# A nonparametric network regression model using partitioning estimators

Jose Olmo*		Marcos Sanso-Navarro†

**Abstract**

This paper extends the spatial lag of the exogenous regressor model (SLX) in two dimensions. First, we consider a nonparametric model in which the spatial effects are modelled as a functional coefficient. This coefficient is approximated using Taylor expansions of arbitrary (finite) order over a set of disjoint intervals covering the support of the spatial variable. Second, by considering the spatial variable to be stochastic and different from geographical distance, we extend the model to a network setting. The model is also extended to incorporate endogenous spatial/network effects in the spirit of nonparametric SAR models. Estimation of the nonparametric SLX model is based on the theory on sieve regression and partitioning estimators. Estimation of the endogenous version is based on GMM. The asymptotic properties of the partitioning estimator of the functional network coefficient for the SLX model are derived. We also propose pointwise and uniform tests for the presence of network effects for this model. The empirical application studies environmental Engel curves and finds strong evidence of neighboring effects in the relationship between households' income and the amount of pollution embodied in the goods and services they consume.

**Key Words:** Network regression, series estimators, interaction matrix, asymptotic theory, Environmental Engel curves.

*Universidad de Zaragoza and University of Southampton. Corresponding address: Departamento de Análisis Económico, Universidad de Zaragoza. Gran Vía 2, 50005, Zaragoza. Spain.

†Departamento de Análisis Económico, Universidad de Zaragoza. Gran Vía 2, 50005, Zaragoza. Spain.
The authors acknowledge financial support from Gobierno de Aragón (ADETRE Research Group, Grant S39-20R). Jose Olmo also acknowledges financial support from project PID2019-104326GB-I00 from Ministerio de Ciencia e Innovación and from Fundación Agencia Aragonesa para la Investigación y el Desarrollo (ARAID).

# 1 Introduction

Spillovers between units in a cross section are of main interest in spatial and network models. They can be defined as the impact of changes to explanatory variables in a particular unit on the dependent variable measured at other units. Spillovers - interpreted as exogenous interactions in the explanatory variables - is one of the three types of interactions across units in a cross section of observations. The other two types are defined by (*ii*) endogenous interactions affecting the dependent variable and (*iii*) 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 considering exogenous interactions is usually specified as a 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 $\beta$ 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 AR model, SAR, 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 $\theta$ captures the spatial correlation between the error terms. The spatial Durbin model combines spatial features in the dependent variable and exogenous regressors. In the network literature a similar specification is the linear-in-means model of peer effects introduced by Manski (1993). In this model agents' outcomes depend on their own characteristics, their peers' characteristics, and their peers' outcomes.

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, Boarnet, Brakman, and Ottaviano (2012).[1] Second, spatial models are characterized by N(N-1) potential relationships among the observations, but only N data observations are available,

[1] 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 across a set of candidates. Lam and Souza (2020) propose to estimate their best linear combination and a sparse adjustment matrix using the least absolute shrinkage and selection operator (LASSO). Higgins and Martellosio (2020) develop a similar approach based on a penalised quasi-maximum likelihood estimator and controlling 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.

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, the spatial econometrics literature has made significant progress in this dimension by developing techniques for the consistent estimation of the model parameters under correct specification of the spatial model and certain assumptions on the weight matrix, see Kelejian and Prucha (1998, 1999), Lee (2004), Bramoullé, Djebbari, and Fortin (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 flexible specification of a SLX model. Our contribution is to model the spatial effect as a functional coefficient that is approximated nonparametrically using a series of Taylor expansions that are applied over disjoint intervals covering a partition of the spatial variable. This approach is nonparametric because the Taylor approximation together with the partition entail a number of regressors that increases with the sample size, see Pinkse, Slade, and Brett (2002), Sun (2016) and Koroglu and Sun (2016), for similar frameworks. The second contribution of this study is to extend the standard SLX model to allow for network effects. The spatial variable - geographical distance - is replaced by a variable that captures network effects between the covariates and the dependent variable measured at different units. This extension, called NLX model in this paper, has nontrivial implications for modelling purposes. Whereas the geographical distance is treated as a nonstochastic variable the network variable indexing the functional coefficient is a random variable, adding another layer of complexity to the model.

Identification of $N(N-1)$ network interactions in a cross section of $N$ observations is possible due to the specification of the functional coefficient characterizing network spillovers. Each pairwise interaction is interpreted as a realization 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 only considering endogenous SAR effects and Koroglu and Sun (2016) 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 network variable. This methodology allows us to estimate the coefficients characterizing the *local* Taylor expansions by minimizing 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, Slade, and Brett (2002) for an application of series estimators to endogenous spatial models) and, more specifically, in the class of partitioning estimators, see Cattaneo and Farrell (2013) and

Cattaneo, Farrell, and Feng (2020), that we extend to the spatial/network literature. The accuracy of our approximation also depends on the order of the Taylor expansion of the functional coefficient that, in contrast to the number of intervals defining the partition, is assumed to be finite.

Although we focus on solving the first two identification issues discussed above, our NLX model can be also extended to incorporate endogenous SAR effects using the approach introduced in Sun (2016). This author considers a functional coefficient SAR model with nonparametric spatial weights that is approximated using a series expansion defined by a sequence of orthonormal basis functions. This model is extended to considering also exogenous spatial effects in Koroglu and Sun (2016). In Section 2.2, we adapt these models to our setting by including endogenous SAR effects in the NLX specification. Estimation of the model parameters is more cumbersome in this context due to the presence of endogeneity. Following Sun (2016) and Koroglu and Sun (2016), Section 3.2 proposes a GMM estimation procedure with instrumental variables that is adapted to our partitioning method.

As an additional contribution, We extend existing theory on partitioning estimators to derive pointwise and uniform convergence of the partitioning estimator of the functional coefficient for the NLX model and leave the analysis of the model with endogenous features for future research. In particular, we derive pointwise estimates based on realizations 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 neighbors and is similar to the square root of $N$ convergence of the partitioning estimator in a nonparametric setting, see Cattaneo and Farrell (2013) and Cattaneo, Farrell, and Feng (2020). We use asymptotic results derived by these authors to extend the convergence of the partitioning estimator to a centered Gaussian process in the functional space. We also develop pointwise t-tests to evaluate the statistical significance of the network effects on specific locations and a uniform test that extends the analysis to the compact support of the network variable. The implementation of the uniform test is not straightforward as it is a composite hypothesis. Under the null, we face Davies (1977, 1987)' problem of 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 covariance kernel that cannot be tabulated. Nevertheless, we follow the theory in Cattaneo, Farrell, and Feng (2020) for deriving the asymptotic distribution of the test, and the use of Wild bootstrap methods in Hansen (1996) to approximate its finite-sample distribution under the null hypothesis.

The finite-sample performance of these tests is evaluated using Monte Carlo methods. Simulations are divided in three exercises. First, we compute the bias and root mean square error (RMSE) of the parameter estimators to demonstrate their consistency. Second, we analyze the size and power of the marginal t-tests and uniform test. The simulations show a very good performance of both tests in terms of power, and a reliable empirical size, that is slightly undersized for the uniform test. In the third simulation exercise, we propose different information criteria to select the optimal tuning parameter based on popular Akaike and Bayesian methods and also on specific criteria developed for series estimators such as Mallows (1973) and Craven and Wahba (1978). Our simulations illustrate how to optimally choose the order of the Taylor expansion and the partition of the grid that determines the number of regressors in our augmented model capturing network effects.

The proposed methodology is illustrated in an empirical application 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 neighboring effects in the relationship between households' income and pollution measures. The network variable is the $L_1$ distance between the pollution content at two different units such that two neighbouring observations are characterized by similar pollution patterns. Building on recent studies about peer effects in household consumption and energy behaviors (Agarwal, Qian, and Zou 2021; De Giorgi, Frederiksen, and Pistaferri 2020; Wolske, Gillingham, and Schultz 2020), we provide strong empirical evidence of neighboring 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 levels of income.

The rest of the paper is structured as follows. Section 2 introduces a nonparametric SLX model with network effects. In Section 3, we propose a nonparametric estimator based on the theory on partitioning estimators. The section also studies GMM methods to estimate the model parameters under the presence of network endogeneity. Section 4 presents the asymptotic theory on the proposed estimators for the NLX model. In particular, we show the consistency and uniform convergence of the network functional parameter estimator obtained by our augmented regression model. The section also derives pointwise convergence results on the asymptotic distribution of the standardized functional coefficient estimator to a Normal distribution. Section 5 presents different hypothesis tests to statistically assess the presence of network effects in the data. The section also discusses model selection and the optimal choice of the tuning parameter. Section 6 presents a Monte Carlo exercise to evaluate parameter estimation, hypothesis testing, and model selection in finite samples. Section 7 contains the empirical application, and Section 8 concludes. An appendix contains the mathematical proofs of the main results of the paper. Tables and figures are collected at the end of this document.

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 $A$ a $m \times n$ 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 Network regression model

This section introduces the nonparametric SLX model based on functional coefficients and discusses its approximation by Taylor expansions over disjoint intervals covering the compact support of the spatial variable characterizing the spatial/network effects. The section also proposes estimators of the slope and network parameters based on the theory on partitioning estimators, see Cattaneo and Farrell (2013) and Cattaneo, Farrell, and Feng (2020), as recent seminal contributions.

## 2.1 The baseline model

We propose the following NLX specification that extends standard SLX models in two dimensions: ($i$) there are network effects that replace spatial effects, ($ii$) network effects are modelled as a functional coefficient. The proposed model is

$$Y = X\lambda + \sum_{j=1}^{M} \mathbb{W}_j(d_j)X_j + \varepsilon, \tag{2.1}$$

with $Y = (Y_1, \ldots, Y_N)'$ a vector that contains the dependent variable evaluated at each unit, $X = (X_1, \ldots, X_M)$ is a $N \times M$ matrix, where $N$ denotes the number of observations and $M$ the number of exogenous regressors; $\lambda$ is a $M \times 1$ vector of coefficients, and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is a zero-mean random vector containing the error term that is assumed to be independent and identically distributed ($iid$). For simplicity, we remove the intercept from the model specification by assuming that $(Y, X)$ are demeaned. The exogenous network effects between the different covariates and the dependent variable are captured by the sequence of $N \times N$ matrices $\mathbb{W}_j(d_j)$. Each of these matrices contains $N(N-1)$ parameters describing the network relationships and is indexed by the distance $d_{j,rs} = f(z_{jr}, z_{js})$, for $j = 1, \ldots, M$, with $\{z_{jr}, z_{js}\}$ realizations of a random variable $Z_j$ evaluated at units $r$ and $s$. This variable characterizes the type of network dependence. The metric $f(\cdot, \cdot)$ can also differ across regressors. For example, the geographical distance is characterized by the Euclidean distance between the geographical coordinates ($Z_i \in R^2$) at two different locations but other metrics are also possible. In the empirical application, we consider, instead, the $L_1$ distance between the regressors, $i.e.$ $d_{rs} = |x_r - x_s|$, with $x_r, x_s$ realizations of the regressor $X$ measured at different units.

To illustrate the model and estimation procedure, we consider one regressor but the methods below can be extended naturally to the case of $M$ regressors, with $M$ finite. The baseline model is

$$Y = X\lambda + \mathbb{W}(d)X + \varepsilon, \tag{2.2}$$

with $X$ an $N \times 1$ vector, $\lambda$ a scalar parameter and $\mathbb{W}(d)$ the corresponding spatial/network weight matrix, with $d$ a spatial/network variable measuring the distance between the different units. The elements of the matrix $\mathbb{W}(d)$ are estimable coefficients $\mathrm{w}(d_{ij})$, for

$i \neq j$, such that the diagonal values satisfy $d_{ii} = 0$ and $\mathrm{w}(d_{ii}) = 0$. The distance between units is defined as $d_{ij} = f(z_i, z_j)$, with $\{z_i, z_j\}$ realizations of the random variable $Z$ evaluated at units $i$ and $j$, and $f(\cdot, \cdot)$ is a function defined over the positive real line and satisfying the properties of a metric.

