TESTING FOR CORRECT SPECIFICATION OF SPATIAL AUTOREGRESSIVE MODELS.

JUNGYOON LEE*, PETER C. B. PHILLIPS† AND FRANCESCA ROSSI‡

Abstract. Spatial autoregressive (SAR) and related models offer flexible yet parsimonious ways to model spatial or network interaction. However, SAR specifications rely on a particular parametric functional form and an exogenous choice of the so-called spatial weight matrix. Moreover, the choice of SAR model over other alternatives, such as Spatial Durbin (SD) or Spatial Lagged X (SLX) models, is often arbitrary and unwarranted. This paper develops a specification test in SAR model that can detect general forms of misspecification including that of the spatial weight matrix, functional form and the model itself. We extend the framework of conditional moment testing of Bierens (1982, 1990) to the general spatial setting. We derive the asymptotic distribution of our test statistic under the null hypothesis of correct SAR specification and show consistency of the test. We also carry out a small Monte Carlo study of finite sample performance.

....

* Royal Holloway, University of London, Egham, Surrey, UK. Email: jungyoon.lee@rhul.ac.uk

† Yale University, 06520 New Haven, United States.

‡ University of Verona, Italy.

NOT FOR CIRCULATION

1. Introduction

The past two decades have seen a remarkable surge in theoretical and empirical literature on a class of spatial econometric models first suggested by Clifford and Ord (1968), see reviews by Arbia (2006), Bivand (2008) and Anselin (2010) for some comprehensive lists. These models are characterized by the use of exogenously given weight matrices, which capture the structure of spatial dependence between units up to a finite number of unknown parameters. Much of theoretical literature has focused on the SAR model, see e.g. Kelejian and Prucha (1998, 1999) and Lee (2002, 2004). While inference on model parameters has attracted much attention, the issue of model specification testing has received relatively little interest. There has been one recent paper by Su and Xi (2016) who extends the nonparametric testing procedure of Fan and Li (1996) to the SAR model in order to test for correct linear functional form specification of SAR model while assuming that the true weight matrix is chosen. The nonparametric nature of their test leads to poor finite sample performance of inference based on the asymptotic distribution, and they suggest a bootstrap scheme.

Our aim in this paper is to consider a testing procedure which can detect general forms of misspecification including that of the model, weight matrix and functional form for the SAR model, using Bieren (1982, 1990)'s testing procedure. We first extend Bieren (1990)' fundamental result Lemma 1 to a general spatial setting, providing justification for using Bieren's methods in spatial models. We then derive the asymptotic null distribution of the test statistic that is based on the Gaussian quasi maximum likelihood estimator of model parameters, prove consistency of the test and present findings of a small Monte Carlo studies.

Throughout the paper, we denote by A_{in} and $A_n^{(i)}$ the vectors formed by the tranpose of the $i-$th row of a generic matrix A_n and A_n^{-1}, respectively, provided that the inverse exists. Moreover, $||.||$ and $||.||_\infty$ indicate the spectral norm and uniform absolute row sum norm, respectively, and $K > 0$ represents an arbitrary finite constant. Let A' be transpose of a matrix A.

2. Set-up

We consider a regression model of the following form

$$Y_{in} = g_{in}(X_n) + \eta_{in},\ E(\eta_{in}|X_n) = 0, \quad i = 1, ..., n, \tag{1}$$

where $X_n = (X_{1n}, \cdots, X_{nn})'$ is $n \times k$ matrix of regressors of *all* sampled units with the true conditional expectation function for ith observation denoted by g_{in}:

$$g_{in}(X_n) = E(Y_{in}|X_n) \quad i = 1, ..., n.$$

By conditioning on the matrix X_n, instead of the individual vector $X_i = X_{in}$, we characterise the above model as a spatial one whereby an individual's outcome is influenced by not only his own characteristics but also by his neighbours'. Unlike Bierens (1990), we allow for possible heterogeneity in the regression function g_{in} across individuals.

On the other hand, the so-called mixed regressive SAR model is given in $n \times 1$ vector form as:

$$Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n, \tag{2}$$

where W_n is a sequence of pre-specified $n \times n$ weight matrices that reflect some notion of distance between units and λ is a scalar parameter that reflects the strength of spatial interaction. Define $S_n(\lambda) = I_n - \lambda W_n$ and $G_n(\lambda) = W_n S_n^{-1}(\lambda)$. The SAR model can be written in its reduced form as follows:

$$Y_n = S_n^{-1}(\lambda)(X_n \beta + \epsilon_n). \tag{3}$$

For $i = 1, ..., n$, the last displayed expression leads to a linear regression relationship of form

$$\begin{aligned} Y_{in} &= m_{in}(X_n, \lambda, \beta) + u_{in} \quad \text{where} \\ m_{in}(X_n, \lambda, \beta) &= S_n^{(i)}(\lambda)' X_n \beta = \sum_{j=1}^{n} s_n^{ij}(\lambda) X'_{jn} \beta, \\ u_{in} &= \sum_{j=1}^{n} s_n^{ij}(\lambda) \epsilon_{jn} \end{aligned} \tag{4}$$

where $s_n^{ij}(\lambda)$ denotes the (i,j)-th element of $S_n^{-1}(\lambda)$ and u_{in} is the reduced form error of the SAR model.

The functions g_{in} and m_{in} and quantities in (1) and (4) are allowed to be triangular arrays, but we will suppress the subscript n for notational simplicity from now on.

In this paper, we are interested in testing whether the regression function $m_i(\cdot)$ of (4) is a correct characterization of the unknown true regression function $g_i(\cdot)$ of (1), i.e. if $g_i = m_i$.
We introduce the following assumptions.

Assumption 1 *For all n, ϵ_i are independent identically distributed (iid) random variables with zero mean and unknown variance σ^2 and, for some $\delta > 0$, $\mathbb{E}|\epsilon_i|^{4+\delta} \leq K$.*

Given the linear representation of m_i in (4), we impose similar structure on the true conditional expectation function g_i.

Assumption 2 For $i = 1, ..., n$, $g_i(\cdot)$ in (1) is given by

$$g_i(X) = \sum_{j=1}^{n} (a_{1ij} \rho_1(X_j) + a_{2ij} \rho_2(X_j)),$$

for some (unknown) $n \times n$ matrices $A_1 = A_{1n} = (a_{1ij})$, $A_2 = A_{2n} = (a_{2ij})$ satisfying $||A_1||_\infty + ||A'_1||_\infty + ||A_2||_\infty + ||A'_2||_\infty \leq K$ for all sufficiently large n, and unknown functions $\rho_1(\cdot), \rho_2(\cdot) : \mathbb{R}^k \Rightarrow \mathbb{R}$ for which $E\rho_1^4(X_1) + E\rho_2^4(X_1) < \infty$ holds.

The form of g_i given in Assumption 2 above covers all special cases of interest that we consider, to be discussed in Section 4. If one wants to relax the form of $g_i(\cdot)$ to be more general, we need to consider other ways of characterizing and controlling dependence across $g_i(X)$, different from Assumption 2. We also impose the following conditions on the error term η_i of the model.

Assumption 3 *For $i = 1, ..., n$, η_i is independent of X. For $i = 1, ..., n$ in (1), $E(\eta_i|X) = 0$, $E(\eta_i^2) < \infty$ and*

$$\max_{1\le i\le n} \sum_{k=1}^{n} |Cov(\eta_i, \eta_k)| = O(1).$$

Denote $\theta = (\lambda, \beta')'$. Let

$$J = J_n := \{i : P\big(g_i(X) = m_i(X, \theta)\big) < 1 \forall\theta \in \Theta\} \tag{5}$$

and $|\cdot|$ denotes the cardinality of a set. We are interested in testing the hypothesis

$$H_0 : P\big(g_i(X) = m_i(X, \theta_0)\big) = 1 \quad \text{for some} \quad \theta_0 \in \Theta, \quad \forall 1 \le i \le n \tag{6}$$

against the fixed alternative hypothesis

$$H_1 : P(g_i(X) = m_i(X, \theta)) < 1 \quad \text{for all } \theta \in \Theta \text{ for sufficiently many } i = 1, ..., n \quad \text{such that } |J_n|/n = O_e(1). \tag{7}$$

We construct our test statistic on the basis of the following result.

Theorem 1 *Let v be a random variable satisfying $E|v| < \infty$ and let $Z = (Z_1, ..., Z_n)$ be a collection of n bounded i.i.d. random variables in $\mathbb{R}^k$. Let*

$$E(v|Z_1, ..., Z_n) = r(Z_1, ..., Z_n) = \sum_{j=1}^{n} r_j(Z_j). \tag{8}$$

If $P(r(Z_1, ..., Z_n) = 0) < 1$ then for each $j = 1, ..., n$ the sets of t's such that v and $e^{t'Z_j}$ are orthogonal,

$$S = \{t \in \mathbb{R}^k : E(ve^{t'Z_j}) = 0\},$$

have Lebesgue measure zero.

The proof of Theorem 1 is given in the Appendix. Theorem 1 is a general result that offers justification for extending the specification testing procedure of Bierens (1982, 1990) in spatial settings where we have $g_i(X_1, \cdots, X_n)$ rather than $g_i(X_i)$, i.e. an individual is affected by his neighbours' covariates as well as his own. Theorem 1 can be used to design specification testing procedure in spatial models other than SAR.

Thus, from Theorem 1 we construct a statistic to test H_0 in (6) under Assumptions 2, 3 and model (4) based on n true moment conditions

$$E\left((Y_i - m_i(X, \lambda_0, \beta_0))e^{t'X_i}\right) = 0 \quad i = 1, ..., n,$$

consolidated into one summation, as

$$\frac{1}{n}\sum_{i=1}^{n} E(Y_i - m_i(X, \lambda_0, \beta_0))e^{t'X_i} = 0, \tag{9}$$

or

$$\frac{1}{n}\sum_{i=1}^{n} E(Y_i - m_i(X, \lambda_0, \beta_0))e^{t'\phi(X_i)} = 0, \tag{10}$$

where $\phi(\cdot) : \mathbb{R}^k \Rightarrow \mathbb{R}^k$ is a one-to-one bounded function. Let $\hat{\theta}$ be a consistent estimator of θ. We replace unknown moments by their sample analogues and define

$$M(\hat{\theta}, t) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} \big(Y_i - m_i(X, \hat{\theta})\big)e^{t'\phi(X_i)}.$$

The probability limit of $M(\hat{\theta}, t)$ is the limit of LHS in (10) under H_0, Assumption 3 and some suitable conditions on the SAR regression function.

