# CONSISTENT SPECIFICATION TESTING UNDER NETWORK DEPENDENCE*

By Abhimanyu Gupta† and Xi Qu

*University of Essex and Shanghai Jiao Tong University*

We propose a series-based nonparametric specification test for a regression function when data are dependent across a network. Network dependence can be parametric, parametric with increasing dimension, semiparametric or any combination thereof, thus covering a vast variety of settings. These include spatial error models of varying types and levels of complexity. Models in which network dependence arises directly in outcome variables are also treated. Under a new smooth network dependence condition, our test statistic is asymptotically standard normal. To prove the latter property, we establish a central limit theorem for quadratic forms in linear processes in an increasing dimension setting. Finite sample performance is investigated in a simulation study and empirical examples illustrate the test with real-world data.

**1. Introduction.** Models for network dependence have recently become the subject of vigorous research in the statistical and econometric literature, see e.g. Kolaczyk (2017), the volume edited by Bramoullé, Galeotti, and Rogers (2016) and the review article by de Paula (2017). This burgeoning interest has roots both in the requirement of theoretical models to understand network formation mechanisms as well as the needs of practitioners who frequently have access to data sets featuring inter-connected cross-sectional units. Motivated by these practical concerns, we propose a specification test for a regression function in a general setup that covers a vast variety of commonly employed network dependence models and permits the complexity of network dependence to increase with sample size.

*We thank Swati Chandna, Miguel Delgado, Emmanuel Guerre, Fernando López Hernandéz (discussant), Hon Ho Kwok, Arthur Lewbel, Daisuke Murakami (discussant), Ryo Okui and Amol Sasane for helpful discussions. We also thank seminar participants at YEAP 2018 (Shanghai University of Finance and Economics), NYU Shanghai, Carlos III Madrid, SEW 2018 (Dijon), Aarhus University, SEA 2018 (Vienna), EcoSta 2018 (Hong Kong), Hong Kong University, AFES 2018 (Cotonou), ESEM 2018 (Cologne) and CFE 2018 (Pisa).

†Research supported by ESRC grant ES/R006032/1.

Primary 62G10, 62G08, 62G20; Secondary 91B72

*Keywords and phrases:* Specification testing, nonparametric regression, networks, spatial dependence, cross-sectional dependence

Our test is consistent, in the sense that a parametric specification is tested with asymptotically unit power against a nonparametric alternative.

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

However, agents may form networks in many ways that do not conform to such a setting. For example, farmers influence the demand of farmers in the same village but not in different villages, as in Case (1991). A similar setting is used by Dray and Jombart (2011) to study a famous 19th century data set on 'moral statistics', such as crime, suicides and literacy in French départements. Likewise, price competition among firms exhibits spatial features (Pinkse, Slade, and Brett (2002)), input-output relations lead to complementarities between sectors (Conley and Dupor (2003)), co-author networks form among scientists (Oettl (2012), Mohnen (2017)), online bloggers form networks by political allegiance (Zhao, Levina, and Zhu (2012)), R&D spillovers occur via technology and product market networks (Bloom, Schankerman, and van Reenen (2013)), children influence the behaviour of their peers (Helmers and Patnam (2014)), networks form due to allegiances in conflicts (König, Rohner, Thoenig, and Zilibotti (2017)), human brain regions may influence each other (Durante, Dunson, and Vogelstein (2017)), overlapping bank portfolios lead to correlated lending decisions (Gupta, Kokas, and Michaelides (2018)). Such examples cannot be studied by simply extending results developed for time series and illustrate the growing need for methods that are valid when agents affect each other through a network.

A very popular model for general network dependence is the spatial autoregressive (SAR) class, due to Cliff and Ord (1973). The key feature of SAR models, and various generalizations such as SARMA (SAR moving average) and matrix exponential spatial specifications (MESS, due to LeSage and Pace (2007) and studied further by Debarsy, Jin, and Lee (2015)), is the presence of one or more spatial weight matrices whose elements characterize the network links between agents. These links may form for a variety of reasons, so the 'spatial' terminology represents a very general notion of space, such as social or economic space. As an example, consider the Add Health dataset. This comprises a sample of schoolchildren from 112 middle and high-schools in the US. Students are asked to name up to five friends of each sex from a list of all other students in the school; see e.g. Moody (2001) for more details on the dataset and Olhede and Wolfe (2014) for an exclusively network oriented analysis. Thus an example of a spatial weight matrix would be a link matrix based on these self-reported friendships. However, the dataset also includes information on the ethnicity of the students, used e.g. by Currarini, Jackson, and Pin (2009) in a study of homophily, as well as the school year (grade) of each student. Each of these two channels generates a new spatial weight matrix, suggesting the use of more than one spatial weight matrix in what are sometimes termed 'higher-order' models. Important papers on the estimation of SAR models and their variants include Kelejian and Prucha (1998) and Lee (2004), but research on various aspects of these is active and ongoing both in econometrics and statistics, see e.g. Kuersteiner and Prucha (2013), Sun, Yan, Zhang, and Lu (2014), Hillier and Martellosio (2018).

Unlike work focusing on independent or time series data, a general drawback of network or spatially oriented research has been the lack of general unified theory. Typically, individual papers have studied specific special cases of various spatial specifications. A strand of the literature has introduced the notion of a cross-sectional, or network, linear-process to help address this problem, and we follow this approach. This representation can accommodate SAR models in the error term (so called spatial error models (SEM)) as a special case, as well as variants like SARMA and MESS, whence its generality is apparent. The linear-process structure shares some similarities with that familiar from the time series literature (see e.g. Hannan (1970)). Indeed, time series versions may be regarded as very special cases but, as stressed before, the features of network dependence must be taken into account in the general formulation. Such a representation was introduced by Robinson (2011) and further examined in other situations by Robinson and Thawornkaiwong (2012) (partially linear regression), Delgado and Robinson (2015) (non-nested correlation testing), Lee and Robinson (2016) (series estimation of nonparametric regression), Peng (2016) and Bandyopadhyay and Maity (2018) (varying coefficient models) and Hidalgo and Schafgans (2017) (cross-sectionally dependent panels).

Examples of nonparametric specification tests with independent data include conditional moments tests (Bierens (1990)), sample splitting tests (Eubank and Spiegelman (1990), Wooldridge (1992), Yatchew (1992)), tests for non-nested models (Delgado and Stengos (1994)), series-based tests (Hong and White (1995), Li, Hsiao, and Zinn (2003)), kernel-based tests (Fan and Li (1996), Zheng (1996)), tests based on general linear smoothers (Ellison and Ellison (2000)), rate-optimal tests (Guerre and Lavergne (2005)), instrumental variables based tests (Horowitz (2006)), tests based on Cramér-von Mises distance (Rothe and Wied (2013)) and tests based on nearest-neighbour estimation (Li, Li, and Liu (2016)). More generally, Tripathi and Kitamura (2003) propose a test of conditional moment restrictions that nests (conditional mean) regression as a special case. With time series dependence, examples include empirical likelihood based testing under $\alpha$-mixing (Chen and Wolfgang (2003)), testing with possible long memory (Delgado, Hidalgo, and Velasco (2005)), continuous-time models with finance applications (Hong and Li (2005)), supremum based testing (Hidalgo (2008)), testing under nonlinearity and nonstationarity (Gao, King, Lu, and Tjøstheim (2009) and Wang and Phillips (2012)), tests for multivariate time series (Yajima and Matsuda (2009)), optimal tests (Hong and Lee (2013)) and tests using marked empirical processes (Escanciano (2018)).

In this paper, we propose a test statistic similar to that of Hong and White (1995), based on estimating the nonparametric specification via series approximations. Assuming an independent and identically distributed sample, their statistic is based on the sample covariance between the residual from the parametric model and the discrepancy between the parametric and nonparametric fitted values. Allowing additionally for network dependence through the form of a linear process as discussed above, our statistic is shown to be asymptotically standard normal, consistent and possessing nontrivial power against local alternatives of a certain type. To prove asymptotic normality, we present a new central limit theorem (CLT) for quadratic forms in linear processes in an increasing dimension setting that may be of independent interest.

Our linear process framework permits network dependence to be parametric, parametric with increasing dimension, semiparametric or any combination thereof, thus covering a vast variety of settings. Our theory covers as special cases SAR, SMA, SARMA, MESS models for the error term. These specifications may be of any fixed order, but our theory also covers the case where they are of increasing order. Thus we permit a more complex model of network dependence as more data become available, which encourages a more flexible approach to modelling such dependence as stressed by

Gupta and Robinson (2015, 2018) in a higher-order SAR context, Huber (1973), Portnoy (1984, 1985) and Anatolyev (2012) in a regression context and Koenker and Machado (1999) for the generalized method of moments setting, amongst others. Our framework is also extended to the situation where network dependence occurs through nonparametric functions of raw distances (these may economic or social distances, say), as in Pinkse et al. (2002), and where once again an increasing number of nonparametric functions may be involved. We also cover the case of higher-order SAR models in the outcome variables, but with network dependence structure in the errors. The specification test of Su and Qu (2017) is a special case. The case of geographical networks, or spatial data, is also covered, for example the important classes of Matern and Wendland (see e.g. Gneiting (2002), Bevilacqua, Faouzi, Furrer, and Porcu (2018)) covariance functions. Finally, we introduce a new notion of smooth network dependence that provides more primitive, and checkable, conditions for certain properties than extant ones in the literature.

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

The next section introduces our setup, while Section 3 discusses estimation and defines the test statistic. In Section 4, we introduce assumptions and the key asymptotic results of the paper. Section 5 examines higher-order SAR models with further network dependence in the error term, while Section 6 deals with nonparametric networks. Sections 7 and 8 contain a study of finite sample performance and the empirical examples respectively. Proofs are contained in supplementary appendices.

## 2. Setup.

We consider the nonparametric regression

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

where $\theta_0(\cdot)$ is an unknown function and $x_i$ is a vector of strictly exogenous explanatory variables with support $\mathcal{X} \subset \mathbb{R}^k$. The $x_i$ can exhibit dependence in arbitrary ways and this does not affect our asymptotic theory, which is established conditional on exogenous covariates. Network dependence is explicitly modeled via the error term $u_i$, which we assume is generated by:

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

where $\varepsilon_s$ are independent random variables, with zero mean and identical variance $\sigma_0^2$. Further conditions on the $\epsilon_s$ will be assumed later. The linear process coefficients $b_{is}$ can depend on $n$, as may the covariates $x_i$. This is generally the case with network/spatial models and implies that asymptotic theory ought to be developed for triangular arrays. There are a number of reasons to permit dependence on sample size. The $b_{is}$ can depend on spatial weight matrices, which are usually normalized for both stability and identification purposes. Such normalizations, e.g. row-standardization or division by spectral norm, may be $n$-dependent. Furthermore, $x_i$ often includes underlying covariates of 'neighbours' defined by spatial weight matrices. For instance, for some $n \times 1$ covariate vector $z$ and spatial weight matrix $W \equiv W_n$, a component of $x_i$ can be $e_i'Wz$, where $e_i$ has unity in the $i$-th position and zeroes elsewhere, which depends on $n$. Thus, subsequently, any spatial weight matrices will also be allowed to depend on $n$. Finally, treating triangular arrays permits re-labelling of quantities that is often required when dealing with network data, due to the lack of natural ordering, see e.g. Robinson (2011). We suppress explicit reference to this $n$-dependence of various quantities for brevity, although mention will be made of this at times to remind the reader of this feature.

Introduce three notational conventions for any parameter $\nu$ for the rest of the paper: $\nu \in \mathbb{R}^{d_\nu}$, $\nu_0$ denotes the true value of $\nu$ and for any scalar, vector or matrix valued function $f(\nu)$, we denote $f \equiv f(\nu_0)$. Now, assume the existence of a $d_\gamma \times 1$ vector $\gamma_0$ such that $b_{is} = b_{is}(\gamma_0)$, with $d_\gamma \to \infty$ as $n \to \infty$, for all $i = 1, \ldots, n$ and $s \geq 1$. Let $u$ be the $n \times 1$ vector with typical element $u_i$, $\varepsilon$ be the infinite dimensional vector with typical element $\varepsilon_s$, and $B$ be an infinite dimensional matrix with typical element $b_{is}$. In matrix form,

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

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

$$H_0 : P[f(x_i, \alpha_0) = \theta_0\left(x_i\right)] = 1, \text{ for some } \alpha_0 \in \mathcal{A}, \tag{2.4}$$

where $\mathcal{A} \subset \mathbb{R}^{d_\alpha}$ against the global alternative $H_1 : P\left[f(x_i, \alpha) \neq \theta_0\left(x_i\right)\right] > 0,\ \text{for all } \alpha \in A$.

We now show how commonly used models for network dependence may be nested in (2.3). Introduce a set of $n \times n$ network adjacency/spatial weight matrices $W_j$, $j = 1, \ldots, m_1 + m_2$. Each $W_j$ can be thought of as representing dependence through a particular network. Recall for example the Add Health dataset discussed in the introduction, where each link may be thought of as a channel of network dependence and a $W_j$ constructed accordingly. Now, denote $H(\gamma) = I_n + \sum_{j=m_1+1}^{m_1+m_2} \gamma_j W_j$ and $K(\gamma) = I_n - \sum_{j=1}^{m_1} \gamma_j W_j$, and consider models with the form $\Sigma(\gamma) = A^{-1}(\gamma) A'^{-1}(\gamma)$. For example, with $\xi$ denoting a vector of i.i.d. disturbances with variance $\sigma_0^2$, the model with SARMA$(m_1, m_2)$ errors is $u = \sum_{j=1}^{m_1} \gamma_j W_j u + \sum_{j=m_1+1}^{m_1+m_2} \gamma_j W_j \xi + \xi, A(\gamma) = H^{-1}(\gamma) K(\gamma)$, The SEM model is obtained by setting $m_2 = 0$ while the the model with SMA errors has $m_1 = 0$. The model with MESS$(m)$ errors (LeSage and Pace (2007), Debarsy et al. (2015)) is $u = \exp\left(\sum_{j=1}^{m} \gamma_j W_j\right) \xi, A(\gamma) = \exp\left(-\sum_{j=1}^{m} \gamma_j W_j\right)$.

In some cases the network under consideration is geographic i.e. the data may be observed at irregular points in Euclidean space. Making the identification $u_i \equiv U(t_i)$, $t_i \in \mathbb{R}^d$ for some $d > 1$, and assuming covariance stationarity, $U(t)$ is said to follow an isotropic model if, for some function $\delta$ on $\mathbb{R}$, the covariance at lag $s$ is $r(s) = E[U(t)U(t+s)] = \delta(\|s\|)$. An important class of parametric isotropic models is that of Matern (1986), which can be parameterized in several ways, see e.g. Stein (1999). Denoting by $\Gamma_f$ the Gamma function and by $\mathcal{K}_{\gamma_1}$ the modified Bessel function of the second kind (Gradshteyn and Ryzhik (1994)), take $\delta(\|s\|, \gamma) = \left(2^{\gamma_1 - 1}\Gamma_f(\gamma_1)\right)^{-1} \left(\gamma_2^{-1}\sqrt{2\gamma_1}\|s\|\right)^{\gamma_1} \mathcal{K}_{\gamma_1}\left(\gamma_2^{-1}\sqrt{2\gamma_1}\|s\|\right)$, with $\gamma_1, \gamma_2 > 0$ and $d_\gamma = 2$. With $d_\gamma = 3$, another model takes $\delta(\|s\|, \gamma) = \gamma_1 \exp\left(-\|s/\gamma_2\|^{\gamma_3}\right)$, see e.g. De Oliveira, Kedem, and Short (1997), Stein (1999). Fuentes (2007) considers this model with $\gamma_3 = 1$, as well as a specific parameterization of the Matern covariance function. Other parameterizations also covered by our setup are available in a number of papers focussed on irregular spatial data, see e.g. Vecchia (1988), Jones and Vecchia (1993), Handcock and Wallis (1994), Gneiting (2002), Stein, Chi, and Welty (2004), Lindgren, Rue, and Lindström (2011), Bevilacqua et al. (2018).

## 3. Test statistic.

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

$n \to \infty$, and a $p \times 1$ vector of coefficients $\beta_0$ such that

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

where $e(\cdot)$ satisfies:

Assumption R.1. *There exists $\mu > 0$ such that $\|e\|_{w_x} = O(p^{-\mu})$, as $p \to \infty$, where $w_x \geq 0$ be the largest value such that $\sup_{x \in \mathcal{X}} E\|x\|^{w_x} < \infty$.*

By Lemma A.4 of Lee and Robinson (2016), this assumption implies that

$$\sup_{x \in \mathcal{X}} E\left(e^2(x)\right) = O\left(p^{-2\mu}\right). \tag{3.2}$$

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

Let $y = (y_1, \ldots, y_n)', \theta_0 = (\theta_0(x_1), \ldots, \theta_0(x_n))', \Psi = (\psi_1', \ldots, \psi_n')'$. We will estimate $\gamma_0$ using a quasi maximum likelihood estimator (QMLE) based on a Gaussian likelihood, although Gaussianity is nowhere assumed. For any admissible values $\beta$, $\sigma^2$ and $\gamma$, the (multiplied by $2/n$) negative quasi log likelihood function based on using the approximation (3.1) is

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

which is minimised with respect to $\beta$ and $\sigma^2$ by

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

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

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

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

We define the QMLE of $\gamma_0$ as $\widehat{\gamma} = \arg\min_{\gamma\in\Gamma}\mathcal{L}(\gamma)$ and the QMLEs of $\beta_0$ and $\sigma_0^2$ as $\widehat{\beta} = \bar{\beta}(\widehat{\gamma})$ and $\widehat{\sigma}^2 = \bar{\sigma}^2(\widehat{\gamma})$. For a given (suppressed) $x$, the series estimate of $\theta_0(x)$ is defined as

$$\widehat{\theta} = \psi(x)'\widehat{\beta}. \tag{3.7}$$

Let $\widehat{\alpha}_n \equiv \widehat{\alpha}$ denote an estimator consistent for $\alpha_0$ under $H_0$. Note that $\widehat{\alpha}$ is consistent only under $H_0$, so we introduce a general probability limit of $\widehat{\alpha}$.

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

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

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

The motivation for such a centering and scaling stems from the fact that for fixed $p$, $n\widehat{m}_n$ has an asymptotic $\chi_p^2$ distribution. Such a distribution has mean $p$ and variance $2p$, and it is a well-known fact that $\left(\chi_p^2 - p\right)/\sqrt{2p} \xrightarrow{d} N(0,1),$ as $p \to \infty$, a type of Wilks phenomenon, see e.g. Fan, Zhang, and Zhang (2001). This motivates our use of (3.8) and explains why we aspire to establish a standard normal distribution under the null hypothesis.

## 4. Asymptotic theory.

4.1. *Consistency of $\widehat{\gamma}$.* We first provide conditions under which our estimator $\widehat{\gamma}$ of $\gamma_0$ is consistent. Such a property is necessary for the results that follow. Let $\overline{\varphi}(A)$ (respectively $\underline{\varphi}(A)$) denote the largest (respectively smallest) eigenvalue of a generic square nonnegative definite matrix $A$. The following assumption is a rather standard type of asymptotic boundedness and full-rank condition on $\Sigma(\gamma)$.

Assumption R.3.

