., 1987. Asymptotic optimality for c_p, c_l, cross-validation and generalized cross-validation: Discrete index set. Annals of Statistics 15, 958–975. doi:10.1214/aos/1176350486.

Li, Q., Racine, J.S., 2007. Nonparametric econometrics: Theory and practice. Princeton University Press, Princeton NJ, United States.

Mallows, C.L., 1973. Some comments on c_p. Technometrics 15, 661–675. doi:10.2307/1267380.

McMillen, D.P., 2012. Perspectives on spatial econometrics: Linear smoothing with structured models. Journal of Regional Science 52, 192–209. doi:10.1111/j.1467-9787.2011.00746.x.

Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168. doi:10.1016/S0304-4076(97)00011-0.

Partridge, M.D., Boarnet, M., Brakman, S., Ottaviano, G., 2012. Introduction: Whither spatial econometrics? Journal of Regional Science 52, 167–172. doi:10.1111/j.1467-9787.2012.00767.x.

Pinkse, J., Slade, M.E., Brett, C., 2002. Spatial price competition: A semiparametric approach. Econometrica 70, 1111–1153. doi:10.1111/1468-0262.00320.

Rose, C.D., 2017. Identification of peer effects through social networks using variance restrictions. The Econometrics Journal 20, S47–S60. doi:10.1111/ectj.12101.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society. Series B (Methodological) 36, 111–147. doi:10.2307/2984809.

Sun, Y., 2016. Functional-coefficient spatial autoregressive models with nonparametric spatial weights. Journal of Econometrics 195, 134–153. doi:10.1016/j.jeconom.2016.07.005.

Wahba, G., 1985. A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. The Annals of Statistics 13, 1378–1402. doi:10.2307/2241360.

Wheeler, D.C., Páez, A., 2010. Geographically weighted regression, in: Handbook of Applied Spatial Analysis, pp. 461–486.

Wolske, K.S., Gillingham, K.T., Schultz, P.W., 2020. Peer influence on household energy behaviours. Nature Energy 5, 202–212. doi:10.1038/s41560-019-0541-9.