Thus, we define our test statistic as

$$\hat{T}(t) = \frac{nM^2(\hat{\theta}, t)}{\hat{a}(t)}, \tag{11}$$

where $\hat{\theta}$ is a consistent estimator of θ and $\hat{a}(t)$ is an estimate of the variance of $\sqrt{n}M(\hat{\theta}, t)$, $a(t) = a_n(t) = Var(\sqrt{n}M(\hat{\theta}, t))$.

3. Limit theory under H_0

We will use the Gaussian quasi maximum likelihood (QML) estimator $\hat{\theta}$ of θ. In order to ensure consistency of $\hat{\theta}$ to θ_0 under H_0 we need to impose some extra conditions (see e.g. Lee (2004)).

Assumption 4 $\lambda_0 \in \Lambda$, *where* Λ *is a closed subset in* $(-1, 1)$.

Assumption 5

(i) *For all* n, $W_{ii} = 0$.
(ii) *For all* n, $||W|| \leq 1$.
(iii) *For all sufficiently large* n, $||W||_\infty + ||W'||_\infty \leq K$.
(iv) *For all sufficiently large* n, *uniformly in* $i, j = 1, ..., n$, $W_{ij} = O(1/h)$, *where* $h = h_n$ *is bounded away from zero for all* n *and* $h/n \to 0$ *as* $n \to \infty$.

Assumption 6 *For all sufficiently large* n, $\sup_{\lambda \in \Lambda} ||S^{-1}(\lambda)||_\infty + ||S^{-1}(\lambda)'||_\infty \leq K$.

Assumption 7 X_i *for* $i = 1, ..., n$ *is a set of iid bounded random variables in* $\mathbb{R}^k$. *There exist a* $(k+1) \times (k+1)$ *nonrandom positive definite matrix* Σ_G *such that*

$$\frac{1}{n}\big(G(\lambda_0)X\beta_0 \quad X\big)'\big(G(\lambda_0)X\beta_0 \quad X\big) \to_p \Sigma_G \tag{12}$$

Conditions given in Assumption 4-7 are standard in SAR literature.

Let $\bar{g}_{ij} := (G(\lambda_0) + G(\lambda_0)')_{ij}/2$ and $\mu_s := E(\epsilon^s)$. We define an $n \times 1$ vector $f_1 = G(\lambda_0)X\beta_0$ and a scalar $f_2 = \sigma^2 tr(G(\lambda_0) + G(\lambda_0)' - 2I_n tr(G(\lambda_0))/n)^2/2$.

For notational simplicity, we abbreviate $G = G(\lambda_0)$ and $S = S(\lambda_0)$. We let Ω be a $(k+1)\times(k+1)$ matrix whose $1-1$th (1×1) block is

$$\omega_{11} = \frac{1}{n\sigma^2}(f_2 + f_1' f_1), \tag{13}$$

$2-2$th block $(k\times k)$ is

$$\omega_{22} = \frac{X'X}{n\sigma^2} \tag{14}$$

and $1-2$th (or the transposed of $2-1$th) block is

$$\omega_{12} = \omega_{21}' = \frac{1}{n\sigma^2} f_1' X. \tag{15}$$

Let ω^{ij} be the $i-j$th block of Ω^{-1} for $i,j = 1,2$, ω^1 the first row of Ω^{-1} and $\omega^2 = (\omega^{21};\ \ \omega^{22})$. Moreover, denote $\underline{e}(t) = (e^{t'\phi(X_1)}, ..., e^{t'\phi(X_n)})'$.

Define the $n-$dimensional row vector $\psi_1(t)$ and the scalar $\psi_2(t)$ as

$$\psi_1(t) = \psi_1(t,\lambda_0,\beta_0,X) = \underline{e}(t)' S^{-1}\left(I_n - \frac{1}{n\sigma^2}\begin{pmatrix} f_1 & X\end{pmatrix}\Omega^{-1}\begin{pmatrix} f_1' \\ X'\end{pmatrix}\right),$$

and

$$\psi_2(t) = \psi_2(t,\lambda_0,\beta_0,X) = -\frac{1}{n\sigma^2}\underline{e}(t)' S^{-1}\begin{pmatrix} GX\beta & X\end{pmatrix}\omega^{1'}. \tag{16}$$

In order to ensure pointwise existence and non-singularity of $a(t)$, as well as existence of $M(\hat\theta, t)$, we impose the following assumption.

Assumption 8 *Conditionally on X, the limits*

$$\lim_{n\to\infty}\psi_1(t), \quad \lim_{n\to\infty}\psi_2(t) \tag{17}$$

exist poitwise in t, and for all sufficiently large n.

We derive the following result.

Theorem 2 *Let Assumptions 1-8 hold. Under H_0 in (6), pointwise in t,*

$$\sqrt{\frac{n}{a(t)}}M(\hat\theta, t) \to_d \mathcal{N}(0,1), \tag{18}$$

conditionally on X where the variance of $\sqrt{n}M(\hat\theta,t)$ is given by

(19)

$$a(t) = a_n(t) = \frac{\sigma^2}{n}\sum_{i=1}^n \psi_{1i}(t)^2 + \frac{2\sigma^4\psi_2(t)^2}{n}\sum_{i,j=1}^n \bar g_{ij}^2 + \frac{\psi_2(t)^2}{n}(\mu^{(4)} - 3\sigma^4)\sum_{i=1}^n \bar g_{ii}^2 + \frac{2\mu^{(3)}}{n}\psi_2(t)\sum_{i=1}^n \psi_{1i}(t)\bar g_{ii}.$$

Since (18) holds for every realisation of X, as long as $p\lim a_n(t)$ exists pointwise in t, Theorem 2 holds also unconditionally. The proof of Theorem 2 is given in the Appendix. Hence, we can construct the test statistic defined in (11), where $\hat a_n(t)$ is obtained by replacing the unknown parameters λ, β, σ^2, μ_3 and μ_4 with their sample versions based on QML estimates $\hat\theta = (\hat\lambda, \hat\beta')'$ and corresponding residuals. From Theorem 2 it is straightforward to conclude that

$$\hat T(t) \to_d \chi_1^2 \tag{20}$$

pointwise in t.

The test in (11) is designed to detect misspecification in the conditional mean function implied by model (3), implicitely assuming that the conditional variance is correctly specified. The analysis of the consequences of conditional variance misspecification and the possiblity of including a second moment-related conditions in our test statistic is currently under investigation (e.g. Wooldridge (1991)).

4. Behaviour of test statistic under misspecification, H_1

Before we discuss properties of testing procedure under misspecification, we first outline some cases of interest that are all covered by the characterization of the true regression function given in Assumption 2.

4.1. Special cases of interest 1-5.

Below are some examples of true regression function where one misspecifies it to take the form (3).

1) Suppose the true weight matrix is given by V, while the practitioner uses W in estimation of the model, i.e. W is misspecified. The true model is

$$Y = \lambda_0 VY + X\beta_0 + \varepsilon.$$

So, in this case, $A_1 = (I - \lambda_0 V)^{-1}$ and $\rho_1(X_i) = X_i'\beta_0$ for A_1 and ρ_1 defined in Assumption 2.

2) One has correctly specified W but not the functional form of $\rho(\cdot)$. The true data generating process is given by

$$Y_i = (\lambda_0 W)_i Y + \rho(X_i) + \varepsilon_i$$

for some function ρ. In this case, $A_1 = (I - \lambda_0 W)^{-1}$ and $\rho_1(\cdot) = \rho(\cdot)$.

3) There is no spatial interaction, i.e. $\lambda = 0$ so that

$$Y_i = X_i'\beta + \varepsilon_i, \quad \text{or} \quad Y_i = \rho(X_i) + \varepsilon_i$$

where we could let ε_i to be dependent across i. In this case, $A_1 = I_n$.

4) The true model is Spatial Durbin (SD) model with weight matrices W_1, W_2:

$$Y = \lambda_0 W_1 Y + X\beta_{10} + W_2 X\beta_{20} + \varepsilon. \tag{21}$$

In reduced form, the model can be written as

$$Y = (I - \lambda_0 W_1)^{-1} X\beta_{10} + (I - \lambda_0 W_1)^{-1} W_2 X\beta_{20} + (I - \lambda_0 W_1)^{-1}\varepsilon,$$

and hence $A_1 = (I - \lambda_0 W_1)^{-1}$, $A_2 = (I - \lambda_0 W_1)^{-1} W_2$, $\rho_1(X_j) = X_j\beta_{10}$ and $\rho_2(X_j) = X_j\beta_{20}$.

5) The true model is spatial lagged X (SLX) model:

$$Y = X\beta_{10} + WX\beta_{20} + \eta \tag{22}$$

where $A_1 = I_n$, $A_2 = W$, $\rho_1(X_j) = X_j\beta_{10}$ and $\rho_2(X_j) = X_j\beta_{20}$.

4.2. **Consistency of the test under fixed alternative.** In order to establish consistency of our test we need to impose some further assumptions.

Assumption 9 *There exists a sequence of deterministic vectors $\theta^\sharp = \theta^\sharp_n$ of order $O(1)$ such that $\hat{\theta} - \theta^\sharp = o_p(1)$ under H_1.*

In line with the previous section, $\theta^\sharp$ should be intended as the value that maximises the (misspecified) pseudo log-likelihood function under H_1 in (7) and thus $\hat{\theta}$ is the QMLE of $\theta^\sharp$ under H_1. Proposition 1 in the Appendix shows that Assumption 9 is satisfied under some additional technical assumption. Clearly, under H_0, $\theta^\sharp = \theta_0$.

Assumption 10 *For all sufficiently large n,*

$$\frac{1}{n}\sum_{i=1}^{n} E(\{g_i(X) - m_i(X,\theta)\}e^{t'\phi(X_i)}) \neq 0, \quad \forall \ \theta \in \Theta \tag{23}$$

Some remarks on the high-level Assumption 10, which characterizes the class of misspecifications that can be detected by our test, are in order here. In the spatial setting it is natural to allow $m_i(\cdot)$ and $g_i(\cdot)$ to vary across i. This in turn gives rise to the possibility that $\text{plim}_{n\to\infty} M_n(\hat{\theta}, t) = 0$ when H_0 is false: even when $E(\{g_i(X) - m_i(X,\theta)\}e^{t'X_i}) \neq 0, \forall\theta \in \Theta$ for every i, one may get $\sum_{i=1}^{n} E(\{g_i(X) - m_i(X_n,\theta)\}e^{t'X_i}) = 0$ if there are cancellations occurring within the summation over i. Assumption 10 precludes such eventuality. A similar problem was discussed by Bierens (1984) where nonstationarity in the time series setting may lead to regression function $g_t(\cdot)$ that varies across time periods t.