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

Assumption R.4. *The $u_i, i = 1, \ldots, n,$ satisfy the representation (2.2). The $\varepsilon_s$, $s \geq 1$, have zero mean, finite third and fourth moments $\mu_3$ and $\mu_4$ respectively and, denoting by $\sigma_{ij}(\gamma)$ the $(i,j)$-th element of $\Sigma(\gamma)$ and defining $b_{is}^* = b_{is}/\sigma_{ii}^{\frac{1}{2}},\ i = 1, \ldots, n,\ n \geq 1, s \geq 1,$ we have*

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

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

Because we often need to consider the difference between values of the matrix-valued function $\Sigma(\cdot)$ at distinct points, it is useful to introduce an appropriate concept of 'smoothness'. This concept occurs in functional analysis and is defined below.

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

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

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

The above definition extends the notion of a derivative that is familiar from real analysis to the functional spaces and allows us to check high-level assumptions that past literature has imposed. To the best of our knowledge, this is the first use of such a concept in the literature on spatial/network models. Denote by $\mathcal{M}^{n\times n}$ the set of real, symmetric and positive semi-definite $n \times n$ matrices. For a generic matrix $A$, denote $\|A\| = [\overline{\varphi}(A'A)]^{1/2}$, i.e. the spectral norm of $A$. Let $\Gamma^o$ be an open subset of $\Gamma$ and consider the Banach spaces $\left(\Gamma, \|\cdot\|_g\right)$ and $(\mathcal{M}^{n\times n}, \|\cdot\|)$, where $\|\cdot\|_g$ is a generic $\ell_p$ norm, $p \geq 1$. Denote by $c$ $(C)$ generic positive constants, independent of $p$, $d_\gamma$ and $n$ and arbitrarily small (big). The following assumption ensures that $\Sigma(\cdot)$ is a 'smooth' function, in the sense of Fréchet-smoothness.

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

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

Assumption R.5 is a functional smoothness condition on network dependence. It has the advantage of being checkable for a variety of commonly employed network models. For example, a first-order SEM has $\Sigma(\gamma) = K^{-1}(\gamma)K'^{-1}(\gamma)$ with $K = I_n - \gamma W$. Corollary CS.1 in the supplementary appendix shows $(D\Sigma(\gamma))\left(\gamma^{\dagger}\right) = \gamma^{\dagger}K^{-1}(\gamma)\left(G'(\gamma)+G(\gamma)\right)K'^{-1}(\gamma)$, at a given point $\gamma \in \Gamma^o$, where $G(\gamma) = WK^{-1}(\gamma)$. Then, taking

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

yields Assumption R.5. Condition (4.4) limits the extent of network dependence and is very standard in the spatial literature; see e.g. Lee (2004) and numerous subsequent papers employing similar conditions. Fréchet derivatives for higher-order SAR, SMA, SARMA and MESS error structures are computed in supplementary appendix B, in Lemmas LS.5-LS.6 and Corollaries CS.1-CS.2. The following proposition is very useful in 'linearizing' perturbations in the $\Sigma(\cdot)$.

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

$$\left\|\Sigma\left(\gamma_1\right) - \Sigma\left(\gamma_2\right)\right\| \le C\left\|\gamma_1 - \gamma_2\right\|. \tag{4.5}$$

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

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

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

We now introduce regularity conditions needed to establish the consistency of $\hat{\gamma}$. For a generic matrix, let $\|A\|_F = \left[tr(AA')\right]^{1/2}$. Define $\sigma^2(\gamma) = n^{-1}\sigma^2 tr\left(\Sigma(\gamma)^{-1}\Sigma\right) = n^{-1}\sigma^2\left\|C(\gamma)C^{-1}\right\|_F^2$, which is nonnegative by definition and bounded by Assumption R.3.

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

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

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

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

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

Assumption R.6 is a boundedness condition originally considered in Gupta and Robinson (2018), while Assumptions R.7 and R.8 are identification conditions. Indeed, Assumption R.7 requires that $\Sigma(\gamma)$ be identifiable in a small neighbourhood around $\gamma_0$. This is apparent on noticing that the ratio in (4.7) is at least one by the inequality between arithmetic and geometric means, and equals one when $\Sigma(\gamma)=\Sigma$. Similar assumptions arise frequently in related literature, see e.g. Lee (2004), Delgado and Robinson (2015). Assumption R.8 is a typical asymptotic boundedness and non-multicollinearity condition, see e.g. Newey (1997) and much other literature on series estimation. By Assumption R.3, it implies $\sup_{\gamma\in\Gamma}\left\{\underline{\varphi}\left(n^{-1}\Psi'\Sigma(\gamma)^{-1}\Psi\right)\right\}^{-1}=O_p(1)$.

THEOREM 4.1. *Under Assumptions R.1-R.8 and* $p^{-1}+d_\gamma^{-1}+\left(d_\gamma+p\right)/n\to 0$ *as* $n\to\infty$, $\|\widehat{\gamma}-\gamma_0\|\xrightarrow{p}0$.

4.2. *Asymptotic properties of the test statistic.* Write $\Sigma_j(\gamma)=\partial\Sigma(\gamma)/\partial\gamma_j$, $j=1,\ldots,d_\gamma$, the matrix differentiated element-wise.

ASSUMPTION R.9. $\overline{\lim}_{n\to\infty}\sup_{j=1,\ldots,d_\gamma}\|\Sigma_j\|<\infty$.

We will later consider the sequence of local alternatives

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

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

ASSUMPTION R.10. *For each* $n\in\mathbb{N}$, *the function* $f:\mathcal{X}\times\mathcal{A}\to\mathbb{R}$ *is such that* $f(\cdot,\alpha)$ *is measurable for each* $\alpha\in\mathcal{A}$, $f(x,\cdot)$ *is a.s. continuous on* $\mathcal{A}$, *with* $f^2(x,\cdot)\le D_n(x)$, *where* $\sup_{n\in\mathbb{N}}D_n(x)$ *is integrable and* $\sup_{\alpha\in\mathcal{A}}\left(\partial f(x,\alpha)/\partial\alpha\right)^2$, $\sup_{\alpha\in\mathcal{A}}\partial^2 f(x,\alpha)/\partial\alpha\partial\alpha'$ *both exist and are bounded in spectral norm by* $D_n(x)$.

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

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

Denote by $\|A\|_R$ the maximum absolute row sum of a generic matrix $A$.

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

Because $\left\|\Sigma^{-1}\right\| \leq \left\|\Sigma^{-1}\right\|_R$, this restriction on network dependence is somewhat stronger than a restriction on spectral norm but is typically imposed for central limit theorems in this type of setting, cf. Lee (2004), Delgado and Robinson (2015), Gupta and Robinson (2018). The next assumption is needed in our proofs to check a Lyapunov condition.

ASSUMPTION R.12. *The $\varepsilon_s$, $s \geq 1$, have finite eighth moment.*

ASSUMPTION R.13. $E\left|\psi\left(x_{jl}\right)\right| < C$.

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

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

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

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

THEOREM 4.4. *Under the conditions of Theorems 4.2 and 4.3, (1) $\mathscr{T} \xrightarrow{d} N(0,1)$ under $H_0$, (2) $\mathscr{T}$ is a consistent test statistic (i.e. it has asymptotically unit power under $H_1$), (3) $\mathscr{T} \xrightarrow{d} N\left(\varkappa, 1\right)$ under local alternatives $H_\ell$.*

**5. Models with SAR structure in responses.** In this section, we consider the model

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

where $W_j$, $j = 1, \ldots, d_\lambda$, are known spatial weight matrices with $i$-th rows denoted $w'_{i,j}$, as discussed earlier, and $\lambda_{0j}$ are unknown parameters measuring the strength of network dependence. The error structure remains the same as in (2.2). Here network dependence arises not only in errors but also responses. For example, this corresponds to a situation where agents in a network influence each other both in their observed and unobserved actions.

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

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

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

at any admissible point $\left(\beta', \phi', \sigma^2\right)'$ with $\phi = (\lambda', \gamma')'$, for nonsingular $S(\lambda)$ and $\Sigma(\gamma)$. For given $\phi = (\lambda', \gamma')'$, (3.3) is minimised with respect to $\beta$ and $\sigma^2$ by

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

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

The QMLE of $\phi_0$ is $\widehat{\phi} = \arg\min_{\phi\in\Phi}\mathcal{L}\left(\phi\right)$, where

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

and $\Phi = \Lambda \times \Gamma$ is taken to be a compact subset of $\mathbb{R}^{d_\lambda + d_\gamma}$. The QMLEs of $\beta_0$ and $\sigma_0^2$ are defined as $\bar{\beta}\left(\widehat{\phi}\right) \equiv \widehat{\beta}$ and $\bar{\sigma}^2\left(\widehat{\phi}\right) \equiv \widehat{\sigma}^2$ respectively. The following assumption controls network dependence and is discussed below equation (4.4).

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

Writing $T(\lambda) = S(\lambda)S^{-1}$ and $\phi = (\lambda', \gamma')'$, define the quantity $\sigma^2(\phi) = n^{-1}\sigma_0^2 tr\left(T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right) = n^{-1}\sigma_0^2\left\|\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)\Sigma^{\frac{1}{2}}\right\|_F^2$, which is non-negative by definition and bounded by Assumptions R.3 and SAR.1. The assumptions below directly extend Assumptions R.6 and R.7 to the present setup.

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

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

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

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

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

Assumption SAR.4. *$\beta_0 \neq 0$, $\lambda_0 \neq 0$ and, for any $\eta > 0$,*

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

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

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

making the identifying nature of the assumption transparent. A similar identifying assumption is used by Gupta and Robinson (2018), and indeed in the context of nonlinear regression by Robinson (1972).

Theorem 5.1. *Under Assumptions R.1-R.5, R.8, SAR.1-SAR.4 and $p^{-1} + d_\gamma^{-1} + (d_\gamma + p)/n \to 0$ as $n \to \infty$, $\left(\widehat{\phi}', \widehat{\sigma}^2\right)' - \left(\phi_0', \sigma_0{}^2\right)' \xrightarrow{p} 0$ as $n \to \infty$.*

Theorem 5.2. *Under Assumptions R.1-R.5, R.8-R.10, SAR.1-SAR.4, $p^{-1} + d_\gamma^{-1} + p\left(p + d_\gamma^2\right)/n + \sqrt{n}/p^{\mu+1/4} + d_\gamma^2/p \to 0$, as $n \to \infty$, and $H_0$, $\mathscr{T} - \left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon - p\right)/\sqrt{2p} = o_p(1)$.*

Theorem 5.3. *Under the conditions of Theorems 4.3, 5.1 and 5.2, (1) $\mathscr{T} \xrightarrow{d} N(0,1)$ under $H_0$, (2) $\mathscr{T}$ is a consistent test statistic (i.e. it has asymptotically unit power under $H_1$), (3) $\mathscr{T} \xrightarrow{d} N(\varkappa, 1)$ under local alternatives $H_\ell$.*

**6. Testing in the presence of nonparametric networks.** In this section we are motivated by settings where network dependence occurs through nonparametric functions of raw distances (this may be geographic, social, economic, or any other type of distance), as is the case in Pinkse et al. (2002), for example. In their kind of setup, $d_{ij}$ is a raw economic distance between units $i$ and $j$ and the corresponding element of the spatial weight matrix is given by $w_{ij} = \zeta(d_{ij})$, where $\zeta(\cdot)$ is an unknown nonparametric function. Pinkse et al. (2002) use such a setup in a SAR model like (5.1), but with a linear regression function. In contrast, in keeping with the focus of this paper we instead model dependence in the errors in this manner. Our formulation is rather general, covering, for example, a specification like $w_{ij} = f(\gamma, \zeta(d_{ij}))$, with $f(\cdot)$ a *known* function, $\gamma$ an *unknown* parameter of possibly increasing dimension, and $\zeta(\cdot)$ an *unknown* nonparametric function.

Let $\Xi$ be a compact space of functions, on which we will specify more conditions later. The linear process coefficients are now $b_{js}(\gamma, \zeta_0(z))$, with $\zeta_0(\cdot) = \left(\zeta_{01}(\cdot), \ldots, \zeta_{0d_\zeta}(\cdot)\right)'$ a fixed-dimensional vector of real-valued nonparametric functions with $\zeta_{0i} \in \Xi$ for each $i = 1, \ldots, d_\zeta$, and $z$ a fixed-dimensional vector of data, independent of the $\varepsilon_s$, $s \geq 1$, with support $\mathcal{Z}$. We base our estimation on approximating each $\zeta_{0i}(z)$, $i = 1, \ldots, d_\zeta$, with the series representation $\delta_{0i}'\varphi_i(z)$, where $\varphi_i(z) \equiv \varphi_i$ is an $r_i \times 1$ ($r_i \to \infty$ as $n \to \infty$) vector of basis functions with typical function $\varphi_{il}$, $l = 1, \ldots, r_i$. The set of linear combinations $\delta_i'\varphi_i(z)$ forms the sequence of sieve spaces $\Phi_{r_i} \subset \Xi$ as $r_i \to \infty$, for any $i = 1, \ldots, d_\zeta$, and

$$\zeta_{0i}(z) = \delta_{0i}'\varphi_i + \nu_i, \tag{6.1}$$

with the following restriction on the function space $\Xi$:

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

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

$$\sup_{z \in \mathcal{Z}} E\left(\nu_i^2\right) = O\left(r_i^{-2\kappa_i}\right), i = 1, \ldots, d_\zeta. \tag{6.2}$$

Thus we now have an infinite-dimensional nuisance parameter $\zeta_0(\cdot)$ and increasing-dimensional nuisance parameter $\gamma$. Writing $\sum_{i=1}^{d_\zeta} r_i = r$ and $\tau = (\gamma', \delta_1', \ldots, \delta_{d_\zeta}')'$, which has increasing dimension $d_\tau = d_\gamma + r$, define $\varsigma(r) = \sup_{z \in \mathcal{Z}; i=1,\ldots,d_\zeta} \|\varphi_i\|$.

For any admissible values $\beta$, $\sigma^2$ and $\tau$, the redefined (multiplied by $2/n$) negative quasi log likelihood function based on using the approximations (3.1) and (6.1) is

$$L(\beta,\sigma^2,\tau) = \ln\left(2\pi\sigma^2\right) + \frac{1}{n}\ln\left|\Sigma\left(\tau\right)\right| + \frac{1}{n\sigma^2}(y-\Psi\beta)'\Sigma\left(\tau\right)^{-1}\left(y-\Psi\beta\right), \tag{6.3}$$

which is minimised with respect to $\beta$ and $\sigma^2$ by

$$\bar{\beta}\left(\tau\right) = \left(\Psi'\Sigma\left(\tau\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\tau\right)^{-1}y, \tag{6.4}$$

$$\bar{\sigma}^2\left(\tau\right) = n^{-1}y'C(\tau)'M(\tau)C(\tau)y, \tag{6.5}$$

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

$$\mathcal{L}(\tau) = \ln(2\pi) + \ln\bar{\sigma}^2(\tau) + \frac{1}{n}\ln\left|\Sigma\left(\tau\right)\right|. \tag{6.6}$$

Again, for compact $\Gamma$ and sieve coefficient space $\Delta$, the QMLE of $\tau_0$ is $\widehat{\tau} = \arg\min_{\tau\in\Gamma\times\Delta}\mathcal{L}(\tau)$ and the QMLEs of $\beta$ and $\sigma^2$ are $\widehat{\beta} = \bar{\beta}\left(\widehat{\tau}\right)$ and $\widehat{\sigma}^2 = \bar{\sigma}^2\left(\widehat{\tau}\right)$. For a given $x$, the series estimate of $\theta_0\left(x\right)$ is defined as $\widehat{\theta}\left(x\right) = \psi(x)'\widehat{\beta}$. Define also the product Banach space $\mathcal{T} = \Gamma\times\Xi^{d_\varsigma}$ with norm $\left\|(\gamma',\varsigma')'\right\|_{\mathcal{T}_w} = \|\gamma\| + \sum_{i=1}^{d_\varsigma}\|\varsigma_i\|_w$, and consider the map $\Sigma:\mathcal{T}^o\to\mathcal{M}^{n\times n}$, where $\mathcal{T}^o$ is an open subset of $\mathcal{T}$.

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

$$\sup_{t\in\mathcal{T}^o}\|D\Sigma(t)\|_{\mathscr{L}(\mathcal{T}^o,\mathcal{M}^{n\times n})} \le C, \tag{6.7}$$

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

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

$$\left\|\Sigma\left(t_1\right) - \Sigma\left(t_2\right)\right\| \le C\varsigma(r)\left\|t_1-t_2\right\|. \tag{6.8}$$

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

$$\overline{\lim_{n\to\infty}}\sup_{t\in\{t:\|t-t^*\|<\eta\}\cap\mathcal{T}^o}\left\|\Sigma(t) - \Sigma\left(t^*\right)\right\| < C\varsigma(r)\eta. \tag{6.9}$$

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

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

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

$$\underline{\lim}_{n\to\infty} \inf_{\tau \in \overline{\mathcal{N}}^{\tau}(\eta)} \frac{n^{-1} tr\left(\Sigma(\tau)^{-1}\Sigma_0\right)}{\left|\Sigma(\tau)^{-1}\Sigma_0\right|^{1/n}} > 1, \tag{6.10}$$

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

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

Theorem 6.1. *Under Assumptions R.1-R.4 (with R.3 and R.4 holding for $t \in \mathcal{T}$ rather than $\gamma \in \Gamma$), R.8, NPN.1-NPN.4 and $p^{-1} + d_\gamma^{-1} + \left(\min_{i=1,\ldots,d_\zeta} r_i\right)^{-1} + \left(d_\gamma + p + \max_{i=1,\ldots,d_\zeta} r_i\right)/n \to 0$ as $n \to \infty$, $\|\widehat{\tau} - \tau_0\| \xrightarrow{p} 0$.*

Theorem 6.2. *Under the conditions of Theorems 4.2 and 6.1, but with $\tau$ and $\mathcal{T}$ replacing $\gamma$ and $\Gamma$ in assumptions prefixed with R, $p \to \infty$ and $d_\gamma \to \infty$,*

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

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

Theorem 6.3. *Let the conditions of Theorems 4.3 and 6.2 hold, but with $\tau$ and $\mathcal{T}$ replacing $\gamma$ and $\Gamma$ in assumptions prefixed with R. Then (1) $\mathscr{T} \xrightarrow{d} N(0,1)$ under $H_0$, (2) $\mathscr{T}$ is a consistent test statistic (i.e. it has asymptotically unit power under $H_1$), (3) $\mathscr{T} \xrightarrow{d} N(\varkappa, 1)$ under local alternatives $H_\ell$.*

**7. Finite sample performance.** We now examine the finite sample performance of our test using Monte Carlo experiments. To study the size behaviour of our test, we generate the null model by DGP1: $\theta(x_i) = 1+x_{1i}+x_{2i} = x_i'\alpha_0$, where $x_{1i} = (z_i+z_{1i})/2$, $x_{2i} = (z_i+z_{2i})/2$, $z_i$, $z_{1i}$, and $z_{2i}$ are i.i.d. $U[0, 2\pi]$. A linear model $f(x_i, \alpha) = x_i'\alpha$ is correctly specified for $\theta(x_i)$ under DGP1 and misspecified under DGP2: $\theta(x_i) = x_i'\alpha_0 + 0.1(z_{1i} - \pi)(z_{2i} - \pi)$ and DGP3: $\theta(x_i) = x_i'\alpha_0 + \sin(x_i'\alpha_0)$.