Our objective is to estimate these parameters from a sample of $N$ observations. To do this, we model the weight function as a functional coefficient such that the elements $w(d_{ij})$ are interpreted as realizations of $w(d)$, with $d \in \mathbb{R}^+$, see Pinkse, Slade, and Brett (2002) and Sun (2016) for similar settings. Then, model (2.2) can be expressed as

$$y_i = \lambda x_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} w(d_{ij}) x_j + \varepsilon_i, \ i = 1, \ldots, N. \tag{2.3}$$

A neighboring unit is determined by the magnitude of $d_{ij}$ and a bandwidth parameter $h$ that defines the width of the intervals covering the support $\chi$ of the network variable, that is assumed to be compact. More specifically, there are $K$ disjoint intervals constructed from a grid of $K$ points $\{z_1, \ldots, z_K\}$. Let $[z_k - h, z_k + h)$ be a generic interval of the partition such that, for $d \in \chi$, $1_k(d)$ is an indicator function with $1_k(d) = 1$ if $d$ belongs to the interval and zero, otherwise. Similarly, let $p_k = P\{d \in [z_k - h, z_k + h)\}$ be the probability of belonging to a given interval and such that $\sum_{k=1}^{K} p_k = 1$.

The following assumptions impose the exogeneity of the regressors and independence of the errors, and introduce regularity conditions on the functional coefficient and the elements of the partition.

**Assumption A.**

(A1) $\{(x_i, z_i, \varepsilon_i)\}$ is an *iid* sequence across index $i$ and $y_i$ is generated from model (2.3). 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 extension of) the compact set $\chi \subset \mathbb{R}^+$, with $q \geq 0$ fixed.

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

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

(A5) Let $K$ denote the number of disjoint intervals covering the compact set $\chi$. Then, we require $K/N \to 0$ and $N/K^{q+1} \to 0$ as $K, N \to \infty$.

(A6) The number of intervals $K$ depends on the tuning parameter $h$ such that $h \asymp K^{-1}$, where for scalars $a$ and $b$, $a \asymp b$ denotes that $C_* b \leq a \leq C^* b$ for positive constants $C_*$ and $C^*$. Similarly, we assume $p_k \asymp K^{-1}$. By construction, $p_k = K^{-1}$ if $p_k$ is exactly the same across intervals.

Assumption A1 imposes the regressors to be *iid* and guarantees the fourth moment of the regressors to be finite. This will be required for proving 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 local Taylor expansions for each interval of the partition of 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 (2.3) conditional on the vector of exogenous covariates $X$ and $D$, with $D$ a vector that contains for each unit the network distance from the rest. 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. Assumption A5 imposes some regularity conditions on the number of intervals characterizing the partition with respect to the order of the Taylor expansion and the sample size. Assumption A6 introduces the asymptotic relationship between the bandwidth parameter $h$ defining the width of the intervals and the probability of an observation belonging to them.

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

$$w(d) = \sum_{k=1}^{K} \sum_{m=0}^{q} \frac{1}{m!} w^{(m)}(z_k)(d - z_k)^m 1_k(d) + R(d), \tag{2.4}$$

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

Local polynomial approximations of functional coefficients are proposed in Fan and Gijbels (1996), Cai and Li (2008), Cai and Xu (2008) and Cai and Xiao (2012), amongst others. However, in contrast to these articles, the approximation proposed below is not based on kernel smoothers of the neighboring 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 network 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. Plugging in the Taylor expansion in equation (2.3), and using the above notation, we obtain the following regression model:

$$y_i = \lambda x_i + \sum_{k=1}^{K}\sum_{m=0}^{q}\gamma_{km}\widetilde{x}_i^{(km)} + \overline{R}_i + \varepsilon_i. \tag{2.5}$$

By applying a Taylor expansion to $w(d_{ij})$ around the different knots in the partition, we reduce the dimension of 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.

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

$$Y = X\lambda + \mathbb{X}\,\Gamma + \overline{R} + \varepsilon, \tag{2.6}$$

where $Y = (Y_1, \ldots, Y_N)'$ and $\mathbb{X} = [\mathbb{X}_1, \ldots, \mathbb{X}_K]$ is a matrix of dimension $N \times K(q+1)$. Each $\mathbb{X}_k$ defines a $N \times (q+1)$ matrix with elements $(\widetilde{x}_i^{(k0)}, \ldots, \widetilde{x}_i^{(kq)})$. Similarly, the vector of coefficients satisfies that $\Gamma = (\Gamma_1', \ldots, \Gamma_K')'$, with $\Gamma_k = (\gamma_{k0}, \ldots, \gamma_{kq})'$. The vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$ is the error term and $\overline{R}$ is a vector with the remainder terms $\overline{R}_i$ for $i = 1, \ldots, N$.

## 2.2 Nonparametric model with endogenous network effects

Spillovers in the SLX model are transmitted only through the covariates to the dependent variable. The presence of endogenous spatial effects has received the attention of researchers for several reasons. McMillen (2012) shows that endogenous spatial effects may be the result of omitted variables and misspecification of the exogenous spatial effects. Brueckner (2006) provides a general framework for a class of theoretical models of spatial interaction among local governments that lead directly to the type of SAR models implemented in the spatial econometrics literature. Sun (2016) uses the latter approach to motivate the presence of endogenous spatial effects as part of a reaction function $y_i = R(y_i^-, x_i)$, in which $y_i$ is an outcome variable from jurisdiction i, $y_i^-$ is a vector containing the observations from the rest of jurisdictions and $x_i$ contains the characteristics of jurisdiction i. See Pinkse, Slade, and Brett (2002) for a similar motivation of empirical endogenous spatial effects on a model of price competition.

In this subsection, we extend the NLX model proposed in (2.3) to accommodate the presence of endogenous network effects. Following Sun (2016) and Koroglu and Sun (2016), these effects are modelled using a functional coefficient $\widetilde{w}(d_{ij})$ such that

$$y_i = \sum_{\substack{j=1\\ j\neq i}}^{N}\widetilde{w}(d_{ij})y_j + \lambda x_i + \sum_{\substack{j=1\\ j\neq i}}^{N} w(d_{ij})x_j + \varepsilon_i,\ \ i = 1, \ldots, N. \tag{2.7}$$

In contrast to these authors, we approximate the functional coefficient using a Taylor expansion over disjoint intervals of a partition of the compact space of the network variable and not as a series expansion of orthonormal basis functions. Following the same steps as above, we obtain

$$\widetilde{w}(d) = \sum_{k=1}^{K}\sum_{m=0}^{q}\frac{1}{m!}\widetilde{w}^{(m)}(z_k)(d-z_k)^m 1_k(d) + \widetilde{R}(d), \tag{2.8}$$

with $\widetilde{w}^{(m)}(z_k)$ the $m^{th}-$derivative of $\widetilde{w}(\cdot)$ evaluated at $z_k$; $\widetilde{w}^{(0)}(z_k) = \widetilde{w}(z_k)$ the functional coefficient evaluated at $z_k$; and $\widetilde{R}(d) = \sum_{k=1}^{K}\widetilde{w}^{(q+1)}(\widetilde{c}_k)(d-z_k)^{q+1}1_k(d)$ the remainder of the Taylor expansion, with $\widetilde{c}_k \in (z_k - h, z_k + h)$. Similarly, let $\widetilde{y}_i^{(km)} = \sum_{\substack{j=1\\j\neq i}}^{N} y_j(d_{ij}-z_k)^m 1_k(d_{ij})$ be the endogenous variables indexed by $k = 1, \ldots, K$ and $m = 0, \ldots, q$, and let $\widetilde{\gamma}_{km} = \frac{1}{m!}\widetilde{w}^{(m)}(z_k)$ be the network coefficients for the endogenous regressors; $\overline{\widetilde{R}}_i = \sum_{\substack{j=1\\j\neq i}}^{N}\widetilde{R}(d_{ij})y_j$ is the aggregate remainder term. Then,

$$y_i = \sum_{k=1}^{K}\sum_{m=0}^{q}\widetilde{\gamma}_{km}\widetilde{y}_i^{(km)} + \lambda x_i + \sum_{k=1}^{K}\sum_{m=0}^{q}\gamma_{km}\widetilde{x}_i^{(km)} + \overline{\overline{R}}_i + \varepsilon_i, \tag{2.9}$$

with $\overline{\overline{R}}_i = \overline{R}_i + \overline{\widetilde{R}}_i$ the approximation error aggregating the exogenous and endogenous error terms. In matrix form, we obtain

$$Y = \mathbb{Y}\,\widetilde{\Gamma} + X\lambda + \mathbb{X}\,\Gamma + \overline{\overline{R}} + \varepsilon, \tag{2.10}$$

with $\mathbb{Y} = [\mathbb{Y}_1, \ldots, \mathbb{Y}_K]$ a matrix of dimension $N \times K(q+1)$, where $\mathbb{Y}_k$ is a $N \times (q+1)$ matrix with elements $(\widetilde{y}_i^{(k0)}, \ldots, \widetilde{y}_i^{(kq)})$. Similarly, the vector of coefficients satisfies that $\widetilde{\Gamma} = (\widetilde{\Gamma}'_1, \ldots, \widetilde{\Gamma}'_K)'$, with $\widetilde{\Gamma}_k = (\widetilde{\gamma}_{k0}, \ldots, \widetilde{\gamma}_{kq})'$.

# 3 Parameter estimation

This section studies the estimation of the NLX model and its extension incorporating endogenous network effects. The first model is estimated using the theory on sieve regression and partitioning estimators, see Newey (1997) and Cattaneo, Farrell, and Feng (2020). The model with endogenous network effects is estimated using GMM methods and is inspired by the models in Sun (2016) and Koroglu and Sun (2016).

## 3.1 Partitioning estimator for NLX model

Using the partitioned inverse, a suitable estimator of $\lambda$ is

$$\widehat{\lambda} = \left(\widehat{X}'_u\widehat{X}_u\right)^{-1}\widehat{X}'_u(Y - \widetilde{Y}), \tag{3.1}$$

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 $\mathbb{X} = [\mathbb{X}_1, \ldots, \mathbb{X}_K]$. This matrix is partitioned in blocks such that each of the network parameters is estimated from the partitioned regression as

$$\widehat{\Gamma}_k = \left( \sum_{i=1}^{N} \mathbb{X}'_{ki} \mathbb{X}_{ki} \right)^{-1} \sum_{i=1}^{N} \mathbb{X}'_{ki} (y_i - x_i \widehat{\lambda}), \tag{3.2}$$

for each $\widehat{\Gamma}_k = (\widehat{\gamma}_{k0}, \ldots, \widehat{\gamma}_{kq})'$, see Cattaneo and Farrell (2013) for details on partitioned regressors. The estimator of $w(d)$ is obtained from the Taylor expansion (2.4) as

$$\widehat{w}(d) = \sum_{k=1}^{K} \sum_{m=0}^{q} \widehat{\gamma}_{km} (d - z_k)^m 1_k(d) = \sum_{k=1}^{K} \widehat{\Gamma}'_k v_k(d) 1_k(d), \tag{3.3}$$

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

Other important quantity for making statistical inference about the network 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,

$$\widehat{V}\left(\widehat{\lambda}\right) = \frac{1}{N} \widehat{\Phi}^{-1} \widehat{\Psi} \widehat{\Phi}^{-1}. \tag{3.4}$$