Theorem 3. *Under H_1 and Assumptions 1-10, for all $c > 0$*

$$Pr\Big(\hat{T}(t) > c\Big) \to 1 \quad \text{as} \quad n \to \infty.$$

5. Monte Carlo study

We carry out a small Monte Carlo experiment to examine finite sample performance of tests for model misspecification based on $\hat{T}(t)$ statistic, examining their Monte Carlo size and power. We generate our data based on the SAR specification of (3), with $X_{1i} = 1$ and *iid* random variables $X_{2i} \sim N(0,1)$, $\epsilon_i \sim N(0,1)$, for $\beta = (1,1)'$, $\lambda = 0.4$, $n = 100, 300, 600$. We use two different weight matrices: Case (1991)'s block diagonal matrix and row normalized exponential distance weight $w_{ij} = exp(-|\ell_i - \ell_j|)1(|\ell_i - \ell_j| < \log n)$ where ℓ_i is location of i along the interval $[0, n]$ which was generated from $Unif[0, n]$. We use $\phi(\cdot) = atan(\cdot)$, and the index $t = 3$ as used by Bierens (1990) in his simulations.

First we examine the size of the test for nominal significance levels $s = 10, 5, 1\%$. For block diagonal matrix there were $10, 15, 20$ blocks with $10, 20, 30$

units in each block, respectively for $n = 100, 300, 600$. Table 1 reports satisfactory size performance of our test.

TABLE 1. Monte Carlo size of test under H_0

	distance weight			block diagonal weight		
$n \backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.101	0.043	0.008	0.074	0.026	0.004
300	0.106	0.055	0.016	0.092	0.052	0.014
600	0.099	0.047	0.011	0.088	0.04	0.007

Next we turn to power of the test. We first try functional form misspecification whereby $r(x) = 1 + x + 0.5x^2$ was used in generating the data and the misspecified linear SAR model was estimated. The power of our test based on $\hat{T}(t)$ is reported to be close to 1 even for $n = 100$, see Table 2.

TABLE 2. Monte Carlo power of test under functional form misspecification

	distance weight			block diagonal weight		
$n \backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.996	0.991	0.975	0.997	0.995	0.988
300	1	1	1	1	1	1
600	1	1	1	1	1	1

We then look at power of test under misspecification of weight matrix. For the block diagonal weight matrix, the true weight matrix has $2, 3, 4$ blocks with $50, 100, 150$ units in each block, whereas the misspecified weight matrix contains $10, 15, 20$ blocks with $10, 20, 30$ units in each block. For the exponential distance weight, we generated two independent sets of locations of individuals which were then used to compute the correct and misspecified weight matrices. The Monte Carlo power, as reported in Tables 3-4, is rather poor, especially for the block diagonal case. We also tried $\lambda = 0.8$ and while the Monte Carlo power improved with the larger λ for the exponential weight matrix case, it remained much the same in the block diagonal case.

TABLE 3. Monte Carlo power for misspecified block diagonal weight

λ	0.4			0.8		
$n \backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.102	0.048	0.009	0.089	0.045	0.007
300	0.109	0.057	0.013	0.108	0.054	0.009
600	0.089	0.045	0.012	0.119	0.056	0.011

Above indicates that while the size of test based on $\hat{T}(t)$ seems to be good, its power may be low under certain misspecifications. Bierens (1990, Theorems 4,5) considered a simple procedure that could improve power of the test as follows. Calculate test statistic at fixed t_0, $\hat{T}(t_0)$, and also obtain test statistics over a range of t, say τ, then set $\hat{t} = argsup_{t \in \tau} \hat{T}(t)$. The range τ may increase with the sample size and Bierens (1990) picked $n/10 - 1$

Table 4. Monte Carlo power for misspecified exponential distance weight

λ	0.4			0.8		
$n\backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.2	0.126	0.037	0.397	0.299	0.132
300	0.212	0.139	0.055	0.409	0.321	0.181
600	0.253	0.169	0.069	0.444	0.368	0.237

number of points in interval $[1, 5]$ using uniform distribution in his Monte Carlo studies. Use $\hat{T}(t_0)$ if $\hat{T}(\hat{t}) - \hat{T}(t_0) < \gamma n^{\rho}$ and $\hat{T}(\hat{t})$ otherwise. Trying this method with $\gamma = 0.5, \rho = 0.5$ as used in Bierens (1990), we obtain the results presented in Table 5 and 6. The improvement in the power is significant only in the block-diagonal case with $\lambda = 0.8$.

Table 5. Monte Carlo power for misspecified block diagonal weight with modified test of Bierens (1990, Theorem 4,5)

λ	0.4			0.8		
$n\backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.1	0.07	0	0.27	0.18	0.03
300	0.13	0.01	0	0.35	0.31	0.08
600	0.11	0.06	0.02	0.36	0.29	0.07

Table 6. Monte Carlo power for misspecified exponential distance weight with modified test of Bierens (1990, Theorem 4,5)

λ	0.4			0.8		
$n\backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.207	0.134	0.036	0.424	0.32	0.134
300	0.217	0.142	0.056	0.442	0.364	0.213
600	0.231	0.143	0.057	0.449	0.359	0.222

Finally we turn to Monte Carlo power against misspecification of the model itself by generating data based on the SD and SLX models. For the SD model case, the same exponential distance weight was used for W_1, W_2 of (21) and misspecified SAR estimation. For the SLX model case, the same exponential distance weight W was used in (22) and estimation of SAR model. We set $\lambda = 0.4$. The power is very low for both models at $n = 100$ but improves rapidly with the sample size, see Table 7.

The Monte Carlo findings presented above indicate that our current test lacks power against misspecification of the weight matrix. This is something we need to address in future works.

Table 7. Monte Carlo power for misspecified model

model	SLX			SD		
$n\backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.04	0.007	0	0.01	0.001	0
300	0.275	0.129	0.008	0.263	0.077	0.001
600	0.585	0.398	0.091	0.695	0.458	0.053

6. Appendix

Proof of Theorem 1. The argument of Bierens (1990), proof of Lemma 1, applies once we can show *for every* $j = 1, ..., n$

$$P(E(v|Z) = 0) < 1 \rightarrow E(ve^{t'Z_j}) \neq 0 \;\; \textit{for at least one } t \in \mathbb{R}^k. \tag{24}$$

Following Bierens (1982) we define the functions

$$r_1(.) = max\{r(.), 0\}, \quad r_2(.) = max\{-r(.), 0\}.$$

We define the expected values

$$c_s = E(r_s(Z_1, ..., Z_n)), \quad s = 1, 2,$$

and probability measures
(25)

$$\nu_{sj}(B_j) = \frac{1}{c_s}\int_{B_j}\int_{\mathbb{R}^k}...\int_{\mathbb{R}^k} r_s(z_1, ..., z_n)dF(z_1)...dF(z_n), \quad s = 1, 2, \quad j = 1, ..., n,$$

where B_j is a Borel set in $\mathbb{R}^k$ and is the range of integration of the variable Z_j, and $F(z)$ is the cumulative distribution function of the *iid* random vectors Z_i for $i \neq j$, $i = 1, ..., n$. The latter is the marginal distribution of component j given the joint probability measure

$$\nu_s(B_1, ..., B_n) = \frac{1}{c_s}\int_{B_1}\int_{B_2}...\int_{B_n} r_s(z_1, ..., z_n)dF(z_1)...dF(z_n). \tag{26}$$

We have

$$E(ve^{t'Z_j}) = E(e^{t'Z_j}r(Z_1, ..., Z_n)) =$$

$$= \int_{\mathbb{R}^k}...\int_{\mathbb{R}^k} r_1(z_1, ..., z_n)e^{t'z_j}dF(z_1)...dF(z_n) - \int_{\mathbb{R}^k}...\int_{\mathbb{R}^k} r_2(z_1, ..., z_n)e^{t'z_j}dF(z_1)...dF(z_n)$$

$$= c_1\int_{\mathbb{R}^k} e^{t'z_j}d\nu_{1j}(z_j) - c_2\int_{\mathbb{R}^k} e^{t'z_j}d\nu_{2j}(z_j),$$

from the definitions of $\nu_{1j}(B_j)$ and $\nu_{2j}(B_j)$. We note that

$$\int_{\mathbb{R}^k} e^{t'z_j}d\nu_{1j}(z_j) \qquad \text{and} \qquad \int_{\mathbb{R}^k} e^{t'z_j}d\nu_{2j}(z_j),$$

are the moment generating functions of the marginal probability measures $\nu_{1j}(B_j)$ and $\nu_{2j}(B_j)$.

We proceed by contradiction. If for each $j = 1, ..., n$ $E(ve^{t'Z_j}) = 0$ for all $t \in \mathbb{R}^k$, by substituting $t = 0$ in

$$c_1\int_{\mathbb{R}^k} e^{t'x_j}d\nu_{1j}(z_j) - c_2\int_{\mathbb{R}^k} e^{t'x_j}d\nu_{2j}(z_j) = 0$$

we easily obtain

$$c_1 = c_2. \tag{27}$$

Thus, for each t

$$\int_{\mathbb{R}^k} e^{t' z_j} d\nu_{1j}(z_j) = \int_{\mathbb{R}^k} e^{t' z_j} d\nu_{2j}(z_j), \tag{28}$$

for any $j = 1, ..., n$, implying

$$\nu_{1j}(B_j) = \nu_{2j}(B_j) \quad \forall B_j \in \mathbb{R}^k, \quad j = 1, ..., n. \tag{29}$$

Now, using additivity of the function $r_s(Z_1, ..., Z_n)$, we obtain

$$\begin{aligned} c_s \nu_{sj}(B_j) &= \int_{B_j} \int_{\mathbb{R}^k} ... \int_{\mathbb{R}^k} \sum_{q=1}^{n} r_{sq}(z_q) dF(z_1)...dF(z_n) \\ &= F(B_j) \sum_{\substack{q=1 \\ q \neq j}}^{n} E(r_{sq}(Z_q)) + \int_{B_j} r_{sj}(z_j) dF(z_j) \quad j = 1, ..., n. \end{aligned}$$

Thus (27) and (29) imply

$$F(B_j) \sum_{\substack{q=1 \\ q \neq j}}^{n} E(r_{1q}(z_q)) + \int_{B_j} r_{1j}(z_j) dF(z_j) = F(B_j) \sum_{\substack{q=1 \\ q \neq j}}^{n} E(r_{2q}(Z_q)) + \int_{B_j} r_{2j}(z_j) dF(z_j). \tag{30}$$