7.1. *Parametric error network structure.* To illustrate different network structure in the error term, for every specification of $\theta(x_i)$, we generate $y$ as:

$$\text{Error SARMA}(m_1,m_2)\text{: } y = \theta(x) + u,\ u = \sum_{k=1}^{m_1} \gamma_{1k} W_{1k} u + \xi + \sum_{l=1}^{m_2} \gamma_{2l} W_{2l} \xi,$$

$$\text{Error MESS}(m)\text{: } y = \theta(x) + u,\ u = \exp\left(\sum_{k=1}^{m} \gamma_k W_k\right) \xi,$$

$$\text{SARSE}(m_1,m_2)\text{: } y = \sum_{k=1}^{m_1} \lambda_k W_{1k} y + \theta(x) + u,\ u = \sum_{l=1}^{m_2} \gamma_l W_{2l} u + \xi,$$

where $\xi \sim N(0, I_n)$. The Error SARMA and Error MESS models add different network structures in the error term, and the SARSE model further considers the spatial autoregressive term in the dependent variable. We use a power series for our test and rates: $p = 9, 14$ (corresponding to power series of degrees 3 and 4 respectively) for $n = 100, 300, 500$.

Table 1 reports the rejection rates using 500 Monte Carlo simulation at the 5% asymptotic level 1.645. Here, the i.i.d. case is a direct replication of the test in Hong and White (1995), so we use it as a benchmark. The spatial weight matrices are generated using LeSage's code *make_neighborsw* from http://www.spatial-econometrics.com, where the row-normalized sparse matrix are generated by choosing a specific number of the closest locations from randomly generated coordinates and we set the number of neighbors to be $n/20$.

The Error SARMA(1,0) corresponds to the commonly used SEM in the literature and this is generated with $\gamma = 0.4$; for the Error SARMA(1,1) and Error MESS(2) models, we choose $\lambda_1 = 0.4$ and $\gamma_1 = 0.5$; for the SARSE(1,1) model, we choose $\gamma_{11} = 0.3$ and $\gamma_{21} = 0.4$; for the SARSE(2,1), we choose $\lambda_1 = 0.3$, $\lambda_2 = 0.2$, and $\gamma_1 = 0.4$. We consider test statistics based on both $n\widehat{m}_n = \widehat{\sigma}^{-2}\widehat{v}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{u}$ and $n\widetilde{m}_n = \widehat{\sigma}^{-2}(\widehat{u}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{u} - \widehat{\eta}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{\eta})$, where $\widehat{\eta} = y - \widehat{\theta}$, i.e., the residual from nonparametric estimation. Analogous to the definition of $\mathscr{T}_n$, define the statistic $\mathscr{T}_n^a = (n\widetilde{m}_n - p)/\sqrt{2p}$.

| $p=14$ | iid | SARMA(1,0) | SARMA(1,1) | MESS(2) | SARSE(1,1) | | SARSE(2,1) | |
|---|---|---|---|---|---|---|---|---|
| $n=100$ | | | | | $\mathscr{T}_n$ | $\mathscr{T}_n^a$ | $\mathscr{T}_n$ | $\mathscr{T}_n^a$ |
| DGP1 | 0.026 | 0.024 | 0.020 | 0.082 | 0.024 | 0.026 | 0.050 | 0.090 |
| DGP2 | 0.234 | 0.240 | 0.606 | 0.656 | 0.308 | 0.280 | 0.378 | 0.392 |
| DGP3 | 0.994 | 0.998 | 0.862 | 0.926 | 0.994 | 0.994 | 0.966 | 0.974 |
| $n=300$ | | | | | | | | |
| DGP1 | 0.030 | 0.012 | 0.026 | 0.032 | 0.034 | 0.034 | 0.030 | 0.034 |
| DGP2 | 0.826 | 0.976 | 0.954 | 0.970 | 0.830 | 0.840 | 0.890 | 0.876 |
| DGP3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $n=500$ | | | | | | | | |
| DGP1 | 0.016 | 0.020 | 0.020 | 0.012 | 0.020 | 0.048 | 0.340 | 0.054 |
| DGP2 | 1 | 0.988 | 0.998 | 0.998 | 1 | 1 | 1 | 1 |
| DGP3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| $p=9$ | iid | SARMA(1,0) | SARMA(1,1) | MESS(2) | SARSE(1,1) | | SARSE(2,1) | |
|---|---|---|---|---|---|---|---|---|
| $n=100$ | | | | | $\mathscr{T}_n$ | $\mathscr{T}_n^a$ | $\mathscr{T}_n$ | $\mathscr{T}_n^a$ |
| DGP1 | 0.016 | 0.020 | 0.038 | 0.078 | 0.012 | 0.014 | 0.028 | 0.074 |
| DGP2 | 0.206 | 0.288 | 0.350 | 0.546 | 0.340 | 0.318 | 0.394 | 0.414 |
| DGP3 | 0.930 | 0.966 | 0.936 | 1 | 0.938 | 0.948 | 0.696 | 0.682 |
| $n=300$ | | | | | | | | |
| DGP1 | 0.016 | 0.016 | 0.020 | 0.028 | 0.014 | 0.012 | 0.020 | 0.032 |
| DGP2 | 0.946 | 0.948 | 0.912 | 0.880 | 0.894 | 0.896 | 0.940 | 0.914 |
| DGP3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $n=500$ | | | | | | | | |
| DGP1 | 0.004 | 0.004 | 0.008 | 0.014 | 0.010 | 0.024 | 0.020 | 0.032 |
| DGP2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| DGP3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

TABLE 1

*Rejection probabilities at 5% asymptotic level, parametric network structures*

In the case of no spatial autoregressive term, and under the power series, $\mathscr{T}_n^a$ and $\mathscr{T}_n$ are numerically identical, as was observed by Hong and White (1995). However, in the SARSE setting a difference arises due to the spatial structure in the response $y$. We show that $\mathscr{T}_n^a - \mathscr{T}_n = o_p(1)$ in Theorem TS.1 in the supplementary appendix, regardless of whether we are in the usual nonparametric regression framework or the SAR setting of Section 5.

From Table 1, we can see the finite performance of our test is comparable to Hong and White (1995), indicated in the column labelled 'iid'. For both values of $p$, power improves as sample size increases for all network structures, although it is lowest under DGP2. This was also noted by Hong and White (1995) in their iid setting. However with the largest sample size (n = 500), power is virtually unity in all cases. Taking $p = 14$ yields more accurate sizes in many cases, but $p = 9$ can give greater power in some cases. An interesting pattern that we observe is that when the SAR term is involved, $\mathscr{T}_n^a$ exhibits better size than $\mathscr{T}_n$. Although the two statistics are asymptotically equivalent, $\mathscr{T}_n^a$ might be preferred with the SAR-type models of Section 5.

7.2. *Nonparametric error network structure.* Now we examine finite sample performance in the setting of Section 5. The three DGPs of $\theta(x)$ are the same as the parametric setting and we generate the $n \times n$ matrix $W^*$ as $w_{ij}^* = \Phi(-d_{ij})I(c_{ij} < 0.05)$ if $i \neq j$, and $w_{ii}^* = 0$, where $\Phi(\cdot)$ is the standard normal cdf, $d_{ij}$ ~i.i.d. $U[-3, 3]$, and $c_{ij}$ ~i.i.d. $U[0, 1]$. From this construction, we ensure that $W^*$ is sparse with no more than 5% elements being nonzero. Then, $y$ is generated from $y = \theta(x) + u$, $u = Wu + \xi$, where $W = W^*/1.2\overline{\varphi}(W^*)$, ensuring the existence of $(I - W)^{-1}$. In estimation, we know the distance $d_{ij}$ and the indicator $I(c_{ij} < 0.05)$, but we do not know the functional form of $d_{ij}$ in $w_{ij}$, so we approximate elements in $W$ by $\widehat{w}_{ij} = \sum_{l=0}^{r} a_l d_{ij}^l I(c_{ij} < 0.05)$ if $i \neq j$; $\widehat{w}_{ii} = 0$. Estimation is carried out using the MLE described in Section 6.

Table 2 reports the rejection rates using 500 Monte Carlo simulation at the 5% asymptotic level 1.645 in this nonparametric error network setting using $r = 2, 3, 4$ and $p = 9, 14$. Our smallest sample size is $n = 150$ rather than $n = 100$ as earlier because two nonparametric functions must be estimated in the nonparametric network setting. We observe a clear pattern of rejection rates approaching the theoretical level as sample size increases. Generally, the power is excellent for all DGPs for $n \geq 300$ and sizes are acceptable for $n = 500$, particularly when $p = 14$. The power against DGP3 is always higher than that against DGP2, as observed in the parametric setup of the previous subsection.

| | $r=2$ | | $r=3$ | | $r=4$ | |
|---|---|---|---|---|---|---|
| $n=150$ | $p=9$ | $p=14$ | $p=9$ | $p=14$ | $p=9$ | $p=14$ |
| DGP1 | 0.086 | 0.202 | 0.118 | 0.206 | 0.142 | 0.224 |
| DGP2 | 0.474 | 0.558 | 0.496 | 0.566 | 0.504 | 0.578 |
| DGP3 | 0.954 | 0.998 | 0.964 | 0.998 | 0.960 | 0.998 |
| $n=300$ | | | | | | |
| DGP1 | 0.082 | 0.096 | 0.088 | 0.108 | 0.106 | 0.110 |
| DGP2 | 0.802 | 0.778 | 0.806 | 0.786 | 0.814 | 0.800 |
| DGP3 | 0.996 | 1 | 0.996 | 1 | 0.996 | 1 |
| $n=500$ | | | | | | |
| DGP1 | 0.028 | 0.042 | 0.026 | 0.040 | 0.036 | 0.048 |
| DGP2 | 0.980 | 0.970 | 0.980 | 0.968 | 0.980 | 0.964 |
| DGP3 | 1 | 1 | 1 | 1 | 1 | 1 |

TABLE 2

*Rejection probabilities at 5% asymptotic level, nonparametric network structure*

## 8. Empirical applications.

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

### 8.1. *Conflict networks.*

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

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

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

where $y$ is a vector of fighting efforts of each group and $X$ includes the current rainfall, rainfall from the last year, and their squares. To consider the spatial correlation in the error term, we consider both the Error SARMA(1,0) and

Error SARMA(0,1) structures. For these, we employ a spatial weight matrix $W^d$, based on the inverse distance between group locations and set to be 0 after 150 km, following König et al. (2017). The idea is that geographical spatial correlation dies out as groups become further apart. We also report results using a nonparametric estimator of the spatial weights, as described in Section 6 and studied in simulations in Section 7. For the nonparametric estimator we take $r = 2$.

In the original dataset, there are 80 groups, but groups 62 and 63 have the same variables and the same locations, so we drop one group and end up with a sample of 79 groups. We use data from 1998 as an example and further use the pooled data from all years as a robustness check. The column IV of Table 3 is from Table 1 of König et al. (2017) based on their panel IV estimation, which we report for the sake of comparison. $H_0$ stands for restricted model where the linear functional form of the regression is imposed, while $H_1$ stands for the unrestricted model where we use basis functions comprising of power series with $p = 9$. In all our specifications, the test statistics are negative, so we cannot reject the null hypothesis that the model is correctly specified. As Table 3 indicates, this failure to reject the null persists when we use pooled data from 13 years, yielding 1027 observations. Thus we conclude that a linear specification is not inappropriate for this setting. Another interesting finding is that allowing for explicit spatial dependence in the disturbances reduces the magnitude of the estimated $\lambda_1$, i.e. the coefficient on $W^A y$, drastically, a feature that is not replicated for the estimate of $\lambda_2$.

8.2. *Innovation networks.* This example is based on the study of the impact of R&D on growth from Bloom et al. (2013). They develop a general framework incorporating two types of spillovers: a positive effect from technology (knowledge) spillovers and a negative 'business stealing' effect from product market rivals. They implement this model using panel data on U.S. firms.

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

$$\ln y = \varphi_1 \ln(R\&D) + \varphi_2 \ln(Sptec) + \varphi_3 \ln(Spsic) + \varphi_4 X + error, \tag{8.1}$$

where $y$ is a vector of sales, $R\&D$ is a vector of R&D stocks, and regressors in $X$ include the log of capital ($Capital$), log of labour ($Labour$), $R\&D$, a dummy for missing values in $R\&D$, a price index, and two spillover terms constructed as the log of $W_{SIC}R\&D$ ($Spsic$) and the log of $W_{TEC}R\&D$ ($Sptec$), where $W_{SIC}$ measures the the product market proximity and $W_{TEC}$ measures the technological proximity. Specifically, they define $W_{SIC,ij} = S_i S_j'/(S_i S_i')^{1/2}(S_j S_j')^{1/2}$ and $W_{TEC,ij} = T_i T_j'/(T_i T_i')^{1/2}(T_j T_j')^{1/2}$, where $S_i =$

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

The column FE of Table 4 is from Table 5 of Bloom et al. (2013) based on their panel estimation and we use it as a baseline for comparison. "SE" is the spatial error model corresponding to the Error SARMA(1,0) in our Monte Carlo setting. We use either $W_{SIC}$ or $W_{TEC}$ in the SE setting and both of these two matrices in the Error SARMA(2,0), Error SARMA(0,2), and Error MESS(2) models. In all of these specifications, the test statistics are larger than 1.645, so we reject the null hypothesis of the linear specification. However, we can say even more as our estimation also sheds light on network effects in the disturbances in (8.1). As before $H_0$ imposes linear functional form of the regressors, while $H_1$ uses the nonparametric series estimate employing power series with $p = 9$. Regardless of the specification of the regression function, the disturbances suggest a strong network effect as the coefficients on $W_{TEC}$ and $W_{SIC}$ are large in magnitude.

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

| | IV | SARMA(1,0) | | | | SARMA(0,1) | | | | Nonparametric | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1998 | | Pooled | | 1998 | | Pooled | | 1998 | | Pooled | |
| | | $H_0$ | $H_1$ | $H_0$ | $H_1$ | $H_0$ | $H_1$ | $H_0$ | $H_1$ | $H_0$ | $H_1$ | $H_0$ | $H_1$ |
| $W^A y$ | -0.218 | -0.005 | -0.003 | 0.013 | 0.013 | 0.001 | 0.011 | 0.013 | 0.013 | -0.052 | -0.011 | 0.033 | 0.033 |
| $W^E y$ | 0.130 | 0.130 | 0.129 | 0.121 | 0.121 | 0.127 | 0.122 | 0.121 | 0.121 | 0.149 | 0.133 | 0.110 | 0.109 |
| $W^d$ | | -0.159 | -0.225 | -0.086 | -0.086 | -0.153 | -0.050 | -0.086 | -0.086 | | | | |
| $\mathscr{T}_n$ | | | -1.921 | | -2.531 | | -1.763 | | -2.421 | | -1.294 | | -2.314 |
| $\mathscr{T}_n^a$ | | | -1.918 | | -2.547 | | -2.349 | | -2.423 | | -1.898 | | -2.530 |

TABLE 3

*The estimates and test statistics for the conflict data*

| Variables | FE | SE $W_{SIC}$ | | SE $W_{TEC}$ | | SARMA(2,0) | | SARMA(0,2) | | Error MESS(2) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $H_0$ | $H_1$ | $H_0$ | $H_1$ | $H_0$ | $H_1$ | $H_0$ | $H_1$ | $H_0$ | $H_1$ |
| ln(*Sptech*) | 0.191 | 0.007 | 0.015 | 0.008 | 0.017 | 0.009 | 0.018 | -0.0002 | 0.013 | 0.002 | 0.014 |
| ln(*Spsic*) | -0.005 | 0.006 | -0.0001 | 0.038 | 0.020 | 0.044 | 0.026 | 0.033 | 0.017 | 0.045 | 0.027 |
| ln(*Capital*) | 0.636 | 0.572 | | 0.571 | | 0.573 | | 0.565 | | 0.569 | |
| ln(*Labor*) | 0.154 | 0.336 | | 0.318 | | 0.315 | | 0.334 | | 0.323 | |
| ln(*R&D*) | 0.043 | 0.814 | | 0.082 | | 0.081 | | 0.076 | | 0.077 | |
| $W_{SIC}$ | | 0.825 | 0.829 | | | 0.696 | 0.693 | 0.624 | 0.728 | 0.775 | 0.836 |
| $W_{TEC}$ | | | | 0.722 | 0.724 | 0.157 | 0.164 | 0.312 | 0.321 | 0.338 | 0.380 |
| $\mathscr{T}_n$ | | | 15.528 | | 10.451 | | 10.485 | | 15.144 | | 12.776 |

TABLE 4

*The estimates and test statistics for the R&D data*

| Variable | $w^*_{ij} = d_{ij}^{-2}$ for $i \neq j$ | | $w^*_{ij} = e^{-2d_{ij}}$ for $i \neq j$ | |
|---|---|---|---|---|
| | estimate | p-value | estimate | p-value |
| Constant | 1.0711 | 0.608 | 0.5989 | 0.798 |
| $\ln(s)$ | 0.8256 | $< 0.001$ | 0.7938 | $< 0.001$ |
| $\ln(n_p + 0.05)$ | $-1.4984$ | 0.008 | $-1.4512$ | 0.009 |
| $W \ln(s)$ | $-0.3159$ | 0.075 | $-0.3595$ | 0.020 |
| $W \ln(n_p + 0.05)$ | 0.5633 | 0.498 | 0.1283 | 0.856 |
| $Wy$ | 0.7360 | $< 0.001$ | 0.6510 | $< 0.001$ |
| Test statistic $\mathscr{T}_n$ | $-1.88$ | | $-2.08$ | |
| Test statistic $\mathscr{T}_n^a$ | $-1.90$ | | $-2.05$ | |
| Restricted regression | | | | |
| Constant | 2.1411 | $< 0.001$ | 2.9890 | $< 0.001$ |
| $\ln(s) - \ln(n + 0.05)$ | 0.8426 | $< 0.001$ | 0.8195 | $< 0.001$ |
| $W[\ln(s) - \ln(n_p + 0.05)]$ | $-0.2675$ | 0.122 | $-0.2589$ | 0.098 |
| $W \ln(y)$ | 0.7320 | $< 0.001$ | 0.6380 | $< 0.001$ |
| $\mathscr{T}_n$ | 0.30 | | 4.04 | |
| $\mathscr{T}_n^a$ | 0.10 | | 4.50 | |

TABLE 5
*The estimates and test statistics of the linear SAR model for the growth data*

In this section, we want to test the linear SAR model specification in Ertur and Koch (2007). Their dataset covers a sample of 91 countries over the period 1960-1995 originally from Heston, Summers, and Aten (2002) obtained from Penn World Tables (PWT version 6.1). The variables in use include per worker income in 1960 ($y60$) and 1995 ($y95$), average rate of growth between 1960 and 1995 ($gy$), average investment rate of this period ($s$), and average rate of growth of working-age population ($n_p$). The dataset can be downloaded from JAE Data Archive at http://qed.econ.queensu.ca/jae/2007-v22.6/.