Similarly, we will show in the following section that an appropriate estimator of the variance of the network parameter estimator is

$$\widehat{V}\left(\widehat{\Gamma}_k\right) = \frac{1}{\alpha_N p_k} \widehat{Q}_k^{-1} \widehat{A}_k \widehat{Q}_k^{-1}, \tag{3.5}$$

with $\alpha_N = N(N-1)$ a standardizing constant for the network coefficient. Note that $Q_k = \frac{1}{(N-1)p_k} E[\mathbb{X}'_{ki} \mathbb{X}_{ki}] = E[\overline{X}'_{k,ij} \overline{X}_{k,ij}]/p_k$, 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 a suitable estimator is $\widehat{Q}_k = \frac{1}{\alpha_N p_k} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \overline{X}'_{k,ij} \overline{X}_{k,ij}$. Similarly, we define $\widehat{A}_k = \frac{1}{\alpha_N p_k} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \overline{X}'_{k,ij} \overline{X}_{k,ij} e_i^2$ as a suitable estimator of $A_k = \frac{1}{(N-1)p_k} E[\mathbb{X}'_{ki} \mathbb{X}_{ki} \varepsilon_i^2] = E[\overline{X}'_{k,ij} \overline{X}_{k,ij} \varepsilon_i^2]/p_k$, under assumption A1. From this expression, a natural estimator of the variance of $\widehat{w}(d)$ in (3.3) is

$$\widehat{V}(\widehat{w}(d)) = \sum_{k=1}^{K} v'_k(d) \widehat{V}(\widehat{\Gamma}_k) v_k(d) 1_k(d). \tag{3.6}$$

## 3.2 GMM estimator for the nonparametric spatial Durbin model

The endogeneity of the spatial Durbin model in (2.10) invalidates the approach proposed for the nonparametric NLX model. In this case a viable estimation approach producing consistent parameter estimates is the application of GMM under a suitable choice of instrumental variables. There are $K(q+1)$ parameters associated to the endogenous regressors and $K(q+1)+1$ parameters associated to the exogenous regressor $X$ evaluated over different elements of the partition. Let $\mathbb{X}_{reg} = [\mathbb{Y}, X, \mathbb{X}]$ a $N \times (2K(q+1)+1)$ matrix, $\Gamma_{All} = [\widetilde{\Gamma}', \lambda, \Gamma']'$ the associated $(2K(q+1)+1) \times 1$ vector and $IV_N$ a $N \times N$ matrix with row elements $IV_{Ni} = (x_i, \overline{d}_i)$ containing the exogenous instruments. The GMM estimate of $\Gamma_{All}$ is obtained from the set of orthogonal moment conditions

$$E[IV_N'(Y - \mathbb{X}_{reg}\Gamma_{All})] = 0,$$

and, more specifically, from minimizing the following objective function

$$\min_{\{\Gamma_{All}\}} \left\{(Y - \mathbb{X}_{reg}\Gamma_{All})' \; IV_N \; IV_N' \; (Y - \mathbb{X}_{reg}\Gamma_{All})\right\}.$$

The solution to this problem yields the following vector of parameter estimates

$$\widehat{\Gamma}_{All} = \left(\mathbb{X}_{reg}' \; IV_N \; IV_N' \; \mathbb{X}_{reg}\right)^{-1} \mathbb{X}_{reg}' \; IV_N \; IV_N' \; Y. \tag{3.7}$$

Sun (2016) and Koroglu and Sun (2016) provide conditions that guarantee the existence and consistency of the GMM estimator. These conditions are similar to the above set of assumptions but also contain certain additional conditions that guarantee the invertibility and boundedness of the matrices $I_N - W_{\mathbb{Y}}$ and $\mathbb{X}_{reg}' IV_N IV_N' \mathbb{X}_{reg}$, with $I_N$ the $N \times N$ identity matrix; $W_{\mathbb{Y}}$ is a $N \times N$ matrix with element $(r, s)$ given by $\widetilde{w}(d_{rs})$ for $r, s = 1, \ldots, N$, see Assumption A1 in Sun (2016).

The above estimator (3.7) is inefficient as also discussed by these authors in a related setting. As in Koroglu and Sun (2016), we do not pursue an efficient estimator of $\Gamma_{All}$ obtained from an optimal choice of instruments due to the potentially large number of regressors in $\mathbb{X}_{reg}$. We simply require that the number of moment conditions $N$ - equal to the number of instrumental variables - is greater than the number of parameters $2K(q+1)+1$ to be estimated. Additionally, the matrix $E[IV_N'\mathbb{X}_{reg}]$ needs to have a full column rank. Under these conditions, the asymptotic properties of the estimator (3.7) follow from extending the asymptotic analysis of the partitioning estimator to a GMM setting. The formal analysis of these properties is, however, beyond the scope of this study. Thus, in what follows, we derive the asymptotic theory for the partitioning estimator corresponding to the NLX approach and leave the study of the asymptotic properties of the GMM estimator for future research.

# 4 Asymptotic convergence of the NLX estimator

This section presents pointwise and uniform convergence results for the functional estimator $\widehat{w}(d)$ in (3.3). The section also presents results necessary to make asymptotic inference 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 $\Phi_0$ and $\Psi_0$ 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 $\|Q_k\| < \infty$ and $\|A_k\| < \infty$, for $k = 1, \ldots, K$.

Conditions in assumption B guarantee the existence and positive definiteness of the population covariance matrices $\Phi_0$, $\Psi_0$, $Q_k$ and $A_k$, for $k = 1, \ldots, K$. This assumption is sufficient to show that $\|\widehat{\Phi} - \Phi_0\| = o_p(1)$. The following result presents convergence results between the network covariance matrices.

**Lemma 1.-** *Under assumptions A and B, for $k = 1, \ldots, K$, it follows that $\|\widehat{Q}_k - Q_k\| = O_p\left(\frac{K^\nu}{\sqrt{N}}\right)$, with $\nu = 0$, if $d_{ij} \neq d_{ji}$, and $\nu = 1/2$ if the network variable is symmetric ($d_{ij} = d_{ji}$, for all $i, j = 1, \ldots N$) as $K, N \to \infty$.*

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

**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 all units for estimating the network functional coefficient. The choice of a partitioning type estimator introduces an additional effect produced by dividing the compact set into $K$ disjoint intervals. The proof of these results is included in the appendix.

**Proposition 2.-** *Under assumptions A and B,*

$$\sqrt{N}(\widehat{\lambda} - \lambda) \xrightarrow{d} N\left(0, \Phi_0^{-1}\Psi_0\Phi_0^{-1}\right). \tag{4.1}$$

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

**Theorem 1.-** *Under assumptions A and B,*

$$\sup_{d\in\chi} |\widehat{w}(d) - w(d)| = O_p\left(\sqrt{K}/N + K^{-(q+1)}\right). \tag{4.2}$$

The uniform convergence is determined by a variance term $\sqrt{K}/N$ given by the estimation of the network 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 2.-** *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 3.-** *Under assumptions A and B, for $k = 1, \ldots, K$, the estimator (3.5) satisfies that*

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

This result can be extended to derive the asymptotic convergence of the variance estimator (3.6). 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 $\widehat{V}_K(d) = \sum_{k=1}^{K} v_k'(d) \widehat{Q}_k^{-1} \widehat{A}_k \widehat{Q}_k^{-1} v_k(d) 1_k(d)/p_k$.[2]

**Proposition 3.-** *Under assumptions A and B, for $k = 1, \ldots, K$ and any $d \in \chi$ fixed, 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, applying the triangular inequality, Proposition 3 shows that

$$|V\left(\sqrt{\alpha_N}(\widehat{w}(d) - w(d))\right) - \widehat{V}_K(d)| = O(N/K^{2(q+1)}) + O_p(K/N) = O_p(K/N), \tag{4.3}$$

[2]Note that knowledge of the probability $p_k$ is not required for obtaining $\widehat{V}_K(d)$. This is so because $p_k$ cancels out with the covariance estimators $\widehat{Q}_k^{-1} \widehat{A}_k \widehat{Q}_k^{-1}$.

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 $d \in \chi$ fixed, it follows that*

$$\sqrt{\alpha_N}\,\frac{\widehat{w}(d) - w(d)}{\widehat{V}_K^{1/2}(d)} \xrightarrow{d} N\left(0,1\right). \tag{4.4}$$

The convergence rate of the estimator reflects the influence of neighboring effects. This result is the basis of pointwise tests for the presence of network effects given by $H_0 : w(d) = 0$ against $H_A : w(d) \neq 0$, for some $d \in \chi$ fixed. 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 $\chi$ equipped with the uniform norm. Nevertheless, the weak convergence of the above process can be obtained adapting the strong approximation results derived in Section 6 of Cattaneo, Farrell, and Feng (2020). We state the following result, the proof of which is obtained from the application of the asymptotic results by these authors.

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

$$\sqrt{\alpha_N}\,\frac{\widehat{w}(d) - w(d)}{\widehat{V}_K^{1/2}(d)} \xrightarrow{w} \mathbb{G}(d), \tag{4.5}$$

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

As a byproduct of this result, the asymptotic distribution of the supremum functional can be obtained as

$$\sqrt{\alpha_N}\sup_{d\in\chi}\left|\frac{\widehat{w}(d) - w(d))}{\widehat{V}_K^{1/2}(d)}\right| \xrightarrow{d} \sup_{d\in\chi}\left|\mathbb{G}(d)\right|, \text{ as } N \to \infty. \tag{4.6}$$

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

# 5 Hypothesis testing and model selection

## 5.1 Hypothesis testing

This subsection exploits the above asymptotic theory to construct different hypothesis tests. Although the focus is on testing for the presence of network effects, we also introduce a framework to statistically assess the functional form of $w(d)$. In our context, testing for the presence of network 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 $H_{0f} : \sup_{d \in \chi} |w(d) - f(d)| = 0$, with $f(\cdot)$ 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

$$T_N = \sqrt{\alpha_N} \sup_{d \in \chi} \left| \frac{\widehat{w}(d) - f(d)}{\widehat{V}_K^{1/2}(d)} \right|, \tag{5.1}$$

where the functional form $f(d)$ depends on the null hypothesis under study. Hypothesis tests involving nuisance parameters under the null 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*

$$T_N \xrightarrow{d} \sup_{d \in \chi} |\mathbb{G}(d)|, \ \textit{as } N \to \infty, \tag{5.2}$$

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

Its proof follows as an application of the asymptotic result (4.6), obtained by replacing $w(d)$ by 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, Farrell, and Feng (2020). We proceed now to discuss a p-value transformation method for testing the null hypothesis of interest. We operate conditionally on a realization of $\{(x_i, y_i)\}_{i=1}^N$, denoted as $\omega_N$.

Expression (8.16) in the mathematical appendix shows that

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

where $e_{0i}$ are the residuals of the data generating process obtained under the null hypothesis, e.g. $e_{0i} = y_i - X_i\widehat{\lambda}$, with $\widehat{\lambda}$ the OLS estimator of the regression model without network effects obtained under the null hypothesis $H_0 : \sup_{d\in\chi} |w(d)| = 0$.

The objective 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,

$$\mathbb{G}_N^\star(d) = \frac{\frac{1}{\sqrt{\alpha_N} p_k}\sum_{i=1}^{N}\sum_{k=1}^{K} \upsilon_k(d)' \widehat{Q}_k^{-1} \mathbb{X}_{ki}' e_{0i}^\star 1_k(d)}{V_k^{1/2}(d)}, \tag{5.3}$$

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

Using the same arguments as in Cattaneo, Farrell, and Feng (2020), we show without proof 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,

$$P_{\omega_N}\left\{T_N^\star > T_N\right\} \to P_{H_0}\left\{T_N > \sup_{d\in\chi} |\mathbb{G}(d)|\right\}, \quad \text{as } N \to \infty, \tag{5.4}$$

with $P_{\omega_N}$ denoting a probability distribution function conditional on the realization of the sample $\omega_N$, and $P_{H_0}$ 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 $\omega_N$. The algorithm 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 characterizes a partition of the set $\chi = [0, C]$ that spans the support of the network variable $d_{ij}$ measuring 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 $h$, we choose the number of intervals $K$ as $K = C/2h$ and satisfying the conditions in assumption A5.

2. Compute the test statistic $T_N = \sqrt{\alpha_N} \sup_{d \in \mathbb{D}} \left| \frac{\widehat{w}(d) - f(d)}{\widehat{V}_K^{1/2}(d)} \right|$, with $f(d)$ denoting the null hypothesis; $\mathbb{D}$ denotes a discrete set of equally-distant points inside $\chi$. This set of points characterizes a finer grid of the interval $\chi$ than $\bar{\mathbb{Z}}$.