Dividing each side of the last displayed equality by $F(B_j)$, summing over j and using (27), we obtain

$$(n-1)c_1 + \sum_{j=1}^{n} \frac{\int_{B_j} r_{1j}(z_j) dF(z_j)}{F(B_j)} = (n-1)c_2 + \sum_{j=1}^{n} \frac{\int_{B_j} r_{2j}(z_j) dF(z_j)}{F(B_j)}$$

and thus

$$\sum_{j=1}^{n} \frac{\int_{B_j} r_{1j}(z_j) dF(z_j)}{F(B_j)} = \sum_{j=1}^{n} \frac{\int_{B_j} r_{2j}(z_j) dF(z_j)}{F(B_j)}. \tag{31}$$

On the other hand

$$c_s \nu_s(B_1, ..., B_n) = \prod_{q=1}^{n} F(B_q) \sum_{j=1}^{n} \frac{\int_{B_j} r_{sj}(z_j) dF(z_j)}{F(B_j)}. \tag{32}$$

Thus, (27) and (31) imply

$$\nu_1(B_1, ..., B_n) = \nu_2(B_1, ..., B_n) \quad \forall (B_1, ..., B_n), \tag{33}$$

and therefore

$$\int_{B_1} \int_{B_2} ... \int_{B_n} r(z_1, ..., z_n) dF(z_1)...dF(z_n) = 0 \quad \forall (B_1, ..., B_n), \tag{34}$$

implying $r(Z_1, ..., Z_n) = 0$ almost surely, as we wanted to show. ■

Proof of Theorem 2. For ease of notation we drop all the subscript n and we denote $G = G(\lambda_0)$, $S^{-1} = S^{-1}(\lambda_0)$ and $\hat{M}(t) = M(\hat{\theta}, t)$. By the mean value theorem (MVT),

$$\sqrt{n}\hat{M}(t) = \sqrt{n}M(t) + \frac{dM(t)}{d\theta}|_{\theta=\bar{\theta}}\sqrt{n}(\hat{\theta}-\theta), \tag{35}$$

with

$$||\bar{\theta}-\theta_0|| < ||\hat{\theta}-\theta_0||. \tag{36}$$

By standard algebra

$$\frac{dM(t)}{d\theta} = -\frac{1}{n}\underline{e}(t)'S^{-1}\begin{pmatrix}GX\beta & X\end{pmatrix}, \tag{37}$$

where $\underline{e}(t) = (e^{t'\phi(X_1)}, ..., e^{t'\phi(X_n)})'$.

From (36) and $||\hat{\theta}-\theta_0|| = o_p(1)$ (Lee(2004)), (35) becomes

$$\sqrt{n}\hat{M}(t) = \sqrt{n}M(t) + \frac{dM(t)}{d\theta}|_{\theta_0}\sqrt{n}(\hat{\theta}-\theta) + o_p(1). \tag{38}$$

Let's now consider $\sqrt{n}(\hat{\theta}-\theta_0)$.

$$\sqrt{n}(\hat{\theta}-\theta_0) = \Omega^{-1}\begin{pmatrix}\frac{1}{\sigma^2\sqrt{n}}f_1'\epsilon \\ \frac{1}{\sigma^2\sqrt{n}}X'\epsilon\end{pmatrix} + \Omega^{-1}\begin{pmatrix}\frac{1}{\sigma^2\sqrt{n}}(\epsilon'G\epsilon-\sigma^2 tr(G)) \\ 0\end{pmatrix} + o_p(1). \tag{39}$$

Collecting (37)- (39), under H_0,

$$\sqrt{n}\hat{M}(t) = \frac{1}{\sqrt{n}}\psi_1(t)\epsilon + \frac{1}{\sqrt{n}}\psi_2(t)(\epsilon'G\epsilon-\sigma^2 tr(G)) + o_p(1) \tag{40}$$

where the $n-$dimensional row vector $\psi_1(t)$ and the scalar $\psi_2(t)$ are, respectvely,

$$\psi_1(t) = \psi_1(t,\lambda_0,\beta_0,X) = \underline{e}(t)'S^{-1}\left(I_n - \frac{1}{n\sigma^2}\begin{pmatrix}f_1 & X\end{pmatrix}\Omega^{-1}\begin{pmatrix}f_1' \\ X'\end{pmatrix}\right),$$

and

$$\psi_2(t) = \psi_2(t,\lambda_0,\beta_0,X) = -\frac{1}{n\sigma^2}\underline{e}(t)'S^{-1}\begin{pmatrix}GX\beta & X\end{pmatrix}\omega^{1'}. \tag{41}$$

Thus

$$\sqrt{n}\hat{M}(t) = \sum_{i=1}^{n} u_i(t), \tag{42}$$

where

$$u_i(t) = u_{in}(t) = \frac{1}{\sqrt{n}}\psi_{1i}(t)\epsilon_i + \frac{\psi_2(t)}{\sqrt{n}}(\epsilon_i^2-\sigma^2)\bar{g}_{ii} + \frac{2\psi_2(t)}{\sqrt{n}}\epsilon_i\sum_{j<i}\bar{g}_{ij}\epsilon_j, \tag{43}$$

and $\bar{g}_{ij} = (G+G')_{ij}/2$. Conditionally on $\{X\}_{i=1}^{\infty}$, $\{u_i(t), i = 1, ..., n; n = 1, 2, ...\}$ for each t is a triangular array of martingale differences with respect to the filtration formed by ϵ_j, $j < i$. Hence

$$Var(\sqrt{n}\hat{M}(t)) \equiv a(t) = \sum_{i=1}^{n} Var(u_i(t)) = \sum_{i=1}^{n} E(u_i(t)^2)$$

$$= \frac{\sigma^2}{n}\sum_{i=1}^{n}\psi_{1i}(t)^2 + \frac{1}{n}\psi_2^2(\mu^{(4)} - \sigma^4)\sum_{i=1}^{n}\bar{g}_{ii}^2 + \frac{4\sigma^4}{n}\psi_2(t)^2\sum_{i=1}^{n}\sum_{j<i}\bar{g}_{ij}^2 + \frac{2}{n}\psi_2(t)\mu^{(3)}\sum_{i=1}^{n}\psi_{i1}(t)\bar{g}_{ii}$$

(44)

$$= \frac{\sigma^2}{n}\sum_{i=1}^{n}\psi_{1i}(t)^2 + \frac{\sigma^4\psi_2^2}{2n}tr\left((G+G')^2\right) + \frac{\psi_2(t)^2}{n}(\mu^{(4)} - 3\sigma^4)\sum_{i=1}^{n}\bar{g}_{ii}^2 + \frac{2\mu^{(3)}}{n}\psi_2(t)\sum_{i=1}^{n}\psi_{1i}(t)\bar{g}_{ii}.$$

The leading term of (44) is the first one, and by Lemma 1 it is non zero as $n \to \infty$.

Let $z_i(t) = z_{in}(t) = a(t)^{-1/2}u_i(t)$. From Scott (1973), if (conditionally on X)

$$\sum_{i=1}^{n} E(z_i(t)^2|\epsilon_j; j < i) \underset{p}{\to} 1, \tag{45}$$

and for each $\zeta > 0$

$$\sum_{i=1}^{n} E(z_i(t)^2 1(|z_i(t) > \zeta|)) \underset{p}{\to} 0, \tag{46}$$

then $\sum_{i=1}^{n} z_i(t) \to_d \mathcal{N}(0,1)$ pointwise in t.

We start by showing (45). We can equivalently show that (conditionally on X), pointwise in t

$$\sum_{i=1}^{n}\left(E(z_i(t)^2|\epsilon_j; j<i) - E(z_i(t)^2)\right) \underset{p}{\to} 0. \tag{47}$$

We have

$$E(z_i(t)^2|\epsilon_j; j<i) = \frac{1}{na(t)}(\psi_{1i}(t)^2\sigma^2 + \psi_2(t)^2\bar{g}_{ii}^2(\mu^{(4)} - \sigma^4) + 4\psi_2(t)^2\sigma^2\sum_{j<i}\sum_{s<i}\bar{g}_{ij}\bar{g}_{is}\epsilon_j\epsilon_s$$
$$+2\psi_{1i}(t)\psi_2(t)\mu^{(3)}\bar{g}_{ii} + 4(\psi_{1i}(t)\psi_2(t)\sigma^2 + \psi_2(t)^2\mu^{(3)}\bar{g}_{ii})\sum_{j<i}\bar{g}_{ij}\epsilon_j).$$

Thus,

$$\sum_{i=1}^{n}\left(E(z_i(t)^2|\epsilon_j; j<i) - E(z_i(t)^2)\right) = \frac{4\psi_2(t)}{na(t)}\left(\psi_2(t)\left(\sigma^2\sum_{i=1}^{n}\sum_{j<i}\sum_{s<i}\bar{g}_{ij}\bar{g}_{is}\epsilon_j\epsilon_s - \sigma^4\sum_{i=1}^{n}\sum_{j<i}\bar{g}_{ij}^2\right)\right.$$
$$\left. + \sum_{i=1}^{n}\psi_{1i}(t)\sum_{j<i}\bar{g}_{ij}\epsilon_j + \psi_2(t)\mu^{(3)}\sum_{i=1}^{n}\bar{g}_{ii}\sum_{j<i}\bar{g}_{ij}\epsilon_j\right). \tag{48}$$

Since $a_n(t) = O(1)$ and non-zero for each t by Lemma 1 and trivially $\psi_2(t) = O(1)$, (45) holds if pointwise in t

$$\frac{1}{n}(\sigma^2\sum_{i=1}^{n}\sum_{j<i}\sum_{s<i}\bar{g}_{ij}\bar{g}_{is}\epsilon_j\epsilon_s - \sigma^4\sum_{i=1}^{n}\sum_{j<i}\bar{g}_{ij}^2) \underset{p}{\to} 0 \tag{49}$$

(50)
$$\frac{1}{n}\sum_{i=1}^{n}\psi_{1i}(t)\sum_{j<i}\bar{g}_{ij}\epsilon_j \underset{p}{\to} 0$$

and

(51)
$$\frac{1}{n}\sum_{i=1}^{n}\bar{g}_{ii}\sum_{j<i}\bar{g}_{ij}\epsilon_j \underset{p}{\to} 0$$

conditionally on X.