Ertur and Koch (2007) consider the model

$$y = \lambda W y + \mathbf{1}_n \beta_0 + X\beta + WX\theta + \varepsilon, \tag{8.2}$$

where the dependent variable is log real income per worker $\ln(y95)$, elements of the explanatory variable $X = (x_1', x_2')$ include log investment rate $\ln(s) = x_1$ and log physical capital effective rate of depreciation $\ln(n_p + 0.05) = x_2$, with corresponding subscripted coefficients $\beta_1, \beta_2, \theta_1, \theta_2$. A restricted regression based on the joint constraints $\beta_1 = -\beta_2$ and $\theta_1 = -\theta_2$ (these constraints are implied by economic theory) is also considered in Ertur and Koch (2007). The model (8.2) can be considered as a special case of the SAR model with a general regressor $X^* = (X, WX)$ and i.i.d. errors, so the test derived in Section 5 can be directly applied here. Denoting by $d_{ij}$ the great-circle distance between the capital cities of countries $i$ and $j$, one construction of $W$ takes $w_{ij} = d_{ij}^{-2}$ while the other takes $w_{ij} = e^{-2d_{ij}}$, following Ertur and Koch (2007). Table 5 presents the estimation and testing results based on the unrestricted model and restricted model under two constructions of $W$ using power series basis functions with $p = 10$. Using our specification test, we cannot reject linearity of the regression function under the unrestricted model. On the other hand, linearity is rejected under the restricted model, which is the preferred specification of Ertur and Koch (2007). Thus, not only can we conclude that the specification of the model is under suspicion we can also infer that such doubts are created by imposing constraints arising from economic theory.

**References.**


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

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

Bandyopadhyay, S. and A. Maity (2018). Asymptotic theory for varying coefficient regression models with dependent data. *Annals of the Institute of Statistical Mathematics 70*, 745–759.

Bevilacqua, M., T. Faouzi, R. Furrer, and E. Porcu (2018). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. Forthcoming: *Annals of Statistics*.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica 58*, 1443–1458.

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

Bramoullé, Y., A. Galeotti, and B. Rogers (Eds.) (2016). *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

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

Chen, S. X. and H. Wolfgang (2003). An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*, 663–678.

Chen, X. (2007). *Large sample sieve estimation of semi-nonparametric models*, Volume 6B, Chapter 76, pp. 5549–5632. North Holland.

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

Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation*. London: Pion.

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

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: homophily, minorities and segregation. *Econometrica 77*, 1003–1045.

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

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

de Paula, A. (2017). Econometrics of network models. In B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*, pp. 268–323.

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

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

Delgado, M. A., J. Hidalgo, and C. Velasco (2005). Distribution free goodness-of-fit tests for linear processes. *Annals of Statistics 33*, 2568–2609.

Delgado, M. A. and T. Stengos (1994). Semiparametric specification testing of non-nested econometric models. *Review of Economic Studies 61*, 291–303.

Dray, S. and T. Jombart (2011). Revisiting Guerry's data: Introducing spatial constraints in multivariate analysis. *Annals of Applied Statistics 5*, 2278–2299.

Durante, D., D. B. Dunson, and J. T. Vogelstein (2017). Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association 112*, 1516–1530.

Ellison, G. and S. F. Ellison (2000). A simple framework for nonparametric specification testing. *Journal of Econometrics 96*, 1–23.

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

Escanciano, J. (2018). Model checks using marked empirical processes. *Statistica Sinica 9*, 475–499.

Eubank, R. L. and C. H. Spiegelman (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association 85*, 387–392.

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

Fan, J., C. Zhang, and J. Zhang (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics 29*, 153–193.

Fan, Y. and Q. Li (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica 64*, 865–890.

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

Gao, J., M. King, Z. Lu, and D. Tjøstheim (2009). Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics 37*, 3893–3928.

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

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

Guerre, E. and P. Lavergne (2005). Data-driven rate-optimal specification testing in regression models. *Annals of Statistics 33*, 840–870.

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

Gupta, A., S. Kokas, and A. Michaelides (2018). Credit market spillovers: Evidence from a syndicated loan market network. Working paper.

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

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

Handcock, M. S. and J. R. Wallis (1994). An approach to statistical spatial-temporal modeling of meteoreological fields. *Journal of the American Statistical Association 89*, 368–390.

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

Helmers, C. and M. Patnam (2014). Does the rotten child spoil his companion? Spatial peer effects among children in rural India. *Quantitative Economics 5*, 67–121.

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

Hidalgo, J. (2008). Specification testing for regression models with dependent data. *Journal of Econometrics 143*, 143–165.

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

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

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

Hong, Y. and Y.-J. Lee (2013). A loss function approach to model specification testing and its relative efficiency. *Annals of Statistics 41*, 1166–1203.

Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies 18*, 37–84.

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

Horowitz, J. L. (2006). Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica 74*, 521–538.

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

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

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

Jones, R. H. and A. V. Vecchia (1993). Fitting continuous ARMA models to unequally spaced spatial data. *Journal of the American Statistical Association 88*, 947–954.

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

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

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

Kolaczyk, E. (2017). *Topics at the Frontier of Statistics and Network Analysis: (Re)Visiting the Foundations*. SemStat Elements. Cambridge University Press.

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

Kuersteiner, G. M. and I. R. Prucha (2013). Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics 174*, 107–126.

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

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

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

LeSage, J. P. and R. Pace (2007). A matrix exponential spatial specification. *Journal of Econometrics 140*, 190–214.

Li, H., Q. Li, and R. Liu (2016). Consistent model specification tests based on k-nearest-neighbor estimation method. *Journal of Econometrics 194*, 187–202.

Li, Q., C. Hsiao, and J. Zinn (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics 112*, 295–325.

Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B 73*, 423–498.

Matern, B. (1986). *Spatial Variation*. Almaenna Foerlaget, Stockholm.

Mohnen, M. (2017). Stars and brokers: peer effects among medical scientists. Mimeo. University of Essex.

Moody, J. (2001). Race, school integration, and friendship segregation in America. *American Journal of Sociology 107*, 679–716.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*, 147–168.

Oettl, A. (2012). Reconceptualizing stars: scientist helpfulness and peer performance. *Management Science 58*, 1122–1140.

Olhede, S. C. and P. J. Wolfe (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences 111*, 14722–14727.

Peng, B. (2016). Inference on modelling cross-sectional dependence for a varying-coefficient model. *Economics Letters 145*, 1–5.

Pinkse, J., M. E. Slade, and C. Brett (2002). Spatial price competition: A semiparametric approach. *Econometrica 70*, 1111–1153.

Portnoy, S. (1984). Asymptotic behavior of $M$-estimators of $p$ regression parameters when $p^2/n$ is large. I. Consistency. *The Annals of Statistics 12*, 1298–1309.

Portnoy, S. (1985). Asymptotic behavior of $M$-estimators of $p$ regression parameters when $p^2/n$ is large; II. Normal approximation. *The Annals of Statistics 13*, 1403–1417.

Robinson, P. M. (1972). Non-linear regression for multiple time-series. *Journal of Applied Probability 9*, 758–768.

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics 165*, 5–19.

Robinson, P. M. and S. Thawornkaiwong (2012). Statistical inference on regression with spatial dependence. *Journal of Econometrics 167*, 521–542.

Rothe, C. and D. Wied (2013). Misspecification testing in a class of conditional distributional models. *Journal of the American Statistical Association 108*, 314–324.

Stein, M. (1999). *Interpolation of Spatial Data*. Springer-Verlag, New York.

Stein, M. L., Z. Chi, and L. J. Welty (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society, Series B 66*, 275–296.

Su, L. and S. Jin (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics 157*, 18–33.

Su, L. and X. Qu (2017). Specification test for spatial autoregressive models. *Journal of Business & Economic Statistics 35*, 572–584.

Sun, Y., H. Yan, W. Zhang, and Z. Lu (2014). A semiparametric spatial dynamic model. *Annals of Statistics 42*, 700–727.

Tripathi, G. and Y. Kitamura (2003). Testing conditional moment restrictions. *Annals of Statistics 31*, 2059–2095.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society, Series B 50*, 297–312.

Wang, Q. and P. C. Phillips (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics 40*, 727–758.

Wooldridge, J. M. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory 8*, 452–475.

Yajima, Y. and Y. Matsuda (2009). On nonparametric and semiparametric testing for multivariate linear time series. *Annals of Statistics 37*, 3529–3554.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory 8*, 435–451.

Zhao, Y., E. Levina, and J. Zhu (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Annals of Statistics 40*, 2266–2292.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics 75*, 263 – 289.

A. Gupta
Department of Economics
University of Essex
Wivenhoe Park, Colchester CO4 3SQ
UK
E-mail: a.gupta@essex.ac.uk

X. Qu
Antai College of Economics and Management
Shanghai Jiao Tong University
Shanghai, 200030
China
E-mail: xiqu@sjtu.edu.cn.

Supplementary appendix to 'Consistent specification testing under network dependence'

A. Gupta and X. Qu

January 11, 2019

# A Proofs of theorems and propositions

*Proof of Proposition 4.1:* Because the map $\Sigma : \Gamma^o \to \mathcal{M}^{n\times n}$ is Fréchet-differentiable on $\Gamma^o$, it is also Gâteaux-differentiable and the two derivative maps coincide. Thus by Theorem 1.8 of Ambrosetti and Prodi (1995), $\|\Sigma(\gamma_1) - \Sigma(\gamma_2)\| \le \sup_{\gamma\in\ell[\gamma_1,\gamma_2]} \|D\Sigma(\gamma)\| \|\gamma_1-\gamma_2\|$, where $\ell[\gamma_1,\gamma_2] = \{t\gamma_1 + (1-t)\gamma_2 : t\in[0,1]\}$. The claim now follows by (4.3) in Assumption 8. $\square$

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

*Proof of Theorem 4.2.* From Corollary 4.1 and Lemma LS.2, $\|\Sigma(\widehat{\gamma}) - \Sigma\| = O_p(\|\widehat{\gamma}-\gamma_0\|) = \sqrt{d_\gamma/n}$, so we have, from Assumption R.3,

$$\left\|\Sigma(\widehat{\gamma})^{-1} - \Sigma^{-1}\right\| \le \left\|\Sigma(\widehat{\gamma})^{-1}\right\| \|\Sigma(\widehat{\gamma}) - \Sigma\| \left\|\Sigma^{-1}\right\| = O_p(\|\widehat{\gamma}-\gamma_0\|) = \sqrt{d_\gamma/n}. \tag{S.1}$$

Similarly,

$$\begin{aligned}
&\left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1} - \left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\| \\
\le\ & \left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1}\right\| \left\|\frac{1}{n}\Psi'\left(\Sigma(\widehat{\gamma})^{-1} - \Sigma^{-1}\right)\Psi\right\| \left\|\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\| \\
\le\ & \sup_{\gamma\in\Gamma}\left\|\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\right\| \left\|\Sigma(\widehat{\gamma})^{-1} - \Sigma^{-1}\right\| \left\|\frac{1}{\sqrt{n}}\Psi\right\|^2 = O_p(\|\widehat{\gamma}-\gamma_0\|) = \sqrt{d_\gamma/n}.
\end{aligned}$$

By Assumption R.2, we have $\widehat{\alpha} - \alpha^* = O_p(1/\sqrt{n})$. Denote by $\theta^*(x) = \psi(x)'\beta^*$, where $\beta^* = \arg\min_\beta E[y_i - \psi(x_i)'\beta)]^2$, and set $\theta_{ni} = \theta(x_i)$, $\theta_{0i} = \theta_0(x_i)$, $\widehat{\theta}_i = \psi_i'\widehat{\beta}$, $\widehat{f}_i = f(x_i,\widehat{\alpha})$, $f_i^* = f(x_i,\alpha^*)$. Then $\widehat{u}_i = y_i - f(x_i,\widehat{\alpha}) = u_i + \theta_{0i} - \widehat{f}_i$. Let $\theta_0 = (\theta_0(x_1),\ldots,\theta_0(x_n))'$ as before, with similar component-wise notation for the $n$-dimensional vectors $\widehat{\theta}$, $\theta^*$, $\widehat{f}$, and $u$. As the approximation error is $e = \theta_0 - \theta^* = \theta_0 - \Psi\beta^*$,

$$\begin{aligned}
\widehat{\theta} - \theta^* &= \Psi(\widehat{\beta} - \beta^*) = \Psi\left(\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(u + \theta_0 - \Psi\beta^*) \\
&= \Psi\left(\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(u + e),
\end{aligned}$$

so that

$$\begin{aligned}
n\widehat{m}_n &= \widehat{\sigma}^{-2}\widehat{v}'\Sigma(\widehat{\gamma})^{-1}\widehat{u} = \widehat{\sigma}^{-2}\left(\widehat{\theta}-\widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\left(y-\widehat{f}\right) \\
&= \widehat{\sigma}^{-2}\left(\widehat{\theta}-\theta^*+\theta^*-\theta_0+\theta_0-\widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\left(u+\theta_0-\widehat{f}\right) \\
&= \widehat{\sigma}^{-2}\left[\Psi\left(\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(u+e)-e+\theta_0-\widehat{f}\right]'\Sigma(\widehat{\gamma})^{-1}\left(u+\theta_0-\widehat{f}\right) \\
&= \widehat{\sigma}^{-2}u'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}u+\widehat{\sigma}^{-2}u'\Sigma(\widehat{\gamma})^{-1}\left(\theta_0-\widehat{f}\right) \\
&- \widehat{\sigma}^{-2}\left(u+\theta_0-\widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\left(I-\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}\right)e \\
&+ \widehat{\sigma}^{-2}\left(\theta_0-\widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}u \\
&+ \widehat{\sigma}^{-2}(\theta_0-\widehat{f})'\Sigma(\widehat{\gamma})^{-1}(\theta_0-\widehat{f}) \\
&= \widehat{\sigma}^{-2}u'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}u+\widehat{\sigma}^{-2}(A_1+A_2+A_3+A_4),
\end{aligned}$$

say. First, for any vector $g$ comprising of conditioned random variables,

$$E\left[\left(u'\Sigma(\gamma)^{-1}g\right)^2\right] = g'\Sigma(\gamma)^{-1}\Sigma\Sigma(\gamma)^{-1}g \le \sup_{\gamma\in\Gamma}\left\|\Sigma(\gamma)^{-1}\right\|^2\|\Sigma\|\,\|g\|^2 = O_p\left(\|g\|^2\right),$$

uniformly in $\gamma\in\Gamma$. Similarly,

$$\begin{aligned}
& E\left[\left(u'\Sigma(\gamma)^{-1}\Psi\left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}g\right)^2\right] \\
=\;& g'\Sigma(\gamma)^{-1}\Psi\left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}\Sigma\Sigma(\gamma)^{-1}\Psi\left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}g \\
\le\;& \sup_{\gamma\in\Gamma}\left\|\Sigma(\gamma)^{-1}\right\|^4\|\Sigma\|\left\|\frac{1}{n}\Psi\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\right\|^2\|g\|^2 = O_p\left(\|g\|^2\right),
\end{aligned}$$

uniformly and, for any $j=1,\ldots,d_\gamma$,

$$\begin{aligned}
E\left[\left(u'\Sigma(\gamma)^{-1}\Sigma_j(\gamma)\Sigma(\gamma)^{-1}g\right)^2\right] &= g'\Sigma(\gamma)^{-1}\Sigma_j(\gamma)\Sigma(\gamma)^{-1}\Sigma\Sigma(\gamma)^{-1}\Sigma_j(\gamma)\Sigma(\gamma)^{-1}g \\
&\le \sup_{\gamma\in\Gamma}\left\|\Sigma(\gamma)^{-1}\right\|^4\|\Sigma_j(\gamma)\|^2\|\Sigma\|\,\|g\|^2 = O_p\left(\|g\|^2\right).
\end{aligned}$$

Let $\Psi_k$ be the $k$-th column of $\Psi$, $k=1,\ldots,p$. Then, we have $\|\Psi_k/\sqrt{n}\| = O_p(1)$ and for any $\gamma\in\Gamma$,

$$\begin{aligned}
E\left\|\frac{1}{\sqrt{n}}u'\Sigma(\gamma)^{-1}\Psi\right\|^2 &\le \sum_{k=1}^{p}E\left(u'\Sigma(\gamma)^{-1}\frac{1}{\sqrt{n}}\Psi_k\right)^2 = O_p(p), \\
E\left\|\frac{1}{\sqrt{n}}u'\Sigma(\gamma)^{-1}\Sigma_j(\gamma)\Sigma(\gamma)^{-1}\Psi\right\|^2 &\le \sum_{k=1}^{p}E\left(u'\Sigma(\gamma)^{-1}\Sigma_j(\gamma)\Sigma(\gamma)^{-1}\frac{1}{\sqrt{n}}\Psi_k\right)^2 = O(p).
\end{aligned}$$

Therefore, by Chebyshev's inequality,

$$\sup_{\gamma\in\Gamma}\left\|\frac{1}{\sqrt{n}}u'\Sigma(\gamma)^{-1}\Psi\right\|=O_p(\sqrt{p})\quad\text{and}\quad\sup_{\gamma\in\Gamma}\left\|\frac{1}{\sqrt{n}}u'\Sigma(\gamma)^{-1}\Sigma_j(\gamma)\Sigma(\gamma)^{-1}\Psi\right\|=O_p(\sqrt{p}).$$

By the decomposition

$$\begin{aligned}
& u'\left(\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}-\Sigma^{-1}\Psi[\Psi'\Sigma^{-1}\Psi]^{-1}\Psi'\Sigma^{-1}\right)u \\
= & u'\left(\Sigma(\widehat{\gamma})^{-1}+\Sigma^{-1}\right)\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\left(\sum_{i=1}^{n}e_{in}e_{in}'\right)\left(\Sigma(\widehat{\gamma})^{-1}-\Sigma^{-1}\right)u \\
& +u'\Sigma^{-1}\Psi\left([\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}-[\Psi'\Sigma^{-1}\Psi]^{-1}\right)\Psi'\Sigma^{-1}u \\
= & u'\left(\Sigma(\widehat{\gamma})^{-1}+\Sigma^{-1}\right)\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\left(\sum_{i=1}^{n}e_{in}e_{in}'\right)\sum_{j=1}^{d_\gamma}\left(\Sigma(\widetilde{\gamma})^{-1}\Sigma_j(\widetilde{\gamma})\Sigma(\widetilde{\gamma})^{-1}\right) \\
\times & u(\widetilde{\gamma}_j-\gamma_{j0})+u'\Sigma^{-1}\Psi\left([\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}-[\Psi'\Sigma^{-1}\Psi]^{-1}\right)\Psi'\Sigma^{-1}u,
\end{aligned}$$

where $e_{in}$ is an $n\times 1$ vector with $i$-th entry one and zeros elsewhere, so $\sum_{i=1}^{n}e_{in}e_{in}'=I_n$, and

$$\begin{aligned}
e_{in}'\left(\Sigma(\widehat{\gamma})^{-1}-\Sigma^{-1}\right)u &= \sum_{j=1}^{d_\gamma}e_{in}'\left(\Sigma(\widetilde{\gamma})^{-1}\Sigma_j(\widetilde{\gamma})\Sigma(\widetilde{\gamma})^{-1}\right)u(\widetilde{\gamma}_j-\gamma_{j0}) \\
&= e_{in}'\sum_{j=1}^{d_\gamma}\left(\Sigma(\widetilde{\gamma})^{-1}\Sigma_j(\widetilde{\gamma})\Sigma(\widetilde{\gamma})^{-1}\right)u(\widetilde{\gamma}_j-\gamma_{j0})
\end{aligned}$$

where $\widetilde{\gamma}$ is a value between $\widehat{\gamma}$ and $\gamma_0$ due to the mean value theorem. We have

$$\begin{aligned}
& \left|u'\left(\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}-\Sigma^{-1}\Psi[\Psi'\Sigma^{-1}\Psi]^{-1}\Psi'\Sigma^{-1}\right)u\right| \\
\leq & 2\sup_{\gamma\in\Gamma}\left\|\frac{1}{\sqrt{n}}u'\Sigma(\gamma)^{-1}\Psi\right\|\left\|\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\right\|\sum_{j=1}^{d_\gamma}\left\|\frac{1}{\sqrt{n}}\Psi'\left(\Sigma(\gamma)^{-1}\Sigma_j(\gamma)\Sigma(\gamma)^{-1}\right)u\right\| \\
\times & |\widetilde{\gamma}_j-\gamma_{j0}|+\left\|\frac{1}{\sqrt{n}}u'\Sigma^{-1}\Psi\right\|^2\left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1}-\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\| \\
= & O_p(\sqrt{p})O_p(d_\gamma\sqrt{p}/\sqrt{n})+O_p(p)O_p(\sqrt{d_\gamma}/\sqrt{n})=O_p(d_\gamma p/\sqrt{n})=o_p(\sqrt{p}),
\end{aligned}$$

where the last equality holds under the conditions of the theorem.