3. For a given realization $\omega_N = \{x_i, y_i\}_{i=1}^N$, execute the following steps for $b = 1, \ldots, B$:

   (a) Generate $\{\epsilon_i^{(b)}\}_{i=1}^N$ $iid(0,1)$ random variables independent of the data to construct the simulated residuals $e_0^{\star(b)} = e_0 \otimes \epsilon^{(b)}$, with $e_0$ the vector of residuals of regression model (2.3) under the null hypothesis $H_0$. Then, compute the simulated process (5.3).

   (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 $\omega_N$ defined by

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

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

## 5.2 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 $\chi$ and, hence, the quality of the approximation of the function $w(d)$. We do not propose a formal selection method for choosing the bandwidth. Alternative methods for partitioning estimators are proposed in Cattaneo and Farrell (2013) and Cattaneo, Farrell, and Feng (2020), as seminal examples. Instead, we suggest off-the-shelf methods for bandwidth selection developed for nonparametric regression models. For illustrative purposes, we consider $\chi \equiv [0, C]$.

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

$$\widehat{h}_M = \arg\min_{\{h\}} \left\{ \widehat{\sigma}_e^2 \left( 1 + \frac{C}{Nh} \right) \right\}, \tag{5.5}$$

with $\widehat{\sigma}_e^2 = \frac{1}{N} \sum_{i=1}^{N} e_i^2$ obtained under conditional homoscedasticity of the error term. Craven and Wahba (1978) propose a generalized cross-validation method[3]. These authors select $\widehat{h}$ such that

$$\widehat{h}_{GCV} = \underset{\{h\}}{\arg\min} \left\{ \frac{\widehat{\sigma}_e^2}{\left(1 - \frac{C}{Nh}\right)^2} \right\}. \tag{5.6}$$

We also want to explore the role of the order of the Taylor expansion $q$ in the approximation of the functional coefficient $w(d)$. This tuning parameter has a non-negligible effect on the accuracy of the approximation because it directly affects the number of regressors in (2.9). To account for this, we adapt the Akaike (AIC) and Bayesian (BIC) information criteria to the present context as

$$\widehat{h}_{AIC} = \underset{\{h,q\}}{\arg\min} \left\{ \ln \widehat{\sigma}_e^2 + 2\frac{(q+1)\lceil C/2h \rceil + 1}{N} \right\}, \tag{5.7}$$

$$\widehat{h}_{BIC} = \underset{\{h,q\}}{\arg\min} \left\{ \ln \widehat{\sigma}_e^2 + \frac{((q+1)\lceil C/2h \rceil + 1) ln\ N}{N} \right\}. \tag{5.8}$$

Another issue to be considered for model selection is the choice of the variable that determines the proximity between units. The network model presented herein can be extended to assume that the network variable $d$ is not known and has to be selected from a set of candidates. A possibility is to be guided by theory and use a variable with a clear network interpretation. For example, in spatial econometrics models, a natural network variable is the geographical distance between observations. More generally, we can rely on statistical techniques to determine the most suitable variable $d$ for model (2.9). A natural approach is to choose the variable that minimizes the mean square error. This analysis goes beyond the scope of this paper and is left for future research.

# 6 Monte-Carlo simulations

This section explores the finite-sample approximation of the asymptotic results using Monte Carlo simulations. We present four different exercises that illustrate i) the consistency of the parameter estimates, ii) the rejection rates associated to the marginal t-tests using the asymptotic distribution in Theorem 2 , iii) the empirical size and power of the uniform tests $H_0$ and $H_{0f}$ obtained from Theorem 3, and iv) the model selection procedure to determine the optimal value of the tuning parameter $h$ and choice of the order of the Taylor expansion $q$.

[3]Other more sophisticated model selection procedures for series estimators can be found in the literature, for example, the leave-one-out cross-validation method of Stone (1974).

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

$$y_i = x_i\lambda + \sum_{\substack{j=1 \\ j\neq i}}^{N} w(d_{ij})x_j + \varepsilon_i, \text{ for } i = 1, \ldots, N, \tag{6.1}$$

with $x_i$ being realizations of a single covariate $X$ distributed as a $N(0,1)$. For simplicity, the regressor also acts as the network variable $Z$ establishing the proximity between units in the cross section, such that $d_{ij} = |x_i - x_j|$. The error term $\varepsilon_i$ is modelled as a $N(0, \log^2|1 + x_i|)$ random variable that is uncorrelated to $X$ but exhibits conditional heteroscedasticity.[4]

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 characterize 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 neighboring 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 $K = 1/2h$. We consider the following values $h = 0.05, 0.075, 0.1$ to assess the sensitivity of the estimates to the choice of tuning parameters. The sample size is equal to $N = 100, 250, 500$ but results for $N = 1000$ are also available upon request. The parameters characterizing the functional form of $w(d)$ are $\beta = 0.1$ and $\theta = 5, 9$. This choice of parameters results in small network effects, however, as shown below, the test statistic (4.4) has considerable power to reject the null hypothesis under the presence of such network effects. We also consider $\theta = 7$ in the study of the asymptotic coverage rate $\alpha$ corresponding to the $(1-\alpha)$-confidence interval of $w(d)$ constructed as

$$\left[\widehat{w}(d) - z_{1-\alpha/2}\widehat{V}_K^{1/2}(d)/\sqrt{\alpha_N}, \widehat{w}(d) + z_{1-\alpha/2}\widehat{V}_K^{1/2}(d)/\sqrt{\alpha_N}\right], \tag{6.2}$$

with $\widehat{V}_K(d)$ the estimator (3.6) for a given $d$, and $z_\alpha$ the critical value of a standard Normal distribution function at an $\alpha$ significance level.

## 6.1 Consistency of the parameter estimates

The consistency of the parameter estimators (3.1) and (3.3) is assessed through the analysis of bias and root mean square error (RMSE). Table 1 reports the bias of the parameter

[4]For the sake of presentation, the simulation exercise only considers one covariate but results for a model with several regressors are available from the authors upon request.

estimator $\widehat{\lambda}$ 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. Unreported simulations also consider the case $\gamma = 0.5$. 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 neighbors. To do this, only the parameter estimates for $d = \{h/2, h, 3h/2, 2h\}$ are reported. Results for the remaining values are available from the authors upon request. The figures displayed in Table 1 do not show evidence of bias in any direction and decreases 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 on the consistency of the parameter estimators of $w(d)$, for different values of $d$ in the interval $[0, 1]$.

## 6.2 Empirical coverage rate and rejection probabilities

This exercise studies the finite-sample coverage probability of the asymptotic confidence intervals for $\lambda$ and $w(d)$, for a discrete grid of values $d = \{h/2, h, 3h/2, 2h\}$, under heteroscedasticity of the error term. To do this, we compute the empirical fraction of times the true parameters $\lambda$ 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 characterized by $\lambda = 1, 0.25$. In line with the previous subsection, the left panel studies the exponential function and the right panel the functional specification of $w(d)$ given by the Gaussian kernel. The simulated results show empirical 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 $w(d)$, different values of $\theta$, and sample sizes. Further, the coverage rates converge to the nominal ones at $\alpha = 0.05$ as the sample size increases.

Tables 5 and 6 study the power of the marginal t-tests obtained from the asymptotic convergence result (4.4) for the pointwise null hypotheses $H_0 : \lambda = 0$ against 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 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 a strong performance of the t-tests to reject the null hypothesis across values of $d$ in the grid and DGPs. The empirical power of the test is large in most instances and achieves values above 0.80 for $N = 500$ in most scenarios.

### 6.3 Size and power of the uniform test

The study of the power of the marginal t-tests confirms empirically their suitability for detecting network effects for specific values of $d$ given by $d = \{h/2, h, 3h/2, 2h\}$.[5] This subsection extends this analysis by evaluating the finite-sample size and power of the uniform test presented in (5.1) and Theorem 3. We consider two different null hypotheses given by $i$) the absence of network effects, and $ii$) a specific functional form for $w(d)$ given by the exponential function.

To assess the presence of network effects, data are generated under the null hypothesis $H_0 : \sup\limits_{[0,1]} |w(d)| = 0$. 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 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 but results for $\theta = 7$ are also available upon request. 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 similar findings as when testing for network effects; i.e., empirical power is extremely high even when the test is slightly undersized.

### 6.4 Model selection

This subsection presents a Monte Carlo exercise that examines the suitability of the loss functions discussed in Subsection 5.2 for model selection under different values of the order of the Taylor expansion. We simulate 500 draws of the data generating process for $\lambda = 1$ with $w(d)$ given by an exponential function, and compute the optimal values of $h$ and $q$ using different information criteria. We consider a grid given by $h = \{0.05, 0.075, 0.10, 0.125, 0.15\}$, and $q = 1, 2$. For simplicity, we consider conditional homoscedasticity of the error term in the DGP (6.1).

Table 8 reports the average optimal $h$ over 500 simulations and its standard deviation for five loss functions. The first column reports the optimal $h$ obtained from minimizing the RMSE in the regression model (2.9). The second and third columns display $h_M$ and $h_{GCV}$ as defined in expressions (5.5) and (5.6), respectively. The last two columns show the optimal values of the tuning parameter according to the Akaike and Bayesian information criteria introduced, respectively, in (5.7) and (5.8). Sample standard deviations are reported in parentheses.

[5] Results for other values of $d \in [0, C]$ are available from the authors upon request.

The results show overwhelming evidence on the suitability of $h = 0.05$ as tuning parameter. The standard deviation is very low, suggesting that this choice is optimal in most simulations and across model selection methods. The Akaike and Bayesian information criteria provide additional value to the model selection exercise. These criteria explicitly consider the parameter $q$ in the loss function, penalizing an increasing order in the Taylor expansion of $w(d)$. Table 8 shows that the quadratic approximation improves over the linear one with respect to both information criteria.

# 7 Empirical application

This section extends the analysis carried out in Levinson and O'Brien (2019) by incorporating neighboring 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 for studying environmental Engel curves (EECs) for the U.S. for each year over the period 1984 and 2012. Levinson and O'Brien (2019) has two main objectives. The first one is to find the shape of the relationship between income and pollution, the magnitude of the slope, and study its curvature. The second aim is to analyze shifts in the EEC 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 the EECs are upward sloping, reflecting that richer households are more pollutant, and that the rate at which pollution increases with income is less than one-for-one. In addition, pollution increases at a decreasing rate with income over time, i.e., EECs are concave. The latter result shows that households consume a basket of goods that are less pollutant, both directly and indirectly, in recent years.

Levinson and O'Brien (2019) construct two types of EECs: one using only income as covariate, and a multivariate model that incorporates households' characteristics (up to 18 regressors[6]). In the present application, we focus on their first model that explains pollution as a function of after-taxed income and its square. These authors estimate separate curves for five major air pollutants – particulates smaller than 10 microns (PM10), volatile organic compounds (VOCs), nitrogen oxides (NOx), sulfur dioxide ($SO_2$), and carbon monoxide (CO) – because they are not measured in the same units, and have different environmental consequences. By adopting this approach, and taking into account the literature on peer effects in household consumption and energy behavior (Agarwal,

[6]See Table 2 in Levinson and O'Brien (2019) for 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 in page 124.

Qian, and Zou 2021; De Giorgi, Frederiksen, and Pistaferri 2020; Wolske, Gillingham, and Schultz 2020), we estimate the following specification:

$$p_{it} = \lambda_{1t} y_{it} + \sum_{\substack{j=1 \\ j \neq i}}^{N} w(d_{ij,t}) y_{jt} + \lambda_{2t} y_{it}^2 + \varepsilon_{it}, \tag{7.1}$$

where $p_{it}$ and $y_{it}$ are pollution and after-tax income, respectively; $\varepsilon_{it}$ is the error term that satisfies $E[\varepsilon_{it} \mid Y_t] = 0$, with $Y_t = (y_{1t}, \ldots, y_{Nt})$, see footnote 2 about the absence of endogeneity in this context. The coefficients are indexed by $t$ because we run separate regressions for each year.