The LHS of (49) can be written as

(52)
$$\frac{\sigma^2}{n}\Big(\sum_{i=1}^{n}\sum_{j<i}\bar{g}_{ij}^2(\epsilon_j^2-\sigma^2)+\sum_{i=1}^{n}\sum_{\substack{j,s<i\\ j\neq s}}\bar{g}_{ij}\bar{g}_{is}\epsilon_j\epsilon_s\Big).$$

The first term of the last displayed expression has mean zero and variance bounded by

(53)
$$\frac{K}{n^2}\sum_{i,k}\sum_{j<i,k}\bar{g}_{ij}^2\bar{g}_{kj}^2 \le \frac{K}{n^2}\sum_{i,j,k}\bar{g}_{ij}^2\bar{g}_{kj}^2 \le \frac{K}{n^2}\left(\max_j\sum_i\bar{g}_{ij}^2\right)\sum_{k,j}\bar{g}_{kj}^2 = O\left(\frac{1}{nh}\right),$$

since

(54)
$$\sum_{k,j}\bar{g}_{kj}^2 = \frac{1}{4}tr((G+G')^2) = O\left(\frac{n}{h}\right)$$

and, denoting by e_j the $n\times 1$ vector with 1 in the $j-$th place and zeros otherwise,

(55)
$$\max_j\sum_i\bar{g}_{ij}^2 = \max_j\frac{1}{4}e_j'(G+G')^2e_j \le K||G+G'||^2 \le K$$

under Assumptions 5 and 6. The second term in (52) has again mean zero and variance bounded by

$$\frac{K}{n^2}|\sum_{i,k}\sum_{j<i,k}\sum_{s<i,k}\bar{g}_{ij}\bar{g}_{is}\bar{g}_{kj}\bar{g}_{ks}| \le \frac{K}{n^2}\sum_{i,k,j,s}|\bar{g}_{ij}\bar{g}_{is}\bar{g}_{kj}\bar{g}_{ks}| \le \frac{K}{n^2}\sum_{i,k,j,s}|\bar{g}_{ij}\bar{g}_{is}|(\bar{g}_{kj}^2+\bar{g}_{ks}^2)$$
$$\le \frac{K}{n^2}\left(\max_j\sum_i|\bar{g}_{ij}|\right)\left(\max_i\sum_s|\bar{g}_{is}|\right)\sum_{k,j}\bar{g}_{kj}^2+\frac{K}{n^2}\left(\max_i\sum_j|\bar{g}_{ij}\right)\left(\max_s\sum_i|\bar{g}_{is}|\right)\sum_{k,s}\bar{g}_{ks}^2$$

(56)
$$= O\left(\frac{1}{nh}\right)$$

under Assumptions 5 and 6 and by (54). Thus, collecting (53) and (56) we conclude (49) by Markov inequality.

The LHS of (50) has mean zero and variance bounded by

$$\frac{K}{n^2}\sum_i\sum_k\sum_{j<i,k}|\psi_{1i}(t)\psi_{ik}(t)\bar{g}_{ij}\bar{g}_{kj}| \leq \frac{K}{n^2}\sum_i\sum_k\sum_j|\bar{g}_{ij}||\bar{g}_{kj}|(\psi_{1i}(t)^2+\psi_{ik}(t)^2)$$

$$\frac{K}{n^2}\big(\max_i\sum_j|\bar{g}_{ij}|\big)\big(\max_j\sum_k|\bar{g}_{kj}|\big)\sum_i\psi_{1i}(t)^2+\frac{K}{n^2}\big(\max_j\sum_i|\bar{g}_{ij}|\big)\big(\max_k\sum_j|\bar{g}_{kj}|\big)\sum_k\psi_{1k}(t)^2$$

$$(57) \quad = O\left(\frac{1}{n}\right)$$

poitwise in t, under Assumptions 5 and 6 and by Lemma 1. (50) follows by Markov inequality.

Similarly, (51) follows by Markov inequality after observing that the LHS has mean zero and variance bounded by

$$\frac{K}{n^2}\sum_i\sum_k\sum_j|\bar{g}_{ij}||\bar{g}_{kj}|(\bar{g}_{ii}^2+\bar{g}_{kk}^2)$$

$$\leq \frac{K}{n^2}\big(\max_i\sum_j|\bar{g}_{ij}|\big)\big(\max_j\sum_k|\bar{g}_{kj}|\big)\sum_i\bar{g}_{ii}^2+\frac{K}{n^2}\big(\max_i\sum_j|\bar{g}_{ij}|\big)\big(\max_k\sum_j|\bar{g}_{kj}|\big)\sum_k\bar{g}_{kk}^2$$

$$(58) \quad = O\left(\frac{1}{nh}\right),$$

since

$$(59) \qquad \sum_i\bar{g}_{ii}^2 \leq \sum_{i,j}\bar{g}_{ij}^2 = O\left(\frac{n}{h}\right).$$

We prove (46) by verifying the sufficient Lyapunov condition poitwise in t

$$(60) \qquad \sum_{i=1}^n E|z_i(t)|^{2+\delta}\to 0$$

conditionally on X, and since $a(t)=O(1)$ and non-zero poitwise in t, we consider equivalently $\sum_i E|u_i(t)|^{2+\delta}$. We use $\sum_i E|u_i(t)|^{2+\delta}=\sum_i E(E|u_i(t)|^{2+\delta}|\epsilon_j, j<i))$. By c_r inequality and since $\psi_2(t)=O(1)$ for each t

$$\sum_{i=1}^n E|u_i(t)|^{2+\delta}\leq\left(\frac{1}{n}\right)^{1+\delta/2}K\sum_i|\psi_{1i}(t)|^{2+\delta}+\left(\frac{1}{n}\right)^{1+\delta/2}K\sum_i|\bar{g}_{ii}|^{2+\delta}$$

$$(61) \qquad +\left(\frac{1}{n}\right)^{1+\delta/2}K\sum_i E|\sum_{j<i}\bar{g}_{ij}\epsilon_j|^{2+\delta}.$$

The first term on the RHS of (61) is bounded by

$$(62) \qquad K\big(\frac{1}{n}\big)^{1+\delta/2}\max_i|\psi_{1i}|^\delta\sum_i\psi_{1i}^2\leq K\frac{1}{n^{\delta/2}}\max_i|\psi_{1i}|^\delta=o(1),$$

since $\sum_i \psi_{1i}^2/n = O(1)$ and non-zero, and for each i $|\psi_{1i}| = O(1)$. Similarly, the second term on the RHS of (61) is bounded by

$$K\frac{1}{n^{\delta/2}}\max_i|\bar{g}_{ii}|^{\delta}\frac{1}{n}\sum_i \bar{g}_{ii}^2 = o(1), \tag{63}$$

since

$$|\bar{g}_{ii}| \leq K||G+G'||_\infty \leq K \tag{64}$$

and by (59). By the Burkholder and von Bahr/Esseen inequalities the last term at the RHS of (61) is bounded by

$$\begin{aligned} &K\Big(\frac{1}{n}\Big)^{1+\delta/2}\sum_i E|\sum_{j<i}\bar{g}_{ij}^2\epsilon_j^2|^{1+\delta/2} \leq K\Big(\frac{1}{n}\Big)^{1+\delta/2}\sum_i\sum_{j<i}|\bar{g}_{ij}|^{2+\delta} \\ &K\Big(\frac{1}{n}\Big)^{1+\delta/2}\sum_i\Big(\sum_{j<i}\bar{g}_{ij}^2\Big)^{1+\delta/2} \leq K\Big(\frac{1}{n}\Big)^{1+\delta/2}\Big(\max_i\sum_j\bar{g}_{ij}^2\Big)^{\delta/2}\sum_{i,j}\bar{g}_{ij}^2 = o(1) \end{aligned} \tag{65}$$

by (54) and (55). Thus, collecting (62), (63) and (65) we conclude that (46) holds pointwise in t. ■

Proof of Theorem 3. One can write

$$\begin{aligned} M(\hat{\theta},t) &= \frac{1}{n}\sum_{i=1}^{n}\big(g_i(X)-m_i(X,\hat{\theta})+\eta_i\big)e^{t'\phi(X_i)} \\ &= \frac{1}{n}\sum_{i=1}^{n}E\Big(\big(g_i(X)-m_i(X,\theta^\sharp)\big)e^{t'\phi(X_i)}\Big)+\frac{1}{n}\sum_{i=1}^{n}\eta_i e^{t'\phi(X_i)} &(66)\\ &+\frac{1}{n}\sum_{i=1}^{n}\Big(g_i(X)e^{t'\phi(X_i)}-E\big(g_i(X)e^{t'\phi(X_i)}\big)\Big) &(67)\\ &-\frac{1}{n}\sum_{i=1}^{n}\Big(m_i(X,\theta^\sharp)e^{t'\phi(X_i)}-E\big(m_i(X,\theta^\sharp)e^{t'\phi(X_i)}\big)\Big) &(68)\\ &-\frac{1}{n}\sum_{i=1}^{n}\big(m_i(X,\hat{\theta})-m_i(X,\theta^\sharp)\big)e^{t'\phi(X_i)} &(69)\\ &=: p_{n,1}+p_{n,2}+p_{n,3}+p_{n,4}+p_{n,5}. &(70) \end{aligned}$$

Under Assumption 10, we have

$$p_{n,1} = \frac{1}{n}\sum_{i=1}^{n}E\Big(\big(g_i(X)-m_i(X,\theta^\sharp)\big)e^{t'\phi(X_i)}\Big) \neq 0,$$

for all n large enough. We will now verify that $p_{n,i} = o_p(1), i = 2,...,5$, which is sufficient to prove Theorem 3.

First we prove that $p_{n,5} = o_p(1)$. By the mean value theorem, one has for some $\bar{\theta}$ such that $\|\bar{\theta}-\theta_0\| < \|\hat{\theta}-\theta_0\|$

$$p_{n,5} = \Big(\frac{1}{n}\sum_{i=1}^{n}e^{t'\phi(X_i)}\frac{\partial m_i(X,\theta)}{\partial\theta}|_{\bar{\theta}}\Big)(\hat{\theta}-\theta^\sharp). \tag{71}$$

Since $(\hat{\theta} - \theta^\sharp) = o_p(1)$ under Assumption 9, it suffices to verify that first term on RHS in (71) is $O_p(1)$. Recall $m_i(X, \theta) = S^{(i)}(\lambda)' X\beta$, where $S^{(i)}$ denotes the transpose of the ith row of S^{-1}. Note

$$\frac{\partial S^{-1}(\lambda)}{\partial \lambda} = -S^{-1}(\lambda)\frac{\partial S(\lambda)}{\partial \lambda}S^{-1}(\lambda) = S^{-1}(\lambda)G(\lambda)$$

and hence

$$\frac{\partial m_i(X,\theta)}{\partial \lambda} = S^{(i)}(\lambda)'G(\lambda)X\beta, \quad \frac{\partial m_i(X,\theta)}{\partial \beta} = S^{(i)}(\lambda)'X.$$

We can therefore write, denoting $n \times 1$ vector $\underline{e}(t) = (e^{t'\phi(X_1)}, \cdots, e^{t'\phi(X_n)})'$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n e^{t'\phi(X_i)}\frac{\partial m_i(X_n,\theta)}{\partial \theta}\Big|_{\bar{\theta}} =& \frac{1}{n}\sum_{i=1}^n e^{t'\phi(X_i)}S^{(i)}(\bar{\lambda})[G(\bar{\lambda})X\bar{\beta}; X] \\ =& \frac{1}{n}[\underline{e}(t)'S^{-1}(\bar{\lambda})G(\bar{\lambda})X\bar{\beta}; \underline{e}(t)'S^{-1}(\bar{\lambda})X] = O_p(1),\end{aligned}$$

which follows from applying Lemma 2 to each element of the $1 \times (k+1)$ vector above, using Assumptions 5,6,7. Hence we have that (69) is $O_p(1)$.