It remains to show that

$$A_i=o_p\left(p^{1/2}\right),i=1,\ldots,4. \tag{S.2}$$

It is convenient to perform the calculations under $H_\ell$, which covers $H_0$ as a particular case. Using the mean value theorem and either $H_0$ or $H_\ell$, we can express

$$\theta_{0i} - \widehat{f}_i = f_i^* - \widehat{f}_i - (p^{1/4}/n^{1/2})h_i = \sum_{j=1}^{d_\alpha} \frac{\partial f(x_i, \widetilde{\alpha})}{\partial \alpha_j}(\alpha_j^* - \widetilde{\alpha}_j) - \frac{p^{1/4}}{n^{1/2}} h_i, \tag{S.3}$$

where $\widetilde{\alpha}_j$ is a value between $\alpha_j^*$ and $\widehat{\alpha}_j$. Then, for any $j = 1, \ldots, d_\alpha$, $\left|\alpha_j^* - \widetilde{\alpha}_j\right| = O_p(1/\sqrt{n})$. Based on

$$\sup_{\gamma\in\Gamma} \left| u'\Sigma(\gamma)^{-1}\Psi\left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}g\right| = O_p\left(\|g\|\right) \text{ and } \sup_{\gamma\in\Gamma}\left|u'\Sigma(\gamma)^{-1}g\right| = O_p\left(\|g\|\right)$$

for any $\gamma \in \Gamma$ and any conditioned vector $g$, if we take $g = \partial f(x, \alpha)/\partial\alpha_j$ or $g = h$, then both satisfy $O_p\left(\|g\|\right) = O_p\left(\sqrt{n}\right)$ and it follows that

$$\begin{aligned} |A_1| &= \left|u'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(\theta_0 - \widehat{f}\right)\right| \le \sup_{\gamma,\alpha}\sum_{j=1}^{d_\alpha}\left\|u'\Sigma(\gamma)^{-1}\frac{\partial f(x,\alpha)}{\partial\alpha_j}\right\|\left|\alpha_j^* - \widetilde{\alpha}_j\right| + \frac{p^{1/4}}{n^{1/2}}\sup_\gamma\left\|u'\Sigma(\gamma)^{-1}h\right\| \\ &= O_p(\sqrt{n})O_p\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{p^{1/4}}{n^{1/2}}\right)O_p(\sqrt{n}) = O_p(p^{1/4}) = o_p(p^{1/2}). \end{aligned}$$

Similarly,

$$\begin{aligned} |A_3| &= \left|u'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(\theta_0 - \widehat{f}\right)\right| \\ &\le \sup_{\gamma,\alpha}\sum_{j=1}^{d_\alpha}\left\|u'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\frac{\partial f(x,\alpha)}{\partial\alpha_j}\right\|\left|\alpha_j^* - \widetilde{\alpha}_j\right| \\ &\quad + \frac{p^{1/4}}{n^{1/2}}\sup_\gamma\left\|u'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}h\right\| \\ &= O_p(1) + O_p(p^{1/4}) = O_p(p^{1/4}) = o_p(p^{1/2}). \end{aligned}$$

Also, by Assumptions R.2 and R.10, we have

$$\left\|\theta_0 - \widehat{f}\right\| \le \sup_\alpha\sum_{j=1}^{d_\alpha}\left\|\frac{\partial f(x,\alpha)}{\partial\alpha_j}\right\|\left|\alpha_j^* - \widetilde{\alpha}_j\right| + \|h\|\frac{p^{1/4}}{n^{1/2}} = O_p(p^{1/4}). \tag{S.4}$$

By (3.2), we have $\|e\| = O(p^{-\mu}n^{1/2})$ and

$$\begin{aligned} |A_2| &= \left|(u + \theta_0 - \widehat{f})'\left(\Sigma\left(\widehat{\gamma}\right)^{-1} - \Sigma\left(\widehat{\gamma}\right)^{-1}\Psi[\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi]^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\right)e\right| \\ &\le \sup_\gamma\left|u'\Sigma\left(\gamma\right)^{-1}e\right| + \sup_\gamma\left|u'\Sigma\left(\gamma\right)^{-1}\Psi[\Psi'\Sigma\left(\gamma\right)^{-1}\Psi]^{-1}\Psi'\Sigma\left(\gamma\right)^{-1}e\right| \end{aligned}$$

$$+\left\|\theta_0-\widehat{f}\right\|\sup_{\gamma}\left(\left\|\Sigma(\gamma)^{-1}\right\|+\left\|\Sigma(\gamma)^{-1}\Psi[\Psi'\Sigma(\gamma)^{-1}\Psi]^{-1}\Psi'\Sigma(\gamma)^{-1}\right\|\right)\|e\|$$
$$= O_p(p^{-\mu}n^{1/2})+O_p(p^{-\mu+1/4}n^{1/2})=O_p(p^{-\mu+1/4}n^{1/2})=o_p(\sqrt{p}).$$

where the last equality holds under the conditions of the theorem. Finally, under $H_\ell$,

$$\begin{aligned}A_4 &= \left(\theta_0-\widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\left(\theta_0-\widehat{f}\right)\\ &= \left(\theta_0-\widehat{f}\right)'\Sigma^{-1}\left(\theta_0-\widehat{f}\right)+\left(\theta_0-\widehat{f}\right)'\left(\Sigma(\widehat{\gamma})^{-1}-\Sigma^{-1}\right)\left(\theta_0-\widehat{f}\right)\\ &= \frac{p^{1/2}}{n}h'\Sigma^{-1}h+o_p(1)+O_p\left(p^{1/2}d_\gamma^{1/2}/n^{1/2}\right)=\frac{p^{1/2}}{n}h'\Sigma^{-1}h+o_p(\sqrt{p}).\end{aligned}$$

Combining these together, we have

$$n\widehat{m}_n=\widehat{\sigma}^{-2}\widehat{v}'\Sigma(\widehat{\gamma})^{-1}\widehat{u}=\frac{1}{\sigma_0^2}\varepsilon'\mathscr{V}\varepsilon+\frac{p^{1/2}}{n}h'\Sigma^{-1}h+o_p(\sqrt{p}),$$

under $H_\ell$ and the same expression holds with $h=0$ under $H_0$.

□

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

$$\frac{\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon-p}{\sqrt{2p}}. \tag{S.5}$$

Writing $q=\sqrt{2p}$, the ratio in (S.5) has zero mean and variance equal to one, and may be written as $\sum_{s=1}^{\infty}w_s$, where $w_s=\sigma_0^{-2}q^{-1}v_{ss}\left(\varepsilon_s^2-\sigma_0^2\right)+2\sigma_0^{-2}q^{-1}\mathbf{1}(s\geq 2)\varepsilon_s\sum_{t<s}v_{st}\varepsilon_t$, with $v_{st}$ the typical element of $\mathscr{V}$, with $s,t=1,2,\ldots,$. We first show that

$$w_*\xrightarrow{p}0, \tag{S.6}$$

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

$$\mathbb{E}w_*^2\xrightarrow{p}0 \tag{S.7}$$

is sufficient to establish (S.6). Notice that $\mathbb{E}w_s^2\leq Cq^{-2}v_{ss}^2+Cq^{-2}\mathbf{1}(s\geq 2)\sum_{t<s}v_{st}^2\leq Cq^{-2}\sum_{t\leq s}v_{st}^2$, so that, writing $\mathscr{M}=\Sigma^{-1}\Psi[\Psi'\Sigma^{-1}\Psi]^{-1}\Psi'\Sigma^{-1}$,

$$\sum_{s=S+1}^{\infty}\mathbb{E}w_s^2\leq Cq^{-2}\sum_{s=S+1}^{\infty}\sum_{t\leq s}v_{st}^2\leq Cq^{-2}\sum_{s=S+1}^{\infty}b_s'M\sum_{t\leq s}b_tb_t'\mathscr{M}b_s$$

$$
\begin{aligned}
&\leq \quad Cq^{-2}\left\|\Sigma\right\| \sum_{s=S+1}^{\infty} b_s'\mathscr{M}^2 b_s \leq Cq^{-2} \sum_{s=S+1}^{\infty} \sum_{i,j,k=1}^{n} b_{is}b_{kt}m_{ij}m_{kj} \\
&\leq \quad Cq^{-2} \sum_{s=S+1}^{\infty} \sum_{i,k=1}^{n} \left|b_{is}^*\right| \left|b_{ks}^*\right| \sum_{j=1}^{n} \left(m_{kj}^2 + m_{ij}^2\right), \qquad \text{(S.8)}
\end{aligned}
$$

where $m_{ij}$ is the $(i,j)$-th element of $\mathscr{M}$ and we have used the inequality $|ab| \leq \left(a^2+b^2\right)/2$ in the last step. Now, denote by $h_i'$ the $i$-th row of the $n \times p$ matrix $\Sigma^{-1}\Psi$. Denoting the elements of $\Sigma^{-1}$ by $\Sigma_{ij}^{-1}$ and $\psi_{jl} = \psi\left(x_{jl}\right)$, $h_i$ has entries $h_{il} = \sum_{j=1}^{n} \Sigma_{ij}^{-1}\psi_{jl}$, $l = 1,\ldots,p$. We have $|h_{il}| = O_p\left(\sum_{j=1}^{n}\left|\Sigma_{ij}^{-1}\right|\right) = O_p\left(\left\|\Sigma^{-1}\right\|_R\right) = O_p(1)$, uniformly, by Assumptions R.11 and R.13. Thus, we have $\|h_i\| = O_p\left(\sqrt{p}\right)$, uniformly in $i$. As a result,

$$
|m_{ij}| = n^{-1}\left|h_i'\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1}h_j\right| = O_p\left(n^{-1}\|h_i\|\|h_j\|\right) = O_p\left(pn^{-1}\right), \qquad \text{(S.9)}
$$

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

$$
\begin{aligned}
\sum_{j=1}^{n} m_{ij}^2 &= n^{-1}h_i'\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1}\left(n^{-1}\Psi'\Sigma^{-2}\Psi\right)\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1}h_i \\
&\leq n^{-1}\|h_i\|^2\left\|\left(n^{-1}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|^2\left\|n^{-1}\Psi'\Sigma^{-2}\Psi\right\| \\
&= O_p\left(pn^{-2}\|\Psi\|^2\left\|\Sigma^{-1}\right\|^2\right) = O_p\left(pn^{-1}\right), \qquad \text{(S.10)}
\end{aligned}
$$

uniformly in $i$. Thus (S.8) is

$$
O_p\left(q^{-2}pn^{-1}\sum_{i=1}^{n}\sum_{s=S+1}^{\infty}\left|b_{is}^*\right|\sum_{t=1}^{n}\left|b_{ks}^*\right|\right) = O_p\left(q^{-2}p\sup_{i=1,\ldots,n}\sum_{s=S+1}^{\infty}\left|b_{is}^*\right|\right), \qquad \text{(S.11)}
$$

by Assumption R.4. By the same assumption, there exists $S_{in}$ such that $\sum_{s=S_{in}+1}^{\infty}\left|b_{is}^*\right| \leq \varepsilon_n$ for any decreasing sequence $\varepsilon_n \to 0$ as $n \to \infty$. Choosing $S = \max_{i=1,\ldots,n} S_{in}$ in $w_S$, we deduce that (S.11) is $O_p\left(q^{-2}p\varepsilon_n\right) = O_p\left(\varepsilon_n\right) = o_p(1)$, proving (S.7). Thus we need only focus on $w_S$, and seek to establish that

$$
w_S \longrightarrow_d N(0,1), \text{ as } n \to \infty. \qquad \text{(S.12)}
$$

From Scott (1973), (S.12) follows if

$$
\sum_{s=1}^{S}\mathbb{E}w_s^4 \xrightarrow{p} 0, \text{ as } n \to \infty, \qquad \text{(S.13)}
$$

and

$$
\sum_{s=1}^{S}\left[\mathbb{E}\left(w_s^2 \mid \epsilon_t, t < s\right) - \mathbb{E}\left(w_s^2\right)\right] \xrightarrow{p} 0, \text{ as } n \to \infty. \qquad \text{(S.14)}
$$

We show (S.13) first. Evaluating the expectation and using (S.10) yields

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

whence

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

by Assumption R.4. Thus (S.13) is established. Notice that $\mathbb{E}\left(w_s^2\middle|\,\epsilon_t, t<s\right)$ equals

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

and $\mathbb{E}w_s^2 = 4q^{-2}\sigma_0^{-4}\left(\mu_4-\sigma_0^4\right)v_{ss}^2 + 4q^{-2}\mathbf{1}(s\ge 2)\sum_{t<s} v_{st}^2$, so that (S.14) is bounded by a constant times

$$q^{-2}\sum_{s=2}^{S}\sum_{t<s} v_{st}v_{ss}\epsilon_t + \left\{\sum_{s=2}^{S}\left(\sum_{t<s} v_{st}\epsilon_t\right)^2 - \sigma_0^2\sum_{t<s} v_{st}^2\right\}. \tag{S.15}$$

By transforming the range of summation, the square of the first term in (S.15) has expectation bounded by

$$Cq^{-4}\mathbb{E}\left(\sum_{t=1}^{S-1}\sum_{s=t+1}^{S} v_{st}v_{ss}\epsilon_t\right)^2 \le Cq^{-4}\sum_{t=1}^{S-1}\left(\sum_{s=t+1}^{S} v_{st}v_{ss}\right)^2, \tag{S.16}$$

where the factor in parentheses on the RHS of (S.16) is

$$\begin{aligned}
&\sum_{s,r=t+1}^{S} b_s'\mathscr{M}b_s b_s'\mathscr{M}b_t b_r'\mathscr{M}b_r b_r'\mathscr{M}b_t \le \sum_{s,r=t+1}^{S}\left|b_s'\mathscr{M}b_s b_r'\mathscr{M}b_r\right|\left|b_s'\mathscr{M}b_t\right|\left|b_r'\mathscr{M}b_t\right| \\
\le\ & C\sum_{s,r=t+1}^{S}\sum_{i,j,k,l=1}^{n}|b_{is}|\,|m_{ij}|\,|b_{jr}|\,|b_{ks}|\,|m_{lk}|\,|b_{kr}|\left|b_s'\mathscr{M}b_t\right|\left|b_r'\mathscr{M}b_t\right| \\
\le\ & C\left(\sup_{i,j}|m_{ij}|\right)^2\left(\sup_{s\ge 1}\sum_{i=1}^{n}|b_{is}^*|\right)^4\sum_{s,r=t+1}^{S}\left|b_s'\mathscr{M}b_t\right|\left|b_r'\mathscr{M}b_t\right|
\end{aligned}$$

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

where we used Assumptions R.4 and (S.9). Now Assumptions R.4, R.11 and (S.9) imply that

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

so (S.16) is $O_p\left(q^{-4}p^4n^{-2}\sup_t\left(\sum_{i=1}^n|b_{it}^*|\right)\left(\sum_{i=1}^n\left(\sum_{t=1}^{S-1}|b_{it}^*|\right)\right)\right)$. By Assumption R.4 the latter is $O_p\left(q^{-4}p^4n^{-1}\right)$ and therefore the first term in (S.15) is $O_p\left(p^2n^{-1}\right)$, which is negligible.