The network structure establishing the proximity between individuals is determined by similaries in after-tax income and captured by the functional coefficient $w(d_{ij,t})$, with $d_{ij,t} = |y_{it} - y_{jt}|$, for $i, j = 1, \ldots, N_t$, with $N_t$ the number of households included in the sample for a given year. For simplicity, we restrict the network effects to the linear relationship between pollution and income. The estimation equation can be extended, at the expense of a larger regression model, by also assuming network effects on the quadratic component. Thus, using the specification presented in expression (2.9), the above model can be approximated by

$$p_{it} = \lambda_{1t} y_{it} + \sum_{k=1}^{K} \sum_{m=0}^{q} \gamma_{km,t} y_{it}^{(km)} + \lambda_{2t} y_{it}^2 + \varepsilon_{it}, \tag{7.2}$$

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

The dynamics of the parameters associated to the relationship between the different pollutants and households' after-tax income ($\lambda_{1t}$) and its square ($\lambda_{2t}$) are plotted in Figure 1. Although these parameters display a different magnitude 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 to after-tax income of 1.95, and a negative coefficient on 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 under study. 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 (7.2) that capture network effects for selected values of $d$, defined in terms of the optimal value of the tuning parameter. The optimality of $h$ is determined using the Bayesian information criterion[7] and slightly varies between 0.225 and 0.25. In this application, the network effects describe the explanatory power of households with similar after-tax income on the pollution generated to produce the goods and services that households consume. These neighboring effects are of small magnitude but statistically significant at the 5% level, according to the p-values of the marginal t-tests for each coefficient, in most periods. The analysis of NOx and $SO_2$ pollutants yields parameter estimates of a larger magnitude. Interestingly, the sign of the functional coefficients varies across $h$: while we find positive neighboring effects for $w(h_{opt}/2)$, they are negative for $w(3h_{opt}/2)$ and $w(2h_{opt})$. The statistical significance of these neighboring effects is illustrated in Figure 3. This chart reports the p-values of the uniform test $T_N$ in expression (5.1) over the evaluation period. Despite their fluctuation, the results provide ample support to the significance of neighboring effects in this context, adding further evidence to that obtained from the marginal t-tests for the different realizations of the function $w(d)$ discussed above.

The analysis is completed by studying the adjusted coefficient of determination ($R^2$) of the network regression model (7.2). The presence of heterogeneity in the explanatory power across models and over time is shown at the top of Figure 4. More specifically, the adjusted $R^2$ fluctuates between 0.25 and 0.45. We should note that these figures are particularly high given that the number of regressors of model (7.2) is $(q+1)K_{opt}+2$, with $K_{opt} = \frac{C}{2h_{opt}}$. This number varies with the choice of the optimal $h$ in each period and model but is between 10 and 15 regressors. To attach a statistical figure to these values we compute the F-test for the difference of the unadjusted $R^2$ between the network regression model and its cross-sectional counterpart 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 4, show overwhelming evidence of the statistical significance of the *augmented* model given by considering the network variables compared to the benchmark model given by the cross-sectional quadratic regression model.

# 8 Conclusions

This paper proposes a network regression model that extends standard 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 so, we avoid model misspecification issues produced by imposing certain parametric structure to the spatial dependence. The second innovation is to extend the SLX model by considering network spillover effects. The geographical distance is re-

[7]Similar results, available from the authors upon request, are obtained when the Akaike information criterion is used.

placed by a broader definition in which neighboring observations are close according to some metric. The technical implications of this model are not trivial because the network variable is stochastic. The NLX model proposed in this paper is also extended to incorporate endogenous spatial effects. This is done nonparametrically by considering local Taylor expansions approximating the unknown functional coefficient capturing network effects in the dependent variable.

These results are formalized by studying their asymptotic properties and proposing a test for assessing statistically the presence of spillover effects. This is done using pointwise hypothesis tests and uniform tests where the presence of network effects is tested over the whole domain of the network variable.

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

# References


Agarwal, Sumit, Wenlan Qian, and Xin Zou. 2021. "Thy neighbor's misfortune: Peer effect on consumption". *American Economic Journal: Economic Policy* 13 (2): 1–25. doi:10.1257/pol.20170634.

Andrews, Donald W. K. 1993. "Tests for parameter instability and structural change with unknown change point". *Econometrica* 61 (4): 821–856. doi:10.2307/2951764.

Andrews, Donald W. K., and Werner Ploberger. 1994. "Optimal tests when a nuisance parameter is present only under the alternative". *Econometrica* 62 (6): 1383–1414. doi:10.2307/2951753.

Anselin, Luc. 1988. *Spatial econometrics: Methods and models*. Studies in Operational Regional Science. Heidelberg, Germany: Springer Verlag. ISBN: 9789024737352.

Bhattacharjee, Arnab, and Chris Jensen-Butler. 2013. "Estimation of the spatial weights matrix under structural constraints". *Regional Science and Urban Economics* 43 (4): 617–634. doi:10.1016/j.regsciurbeco.2013.03.005.

Bramoullé, Y., H. Djebbari, and B. Fortin. 2009. "Identification of Peer Effects through Social Networks". *Journal of Econometrics* 150:41–55.

Brueckner, J. K. 2006. "Strategic Interactions among Governments". In *A Companion to Urban Economics*, ed. by Richard J. Arnott and Daniel P. McMillen, 332–347. Malden, MA: Blackwell.

Cai, Z., and Q. Li. 2008. "Nonparametric estimation of varying coefficient dynamic panel data models". *Econometric Theory* 24:1321–1342.

Cai, Zongwu, and Zhijie Xiao. 2012. "Semiparametric quantile regression estimation in dynamic models with partially varying coefficients". *Journal of Econometrics* 167 (2): 413–425. doi:10.1016/j.jeconom.2011.09.025.

Cai, Zongwu, and Xiaoping Xu. 2008. "Nonparametric quantile estimations for dynamic smooth coefficient models". *Journal of the American Statistical Association* 103 (484): 1595–1608. doi:10.1198/016214508000000977.

Cattaneo, Matias D., and Max H. Farrell. 2013. "Optimal convergence rates, Bahadur representation, and asymptotic normality of partitioning estimators". *Journal of Econometrics* 174 (1): 127–143. doi:10.1016/j.jeconom.2013.02.002.

Cattaneo, Matias D., Max H. Farrell, and Yingjie Feng. 2020. "Large sample properties of partitioning-based series estimators". *The Annals of Statistics* 48 (3): 1718–1741. doi:10.1214/19-AOS1865.

Cleveland, William S., and Susan J. Devlin. 1988. "Locally weighted regression: An approach to regression analysis by local fitting". *Journal of the American Statistical Association* 83 (403): 596–610. doi:10.1080/01621459.1988.10478639.

Corrado, Luisa, and Bernard Fingleton. 2012. "Where is the Economics in Spatial Econometrics?" *Journal of Regional Science* 52 (2): 210–239.

Craven, Peter, and Grace Wahba. 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, Robert B. 1977. "Hypothesis testing when a nuisance parameter is present only under the alternative". *Biometrika* 64 (2): 247–254. doi:10.1093/biomet/64.2.247.

— . 1987. "Hypothesis testing when a nuisance parameter is present only under the alternatives". *Biometrika* 74 (1): 33–43. doi:10.2307/2336019.

De Giorgi, Giacomo, Anders Frederiksen, and Luigi Pistaferri. 2020. "Consumption network effects". *The Review of Economic Studies* 87 (1): 130–163. doi:10.1093/restud/rdz026.

Fan, Jianqing, and Irene Gijbels. 1996. *Local polynomial modelling and its applications*. Monographs on Statistics and Applied Probability. London, United Kingdom: Chapman & Hall/CRC Press. ISBN: 9780412983214.

Fischer, Manfred M., and Jinfeng Wang. 2011. *Spatial data analysis: Models, methods and techniques*. Springer Briefs in Regional Science. Berlin, Germany: Springer-Verlag. ISBN: 9783642217203. doi:10.1007/978-3-642-21720-3.

Gibbons, S., and H. Overman. 2012. "Mostly Pointless Econometrics?" *Journal of Regional Science* 52 (2): 172–191.

Halleck Vega, Solmaria, and J. Paul Elhorst. 2015. "The SLX model". *Journal of Regional Science* 55 (3): 339–363. doi:10.1111/jors.12188.

Hansen, Bruce E. 1996. "Inference when a nuisance parameter is not identified under the null hypothesis". *Econometrica* 64 (2): 413–430. doi:10.2307/2171789.

Higgins, Ayden, and Federico Martellosio. 2020. *Shrinkage estimation of network spillovers with factor structured errors.* arXiv: 1909.02823v2 [econ.EM].

Kelejian, H. H., and I. Prucha. 1998. "A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances". *Journal of Real Estate Finance and Economics* 17:99–121.

— . 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model". *International Economic Review* 40:509–533.

Kelejian, Harry 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 (1): 3–11. doi:10.1007/s12076-008-0001-9.

Kelejian, Harry H., and Gianfranco Piras. 2011. "An extension of Kelejian's J-test for non-nested spatial models". *Regional Science and Urban Economics* 41 (3): 281–292. doi:10.1016/j.regsciurbeco.2011.01.002.

Koroglu, Mustafa, and Yiguo Sun. 2016. "Functional-coefficient spatial Durbin models with nonparametric spatial weights: An application to economic growth". *Econometrics* 4 (1): 6. doi:10.3390/econometrics4010006.

Lam, Clifford, and Pedro C. L. Souza. 2020. "Estimation and selection of spatial weight matrix in a spatial lag model". *Journal of Business & Economic Statistics* 38 (3): 693–710. doi:10.1080/07350015.2019.1569526.

Lee, Lung-Fei. 2004. "Asymptotic Distribution of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models". *Econometrica* 72:1899–1925.

Levinson, Arik, and James O'Brien. 2019. "Environmental Engel curves: Indirect emissions of common air pollutants". *The Review of Economics and Statistics* 101 (1): 121–133. doi:10.1162/rest_a_00736.

Li, Ker-Chau. 1987. "Asymptotic optimality for $C_p, C_L$, cross-validation and generalized cross-validation: Discrete index set". *Annals of Statistics* 15 (3): 958–975. doi:10.1214/aos/1176350486.

Li, Qi, and Jeffrey S. Racine. 2007. *Nonparametric econometrics: Theory and practice.* Princeton NJ, United States: Princeton University Press. ISBN: 9780691121611.

Mallows, Colin L. 1973. "Some comments on $C_p$". *Technometrics* 15 (4): 661–675. doi:10.2307/1267380.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem". *The Review of Economic Studies* 60 (3): 531–542. doi:10.2307/2298123.

McMillen, Daniel P. 2012. "Perspectives on spatial econometrics: linear smoothing with structured models". *Journal of Regional Science* 52 (2): 192–209.

Newey, Whitney K. 1997. "Convergence rates and asymptotic normality for series estimators". *Journal of Econometrics* 79 (1): 147–168. doi:10.1016/S0304-4076(97)00011-0.

Partridge, M. D., M. G. Boarnet, S. Brakman, and G. Ottaviano. 2012. "Introduction: Whither Spatial Econometrics?" *Journal of Regional Science* 52 (2): 167–172.

Pinkse, Joris, Margaret E. Slade, and Craig Brett. 2002. "Spatial price competition: A semiparametric approach". *Econometrica* 70 (3): 1111–1153. doi:10.1111/1468-0262.00320.

Rose, Christiern D. 2017. "Identification of peer effects through social networks using variance restrictions". *The Econometrics Journal* 20 (3): S47–S60. doi:10.1111/ectj.12101.

Stone, Mervyn. 1974. "Cross-validatory choice and assessment of statistical predictions". *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2): 111–147. doi:10.2307/2984809.

Sun, Yiguo. 2016. "Functional-coefficient spatial autoregressive models with nonparametric spatial weights". *Journal of Econometrics* 195 (1): 134–153. doi:10.1016/j.jeconom.2016.07.005.

Wahba, Grace. 1985. "A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem". *The Annals of Statistics* 13 (4): 1378–1402. doi:10.2307/2241360.