Now, we verify that $p_{n,i} = o_p(1), i = 3, 4$. Denote mean-differenced quantities $\bar{e}^{t'\phi(X_i)} := e^{t'\phi(X_i)} - E(e^{t'\phi(X_i)})$ and $\bar{r}(X_i) := r(X_i) - E(r(X_i))$. We can write

$$\begin{aligned} np_{n,3} =& \sum_{i,j=1:j\neq i}^n a_{1ij}\bar{e}^{t'\phi(X_j)}\bar{r}_1(X_i) + \sum_{i,j=1:j\neq i}^n a_{2ij}\bar{e}^{t'\phi(X_j)}\bar{r}_2(X_i) \\ &+ \sum_{i,j=1:j\neq i}^n a_{1ij}\bar{e}^{t'\phi(X_j)}E(r_1(X_1)) + \sum_{i,j=1:j\neq i}^n a_{2ij}\bar{e}^{t'\phi(X_j)}E(r_2(X_1)) + \sum_{i,j=1:j\neq i}^n a_{1ij}\bar{r}_1(X_i)E(e^{t'\phi(X_1)}) \\ &+ \sum_{i,j=1:j\neq i}^n a_{2ij}\bar{r}_2(X_i)E(e^{t'\phi(X_1)}) + \sum_{i=1}^n a_{1ii}\big(r_1(X_i)e^{t'\phi(X_i)} - E(r_1(X_i)e^{t'\phi(X_i)})\big) \\ &+ \sum_{i=1}^n a_{2ii}\big(r_2(X_i)e^{t'\phi(X_i)} - E(r_2(X_i)e^{t'\phi(X_i)})\big). \end{aligned}$$

Similarly, denote $\bar{X}_i := X_i - E(X_i)$, and decompose

$$\begin{aligned} np_{n,4} =& \sum_{i,j=1:j\neq i}^n s^{ij}\bar{e}^{t'\phi(X_j)}\bar{X}_i\beta^\sharp + \sum_{i,j=1:j\neq i}^n s^{ij}\bar{e}^{t'\phi(X_j)}E(X_i\beta^\sharp) \\ &+ \sum_{i,j=1:j\neq i}^n s^{ij}\bar{X}_i\beta^\sharp E(e^{t'\phi(X_1)}) + \sum_{i=1}^n s^{ii}\big(X_i\beta^\sharp e^{t'\phi(X_i)} - E(X_i\beta^\sharp e^{t'\phi(X_i)})\big). \end{aligned}$$

To verify that $p_{n,3}, p_{n,4} = o_p(1)$, note all terms in the expansion of $p_{n,3}$ and $p_{n,4}$ above have mean zero. Their variances are of order $o(1)$: they can be decomposed as in line (79) of the proof of Lemma 2 and shown to be $o(1)$ under Assumptions 2, 6 and 7, following the same steps as given there.

Finally we show $p_{n,2} = o_p(1)$. Assumption 3 implies $E(\eta_1 e^{t'\phi(X_1)}) = 0$, so it suffice to verify the variance is $o(1)$:

$$Var(\frac{1}{n}\sum_{i=1}^{n}\eta_i e^{t'\phi(X_i)}) \leq \frac{C}{n^2}\sum_{i,j=1}^{n}|Cov(\eta_i,\eta_j)| = o(1)$$

under Assumption 3 and noting $\phi(\cdot)$ is a bounded function.

We have now shown that

$$\hat{M}(t) = \frac{1}{n}\sum_{i=1}^{n} E\Big(\{g_i(X) - m_i(X,\theta^\sharp)\}e^{t'\phi(X_i)}\Big) + o_p(1),$$

proving Theorem 3. ■

7. Lemma and Proposition

Let $||\cdot||_e$ denote the Euclidean norm of a $n\times 1$ vector.

Lemma 1 *Under (3) and Assumptions 5-8, $||\psi_1(t)||_e^2/n \geq c > 0$ as $n\to\infty$ for all $t\in\mathbb{R}^k$.*

Proof of Lemma 1. Under Assumptions 5-7,

$$\Omega = \frac{1}{n\sigma^2}\begin{pmatrix} f_1'f_1 & f_1'X \\ X'f_1 & X'X\end{pmatrix} + O\left(\frac{1}{h}\right) \equiv \tilde{\Omega} + \left(\frac{1}{h}\right). \tag{72}$$

We have $\Omega = \tilde{\Omega}(I_k + \tilde{\Omega}^{-1}O(1/h))$ and therefore

$$\Omega^{-1} = \tilde{\Omega}^{-1} + O\left(\frac{1}{h}\right), \tag{73}$$

since $\tilde{\Omega} = O(1)$ and non-singular by Assumption 8.

Since

$$\frac{1}{n}\underline{e}(t)'S^{-1}S^{-1'}\underline{e}(t) = O(1), \tag{74}$$

we derive

$$\frac{\psi_1'\psi_1}{n} = \underline{e}(t)'S^{-1}\left(I_n - \frac{1}{n\sigma^2}\begin{pmatrix} f_1 & X\end{pmatrix}\tilde{\Omega}^{-1}\begin{pmatrix} f_1' \\ X'\end{pmatrix}\right)^2 S^{-1'}\underline{e}(t) + O\left(\frac{1}{h}\right). \tag{75}$$

The $n\times n$ matrix

$$I_n - \frac{1}{n\sigma^2}\begin{pmatrix} f_1 & X\end{pmatrix}\tilde{\Omega}^{-1}\begin{pmatrix} f_1' \\ X'\end{pmatrix}$$

is symmetric and idempotent with rank $n-k-1$ under Assumption 7, so

$$\begin{aligned}\frac{\psi_1'\psi_1}{n} =& \frac{1}{n}\underline{e}(t)'S^{-1}QDQ'S^{-1'}\underline{e}(t) \\ =& \frac{1}{n}\sum_{i=1}^{n-k-1}(\underline{e}(t)'S^{-1}Q)_i^2, \end{aligned} \tag{76}$$

where D is the $n \times n$ diagonal matrix of eigenvalues with the first $n-k-1$ diagonal entries equal to one and the others equal to zero, and Q is the orthogonal matrix of the corresponding eigenvectors. As $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n-k-1}(\underline{e}'(t)S^{-1}Q)_i^2 \to \lim_{n\to\infty}\frac{||\underline{e}(t)'S^{-1}Q||_e^2}{n} = \lim_{n\to\infty}\frac{1}{n}\underline{e}(t)'S^{-1}S^{-1\prime}\underline{e}(t) \tag{77}$$

pointwise in t, bearing in mind last k terms omitted from summation in the first term is negligible as $n \to \infty$.

Since elements of $\underline{e}(t)$ are non-zero for all $t \in \mathbb{R}^k$, a sufficient condition to guarantee the last limit in (77) to be non-zero is non-singularity of $S^{-1}(\lambda_0)$ for n large enough, which is guaranteed under Assumptions 4 and 5.■

Lemma 2 *Let $A = A_n$ be an $n \times n$ matrix satisfying $\|A\|_\infty + \|A'\|_\infty < K$. Functions $d_1(\cdot), d_2(\cdot) : \mathbb{R}^K \to \mathbb{R}^K$ satisfy $Ed_1^4(X_1) + Ed_2^4(X_1) < \infty$ for iid X_i. Denoting $\underline{\mathrm{d}}_1 = (d_1(X_1), \cdots, d_1(X_n))', \underline{\mathrm{d}}_2 = (d_2(X_1), \cdots, d_2(X_n))'$,*

$$\frac{1}{n}\underline{\mathrm{d}}_1' A \underline{\mathrm{d}}_2 - v_n = o_p(1), \tag{78}$$

where

$$v_n = \frac{1}{n}E(\underline{\mathrm{d}}_1' A \underline{\mathrm{d}}_2) = E(d_1(X_1))E(d_2(X_1))\frac{1}{n}\sum_{i,j=1:i\neq j}^{n} a_{ij} + E(d_1(X_1)d_2(X_1))\frac{1}{n}\sum_{i=1}^{n} a_{ii} = O(1).$$

If $v_n = 0$, then $\underline{\mathrm{d}}_1' A \underline{\mathrm{d}}_2/\sqrt{n} = O_p(1)$.

Proof of Lemma 2. Firstly,

$$v_n \leq CE|d_1(X_1)|E|d_2(X_1)|\frac{1}{n}n\max_j \sum_{i=1}^{n}|a_{ij}| + (Ed_1^2(X_1)Ed_2^2(X_1))^{1/2}\frac{1}{n}\sum_{i=1}^{n}|a_{ii}| = O(1).$$

To show (78), we verify $Var\big(\underline{\mathrm{d}}_1' A \underline{\mathrm{d}}_2/n\big) = o(1)$ below:

$$\begin{aligned} Var\Big(\frac{1}{n}\sum_{i,j=1}^{n} a_{ij}d_1(X_i)d_2(X_j)\Big) &= \frac{1}{n^2}\sum_{\substack{i,j,j'=1 \\ i\neq j, j\neq j', j'\neq i}}^{n}[C_1 a_{ij}a_{ij'} + C_2(a_{ij}a_{j'i} + a_{ij}a_{jj'}) + C_3 a_{ij}a_{j'j}] \\ &+ \frac{1}{n^2}\sum_{i,j=1:i\neq j}^{n}[C_4(a_{ij})^2 + C_5 a_{ij}a_{ji}] + \frac{1}{n^2}\sum_{i=1}^{n} C_6 a_{ii}^2 \end{aligned} \tag{79}$$

where denoting $Ed_1(X_1) = \alpha_1, Ed_2(X_1) = \alpha_2$, set

$$C_1 = E(d_1^2(X_1))\alpha_2^2 - \alpha_1^2\alpha_2^2,\ C_2 = \alpha_1\alpha_2 E(d_1(X_1)d_2(X_1)) - \alpha_1^2\alpha_2^2,$$
$$C_3 = \alpha_1^2 E(d_2^2(X_1)) - \alpha_1^2\alpha_2^2,\ C_4 = E(d_1^2(X_1))E(d_2^2(X_1)) - \alpha_1^2\alpha_2^2,$$
$$C_5 = E^2(d_1(X_1)d_2(X_1)) - \alpha_1^2\alpha_2^2,\ C_6 = E(d_1^2(X_1)d_2^2(X_1)) - \alpha_1^2\alpha_2^2.$$