Once again transforming the summation range and using the inequality $|a+b|^2 \le C\left(a^2+b^2\right)$, we can bound the square of the second term in (S.15) by a constant times

$$\left(\sum_{t=1}^{S-1}\sum_{s=t+1}^{S}v_{st}^2\left(\epsilon_t^2-\sigma_0^2\right)\right)^2 + \left(2\sum_{t=1}^{S-1}\sum_{r=1}^{t-1}\sum_{s=t+1}^{S}v_{st}v_{sr}\epsilon_t\epsilon_r\right)^2. \tag{S.17}$$

Using Assumption R.4, the expectations of the two terms in (S.17) are bounded by a constant times $\alpha_1$ and a constant times $\alpha_2$, respectively, where $\alpha_1 = \sum_{t=1}^{S-1}\left(\sum_{s=t+1}^{S}v_{st}^2\right)^2, \alpha_2 = \sum_{t=1}^{S-1}\sum_{r=1}^{t-1}\left(\sum_{s=t+1}^{S}v_{st}v_{sr}\right)^2$. Thus (S.17) is $O_p\left(\alpha_1+\alpha_2\right)$. Now by (S.9), Assumptions R.4, R.11 and elementary inequalities $\alpha_2$ is bounded by

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

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

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

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

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

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

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

where we used (S.10) in the last step. A similar use of the conditions of the theorem and (S.9) implies that $\alpha_1$ is

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

proving (S.14), as $p^2/n \to 0$ by the conditions of the theorem. □

*Proof of Theorem 4.4.* (1) Follows from Theorems 4.2 and 4.3. (2) Following reasoning analogous to the proofs of Theorems 4.2 and 4.3, it can be shown that under $H_1$, $\widehat{m}_n = n^{-1}\sigma^{*-2}(\theta_0 - f^*)'\Sigma\left(\gamma^*\right)^{-1}(\theta_0 - f^*) + o_p(1)$. Then,

$$
\mathscr{T}_n = \left(n\widehat{m}_n - p\right)/\sqrt{2p} = \left(n/\sqrt{p}\right)(\theta_0 - f^*)'\Sigma\left(\gamma^*\right)^{-1}(\theta_0 - f^*)/\left(\sqrt{2}n\sigma^{*2}\right) + o_p\left(n/\sqrt{p}\right)
$$

and for any nonstochastic sequence $\{C_n\}$, $C_n = o(n/p^{1/2})$, $P(\mathscr{T}_n > C_n) \to 1$, so that consistency follows. (3) Follows from Theorems 4.2 and 4.3. □

*Proof of Theorem 5.1.* We show $\widehat{\phi} \xrightarrow{p} \phi_0$, whence $\widehat{\beta} \xrightarrow{p} \beta_0$ and $\widehat{\sigma}^2 \xrightarrow{p} \sigma_0^2$ follow from (5.3) and (5.4) respectively. First note that

$$\mathcal{L}(\phi) - \mathcal{L} = \log \overline{\sigma}^2(\phi)/\overline{\sigma}^2 - n^{-1}\log\left|T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right| = \log\overline{\sigma}^2(\phi)/\sigma^2(\phi) - \log\overline{\sigma}^2/\sigma_0^2 + \log r(\phi), \tag{S.18}$$

where recall that $\sigma^2(\phi) = n^{-1}\sigma_0^2 tr\left(T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right)$, $\overline{\sigma}^2 = \overline{\sigma}^2(\phi_0) = n^{-1}u'\Sigma'^{-\frac{1}{2}}M\Sigma^{-\frac{1}{2}}u$, using (5.4) and also $r(\phi) = n^{-1}tr\left(T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right)/\left|T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma\right|^{1/n}$. We have $\overline{\sigma}^2(\phi) = n^{-1}\left\{S^{-1'}(\Psi\beta_0+u)\right\}' S'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}S(\lambda)S^{-1}(\Psi\beta_0+u) = c_1(\phi)+c_2(\phi)+c_3(\phi)$, where

$$\begin{aligned}
c_1(\phi) &= n^{-1}\beta_0'\Psi'T'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)\Psi\beta_0,\\
c_2(\phi) &= n^{-1}\sigma_0^2 tr\left(T'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)\Sigma\right),\\
c_3(\phi) &= n^{-1}tr\left(T'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)\left(uu'-\sigma_0^2\Sigma\right)\right)\\
&+ 2n^{-1}\beta_0'\Psi'T'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)u.
\end{aligned}$$

Note that in the particular cases of Theorems 4.1 and 6.1, where $T(\lambda) = I_n$, the $c_1$ term vanishes because $M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}\Psi = 0$ and $M(\tau)\Sigma(\tau)^{-\frac{1}{2}}\Psi = 0$. Proceeding with the current, more general proof

$$\begin{aligned}
\log\frac{\overline{\sigma}^2(\phi)}{\sigma^2(\phi)} &= \log\frac{\overline{\sigma}^2(\phi)}{(c_1(\phi)+c_2(\phi))} + \log\frac{c_1(\phi)+c_2(\phi)}{\sigma^2(\phi)}\\
&= \log\left(1+\frac{c_3(\phi)}{c_1(\phi)+c_2(\phi)}\right) + \log\left(1+\frac{c_1(\phi)-f(\phi)}{\sigma^2(\phi)}\right),
\end{aligned}$$

where $f(\phi) = n^{-1}\sigma_0^2 tr\left(\Sigma'^{\frac{1}{2}}T'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}(I_n - M(\gamma))\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)\Sigma^{\frac{1}{2}}\right)$. Then (S.18) implies

$$\begin{aligned}
P\left(\left\|\widehat{\phi}-\phi_0\right\| \in \overline{\mathcal{N}}^{\phi}(\eta)\right) &= P\left(\inf_{\phi\in\overline{\mathcal{N}}^{\phi}(\eta)} \mathcal{L}(\phi) - \mathcal{L} \le 0\right)\\
&\le P\left(\log\left(1+\sup_{\phi\in\overline{\mathcal{N}}^{\phi}(\eta)}\left|\frac{c_3(\phi)}{c_1(\phi)+c_2(\phi)}\right|\right) + \left|\log\left(\overline{\sigma}^2/\sigma_0^2\right)\right|\right.\\
&\qquad \left.\ge \inf_{\phi\in\overline{\mathcal{N}}^{\phi}(\eta)}\left(\log\left(1+\frac{c_1(\phi)-f(\phi)}{\sigma^2(\phi)}\right)+\log r(\phi)\right)\right),
\end{aligned}$$

where recall that $\overline{\mathcal{N}}^{\phi}(\eta) = \Phi\backslash\mathcal{N}^{\phi}(\eta)$, $\mathcal{N}^{\phi}(\eta) = \{\phi : \|\phi-\phi_0\| < \eta\}\cap\Phi$. Because $\overline{\sigma}^2/\sigma_0^2 \xrightarrow{p} 1$, the property $\log(1+x) = x + o(x)$ as $x \to 0$ implies that it is sufficient to show that

$$\sup_{\phi\in\overline{\mathcal{N}}^{\phi}(\eta)}\left|\frac{c_3(\phi)}{c_1(\phi)+c_2(\phi)}\right| \xrightarrow{p} 0, \tag{S.19}$$

$$\sup_{\phi\in \overline{\mathcal{N}}^{\phi}(\eta)} \left|\frac{f(\phi)}{\sigma^2(\phi)}\right| \xrightarrow{p} 0, \tag{S.20}$$

$$P\left(\inf_{\phi\in \overline{\mathcal{N}}^{\phi}(\eta)} \left\{\frac{c_1(\phi)}{\sigma^2(\phi)} + \log r(\phi)\right\} > 0\right) \longrightarrow 1. \tag{S.21}$$

Because $\overline{\mathcal{N}}^{\phi}(\eta) \subseteq \left\{\Lambda \times \overline{\mathcal{N}}^{\gamma}(\eta/2)\right\} \cup \left\{\overline{\mathcal{N}}^{\lambda}(\eta/2) \times \Gamma\right\}$, we have

$$\begin{aligned} P\left(\inf_{\phi\in \overline{\mathcal{N}}^{\phi}(\eta)} \left\{\frac{c_1(\phi)}{\sigma^2(\phi)} + \log r(\phi)\right\} > 0\right) &\geq P\left(\min\left\{\inf_{\Lambda\times\overline{\mathcal{N}}^{\gamma}(\eta/2)} \frac{c_1(\phi)}{\sigma^2(\phi)}, \inf_{\overline{\mathcal{N}}^{\lambda}(\eta/2)} \log r(\phi)\right\} > 0\right) \\ &\geq P\left(\min\left\{\inf_{\Lambda\times\overline{\mathcal{N}}^{\gamma}(\eta/2)} \frac{c_1(\phi)}{C}, \inf_{\overline{\mathcal{N}}^{\lambda}(\eta/2)} \log r(\phi)\right\} > 0\right), \end{aligned}$$

from Assumption SAR.2, whence Assumptions SAR.3 and SAR.4 imply (S.21). Again using Assumption SAR.2, uniformly in $\phi$, $\left|f(\phi)/\sigma^2(\phi)\right| = O_p(|f(\phi)|)$ and

$$\begin{aligned} |f(\phi)| &= O_p\left(tr\left(\Sigma'^{\frac{1}{2}}T'(\lambda)\Sigma(\gamma)^{-1}\Psi\left(\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}T(\lambda)\Sigma^{\frac{1}{2}}\right)/n\right) \\ &= O_p\left(tr\left(\Sigma'^{\frac{1}{2}}T'(\lambda)\Sigma(\gamma)^{-1}\Psi\Psi'\Sigma(\gamma)^{-1}T(\lambda)\Sigma^{\frac{1}{2}}\right)/n^2\right) = O_p\left(\left\|\Psi'\Sigma(\gamma)^{-1}T(\lambda)\Sigma^{\frac{1}{2}}/n\right\|_F^2\right) \\ &= O_p\left(\|\Psi/n\|_F^2\,\overline{\varphi}^2\left(\Sigma(\gamma)^{-1}\right)\|T(\lambda)\|^2\left\|\Sigma^{\frac{1}{2}}\right\|^2\right) = O_p\left(\|\Psi/n\|_F^2\,\|T(\lambda)\|^2\,\overline{\varphi}(\Sigma)/\underline{\varphi}^2(\Sigma(\gamma))\right) \\ &= O_p\left(\|T(\lambda)\|^2/n\right), \end{aligned} \tag{S.22}$$

where we have twice made use of the inequality

$$\|AB\|_F \leq \|A\|_F\,\|B\| \tag{S.23}$$

for generic multiplication compatible matrices $A$ and $B$. (S.20) now follows by Assumption SAR.1 and compactness of $\Lambda$ because $T(\lambda) = I_n + \sum_{j=1}^{d_\lambda}(\lambda_{0j} - \lambda_j)\,G_j$. Finally consider (S.19). We first prove pointwise convergence. For any fixed $\phi \in \overline{\mathcal{N}}^{\phi}(\eta)$ and large enough $n$, Assumptions SAR.2 and SAR.4 imply

$$\begin{aligned} \{c_1(\phi)\}^{-1} &= O_p\left(\|\beta_0\|^{-2}\right) = O_p(1) \\ \{c_2(\phi)\}^{-1} &= O_p(1), \end{aligned} \tag{S.24, S.25}$$

because $\left\{n^{-1}\sigma_0^2 tr\left(\Sigma'^{\frac{1}{2}}T'(\lambda)\Sigma(\gamma)^{-1}T(\lambda)\Sigma^{\frac{1}{2}}\right)\right\}^{-1} = O_p(1)$ and, proceeding like in the bound for $|f(\phi)|$, $tr\left(\Sigma'^{\frac{1}{2}}T'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}(I - M(\gamma))\,\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)\Sigma^{\frac{1}{2}}\right) = O_p\left(\|T(\lambda)\|^2/n\right) = O_p(1/n)$. In fact it is worth noting for the equicontinuity argument presented later that Assumptions SAR.2 and SAR.4 actually imply that (S.24) and (S.25) hold uniformly over $\overline{\mathcal{N}}^{\phi}(\eta)$, a property not needed for the present pointwise arguments. Thus $c_3(\phi)/(c_1(\phi)+c_2(\phi)) = O_p(|c_3(\phi)|)$ where, writing $\mathfrak{B}(\phi) = T'(\lambda)\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)$ with typical element $\mathfrak{b}_{rs}(\phi)$, $r,s=1,\ldots,n$, $c_3(\phi)$ has mean 0 and variance

$$O_p\left(\frac{\|\mathfrak{B}(\phi)\Sigma\|_F^2}{n^2} + \frac{\sum_{r,s,t,v=1}^{n}\mathfrak{b}_{rs}(\phi)\mathfrak{b}_{tv}(\phi)\kappa_{rstv}}{n^2} + \frac{\left\|\beta_0'\Psi'\mathfrak{B}(\phi)\Sigma^{\frac{1}{2}}\right\|^2}{n^2}\right), \tag{S.26}$$

with $\kappa_{rstv}$ denoting the fourth cumulant of $u_r, u_s, u_t, u_v$, $r,s,t,v=1,\ldots,n$. Under the linear process assumed in Assumption R.4 it is known that

$$\sum_{r,s,t,v=1}^{n}\kappa_{rstv}^2 = O(n). \tag{S.27}$$

Using (S.23) and Assumptions SAR.1 and R.3, the first term in parentheses in (S.26) is

$$\begin{aligned} O_p\left(\|\mathfrak{B}(\phi)\|_F^2\,\overline{\varphi}^2(\Sigma)/n^2\right) &= O_p\left(\|T(\lambda)\|_F^2\left\|\Sigma(\gamma)^{-\frac{1}{2}}\right\|^4\|M(\gamma)\|^2\|T(\lambda)\|^2/n^2\right) \\ &= O_p\left(\|T(\lambda)\|^4/n\underline{\varphi}^2(\Sigma(\gamma))\right) = O_p\left(\|T(\lambda)\|^4/n\right), \end{aligned} \tag{S.28}$$

while the second is similarly

$$O_p\left\{\left(\|\mathfrak{B}(\phi)\|_F^2/n\right)\left(\sum_{r,s,t,v=1}^{n}\kappa_{rstv}^2/n^2\right)^{\frac{1}{2}}\right\} = o_p\left(\|T(\lambda)\|^4\right), \tag{S.29}$$

using (S.27). Finally, the third term in parentheses in (S.26) is

$$O_p\left(\|\mathfrak{B}(\phi)\|^2/n\right) = O_p\left(\|T(\lambda)\|^4/n\right). \tag{S.30}$$

By compactness of $\Lambda$ and Assumption SAR.1, (S.28), (S.29) and (S.30) are negligible, thus pointwise convergence is established.

Uniform convergence will follow from an equicontinuity argument. First, for arbitrary $\varepsilon > 0$ we can find points $\phi_* = (\lambda_*', \gamma_*')'$, possibly infinitely many, such that the neighbourhoods $\|\phi - \phi^*\| < \varepsilon$ form an open cover of $\overline{\mathcal{N}}^{\phi}(\eta)$. Since $\Phi$ is compact any open cover has a finite subcover and thus we may in fact choose finitely many $\phi_* = (\lambda_*', \gamma_*')'$, whence it suffices to prove

$$\sup_{\|\phi-\phi_*\|<\varepsilon}\left|\frac{c_3(\phi)}{c_1(\phi)+c_2(\phi)} - \frac{c_3(\phi_*)}{c_1(\phi_*)+c_2(\phi_*)}\right| \xrightarrow{p} 0.$$

Proceeding as in Gupta and Robinson (2018), we denote the two components of $c_3(\phi)$ by $c_{31}(\phi)$,

$c_{32}(\phi)$, and are left with establishing the negligibility of

$$\begin{aligned} & \frac{|c_{31}(\phi)-c_{31}(\phi_*)|}{c_2(\phi)} + \frac{|c_{32}(\phi)-c_{32}(\phi_*)|}{c_1(\phi)} + \frac{|c_3(\phi_*)|}{c_1(\phi)\,c_1(\phi_*)}\,|c_1(\phi_*)-c_1(\phi)| \\ + \; & \frac{|c_3(\phi_*)|}{c_2(\phi)\,c_2(\phi_*)}\,|c_2(\phi_*)-c_2(\phi)|, \end{aligned} \tag{S.31}$$

uniformly on $\|\phi-\phi_*\|<\varepsilon$. By the fact that (S.24) and (S.25) hold uniformly over $\Phi$, we first consider only the numerators in the first two terms in (S.31). As in the proof of Theorem 1 of Delgado and Robinson (2015), (S.23) implies that $\mathbb{E}\left(\sup_{\|\phi-\phi_*\|<\varepsilon}|c_{31}(\phi)-c_{31}(\phi_*)|\right)$ is bounded by

$$n^{-1}\left(\mathbb{E}\|u\|^2+\sigma_0^2 tr\Sigma\right)\sup_{\|\phi-\phi_*\|<\varepsilon}\|\mathfrak{B}(\phi)-\mathfrak{B}(\phi_*)\| = O_p\left(\sup_{\|\phi-\phi_*\|<\varepsilon}\|\mathfrak{B}(\phi)-\mathfrak{B}(\phi_*)\|\right),$$

because $\mathbb{E}\|u\|^2=O(n)$ and $tr\Sigma=O(n)$. $\mathfrak{B}(\phi)-\mathfrak{B}(\phi_*)$ can be written as

$$\begin{aligned} & (T(\lambda)-T(\lambda_*))'\,\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}T(\lambda)+T(\lambda_*)'\Sigma'(\gamma_*)M(\gamma_*)\Sigma(\gamma_*)^{-\frac{1}{2}}\,(T(\lambda)-T(\lambda_*)) \\ + \; & T'(\lambda_*)\left(\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}-\Sigma(\gamma_*)'^{-\frac{1}{2}}M(\gamma_*)\Sigma(\gamma_*)^{-\frac{1}{2}}\right)T(\lambda), \end{aligned} \tag{S.32}$$

which, by the triangle inequality, has spectral norm bounded by

$$\begin{aligned} & \|T(\lambda)-T(\lambda_*)\|\left(\left\|\Sigma(\gamma)^{-\frac{1}{2}}\right\|^2\|T(\lambda)\|+\left\|\Sigma(\gamma_*)^{-\frac{1}{2}}\right\|^2\|T(\lambda_*)\|\right) \\ + \; & \|T(\lambda_*)\|\left\|\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}-\Sigma(\gamma_*)'^{-\frac{1}{2}}M(\gamma_*)\Sigma(\gamma_*)^{-\frac{1}{2}}\right\|\|T(\lambda)\| \\ = \; & O_p\left(\|T(\lambda)-T(\lambda_*)\|+\left\|\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}}-\Sigma(\gamma_*)'^{-\frac{1}{2}}M(\gamma_*)\Sigma(\gamma_*)^{-\frac{1}{2}}\right\|\right). \end{aligned} \tag{S.33}$$

By Assumption SAR.1 the first term in parentheses on the right side of (S.33) is bounded uniformly on $\|\phi-\phi_*\|<\varepsilon$ by

$$\sum_{j=1}^{d_\lambda}|\lambda_j-\lambda_{*j}|\,\|G_j\|\le\max_{j=1,\ldots,d_\lambda}\|G_j\|\,\|\lambda-\lambda_*\|=O_p(\varepsilon), \tag{S.34}$$

while because $\Sigma(\gamma)'^{-\frac{1}{2}}M(\gamma)\Sigma(\gamma)^{-\frac{1}{2}} = n^{-1}\Sigma(\gamma)^{-1}\Psi\left(n^{-1}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\Psi'\Sigma(\gamma)^{-1}$ for any $\gamma\in\Gamma$, the second one can be decomposed into terms with bounds typified by

$$\begin{aligned} & n^{-1}\left\|\Sigma(\gamma)^{-1}-\Sigma(\gamma_*)^{-1}\right\|\|\Psi\|^2\left\|\left(n^{-1}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\right\|\left\|\Sigma(\gamma)^{-1}\right\|^2 \\ \le \; & n^{-1}\|\Sigma(\gamma)-\Sigma(\gamma_*)\|\,\|\Psi\|^2\left\|\left(n^{-1}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\right\|\left\|\Sigma(\gamma)^{-1}\right\|^3\left\|\Sigma(\gamma_*)^{-1}\right\| \\ = \; & O_p\left(\|\Sigma(\gamma)-\Sigma(\gamma_*)\|\right)=O_p(\varepsilon), \end{aligned}$$

uniformly on $\|\phi-\phi_*\|<\varepsilon$, by Assumptions R.3 and R.8, Proposition 4.1 and the inequality $\|A\| \leq \|A\|_F$ for a generic matrix $A$, so that

$$\sup_{\|\phi-\phi_*\|<\varepsilon} \|\mathfrak{B}(\phi)-\mathfrak{B}(\phi_*)\| = O_p(\varepsilon). \tag{S.35}$$

Thus equicontinuity of the first term in (S.31) follows because $\varepsilon$ is arbitrary. The equicontinuity of the second term in (S.31) follows in much the same way. Indeed $\sup_{\|\phi-\phi_*\|<\varepsilon} c_{32}(\phi) - c_{32}(\phi_*) = 2n^{-1}\beta_0'\Psi' \sup_{\|\phi-\phi_*\|<\varepsilon} (\mathfrak{B}(\phi)-\mathfrak{B}(\phi_*))\, u = O_p\left(\sup_{\|\phi-\phi_*\|<\varepsilon}\|\mathfrak{B}(\phi)-\mathfrak{B}(\phi_*)\|\right) = O_p(\varepsilon)$, using earlier arguments and (S.35). Because $c_1(\phi)$ is bounded and bounded away from zero in probability (see S.24) for sufficiently large $n$ and all $\phi \in \overline{\mathcal{N}}^{\phi}(\eta)$, the third term in (S.31) may be bounded by $|c_3(\phi_*)|/c_1(\phi_*)\,(1+c_1(\phi_*)/c_1(\phi)) \xrightarrow{p} 0$, convergence being uniform on $\|\phi-\phi_*\|<\varepsilon$ by pointwise convergence of $c_3(\phi)/\left(c_1(\phi)+c_2(\phi)\right)$, cf. Gupta and Robinson (2018). The uniform convergence to zero of the fourth term in (S.31) follows in identical fashion, because $c_2(\phi)$ is bounded and bounded away from zero (see (S.25)) in probability for sufficiently large $n$ and all $\phi \in \overline{\mathcal{N}}^{\phi}(\eta)$. This concludes the proof. □