Wheeler, David C., and Antonio Páez. 2010. "Geographically weighted regression". In *Handbook of Applied Spatial Analysis*, 461–486. Berlin, Germany: Springer-Verlag. ISBN: 9783642036460. doi:10.1007/978-3-642-03647-7_22.

Wolske, Kimberly S., Kenneth T. Gillingham, and P. Wesley Schultz. 2020. "Peer influence on household energy behaviours". *Nature Energy* 5:202–212. doi:10.1038/s41560-019-0541-9.

# Mathematical Proofs

**Proof of Lemma 1.** Let $Q_k = E[\overline{X}'_{k,ij}\overline{X}_{k,ij}]/p_k$, for each $k = 1, \ldots, K$, as $K, N \to \infty$, and, under assumption A1, 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 $\overline{X}_{k,ij} = (x_j 1_k(d_{ij}), x_j(d_{ij}-z_k)1_k(d_{ij}), \ldots, x_j(d_{ij}-z_k)^q 1_k(d_{ij}))$. To show the asymptotic convergence of $\widehat{Q}_k$ to $Q_k$ as stated in Lemma 1, it is sufficient to show that $E[||\widehat{Q}_k - Q_k||^2] = o_p(1)$. Thus,

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

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

Under assumptions A1, A5 and A6, the covariance terms $cov(x_i^r, (d_{ij}-z_k)^s 1_k(d_{ij}))$ converge to zero for $r \leq 4$ and $s \leq q$, as $h \to 0$, with $K, N \to \infty$. Then, it is sufficient to study the above convergence result assuming that the variables are asymptotically mean independent. Thus after tedious algebra, the above expression can be written as

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

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

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

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

The *iid* cross-sectional assumption in A1 entails the condition $E[x_j^2x_l^2] = E[x_j^2]E[x_l^2]$, such that

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

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

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

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

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

$$
= O\left(\frac{K}{\alpha_N} + \frac{1}{N}\right),
$$

with $C_0$ some positive constant that reflects the finite character of the fourth moment of $x_i$ imposed in A1, $p_k \asymp K^{-1}$ under assumption A6, and $\sum_{r=0}^{q}\sum_{s=0}^{q} h^{2(r+s)} = \sum_{r=0}^{q} h^{2r}\sum_{s=0}^{q} h^{2s} = \left(\frac{1-h^{2(q+1)}}{1-h^2}\right)^2$. Therefore, we obtain $E[||\widehat{Q}_k - Q_k||^2] = O\left(\frac{1}{N}\right)$ such that $||\widehat{Q}_k - Q_k|| = O_p\left(\frac{1}{\sqrt{N}}\right)$.
The above expression simplifies under symmetry of the network variable. In this case $d_{ij} = d_{ji}$ for all $i, j = 1, \ldots, N$, such that

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

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

$$
\leq \frac{C_0}{N}\left(\frac{1-h^{2(q+1)}}{1-h^2}\right)^2/p_k = O\left(\frac{K}{N}\right).
$$

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

$$
\widehat{\lambda} - \lambda = \widehat{\Phi}^{-1}\frac{1}{N}X'M_{\mathbb{X}}\varepsilon + \widehat{\Phi}^{-1}\frac{1}{N}X'M_{\mathbb{X}}\overline{R}. \tag{8.1}
$$

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 $\varepsilon_i$. More formally, $\frac{1}{N^2}E\left[\left(\sum_{i=1}^{N} X_i' M_{\mathbb{X}i}\varepsilon_i\right)^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[X_i' M_{\mathbb{X}i}\varepsilon_i X_j' M_{\mathbb{X}j}\varepsilon_j\right] = \frac{1}{N}E\left[X_i' M_{\mathbb{X}i} X_i \varepsilon_i^2\right]$, 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 $\frac{1}{N^2}E\left[\left(\sum_{i=1}^{N} X_i' M_{\mathbb{X}i}\overline{R}_i\right)^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[X_i' M_{\mathbb{X}i}\overline{R}_i X_j' M_{\mathbb{X}j}\overline{R}_j\right]$. In contrast to the preceding case, there is cross-sectional dependence between the observations such that, applying the Cauchy-Schwarz inequality,

$$E\left[X_i' M_{\mathbb{X}i}\overline{R}_i X_j' M_{\mathbb{X}j}\overline{R}_j\right] \leq E\left[(X_i' M_{\mathbb{X}i} M_{\mathbb{X}j}' X_j)^2\right]^{1/2} E\left[\overline{R}_i^2 \overline{R}_j^2\right]^{1/2}. \tag{8.2}$$

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

$$E[\overline{R}_i^2 \overline{R}_j^2] = E\left[\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} R(d_{ij})x_j\right)^2 \left(\sum_{\substack{l=1 \\ l\neq j}}^{N} R(d_{jl})x_l\right)^2\right] \tag{8.3}$$

$$= \sum_{\substack{j=1 \\ j\neq i}}^{N}\sum_{\substack{l=1 \\ l\neq i}}^{N}\sum_{\substack{j'=1 \\ j'\neq j}}^{N}\sum_{\substack{l'=1 \\ l'\neq j}}^{N} E[x_j x_l x_{j'} x_{l'} R(d_{ij})R(d_{il})R(d_{jj'})R(d_{jl'})]. \tag{8.4}$$

Under assumption A1 imposing the independence between the different units, and using similar algebra to the proof of Lemma 1, the leading term is $(N-1)^2 E[x_j^2]E[x_l^2]E[R(d_{ij})^2]E[R(d_{il})^2]$, 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

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

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 $h \asymp K^{-1}$, under assumption A6. Therefore, $E\left[X_i' M_{\mathbb{X}i}\overline{R}_i X_j' M_{\mathbb{X}j}\overline{R}_j\right] = O\left(N/K^{2q+5/2}\right)$ such that $\|\frac{1}{N}X' M_{\mathbb{X}}\overline{R}\| = O_p\left(\sqrt{N}/K^{q+5/4}\right)$. Then,

$$|\widehat{\lambda} - \lambda| = O_p\left(\frac{1}{\sqrt{N}} + \frac{\sqrt{N}}{K^{q+5/4}}\right) = O_p\left(\frac{1}{\sqrt{N}}\right) = o_p(1), \tag{8.5}$$

under assumption A5.

For the second part of the proof, expression (3.2) implies that

$$\widehat{\Gamma}_k - \Gamma_k = \widehat{Q}_k^{-1}\frac{1}{\alpha_N}\mathbb{X}_k' X(\lambda - \widehat{\lambda})/p_k + \widehat{Q}_k^{-1}\frac{1}{\alpha_N}\mathbb{X}_k'\overline{R}/p_k + \widehat{Q}_k^{-1}\frac{1}{\alpha_N}\mathbb{X}_k'\varepsilon/p_k. \tag{8.6}$$

We study each right hand side term in (8.6) separately. First, using Lemma 1, $\|\widehat{Q}_k^{-1}\| = O_p(1)$. Also,

$$\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}' x_i(\lambda - \widehat{\lambda})/p_k\| = |\widehat{\lambda} - \lambda|\,\|\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}' x_i/p_k\|. \tag{8.7}$$

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, $E[x_j x_k] = 0$ and $E[x_j x_k x_i^2] = 0$, 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 $\|\widehat{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}_{ki}' X_i(\lambda - \widehat{\lambda})/p_k\| = O_p\left(\frac{\sqrt{K}}{N^{3/2}}\right)$.

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

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

Using Lemma 1, $\|\widehat{Q}_k - Q_k\| = O_p(1/\sqrt{N})$ 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 (8.3) 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,

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

Finally, we prove that $\|\widehat{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}'_{ki}\varepsilon_i/p_k\| = O_p\left(\sqrt{K}/N\right)$. 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 $\|\widehat{Q}_k^{-1}\frac{1}{\alpha_N}\sum_{i=1}^{N}\mathbb{X}'_{ki}\varepsilon_i/p_k\| = O_p\left(\frac{\sqrt{K}}{N}\right)$.

Putting together the different expressions, and for $q \geq 1$ fixed, we obtain

$$\|\widehat{\Gamma}_k - \Gamma_k\| = O_p\left(\frac{\sqrt{K}}{N^{3/2}}\right) + O_p\left(K^{-q-3/4}\right) + O_p\left(\frac{\sqrt{K}}{N}\right) = O_p\left(\frac{\sqrt{K}}{N}\right), \tag{8.8}$$

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

$$\widehat{\lambda} - \lambda = \widehat{\Phi}^{-1}\frac{1}{N}\widehat{X}'_u\varepsilon + \widehat{\Phi}^{-1}\frac{1}{N}\widehat{X}'_u\overline{R}, \tag{8.9}$$

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 (8.9) 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

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i' M_{\mathbb{X}} \varepsilon_i \to N(0, \Psi_0), \tag{8.10}$$

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:

$$\sqrt{N}\left(\widehat{\lambda} - \lambda\right) \to N(0, \Phi_0^{-1}\Psi_0\Phi_0^{-1}). \tag{8.11}$$

■

**Proof of Theorem 1.** To prove this result we combine expressions (2.4) and (3.3), 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

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

Furthermore, applying the triangular inequality and Proposition 1, this quantity is bounded by

$$\begin{aligned}\max_{\{k=1,\ldots,K\}}\left\{\|\widehat{\Gamma}_k - \Gamma_k\| \sup_{d\in\chi}|v_k(d)1_k(d)|\right\} &\leq O_p\left(\frac{\sqrt{K}}{\sqrt{N}}\right)\sum_{m=0}^{q} h^m = O_p\left(\frac{\sqrt{K}}{\sqrt{N}}\right)\frac{1-h^{q+1}}{1-h}\\ &= O_p\left(\sqrt{K}/\sqrt{N}\right).\end{aligned}$$

For the second term, $R(d) = \sum_{k=1}^{K} w^{(q+1)}(c_k)(d - z_k)^{q+1} 1_k(d)$. Then, $\sup_{d\in\chi}|R(d)| \leq C_0 \max_{\{k=1,\ldots,K\}}\{h^{q+1}\}$, 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,

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

■

**Proof of Lemma 2.** 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 1, 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, $E[x_l^2 x_{l'}^2 \varepsilon_i^2 \varepsilon_j^2] = E[x_l^2 \varepsilon_i^2] E[x_{l'}^2 \varepsilon_j^2]$. Then, the above expression is equal to

$$\frac{1}{N^2(N-1)^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N} E[x_j^4\varepsilon_i^4]E[(d_{ij}-z_k)^{2(r+s)}1_k(d_{ij})]/p_k^2\right)$$
$$+\frac{1}{N^2(N-1)^2}\sum_{r=0}^{q}\sum_{s=0}^{q}\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\sum_{\substack{l=1\\l\neq i,j}}^{N} E[x_j^2 x_l^2\varepsilon_i^4]E[(d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s}1_k(d_{ij})1_k(d_{il})]/p_k^2\right)$$
$$=\frac{1}{N(N-1)}\sum_{r=0}^{q}\sum_{s=0}^{q}E[x_j^4\varepsilon_i^4]E[(d_{ij}-z_k)^{2(r+s)} \mid 1_k(d_{ij})1_k(d_{il})]/p_k$$
$$+\frac{1}{N}\sum_{r=0}^{q}\sum_{s=0}^{q}E[x_j^2x_l^2\varepsilon_i^4]E[(d_{ij}-z_k)^{2r}(d_{il}-z_k)^{2s} \mid 1_k(d_{ij})1_k(d_{il})]$$
$$\leq O\left(\frac{K}{N^2}\right)+\frac{C_0}{N}\sum_{r=0}^{q}\sum_{s=0}^{q}h^{2(r+s)} = O\left(\frac{K}{N^2}\right)+\frac{C_0}{N}\left(\frac{1-h^{2(q+1)}}{1-h^2}\right)^2 = O\left(\frac{1}{N}\right),$$

with $C_0$ some positive constant that reflects the finite character of the first four moments of $x_i$ and $\varepsilon_i$ imposed in A1 and A4. Note also that $p_k \asymp K^{-1}$ under assumption A6. 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 3.** Expression (8.6) implies that