Under assumptions of the lemma, $C_i < \infty, i = 1, ..., 6$. To see that the first term on RHS of (79) is $o(1)$, note

$$\frac{1}{n^2}\sum_{i,j,j'=1}^{n} a_{ij}a_{ij'} \leq \frac{1}{n}(\max_i \sum_{j=1}^{n}|a_{ij}|)^2 = O(1/n) = o(1).$$

Proceeding similarly for all terms of (79) completes the proof of the lemma.
■

Lemma 3 *Let $A = A_n$ be a $n \times n$ matrix satisfying $\|A\|_\infty + \|A'\|_\infty < K$. Let X_i be sequence of iid random variables and function $d(\cdot) : \mathbb{R}^k \to \mathbb{R}$ satisfies $Ed^2(X_1) < \infty$. Let η_i be zero mean process which is independent of X and satisfies $\max_{1 \le i \le n} \sum_{k=1}^{n} |Cov(\eta_i, \eta_k)| = O(1)$. Denote $\underline{\mathrm{d}} = (d(X_1), \cdots, d(X_n))'$. Then,*

$$\frac{1}{\sqrt{n}} \underline{\mathrm{d}}' A \eta = O_p(1). \tag{80}$$

Proof of Lemma 3. Since $E(\underline{\mathrm{d}}' A\eta) = 0$, it suffices to verify that $Var(\underline{\mathrm{d}}' A\eta/\sqrt{n}) = O(1)$. We have

$$\begin{aligned}
Var(\frac{1}{\sqrt{n}} \underline{\mathrm{d}}' A\eta) &= \frac{1}{n} \sum_{i,j,i',j'=1}^{n} a_{ij} a_{i'j'} E(d(X_i) d(X_{i'})) E(\eta_j \eta_{j'}) \\
&\le E(d^2(X_1)) \frac{1}{n} \sum_{i,j,i',j'=1}^{n} |a_{ij} a_{i'j'} Cov(\eta_j, \eta_{j'})| \\
&\le \frac{C}{n} n \max_j \Big[\sum_{j'=1}^{n} |Cov(\eta_j, \eta_{j'})| \Big] \Big[\max_j \Big(\sum_{i=1}^{n} |a_{ij}| \Big) \Big]^2 = O(1). \quad \blacksquare
\end{aligned}$$

Below are the additional technical assumptions that are required for Proposition.

Let $\mathcal{N}_\delta = \{\lambda : |\lambda - \lambda^\sharp| < \delta\}$ and $\bar{\mathcal{N}_\delta} = \Lambda / \mathcal{N}_\delta$ for some $\delta > 0$. Define $\hat{\sigma}^2(\lambda) = \frac{1}{n} y' S(\lambda)' M_X S(\lambda) y$, $M_X = I - X(X'X)^{-1} X'$ and $\tilde{\sigma}^2 := E_1(\hat{\sigma}^2(\lambda))$, with $E_1(\cdot)$ denoting expectation under H_1.

Assumption A *For all sufficiently large n, $\lambda^\sharp \in \Lambda$ and there exists $\delta > 0$ such that*

$$\underline{\lim}_{n \to \infty} \inf_{\lambda \in \bar{\mathcal{N}_\delta}} \Big(\frac{\tilde{\sigma}^2(\lambda)}{\tilde{\sigma}^2(\lambda^\sharp)} |S(\lambda^\sharp)' S^{-1}(\lambda)' S^{-1}(\lambda) S(\lambda^\sharp)|^{1/n} \Big) > 1$$

Assumption A is an identification condition on $\lambda^\sharp$ under H_1, akin to Assumption 5 of Delgado and Robinson (2015).

Denote $g = (g_1(X), \cdots, g_n(X))'$.

Assumption B

$$\underline{\lim}_{n \to \infty} \inf_{\Lambda} \tilde{\sigma}^2(\lambda) = \underline{\lim}_{n \to \infty} \inf_{\Lambda} \frac{1}{n} \big[E(g' S(\lambda)' M_X S(\lambda) g) + E(\eta' S(\lambda)' M_X S(\lambda) \eta) \big] > 0$$

Denoting by $\kappa(u_i, u_j, u_k, u_\ell)$ the fourth cumulant of process $\{u_i\}$, the following assumption is satisfied by weakly dependent η_i:

Assumption C *As $n \to \infty$,*

$$\sum_{j,\ell=1}^{n} \max_{1 \le i,k \le n} |\kappa(\eta_i, \eta_j, \eta_k, \eta_\ell)| = o(n^2).$$

For $N(\lambda) := S(\lambda) M_X S(\lambda)'$, we need the following assumption:

Assumption D *For any $\lambda^\dagger \in \Lambda$ and any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\text{plim}_{n\to\infty} \sup_{\lambda:\|\lambda-\lambda^\dagger\|<\delta;\lambda\in\Lambda} \|N(\lambda) - N(\lambda^\dagger)\| < \varepsilon.$$

Sufficient condition for Assumption D are: $\|W\|, sup_\lambda\|S(\lambda)\| < K$ which follow from Assumption 5 (ii), and $\|M_X\| = O_p(1)$. To see that these are sufficient, note we have $dN(\lambda)/d\lambda = -WM_XS(\lambda)' - S(\lambda)M_XW'$ and by the mean value theorem for some $\bar{\lambda}$ such that $|\bar{\lambda} - \lambda^\dagger| < |\lambda - \lambda^\dagger|$:

$$\|N(\lambda) - N(\lambda^\dagger)\| = \|(\lambda - \lambda^\dagger)\frac{dN(\lambda)}{d\lambda}|_{\lambda=\bar{\lambda}}\| \leq |\lambda - \lambda^\dagger|\|\frac{dN(\lambda)}{d\lambda}|_{\lambda=\bar{\lambda}}\|,$$

$$\|\frac{dN(\lambda)}{d\lambda}|_{\lambda=\bar{\lambda}}\| = 2\|S(\bar{\lambda})M_XW'\| \leq 2\|S(\bar{\lambda})\|\|M_X\|\|W\| < K.$$

Therefore $\|N(\lambda) - N(\lambda^\dagger)\| \leq K|\lambda - \lambda^\dagger|$ and for any $\varepsilon > 0$, one can set $\delta = \varepsilon/K$ to satisfy Assumption D.

Assumption E *The square of the Euclidean norm of the function $g(\cdot)$ has fourth moment of order $o(n^2)$, i.e.*

$$E\left(\sum_i g_i^2\right)^4 = o(n^2). \tag{81}$$

Proposition 1 *Define a sequence of pseudo true values as $\lambda_n^\sharp = \lambda^\sharp := \arg\min_{\lambda\in\Lambda} \tilde{Q}(\lambda)$ where*

$$\tilde{Q}(\lambda) = \log(\tilde{\sigma}^2(\lambda)) + \frac{1}{n}\log|S^{-1}(\lambda)'S^{-1}(\lambda)|. \tag{82}$$

Under Assumptions A-E and 1-7, $\hat{\theta} - \theta^\sharp = o_p(1)$ under H_1.

Proof of Proposition 1.

We follow Delgado and Robinson (2015)'s arguments given in the proof of their Theorem 1. As usual we proceed with concentrated likelihood as

$$\hat{\lambda} = \arg\min_{\lambda\in\Lambda} (Q(\lambda)) \tag{83}$$

where

$$Q(\lambda) = \log(\hat{\sigma}^2(\lambda)) + \frac{1}{n}\log|S^{-1}(\lambda)'S^{-1}(\lambda)|. \tag{84}$$

Since $E(g'\eta) = 0$ and $Y = g(X) + \eta = g + \eta$ under H_1, we have

$$\tilde{\sigma}^2(\lambda) = E_1(\hat{\sigma}^2(\lambda)) = \frac{1}{n}E(g'S(\lambda)'M_XS(\lambda)g) + \frac{1}{n}E(\eta'S(\lambda)'M_XS(\lambda)\eta).$$

One can write

$$\begin{aligned}
\tilde{Q}(\lambda)-\tilde{Q}(\lambda^{\sharp}) &= \log\big(\frac{\tilde{\sigma}^2(\lambda)}{\tilde{\sigma}^2(\lambda^{\sharp})}|S(\lambda^{\sharp})S(\lambda^{\sharp})'S^{-1}(\lambda)'S^{-1}(\lambda)|^{1/n}\big)\\
&= \log\big(\frac{\tilde{\sigma}^2(\lambda)}{\tilde{\sigma}^2(\lambda^{\sharp})}|S(\lambda^{\sharp})'S^{-1}(\lambda)'S^{-1}(\lambda)S(\lambda^{\sharp})|^{1/n}\big),\\
Q(\lambda)-\tilde{Q}(\lambda) &= \log\hat{\sigma}^2(\lambda)-\log\tilde{\sigma}^2(\lambda)=\log\frac{\hat{\sigma}^2(\lambda)}{\tilde{\sigma}^2(\lambda)}.
\end{aligned}$$

Let P_1 denote probability under H_1. The following chain of inequalities hold:

$$P_1(\hat{\lambda}\in\bar{\mathcal{N}}_\delta)\le P_1\big(\inf_{\bar{\mathcal{N}}_\delta}Q(\lambda)<Q(\lambda^{\sharp})\big)\le P_1\big(\sup_{\Lambda}|Q(\lambda)-\tilde{Q}(\lambda)|\ge\inf_{\bar{\mathcal{N}}_\delta}|\tilde{Q}(\lambda)-\tilde{Q}(\lambda^{\sharp})|\big).$$

To see the last step, note that from the definition of $\lambda^{\sharp}$, it holds that $\inf_{\bar{\mathcal{N}}_\delta}\tilde{Q}(\lambda)>\tilde{Q}(\lambda^{\sharp})$. Therefore, for $\inf_{\bar{\mathcal{N}}_\delta}Q(\lambda)\le Q(\lambda^{\sharp})$ to hold, it must be that at $\lambda^*=\operatorname{argmin}_{\bar{\mathcal{N}}_\delta}Q(\lambda)$, the magnitude of $|Q(\lambda^*)-\tilde{Q}(\lambda^*)|$ dominates that of $|\tilde{Q}(\lambda^*)-\tilde{Q}(\lambda^{\sharp})|$, implying

$$\sup_{\bar{\mathcal{N}}_\delta}|Q(\lambda)-\tilde{Q}(\lambda)|\ge\inf_{\bar{\mathcal{N}}_\delta}|\tilde{Q}(\lambda)-\tilde{Q}(\lambda^{\sharp})|,$$

which in turn implies

$$\sup_{\Lambda}|Q(\lambda)-\tilde{Q}(\lambda)|\ge\inf_{\bar{\mathcal{N}}_\delta}|\tilde{Q}(\lambda)-\tilde{Q}(\lambda^{\sharp})|.$$

Hence, to complete the proof of Proposition, it suffices to verify the following two statements:

(85) $\inf_{\bar{\mathcal{N}}_\delta}\big(\tilde{Q}(\lambda)-\tilde{Q}(\lambda^{\sharp})\big)>\epsilon$, for all sufficiently large n and for some $\epsilon>0$,

(86) $\sup_{\Lambda}|Q(\lambda)-\tilde{Q}(\lambda)|\to 0, \quad \text{as} \quad n\to\infty.$

(85) follows from Assumption A.