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

$$\begin{aligned}
A_5 &= \sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j) y' W_j' \Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j) W_j y,\\
A_6 &= \sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j) y' W_j' \Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(u+\theta_0-\widehat{f}),\\
A_7 &= \left(\Psi\left(\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi\right)^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}(u+e) - e + \theta_0 - \widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j) W_j y.
\end{aligned}$$

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

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

Uniformly in $\gamma$ and $j$,

$$\begin{aligned}
&E\left(u'S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi[\Psi'\Sigma(\gamma)^{-1}\Psi]^{-1}\Psi'\Sigma(\gamma)^{-1}u\right)\\
=\;&Etr\left(\left(\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Psi\right)^{-1}\frac{1}{n}\Psi'\Sigma(\gamma)^{-1}\Sigma S^{-1\prime}W_j'\Sigma(\gamma)^{-1}\Psi\right)=O_p(p)
\end{aligned}$$

and

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

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

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

and

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

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

$$\left|\sum_{j=1}^{d_\lambda}(\lambda_{j0}-\widehat{\lambda}_j)y'W_j'\Sigma(\widehat{\gamma})^{-1}\Psi[\Psi'\Sigma(\widehat{\gamma})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\gamma})^{-1}e\right|$$

$$
\begin{aligned}
&\leq \; d_\lambda \left\| \lambda_0 - \widehat{\lambda} \right\| \|y\| \sup_j \|W_j\| \sup_\gamma \left\| \frac{1}{n} \Psi \left( \frac{1}{n} \Psi' \Sigma(\gamma)^{-1} \Psi \right)^{-1} \Psi \right\| \sup_\gamma \left\| \Sigma(\gamma)^{-1} \right\|^2 \|e\| \\
&= \; O_p\left(\sqrt{d_\gamma/n}\right) O_p\left(\sqrt{n}\right) O_p\left(p^{-\mu}\sqrt{n}\right) = O_p\left(\sqrt{d_\gamma} p^{-\mu}\sqrt{n}\right) = o_p(\sqrt{p}),
\end{aligned}
$$

and

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

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

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

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

$$
\|\Sigma(t_1) - \Sigma(t_1)\| \leq \sup_{t\in\mathcal{T}^o} \|D\Sigma(t)\|_{\mathscr{L}(\mathcal{T}^o, \mathcal{M}^{n\times n})} \left( \|\gamma_1 - \gamma_2\| + \sum_{i=1}^{d_\zeta} \left\| (\delta_{i1} - \delta_{i2})' \varphi_i \right\|_w \right), \tag{S.36}
$$

where

$$
\begin{aligned}
\sum_{i=1}^{d_\zeta} \left\| (\delta_{i1} - \delta_{i2})' \varphi_i \right\|_w &= \sum_{i=1}^{d_\zeta} \sup_{z\in\mathcal{Z}} \left| (\delta_{i1} - \delta_{i2})' \varphi_i \right| \left( 1 + \|z\|^2 \right)^{-w/2} \\
&\leq \sum_{i=1}^{d_\zeta} \|\delta_{i1} - \delta_{i2}\| \sup_{z\in\mathcal{Z}} \|\varphi_i\| \left( 1 + \|z\|^2 \right)^{-w/2} \\
&\leq C\varsigma(r) \sum_{i=1}^{d_\zeta} \|\delta_{i1} - \delta_{i2}\| \leq C\varsigma(r) \|t_1 - t_2\|.
\end{aligned}
$$

The claim now follows by (6.7) in Assumption NPN.2, because $\|\gamma_1 - \gamma_2\| \leq C\varsigma(r)\|t_1 - t_2\|$ for some suitably chosen $C$. □

*Proof of Theorem 6.1.* The proof is omitted as it is entirely analogous to that of Theorem 5.1, with the exception of one difference when proving equicontinuity. In the setting of Section 6, we obtain via Proposition 6.1 $\|\Sigma(\tau)-\Sigma(\tau^*)\| = O_p(\varsigma(r)\varepsilon)$. Because $\varepsilon>0$ is arbitrarily small we may choose it smaller than $\varepsilon'/\varsigma(r)$, for some arbitrary $\epsilon'>0$. □

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

$$\begin{aligned}\left\|\Sigma(\widehat{\tau})-\Sigma\right\| = O_p\left(\varsigma(r)\left(\|\widehat{\gamma}-\gamma_0\|+\left\|\widehat{\delta}-\zeta_0\right\|\right)\right) &= O_p\left(\varsigma(r)\left(\|\widehat{\tau}-\tau_0\|+\|\nu\|\right)\right)\\ &= O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n},\sqrt{\sum_{i=1}^{d_\zeta} r_i^{-2\kappa_i}}\right\}\right),\end{aligned}$$

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

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

And similarly,

$$\begin{aligned}&\left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\tau})^{-1}\Psi\right)^{-1}-\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|\\ \le\;& \left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\tau})^{-1}\Psi\right)^{-1}\right\|\left\|\frac{1}{n}\Psi'\left(\Sigma(\widehat{\tau})^{-1}-\Sigma^{-1}\right)\Psi\right\|\left\|\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|\\ =\;& O_p\left(\left\|\Sigma(\widehat{\tau})^{-1}-\Sigma^{-1}\right\|\right)=O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n},\sqrt{\sum_{i=1}^{d_\zeta} r_i^{-2\kappa_i}}\right\}\right).\end{aligned}$$

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

$$\begin{aligned}&\left|u'\left(\Sigma(\widehat{\tau})^{-1}\Psi[\Psi'\Sigma(\widehat{\tau})^{-1}\Psi]^{-1}\Psi'\Sigma(\widehat{\tau})^{-1}-\Sigma^{-1}\Psi[\Psi'\Sigma^{-1}\Psi]^{-1}\Psi'\Sigma^{-1}\right)u\right|\\ \le\;& 2\left(\sup_t\left\|\frac{1}{\sqrt{n}}u'\Sigma(t)^{-1}\Psi\right\|\left\|\left(\frac{1}{n}\Psi'\Sigma(t)^{-1}\Psi\right)^{-1}\right\|\right)\sum_{j=1}^{d_\tau}\left\|\frac{1}{\sqrt{n}}\Psi'\left(\Sigma(\widetilde{\tau})^{-1}\Sigma_j(\widetilde{\tau})\Sigma(\widetilde{\tau})^{-1}\right)u\right\|\\ \times\;& |\widetilde{\tau}_j-\tau_{j0}|+2\sup_t\left\|\frac{1}{\sqrt{n}}u'\Sigma(t)^{-1}\Psi\right\|\left\|\left(\frac{1}{n}\Psi'\Sigma(t)^{-1}\Psi\right)^{-1}\right\|\left\|\frac{1}{\sqrt{n}}\Psi'(\Sigma_0-\Sigma)u\right\|\\ &+\left\|\frac{1}{\sqrt{n}}u'\Sigma^{-1}\Psi\right\|^2\left\|\left(\frac{1}{n}\Psi'\Sigma(\widehat{\tau})^{-1}\Psi\right)^{-1}-\left(\frac{1}{n}\Psi'\Sigma^{-1}\Psi\right)^{-1}\right\|\end{aligned}$$

$$
\begin{aligned}
&= O_p(\sqrt{p})O_p(d_\tau\sqrt{p}\varsigma(r)/\sqrt{n}) + O_p(\sqrt{p})O_p\left(\sqrt{p}\varsigma(r)\sqrt{\sum_{i=1}^{d_\zeta} r_i^{-2\kappa_i}}\right) \\
&+ O_p(p)O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n}, \sqrt{\sum_{i=1}^{d_\zeta} r_i^{-2\kappa_i}}\right\}\right) \\
&= O_p\left(p\varsigma(r)\max\left\{d_\tau/\sqrt{n}, \sqrt{\sum_{i=1}^{d_\zeta} r_i^{-2\kappa_i}}\right\}\right) = o_p(\sqrt{p}),
\end{aligned}
$$

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

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

However, $A_4$ has a different order. Under $H_\ell$,

$$
\begin{aligned}
A_4 &= \left(\theta_0 - \widehat{f}\right)'\Sigma(\widehat{\gamma})^{-1}\left(\theta_0 - \widehat{f}\right) \\
&= \left(\theta_0 - \widehat{f}\right)'\Sigma_0^{-1}\left(\theta_0 - \widehat{f}\right) + \left(\theta_0 - \widehat{f}\right)'\left(\Sigma(\widehat{\tau})^{-1} - \Sigma^{-1}\right)\left(\theta_0 - \widehat{f}\right) \\
&= \frac{p^{1/2}}{n}h'\Sigma_0^{-1}h + o_p(1) + O_p\left(p^{1/2}\right)O_p\left(\varsigma(r)\max\left\{\sqrt{d_\tau/n}, \sqrt{\sum_{i=1}^{d_\zeta} r_i^{-2\kappa_i}}\right\}\right)
\end{aligned}
$$

$$= \quad \frac{p^{1/2}}{n} h' \Sigma_0^{-1} h + o_p(\sqrt{p}),$$

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

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

# B Lemmas

**Lemma LS.1.** *Under the conditions of Theorem 4.1,* $c_1(\gamma) = n^{-1}\beta'\Psi'C'(\gamma)M(\gamma)C(\gamma)\Psi\beta + o_p(1)$.

*Proof.* First,

$$c_1(\gamma) = n^{-1}\beta'\Psi'C'(\gamma)M(\gamma)C(\gamma)\Psi\beta + c_{12}(\gamma) + c_{13}(\gamma).$$

with $c_{12}(\gamma) = 2n^{-1}e'C'(\gamma)M(\gamma)C(\gamma)\Psi\beta$ and $c_{13}(\gamma) = n^{-1}e'C'(\gamma)M(\gamma)C(\gamma)e$. It is readily seen that $c_{12}(\gamma)$ and $c_{13}(\gamma)$ are negligible. □

**Lemma LS.2.** *Under the conditions of Theorem 4.2 or Theorem 5.2,* $\|\widehat{\gamma} - \gamma_0\| = O_p\left(\sqrt{d_\gamma/n}\right)$.

*Proof.* We show the details for the setting of Theorem 4.2 and omit the details for the setting of Theorem 5.2. Write $l = \partial L(\beta_0, \gamma_0)/\partial\gamma$. By Robinson (1988), we have $\|\widehat{\gamma} - \gamma_0\| = O_p(\|l\|)$. Now $l = \left(l_1, \ldots, l_{d_\gamma}\right)'$, with $l_j = n^{-1}tr\left(\Sigma^{-1}\Sigma_j\right) - n^{-1}\sigma_0^{-2}u'\Sigma^{-1}\Sigma_j\Sigma^{-1}u$. Next, $E\|l\|^2 = \sum_{j=1}^{d_\gamma} E\left(l_j^2\right)$ and

$$E\left(l_j^2\right) = \frac{1}{n^2\sigma_0^4} var\left(u'\Sigma^{-1}\Sigma_j\Sigma^{-1}u\right) = \frac{1}{n^2\sigma_0^4} var\left(\varepsilon'B'\Sigma^{-1}\Sigma_j\Sigma^{-1}B\varepsilon\right) = \frac{1}{n^2\sigma_0^4} var\left(\varepsilon'D_j\varepsilon\right), \quad \text{(S.1)}$$

say. But, writing $d_{j,st}$ for a typical element of the infinite dimensional matrix $D_j$, we have

$$var\left(\varepsilon'D_j\varepsilon\right) = \left(\mu_4 - 3\sigma_0^4\right)\sum_{s=1}^{\infty} d_{j,ss}^2 + 2\sigma_0^4 tr\left(D_j^2\right) = \left(\mu_4 - 3\sigma_0^4\right)\sum_{s=1}^{\infty} d_{j,ss}^2 + 2\sigma_0^4\sum_{s,t=1}^{\infty} d_{j,st}^2. \quad \text{(S.2)}$$

Next, by Assumptions R.4, R.3 and R.9

$$\sum_{s=1}^{\infty} d_{j,ss}^2 = \sum_{s=1}^{\infty}\left(b_s'\Sigma^{-1}\Sigma_j\Sigma^{-1}b_s\right)^2 \leq \left(\sum_{s=1}^{\infty}\|b_s\|^2\right)\left\|\Sigma^{-1}\right\|^2\|\Sigma_j\| = O\left(\sum_{j=1}^{n}\sum_{s=1}^{\infty} b_{js}^{*2}\right) = O(n). \quad \text{(S.3)}$$

Similarly,

$$\sum_{s,t=1}^{\infty} d_{j,st}^2 = \sum_{s=1}^{\infty} b_s'\Sigma^{-1}\Sigma_j\Sigma^{-1}\left(\sum_{t=1}^{\infty} b_t b_t'\right)\Sigma^{-1}\Sigma_j\Sigma^{-1}b_s = \sum_{s=1}^{\infty} b_s'\Sigma^{-1}\Sigma_j\Sigma^{-1}\Sigma_j\Sigma^{-1}b_s = O(n). \quad \text{(S.4)}$$

Using (S.3) and (S.4) in (S.2) implies that $E\left(l_j^2\right) = O\left(n^{-1}\right)$, by (S.1). Thus we have $E\left\|l\right\|^2 = O\left(d_\gamma/n\right)$, and thus $\|l\| = O_p\left(\sqrt{d_\gamma/n}\right)$, by Markov's inequality, proving the lemma. □

**Lemma LS.3.** *Under the conditions of Theorem 4.3,* $E\left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon\right) = p$ *and* $Var\left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon\right)/2p \to 1$.

*Proof.* As $E\left(\sigma_0^{-2}\varepsilon'\mathscr{V}\varepsilon\right) = tr\left(E[B'\Sigma^{-1}\Psi(\Psi'\Sigma^{-1}\Psi)^{-1}\Psi'\Sigma^{-1}B]\right) = p$, and

$$Var\left(\frac{1}{\sigma_0^2}\varepsilon'\mathscr{V}\varepsilon\right) = \left(\frac{\mu_4}{\sigma_0^4} - 3\right)\sum_{s=1}^{\infty} E(v_{ss}^2) + E[tr(\mathscr{V}\mathscr{V}') + tr(\mathscr{V}^2)] = \left(\frac{\mu_4}{\sigma_0^4} - 3\right)\sum_{s=1}^{\infty} v_{ss}^2 + 2p, \quad \text{(S.5)}$$

it suffices to show that

$$(2p)^{-1}\sum_{s=1}^{\infty} v_{ss}^2 \xrightarrow{p} 0. \quad \text{(S.6)}$$

Because $v_{ss} = b_s'\mathscr{M}b_s$, we have $v_{ss}^2 = \left(\sum_{i,j=1}^{n} b_{is}b_{js}m_{ij}\right)^2$. Thus, using Assumption R.4 and (S.9), we have

$$\begin{aligned}\sum_{s=1}^{\infty} v_{ss}^2 &\leq \left(\sup_{i,j}|m_{ij}|\right)^2 \sum_{s=1}^{\infty}\left(\sum_{i,j=1}^{n}|b_{is}^*|\left|b_{js}^*\right|\right)^2 = O_p\left(p^2 n^{-2}\left(\sup_s \sum_{i=1}^{n}|b_{is}^*|\right)^3 \sum_{i=1}^{n}\sum_{s=1}^{\infty}|b_{is}^*|\right) \\ &= O_p\left(p^2 n^{-1}\right), \end{aligned} \quad \text{(S.7)}$$

establishing (S.6) because $p^2/n \to 0$. □

**Lemma LS.4.** *Under the conditions of Theorem 6.2,* $\|\widehat{\tau} - \tau_0\| = O_p\left(\sqrt{d_\tau/n}\right)$.

*Proof.* The proof is similar to that of Lemma LS.2 and is omitted. □

Let $G_j(\gamma) = W_j K^{-1}(\gamma)$, $j = 1, \ldots, m_1$, $T_j = H^{-1}(\gamma)W_j$, $j = m_1 + 1, \ldots, m_1 + m_2$ and, for a generic matrix $A$, denote $\overline{A} = A + A'$. Our final conditions may differ according to whether the $W_j$ are of general form or have 'single nonzero diagonal block structure', see e.g Gupta and Robinson (2015). To define these, denote by $V$ an $n \times n$ block diagonal matrix with $i$-th block $V_i$, a $s_i \times s_i$ matrix, where $\sum_{i=1}^{m_1+m_2} s_i = n$, and for $i = 1, ..., m_1 + m_2$ obtain $W_j$ from $V$ by replacing each $V_j$, $j \neq i$, by a matrix of zeros. Thus $V = \sum_{i=1}^{m_1+m_2} W_j$.

**Lemma LS.5.** *For the spatial error model with SARMA$(p, q)$ errors, if*

$$\sup_{\gamma\in\Gamma^o}\left(\left\|K^{-1}(\gamma)\right\| + \left\|K'^{-1}(\gamma)\right\| + \left\|H^{-1}(\gamma)\right\| + \left\|H'^{-1}(\gamma)\right\|\right) + \max_{j=1,\ldots,m_1+m_2}\|W_j\| < C, \quad \text{(S.8)}$$

*then*

$$\left(D\Sigma(\gamma)\right)\left(\gamma^\dagger\right) = A^{-1}(\gamma)\left(\sum_{j=1}^{m_1}\gamma_j^\dagger\overline{H^{-1}(\gamma)G_j(\gamma)} + \sum_{j=m_1+1}^{m_1+m_2}\gamma_j^\dagger\overline{T_j(\gamma)}\right)A'^{-1}(\gamma).$$

*Proof.* We first show that $D\Sigma \in \mathscr{L}(\Gamma^o, \mathcal{M}^{n\times n})$. Clearly, $D\Sigma$ is a linear map and (S.8)

$$\left\|(D\Sigma(\gamma))\left(\gamma^{\dagger}\right)\right\| \leq C\left\|\gamma^{\dagger}\right\|_1,$$

in the general case and

$$\left\|(D\Sigma(\gamma))\left(\gamma^{\dagger}\right)\right\| \leq C \max_{j=1,\ldots,m_1+m_2}\left|\gamma_j^{\dagger}\right|,$$

in the 'single nonzero diagonal block' case. Thus $D\Sigma$ is a bounded linear operator between two normed linear spaces, i.e. it is a continuous linear operator.