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

Then,

$$E[(\widehat{\Gamma}_k - \Gamma_k)^2] = \frac{1}{p_k}\widehat{Q}_k^{-1}\left(\frac{1}{\alpha_N^2}\sum_{i=1}^{N}E\left[\mathbb{X}_{ki}'\mathbb{X}_{ki}\varepsilon_i^2\right]/p_k\right)\widehat{Q}_k^{-1} \tag{8.12}$$
$$+\frac{1}{p_k}\widehat{Q}_k^{-1}\left(\frac{1}{\alpha_N^2}\sum_{i=1}^{N}E\left[\mathbb{X}_{ki}'x_i(\widehat{\lambda}-\lambda)^2x_i'\mathbb{X}_{ki}\right]/p_k\right)\widehat{Q}_k^{-1} \tag{8.13}$$
$$-\frac{1}{p_k}\widehat{Q}_k^{-1}\left(\frac{1}{\alpha_N^2}\sum_{i=1}^{N}E\left[\mathbb{X}_{ki}'\varepsilon_i(\widehat{\lambda}-\lambda)x_i'\mathbb{X}_{ki}\right]/p_k\right)\widehat{Q}_k^{-1} + O_p\left(K^{-2q-3/2}\right). \tag{8.14}$$

Using Lemma 1 and the definition of $A_k$, expression (8.12) 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 (8.13) 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 $\Phi_0$ 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,

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

The asymptotic convergence of expression (8.14) is studied in a similar fashion. More specifically, replacing expression (8.9):

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

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 network 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 1 and 2 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 (3.6), 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 1 and 2, 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 (2.4) and (3.3), and obtain

$$\sqrt{\alpha_N}(\widehat{w}(d) - w(d)) = \sum_{k=1}^{K}\sqrt{\alpha_N}\left(\widehat{\Gamma}_k - \Gamma_k\right)'v_k(d)1_k(d) - \sqrt{\alpha_N}R(d). \tag{8.15}$$

Using the result in Lemma 3, 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

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

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

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 (8.6) and (8.8) 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 (8.15), we obtain

$$\sqrt{\alpha_N}(\widehat{w}(d)-w(d)) = \frac{1}{\sqrt{\alpha_N}}\sum_{i=1}^{N}\sum_{k=1}^{K} v_k(d)'\widehat{Q}_k^{-1}\mathbb{X}_{ki}'\varepsilon_i 1_k(d)/p_k + O_p\left(\frac{\sqrt{K}}{\sqrt{N}}\right) - \sqrt{\alpha_N}R(d). \tag{8.16}$$

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 $V_K = \sum_{k=1}^{K} v_k(d)'Q_k^{-1}A_kQ_k^{-1}v_k(d)1_k(d)/p_k$. The process $\{z_{iN}(d)\}_{i=1}^N$ inherits the properties of the error term $\varepsilon_i$, 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

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

■

Table 1: Bias of estimators of $\lambda$ and $w(d)$ in (3.1) and (3.3).

| | | | | Model 1: Exponential function | | | | | Model 2: Gaussian kernel function | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | h | $\theta$ | N | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ |
| 1 | 0.05 | 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 |
| | 0.1 | 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 | 0.05 | 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 |
| | 0.1 | 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.

Table 2: Root mean square error of estimators of $\lambda$ and $w(d)$ in (3.1) and (3.3).

| | | | | Model 1: Exponential function | | | | | Model 2: Gaussian kernel function | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | h | $\theta$ | N | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ |
| 1 | 0.05 | 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 |
| | 0.1 | 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 | 0.05 | 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 |
| | 0.1 | 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.

Table 3: Empirical coverage rates for confidence interval (6.2) at an $\alpha = 0.05$ significance level (coverage rate).

| | | | Model 1: Exponential function | | | | | Model 2: Gaussian kernel function | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| h | $\theta$ | N | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ |
| 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 $w(d)$ considering 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.

Table 4: Empirical coverage rates for confidence interval (6.2) at an $\alpha = 0.05$ significance level (coverage rate).

| | | | Model 1: Exponential function | | | | | Model 2: Gaussian kernel function | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| h | $\theta$ | N | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ |
| 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 $w(d)$ considering 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.

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

| | | | Model 1: Exponential function | | | | | Model 2: Gaussian kernel function | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| h | $\theta$ | N | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ |
| 0.05 | 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 |
| 0.10 | 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 rejection rates of marginal t-tests for the null hypothesis $H_0 : w(d) = 0$ against the alternative $H_A : w(d) \neq 0$. The data has been generated under the alternative hypothesis considering two specifications for the functional parameter: Exponential and Gaussian functions, for $d \in [0, C]$, with $C = 1$. The number of intervals characterizing the partition is $K = 1/2h$. The DGP is given by $\gamma = 1$ and the number of simulations is 500.

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

| | | | Model 1: Exponential function | | | | | Model 2: Gaussian kernel function | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| h | $\theta$ | N | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ | $\lambda$ | $w(h/2)$ | $w(h)$ | $w(3h/2)$ | $w(2h)$ |
| 0.05 | 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 |
| 0.10 | 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 rejection rates of marginal t-tests for the null hypothesis $H_0 : w(d) = 0$ against the alternative $H_A : w(d) \neq 0$. The data has been generated under the alternative hypothesis considering two specifications for the functional parameter: Exponential and Gaussian functions, for $d \in [0, C]$, with $C = 1$. The number of intervals characterizing the partition is $K = 1/2h$. The DGP is given by $\gamma = 0.25$ and the number of simulations is 500.

Table 7: Empirical size and power of the uniform test (5.1).

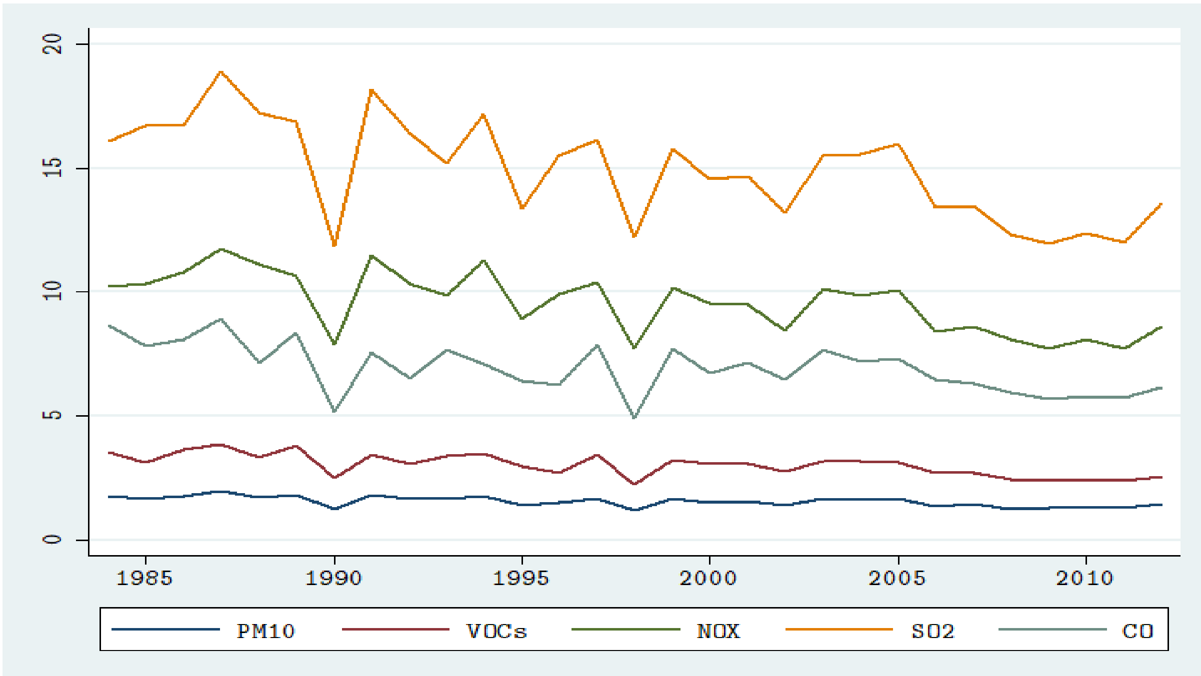
| | | $H_0$: No network structure | | | | | |
|---|---|---|---|---|---|---|---|
| | | Size | | | Power | | |
| $\theta$ | N | h=0.05 | 0.075 | 0.1 | h=0.05 | 0.075 | 0.1 |
| 5 | 100 | 0.071 | 0.056 | 0.065 | 0.998 | 0.998 | 1.000 |
| | 250 | 0.049 | 0.052 | 0.036 | 1.000 | 1.000 | 1.000 |
| | 500 | 0.073 | 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 | | |
| $\theta$ | N | h=0.05 | 0.075 | 0.1 | h=0.05 | 0.075 | 0.1 |
| 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 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.

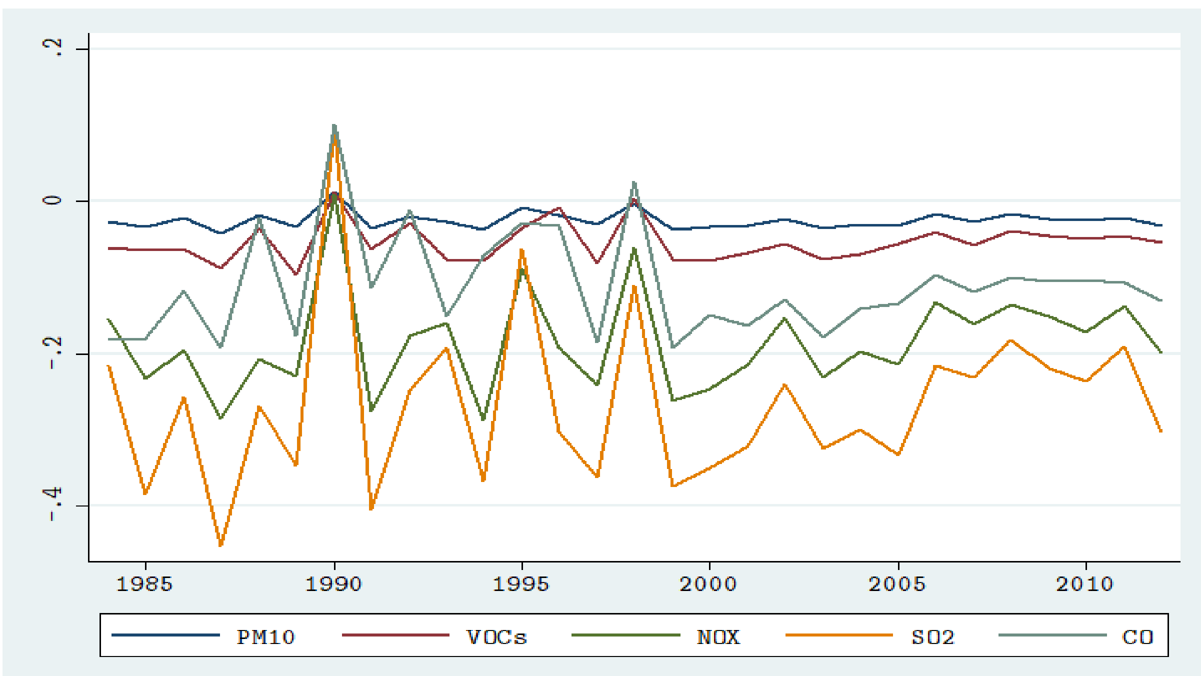
Table 8: Optimal choice of the tuning parameter $h$.

| $\theta$ | N | RMSE | M | GCV | AIC | BIC |
|---|---|---|---|---|---|---|
| 5 | 100 | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| | | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| | 250 | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| | | (0.006) | (0.006) | (0.006) | (0.006) | (0.005) |
| | 500 | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| | | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| 9 | 100 | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| | | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| | 250 | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| | | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| | 500 | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| | | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |

Note: This table reports the optimal value of the tuning parameter $h$ under the different criteria described in Subsection 5.2. The network regression model is (2.9), with $\gamma = 1$, $w(d)$ given by $\beta = 0.1$, and an exponential function. Standard errors are shown in parentheses. The number of simulations is 500.