Noting LHS of (86) is bounded by

$$\sup_{\Lambda}\log\frac{\hat{\sigma}^2(\lambda)}{\tilde{\sigma}^2(\lambda)}\le\sup_{\Lambda}|\hat{\sigma}^2(\lambda)-\tilde{\sigma}^2(\lambda)|/\inf_{\Lambda}\tilde{\sigma}^2(\lambda),$$

to show (86), we need to establish

$$\sup_{\Lambda}|\hat{\sigma}^2(\lambda)-\tilde{\sigma}^2(\lambda)|\to 0, \tag{87}$$

$$\underline{\lim}_{n\to\infty}\inf_{\Lambda}\tilde{\sigma}^2(\lambda)>0. \tag{88}$$

Assumption B implies (88).

To establish (87), we first verify pointwise convergence of $\hat{\sigma}^2(\lambda)-\tilde{\sigma}^2(\lambda)$ to zero in probability by showing

$$Var(\hat{\sigma}^2(\lambda)-\tilde{\sigma}^2(\lambda))\to_p 0, \text{ pointwise for given } \lambda. \tag{89}$$

Under H_1, $\hat{\sigma}^2(\lambda) - \tilde{\sigma}^2(\lambda)$ has zero mean and variance bounded by

$$
\begin{aligned}
& KE\left(\frac{1}{n}\eta' N(\lambda)\eta - E\left(\frac{1}{n}\eta' N(\lambda)\eta\right)\right)^2 + KE\left(\frac{1}{n}g' N(\lambda)g - E\left(\frac{1}{n}g' N(\lambda)g\right)\right)^2 \\
+ \; & KE\left(\frac{2}{n}g' N(\lambda)\eta)\right)^2, \qquad (90)
\end{aligned}
$$

by the c$_r$ inequality. Let $E(\eta\eta'|X) = E(\eta\eta') = \Omega_\eta$ and denote below $\omega_{ij} = Cov(\eta_i, \eta_j)$. The first term in (90) is bounded by

$$
\frac{K}{n^2}E((\eta' N(\lambda)\eta)^2) = \frac{K}{n^2}E\left(E((\eta' N(\lambda)\eta)^2|X)\right) \leq \frac{K}{n^2}E((tr(N(\lambda)\Omega_\eta))^2), \qquad (91)
$$

where in the last step we calculated the inner expectation by standard formulae for moments of quadratic forms and retained only dominant terms. We write the RHS of (91) as

$$
\begin{aligned}
\frac{1}{n^2}E|\sum_{i,j,s,t} n_{ij}n_{st}\omega_{ji}\omega_{ts}| \;&\leq\; \frac{K}{n^2}E\left(\sum_{i,j,s,t}|\omega_{ji}||\omega_{ts}|(n_{ij}^2 + n_{st}^2)\right) \\
&\leq\; \frac{K}{n}\sup_j\sum_i|\omega_{ji}|\sup_t\sum_s|\omega_{ts}|E\left(\sup_i\sum_j n_{ij}^2\right) \\
&+\; \frac{K}{n}\sup_j\sum_i|\omega_{ji}|\sup_t\sum_s|\omega_{ts}|E\left(\sup_s\sum_t n_{st}^2\right) = O\left(\frac{1}{n}\right) \qquad (92)
\end{aligned}
$$

under Assumption 3, and since

$$
\sup_i\sum_j n_{ij}^2 = \sup_i(N(\lambda)^2)_{ii} \leq ||N(\lambda)||^2 = O_p(1) \qquad (93)
$$

under Assumptions 4 and 5, as long as $E||M_X||^2$ exists.

The second term in (90) is bounded by

$$\frac{K}{n^2}E\left(\sum_{i,j,k,t} g_ig_jg_kg_tn_{ij}n_{kt}\right) \leq \frac{K}{n^2}E\left(\sum_{i,j,k,t}|g_ig_jg_kg_t|(n_{ij}^2+n_{kt}^2)\right)$$

$$\leq \frac{K}{n^2}E\left(\sum_{i,j,k,t}(g_i^2g_k^2+g_j^2g_t^2)(n_{ij}^2+n_{kt}^2)\right)$$

$$= \frac{K}{n}\left(E\left(\sum_{i,k,j}g_i^2g_k^2n_{ij}^2\right)+E\left(\sum_{j,t,k}g_j^2g_t^2n_{kt}^2\right)+E\left(\sum_{i,k,t}g_i^2g_k^2n_{kt}^2\right)+E\left(\sum_{i,t,j}g_j^2g_t^2n_{ij}^2\right)\right)$$

$$\leq \frac{K}{n}E\left(\sup_i\sum_j n_{ij}^2\sum_{i,k}g_i^2g_k^2\right)+\frac{K}{n}E\left(\sup_t\sum_k n_{kt}^2\sum_{j,t}g_j^2g_t^2\right)+\frac{K}{n}E\left(\sup_k\sum_t n_{kt}^2\sum_{i,k}g_k^2g_i^2\right)$$

$$+ \frac{K}{n}E\left(\sup_j\sum_i n_{ij}^2\sum_{j,t}g_j^2g_t^2\right).$$

. (94)

Now, each term of the last displayed expression is bounded by

(95)

$$\frac{K}{n}E\left(||N(\lambda)||^2\left(\sum_i g_i^2\right)^2\right) \leq \frac{K}{n}\left(E\left(||N(\lambda)||^4\right)\right)^{1/2}\left(E\left(\sum_i g_i^2\right)^4\right)^{1/2} = o(1),$$

under Assumption E and provided that $E||M_X||^4$ exists.

Similarly, the last term of (90) is bounded by

$$\frac{K}{n^2}\sup_{j,k}|\omega_{jk}|E\left(\sum_{i,j,k,t}|g_i||g_k|(n_{ij}^2+n_{kt}^2)\right)$$

$$\leq \frac{K}{n}E\left(\sup_i\sum_j n_{ij}^2\left(\sum_i|g_i|\right)^2\right)+\frac{K}{n}E\left(\sup_k\sum_t n_{kt}^2\left(\sum_i|g_i|\right)^2\right).$$

(96)

Each term of the last displayed expression is bounded by

$$\frac{k}{n}\left(E\left(||N(\lambda)||^4\right)\right)^{1/2}\left(E\left(\sum_i|g_i|\right)^4\right)^{1/2}, \tag{97}$$

which is $o(1)$ by the same argument outlined above after observing that

$$\left(E\left(\sum_i|g_i|\right)^4\right)^{1/2} \leq \left(E\left(\sum_i g_i^2\right)^2\right)^{1/2} = o(n) \tag{98}$$

under Assumption E.

Hence, we have established (89). The uniform convergence in (87) follows from compactness of Λ and noting that for any $\lambda^{\dagger} \in \Lambda$ and small enough

$\varepsilon > 0$, we can find $\delta > 0$ such that for $\mathcal{N}_{\dagger\delta} = \{\lambda : |\lambda - \lambda^\dagger| \leq \delta\}$:

$$\begin{aligned}
&E_1 \sup_{\mathcal{N}_{\dagger\delta}} |(\hat{\sigma}^2(\lambda) - \tilde{\sigma}^2(\lambda)) - (\hat{\sigma}^2(\lambda^\dagger) - \tilde{\sigma}^2(\lambda^\dagger))| \\
&= \frac{1}{n} E_1 \sup_{\mathcal{N}_{\dagger\delta}} |y'(N(\lambda) - N(\lambda^\dagger))y - E_1 y'(N(\lambda) - N(\lambda^\dagger))y| \\
&= \frac{1}{n} E_1 \sup_{\mathcal{N}_{\dagger\delta}} |\mathrm{tr}\big((\mathrm{N}(\lambda) - \mathrm{N}(\lambda^\dagger))(\mathrm{yy}' - \mathrm{E_1yy}')\big)| \\
&\leq \frac{1}{n} \big(E_1 \|y\|_e^2 + \mathrm{tr}(\mathrm{E_1yy}')\big) \sup_{\mathcal{N}_{\dagger\delta}} \|\mathrm{N}(\lambda) - \mathrm{N}(\lambda^\dagger)\| = \mathrm{O}(\varepsilon).
\end{aligned}$$

$\big(E_1\|y\|_e^2 + \mathrm{tr}(\mathrm{E_1yy}')\big)/\mathrm{n} = \mathrm{O}(1)$ from Assumptions 2 and 3.

References

[1] Bierens, H. J. (1982), "Consistent model specification tests ", *Journal of Econometrics*, 20, 105-134.

[2] Bierens, H. J. (1990), "A consistent conditional moment test of functional form", *Econometrica*, 58, 1443-1458.

[3] Case, A.C. (1991), "Spatial patterns in household demand", *Econometrica*, 59, 953-965.

[4] Cliff, A. and J.K. Ord, (1968), "The problem of spatial autocorrelation", *Joint Discus- sion Paper, University of Bristol: Department of Economics*, 26, *Department of Geography, Series A*, 15.

[5] Delgado,M. A. and P. M. Robinson (2015), "Non-nested testing of spatial correlation", *Journal of Econometrics*, 187, 385-401.

[6] Fan, Y. and Q. Li, (1996), "Consistent model specification tests: omitted variables and semiparametric functional forms", *Econometrica*, 64, 865-890.

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

[8] Kelejian, H.H. and Prucha, I.R. (1999), "A generalized moments estimator for the autoregressive parameter in a spatial model", *International Economic Review*, 40, 509-533.

[9] Lee, L.F. (2002), "Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models", *Econometric Theory*, 18, 252-277.

[10] Lee, L.F. (2004), "Asymptotic distribution of quasi-maximum likelihood estimates for spatial autoregressive models", *Econometrica*, 72, 1899-1925.

[11] Su, L. and Xi, Q. (2016), "Specification test for spatial autoregressive models", forthcoming in *Journal of Business and Economic Statistics.*

[12] Wooldridge, J.M. (1991), "Specification testing and quasi maximum likelihood estimation", *Journal of Econometrics*, 48, 29-55.