With $A(\gamma) = H^{-1}(\gamma)K(\gamma)$, we now show that

$$\frac{A^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}\left(\gamma+\gamma^{\dagger}\right) - A^{-1}(\gamma)A'^{-1}(\gamma) - (D\Sigma(\gamma))\left(\gamma^{\dagger}\right)}{\|\gamma^{\dagger}\|_g} \to 0, \text{ as } \left\|\gamma^{\dagger}\right\|_g \to 0, \tag{S.9}$$

where $\|\cdot\|_g$ is either the 1-norm or the max norm on $\Gamma$. First, note that

$$\begin{aligned}
& A^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}\left(\gamma+\gamma^{\dagger}\right) - A^{-1}(\gamma)A'^{-1}(\gamma) \\
= \; & A^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(A^{-1}\left(\gamma+\gamma^{\dagger}\right) - A^{-1}(\gamma)\right)' + \left(A^{-1}\left(\gamma+\gamma^{\dagger}\right) - A^{-1}(\gamma)\right)A^{-1}(\gamma) \\
= \; & -A^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(A\left(\gamma+\gamma^{\dagger}\right) - A(\gamma)\right)'A'^{-1}(\gamma) \\
- \; & A^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(A\left(\gamma+\gamma^{\dagger}\right) - A(\gamma)\right)A^{-1}(\gamma)A'^{-1}(\gamma).
\end{aligned} \tag{S.10}$$

Next,

$$\begin{aligned}
A\left(\gamma+\gamma^{\dagger}\right) - A(\gamma) &= H^{-1}\left(\gamma+\gamma^{\dagger}\right)K\left(\gamma+\gamma^{\dagger}\right) - H^{-1}(\gamma)K(\gamma) \\
&= H^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(K\left(\gamma+\gamma^{\dagger}\right) - K(\gamma)\right) \\
&+ H^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(H(\gamma) - H\left(\gamma+\gamma^{\dagger}\right)\right)H^{-1}(\gamma)K(\gamma) \\
&= -H^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(\sum_{j=1}^{m_1}\gamma_j^{\dagger}W_j + \sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}W_jH^{-1}(\gamma)K(\gamma)\right).
\end{aligned} \tag{S.11}$$

Substituting (S.11) in (S.10) implies that

$$A^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}\left(\gamma+\gamma^{\dagger}\right) - A^{-1}(\gamma)A'^{-1}(\gamma) = \Delta_1\left(\gamma,\gamma^{\dagger}\right) + \Delta_2\left(\gamma,\gamma^{\dagger}\right) = \Delta\left(\gamma,\gamma^{\dagger}\right), \tag{S.12}$$

say, where

$$
\begin{aligned}
\Delta_1\left(\gamma,\gamma^{\dagger}\right) &= A^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(\sum_{j=1}^{m_1}\gamma_j^{\dagger}W_j'+K'(\gamma)H'^{-1}(\gamma)\sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}W_j'\right)\\
&\times\ H'^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}(\gamma),\\
\Delta_2\left(\gamma,\gamma^{\dagger}\right) &= A^{-1}\left(\gamma+\gamma^{\dagger}\right)H^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(\sum_{j=1}^{m_1}\gamma_j^{\dagger}W_j+\sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}W_jH^{-1}(\gamma)K(\gamma)\right)\\
&\times\ A^{-1}(\gamma)A'^{-1}(\gamma).
\end{aligned}
$$

From the definitions above and recalling that $A(\gamma)=H^{-1}(\gamma)K(\gamma)$, we can write

$$
\Delta\left(\gamma,\gamma^{\dagger}\right)=A^{-1}\left(\gamma+\gamma^{\dagger}\right)\Upsilon\left(\gamma,\gamma^{\dagger}\right)A'^{-1}(\gamma), \tag{S.13}
$$

with

$$
\begin{aligned}
\Upsilon\left(\gamma,\gamma^{\dagger}\right) &= \sum_{j=1}^{m_1}\gamma_j^{\dagger}G_j'\left(\gamma+\gamma^{\dagger}\right)H'^{-1}\left(\gamma+\gamma^{\dagger}\right)+A'^{-1}\left(\gamma+\gamma^{\dagger}\right)A'(\gamma)\sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}T_j'\left(\gamma+\gamma^{\dagger}\right)\\
&+\ \sum_{j=1}^{m_1}\gamma_j^{\dagger}H^{-1}\left(\gamma+\gamma^{\dagger}\right)G_j(\gamma)+\sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}T_j\left(\gamma+\gamma^{\dagger}\right).
\end{aligned}
$$

Then (S.12) implies that

$$
\begin{aligned}
& A^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}\left(\gamma+\gamma^{\dagger}\right)-A^{-1}(\gamma)A'^{-1}(\gamma)-(D\Sigma(\gamma))\left(\gamma^{\dagger}\right)\\
=\ & A^{-1}\left(\gamma+\gamma^{\dagger}\right)A'^{-1}\left(\gamma+\gamma^{\dagger}\right)-A^{-1}(\gamma)A'^{-1}(\gamma)-\Delta\left(\gamma,\gamma^{\dagger}\right)-(D\Sigma(\gamma))\left(\gamma^{\dagger}\right)+\Delta\left(\gamma,\gamma^{\dagger}\right)\\
=\ & \Delta\left(\gamma,\gamma^{\dagger}\right)-(D\Sigma(\gamma))\left(\gamma^{\dagger}\right),
\end{aligned} \tag{S.14}
$$

so to prove (S.9) it is sufficient to show that

$$
\frac{\Delta\left(\gamma,\gamma^{\dagger}\right)-(D\Sigma(\gamma))\left(\gamma^{\dagger}\right)}{\left\|\gamma^{\dagger}\right\|_g}\to 0 \text{ as } \left\|\gamma^{\dagger}\right\|_g\to 0. \tag{S.15}
$$

The numerator in (S.15) can be written as $\sum_{i=1}^{7}\Pi_i\left(\gamma,\gamma^{\dagger}\right)A'^{-1}(\gamma)$ by adding, subtracting and grouping terms, where (omitting the argument $\left(\gamma,\gamma^{\dagger}\right)$)

$$
\Pi_1 = A^{-1}\left(\gamma+\gamma^{\dagger}\right)\sum_{j=1}^{m_1}\gamma_j^{\dagger}G_j'\left(\gamma+\gamma^{\dagger}\right)H'^{-1}(\gamma)\left(H(\gamma)-H\left(\gamma+\gamma^{\dagger}\right)\right)'H'^{-1}\left(\gamma+\gamma^{\dagger}\right),
$$

$$
\begin{aligned}
\Pi_2 &= A^{-1}\left(\gamma+\gamma^{\dagger}\right)\sum_{j=1}^{m_1}\gamma_j^{\dagger}H^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(H(\gamma)-H\left(\gamma+\gamma^{\dagger}\right)\right)H^{-1}(\gamma)G_j\left(\gamma\right),\\
\Pi_3 &= A^{-1}\left(\gamma+\gamma^{\dagger}\right)\sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}\left(A^{-1}\left(\gamma+\gamma^{\dagger}\right)-A^{-1}\left(\gamma\right)\right)T_j'\left(\gamma+\gamma^{\dagger}\right),\\
\Pi_4 &= \left(A^{-1}\left(\gamma+\gamma^{\dagger}\right)-A^{-1}\left(\gamma\right)\right)\sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}\overline{T_j\left(\gamma+\gamma^{\dagger}\right)},\\
\Pi_5 &= A^{-1}(\gamma)\sum_{j=m_1+1}^{m_1+m_2}\gamma_j^{\dagger}\overline{H^{-1}\left(\gamma+\gamma^{\dagger}\right)\left(H(\gamma)-H\left(\gamma+\gamma^{\dagger}\right)\right)H^{-1}(\gamma)W_j},\\
\Pi_6 &= \Delta\left(\gamma,\gamma^{\dagger}\right)\sum_{j=1}^{m_1}\gamma_j^{\dagger}W_j'H'^{-1}(\gamma),\\
\Pi_7 &= \left(A^{-1}\left(\gamma+\gamma^{\dagger}\right)-A^{-1}\left(\gamma\right)\right)\sum_{j=1}^{m_1}\gamma_j^{\dagger}H^{-1}(\gamma)G_j(\gamma).
\end{aligned}
$$

By (S.8), (S.13) and replication of earlier techniques, we have

$$
\max_{i=1,\ldots,7}\sup_{\gamma\in\Gamma^o}\left\|\Pi_i\left(\gamma,\gamma^{\dagger}\right)A^{-1}(\gamma)\right\|\le C\left\|\gamma^{\dagger}\right\|_g^2, \tag{S.16}
$$

where the norm used on the RHS of (S.16) depends on whether we are considering the general case or the 'single nonzero diagonal block' case. Thus

$$
\frac{\left\|\Delta\left(\gamma,\gamma^{\dagger}\right)-\left(D\Sigma(\gamma)\right)\left(\gamma^{\dagger}\right)\right\|}{\left\|\gamma^{\dagger}\right\|_g}\le C\left\|\gamma^{\dagger}\right\|_g\to 0 \text{ as } \left\|\gamma^{\dagger}\right\|_g\to 0,
$$

proving (S.15) and thus (S.9). □

**Corollary CS.1.** *For the spatial error model with SAR(p) errors,*

$$
\left(D\Sigma(\gamma)\right)\left(\gamma^{\dagger}\right)=K^{-1}(\gamma)\sum_{j=1}^{m_1}\gamma_j^{\dagger}\overline{G_j(\gamma)}K'^{-1}(\gamma).
$$

*Proof.* Taking $m_2=0$ in Lemma LS.5, the elements involving sums from $m_1+1$ to $m_1+m_2$ do not arise and $H(\gamma)=I_n$, proving the claim. □

**Corollary CS.2.** *For the spatial error model with SMA($m_2$) errors,*

$$
\left(D\Sigma(\gamma)\right)\left(\gamma^{\dagger}\right)=H(\gamma)\sum_{j=1}^{m_2}\gamma_j^{\dagger}\overline{T_j(\gamma)}H'(\gamma).
$$

*Proof.* Taking $m_1=0$ in Lemma LS.5, the elements involving sums from 1 to $m_1$ do not arise and

$K(\gamma) = I_n$, proving the claim. □

**Lemma LS.6.** *For the spatial error model with MESS(p) errors, if*

$$\max_{j=1,\ldots,m_1} \left(\|W_j\| + \left\|W_j'\right\|\right) < 1, \tag{S.17}$$

*then*

$$(D\Sigma(\gamma))\left(\gamma^\dagger\right) = \exp\left(\sum_{j=1}^{m_1} \gamma_j \left(W_j + W_j'\right)\right) \sum_{j=1}^{m_1} \gamma_j^\dagger \left(W_j + W_j'\right).$$

*Proof.* Clearly $D\Sigma \in \mathscr{L}\left(\Gamma^o, \mathcal{M}^{n\times n}\right)$. Next,

$$\begin{aligned}
& \left\|A^{-1}\left(\gamma+\gamma^\dagger\right)A'^{-1}\left(\gamma+\gamma^\dagger\right) - A^{-1}(\gamma)A'^{-1}(\gamma) - (D\Sigma(\gamma))\left(\gamma^\dagger\right)\right\| \\
= & \left\|\exp\left(\sum_{j=1}^{m_1}\left(\gamma_j+\gamma_j^\dagger\right)\left(W_j+W_j'\right)\right) - \exp\left(\sum_{j=1}^{m_1}\gamma_j\left(W_j+W_j'\right)\right) - (D\Sigma(\gamma))\left(\gamma^\dagger\right)\right\| \\
= & \left\|\exp\left(\sum_{j=1}^{m_1}\gamma_j\left(W_j+W_j'\right)\right)\left(\exp\left(\sum_{j=1}^{m_1}\gamma_j^\dagger\left(W_j+W_j'\right)\right) - I_n - \sum_{j=1}^{m_1}\gamma_j^\dagger\left(W_j+W_j'\right)\right)\right\| \\
\leq & \left\|\exp\left(\sum_{j=1}^{m_1}\gamma_j\left(W_j+W_j'\right)\right)\right\|\left\|\exp\left(\sum_{j=1}^{m_1}\gamma_j^\dagger\left(W_j+W_j'\right)\right) - I_n - \sum_{j=1}^{m_1}\gamma_j^\dagger\left(W_j+W_j'\right)\right\| \\
\leq & \; C\left\|I_n + \sum_{j=1}^{p}\gamma_j^\dagger\left(W_j+W_j'\right) + \sum_{k=2}^{\infty}\left\{\sum_{j=1}^{m_1}\gamma_j^\dagger\left(W_j+W_j'\right)\right\}^k - I_n - \sum_{j=1}^{m_1}\gamma_j^\dagger\left(W_j+W_j'\right)\right\| \\
\leq & \; C\left\|\sum_{k=2}^{\infty}\left\{\sum_{j=1}^{m_1}\gamma_j^\dagger\left(W_j+W_j'\right)\right\}^k\right\| \leq C\sum_{k=2}^{\infty}\sum_{j=1}^{m_1}\left|\gamma_j^\dagger\right|\left\|\left(W_j+W_j'\right)\right\|^k \\
\leq & \; C\sum_{k=2}^{\infty}\left\|\gamma^\dagger\right\|_g^k,
\end{aligned} \tag{S.18}$$

by (S.17), without loss of generality, and again the norm used in (S.18) depending on whether we are in the general or the 'single nonzero diagonal block' case. Thus

$$\frac{\left\|A^{-1}\left(\gamma+\gamma^\dagger\right)A'^{-1}\left(\gamma+\gamma^\dagger\right) - A^{-1}(\gamma)A'^{-1}(\gamma) - (D\Sigma(\gamma))\left(\gamma^\dagger\right)\right\|}{\left\|\gamma^\dagger\right\|_g} \leq C\sum_{k=2}^{\infty}\left\|\gamma^\dagger\right\|_g^{k-1} \to 0,$$

as $\left\|\gamma^\dagger\right\|_g \to 0$, proving the claim. □

**Theorem TS.1.** *Under the conditions of Theorem 4.4 or 5.3, $\mathscr{T}_n - \mathscr{T}_n^a = o_p(1)$ as $n \to \infty$.*

*Proof.* It suffices to show that $n\widetilde{m}_n = n\widehat{m}_n + o_p(\sqrt{p})$. As $\widehat{\eta} = y - \widehat{\theta}$, $\widehat{u} = y - \widehat{f}$, and $\widehat{v} = \widehat{\theta} - \widehat{f}$, we have $\widehat{u} = \widehat{\eta} + \widehat{v}$ and

$$\begin{aligned}
n\widetilde{m}_n &= \widehat{\sigma}^{-2}\left(\widehat{u}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{u} - \widehat{\eta}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{\eta}\right) = \widehat{\sigma}^{-2}\left(2\widehat{u}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{v} - \widehat{v}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{v}\right) \\
&= 2n\widehat{m}_n - \widehat{\sigma}^{-2}\left[\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}(u+e) - e + \theta_0 - \widehat{f}\right]' \\
&\quad \Sigma\left(\widehat{\gamma}\right)^{-1}\left[\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}(u+e) - e + \theta_0 - \widehat{f}\right] \\
&= 2n\widehat{m}_n - \widehat{\sigma}^{-2}u'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}u - \widehat{\sigma}^{-2}\left(\theta_0 - \widehat{f}\right)'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(\theta_0 - \widehat{f}\right) \\
&\quad + \widehat{\sigma}^{-2}\left(2(\theta_0 - \widehat{f}) - e\right)'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(I - \Psi[\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi]^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\right)e \\
&\quad - 2\widehat{\sigma}^{-2}\left(\theta_0 - \widehat{f}\right)'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}u \\
&= 2n\widehat{m}_n - \left(n\widehat{m}_n - \widehat{\sigma}^{-2}(A_1 + A_2 + A_3 + A_4)\right) - \widehat{\sigma}^{-2}A_4 \\
&\quad + \widehat{\sigma}^{-2}\left(2(\theta_0 - \widehat{f}) - e\right)'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(I - \Psi[\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi]^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\right)e - 2\widehat{\sigma}^{-2}A_3 \\
&= n\widehat{m}_n + \widehat{\sigma}^{-2}(A_1 + A_2 - A_3) \\
&\quad + \widehat{\sigma}^{-2}\left(2(\theta_0 - \widehat{f}) - e\right)'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(I - \Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\right)e. \qquad \text{(S.19)}
\end{aligned}$$

In the proof of Theorem 4.2, we have shown that

$$\left|\left(\theta_0 - \widehat{f}\right)'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(I - \Psi[\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi]^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\right)e\right| = o_p(\sqrt{p})$$

in the process of proving $|A_2| = o_p(\sqrt{p})$. Along with

$$\begin{aligned}
&\left|e'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(I - \Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\right)e\right| \\
\leq\ & \left|e'\Sigma\left(\widehat{\gamma}\right)^{-1}e\right| + \left|e'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}e\right| \\
\leq\ & \|e\|^2\sup_{\gamma\in\Gamma}\left\|\Sigma\left(\gamma\right)^{-1}\right\| + \|e\|^2\sup_{\gamma\in\Gamma}\left\|\Sigma\left(\gamma\right)^{-1}\right\|^2\left\|\frac{1}{n}\Psi\left(\frac{1}{n}\Psi'\Sigma\left(\gamma\right)^{-1}\Psi\right)^{-1}\Psi'\right\| \\
=\ & O_p\left(\|e\|^2\right) = O_p\left(p^{-2\mu}n\right) = o_p(\sqrt{p}),
\end{aligned}$$

we complete the proof that $n\widetilde{m}_n = n\widehat{m}_n + o_p(\sqrt{p})$. In the SAR setting of Section 5,

$$\begin{aligned}
n\widetilde{m}_n &= \widehat{\sigma}^{-2}\left(\widehat{u}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{u} - \widehat{\eta}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{\eta}\right) = \widehat{\sigma}^{-2}\left(2\widehat{u}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{v} - \widehat{v}'\Sigma\left(\widehat{\gamma}\right)^{-1}\widehat{v}\right) \\
&= 2n\widehat{m}_n - \widehat{\sigma}^{-2}\left[\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(u + e + \sum_{j=1}^{d_\lambda}(\lambda_{j0} - \widehat{\lambda}_j)W_j y\right) - e + \theta_0 - \widehat{f}\right]'
\end{aligned}$$

$$\Sigma\left(\widehat{\gamma}\right)^{-1}\left[\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(u+e+\sum_{j=1}^{d_\lambda}(\lambda_{j_0}-\widehat{\lambda}_j)W_jy\right)-e+\theta_0-\widehat{f}\right].$$

Compared to the expression in (S.19), we have the additional terms

$$-\widehat{\sigma}^{-2}\sum_{j=1}^{d_\lambda}(\lambda_{j_0}-\widehat{\lambda}_j)W_jy'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\sum_{j=1}^{d_\lambda}(\lambda_{j_0}-\widehat{\lambda}_j)W_jy$$

and

$$-2\widehat{\sigma}^{-2}\sum_{j=1}^{d_\lambda}(\lambda_{j_0}-\widehat{\lambda}_j)W_jy'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\left(\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\Psi\right)^{-1}\Psi'\Sigma\left(\widehat{\gamma}\right)^{-1}\left(u+\theta_0-\widehat{f}\right).$$

Both terms are $o_p(\sqrt{p})$ from the orders of $A_5$ and $A_6$ in the proof of Theorem 5.2. Hence, in the SAR setting, $n\widetilde{m}_n = n\widehat{m}_n + o_p(\sqrt{p})$ also holds.

□

# References


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

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

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

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

Robinson, P. M. (1988). The stochastic difference between econometric statistics. *Econometrica 56*, 531–548.