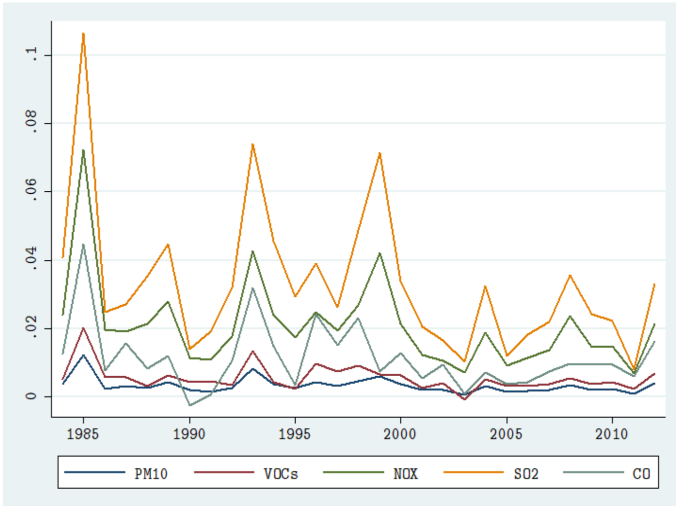



(a) Coefficient for after-tax household income ($\gamma_{1t,net}$).

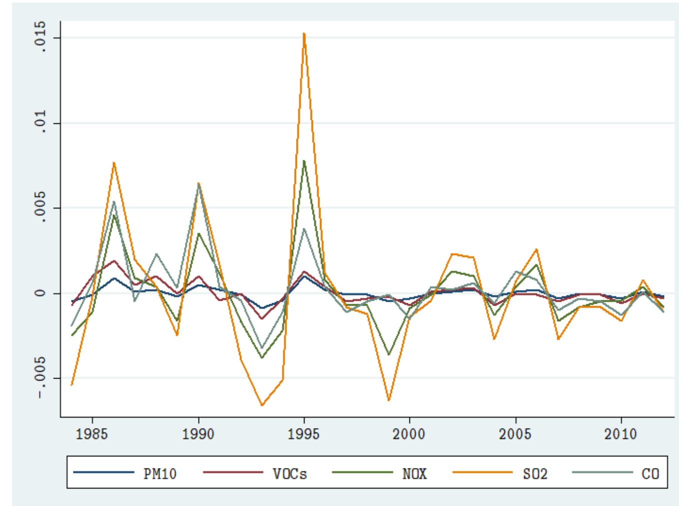



(b) Coefficient for squared after-tax household income ($\gamma_{2t,net}$).

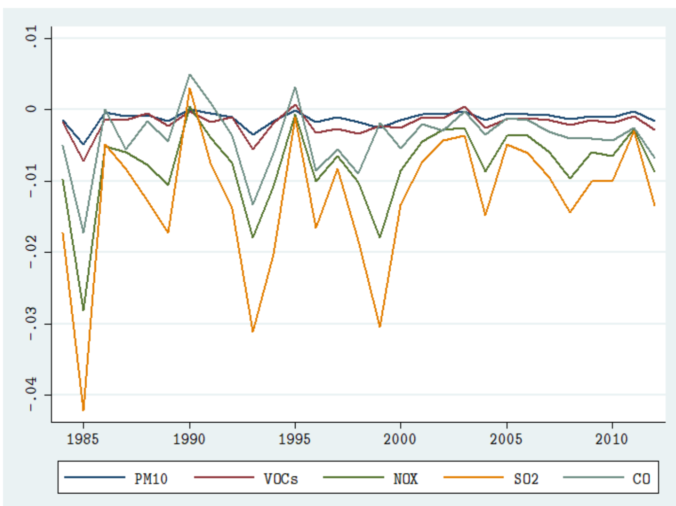
Figure 1: Network regression estimation of environmental Engel curves (EECs) in the U.S., 1984-2012.



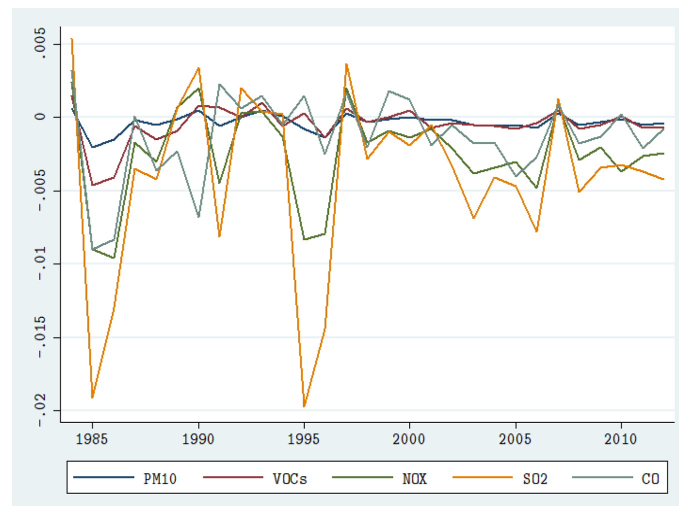

(a) $w(h/2)$




(b) $w(h)$




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




(d) $w(2h)$

Figure 2: Network regression estimation of EECs in the U.S., 1984-2012. Functional coefficient.

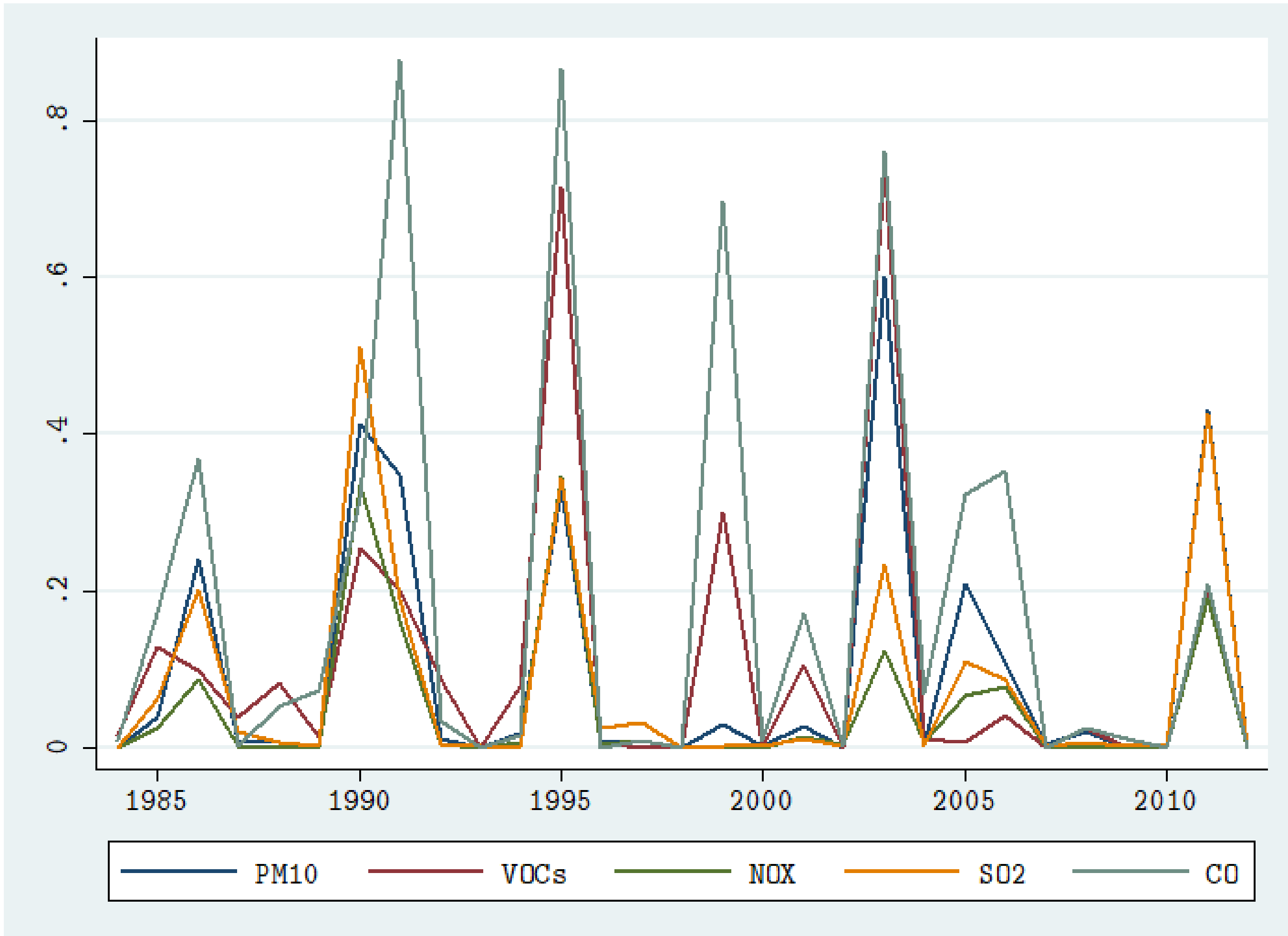
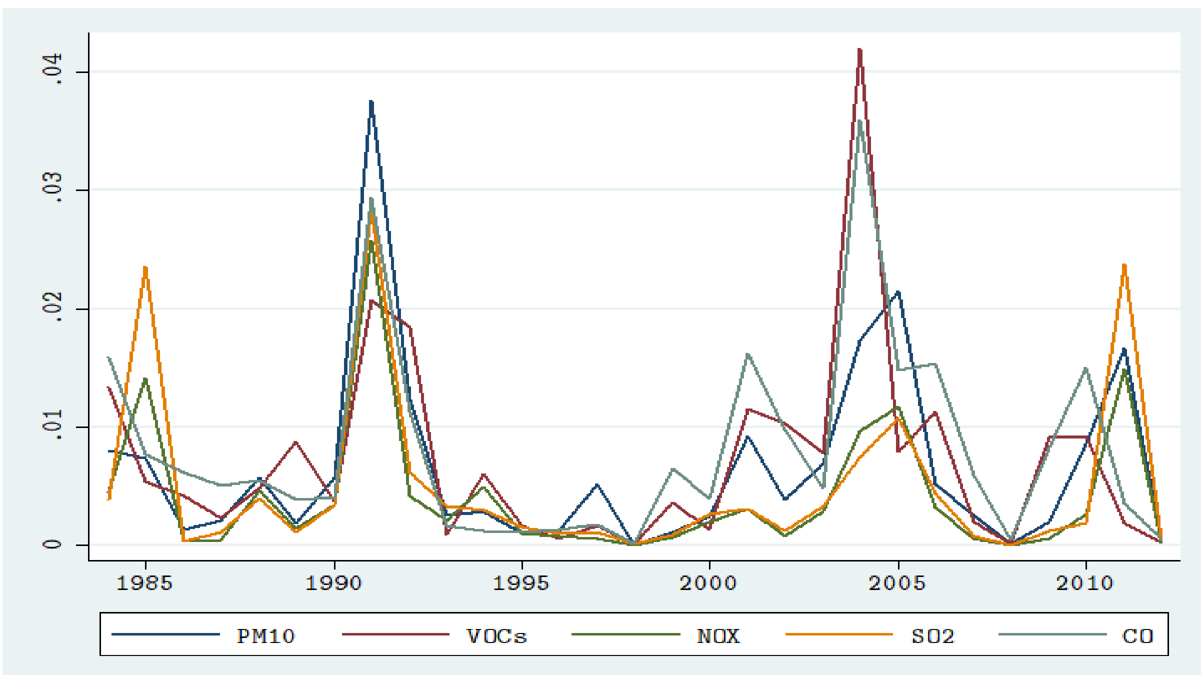



Figure 3: Network regression estimation of the functional coefficient. Uniform test statistic, p-values.


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




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

Figure 4: Network regression estimation of EECs in the U.S., 1984-2012.