8-28-2020

Consistent Misspecification Testing in Spatial Autoregressive Models

Jungyoon Lee

Peter C. B. Phillips
Yale University

Francesca Rossi

CONSISTENT MISSPECIFICATION TESTING
IN SPATIAL AUTOREGRESSIVE MODELS

By

Jungyoon Lee, Peter C. B. Phillips, and Francesca Rossi

August 2020

COWLES FOUNDATION DISCUSSION PAPER NO. 2256

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Consistent Misspecification Testing in Spatial Autoregressive Models. *

Jungyoon Lee*, Peter C. B. Phillips† and Francesca Rossi‡

August 27, 2020

Abstract

Spatial autoregressive (SAR) and related models offer flexible yet parsimonious ways to model spatial or network interaction. SAR specifications typically rely on a particular parametric functional form and an exogenous choice of the so-called spatial weight matrix with only limited guidance from theory in making these specifications. The choice of a SAR model over other alternatives, such as spatial Durbin (SD) or spatial lagged X (SLX) models, is often arbitrary, raising issues of potential specification error. To address such issues, this paper develops an omnibus specification test within the SAR framework that can detect general forms of misspecification including that of the spatial weight matrix, functional form and the model itself. The approach extends 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. A Monte Carlo study is conducted to study finite sample performance of the test. An empirical illustration on the performance of our test in the modelling of tax competition in Finland and Switzerland is included.

Keywords: Conditional moment test, Functional form misspecification, Misspecification test, Omnibus testing, Spatial autoregressions, Weight matrix misspecification.

JEL Classification: C21, C23

....

* Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK.
e-mail: Jungyoon.Lee@rhul.ac.uk

† Yale University, 06520 New Haven, United States. e-mail: peter.phillips@yale.edu

‡ University of Verona, via Cantarane 24, 37129, Verona, Italy.
e-mail: francesca.rossi_02@univr.it

*Phillips acknowledges research support from the NSF under Grant No. SES 18-50860 and the Kelly Fund at the University of Auckland. Lee and Rossi acknowledge research support from the ESRC Grant No. ES/P011705/1. We are indebted to Teemu Lyytikäinen and Raphaël Parchet for sharing their datasets.

1 Introduction

The past two decades have seen a remarkable surge in both the theoretical and empirical literatures on the class of spatial econometric models known as spatial autoregressions (SARs, henceforth). These models were first suggested by Cliff and Ord (1968) and have since been widely extended in directions to suit applied research. In their various specifications, SAR models are typically characterized by parsimonious and intuitive functional forms that employ exogenously assigned weight matrices intended to capture the structure of spatial dependence between units up to a finite number of unknown parameters. Much of theoretical literature has focused on parameter estimation in these models. Standard methods, such as instrumental variables/two-stage least squares (e.g. Kelejian and Prucha (1998)), Gaussian maximum likelihood/quasi-maximum likelihood estimation (e.g. Ord (1975) and Lee (2004)) and generalized methods of moments (e.g. Kelejian and Prucha (1999) and Lee (2007)) have been developed to address the endogeneities inherent in SAR specifications and extended to accommodate increasingly more complex models and data structures. At the same time, a large body of the literature has focused on the derivation of the asymptotic theory of various tests for lack of spatial correlation and/or for joint significance of the model parameters. These tests have employed common approaches such as Wald, Lagrange Multiplier or Likelihood Ratio methods in the spatial setting. Among many others, see Burridge (1980), Cliff and Ord (1981), Kelejian and Prucha (2001), Anselin (2001), Robinson (2008), Lee and Yu (2012), Martellosio (2012) and Delgado and Robinson (2015).

More general specification assessment, in addition to significance testing, is of obvious importance in this class of models, more especially in view of the extensive use of exogenously chosen weight matrices and alternative model forms. Detection of misspecification in one prespecified aspect of the model while assuming the remainder of the model is correctly chosen has often been considered in the literature. For instance Baltagi and Li (2001) offer a test for the correct specification of a (log-) linear functional form in spatial error models against the alternative of a Box-Cox transformation. Su and Qu (2016) extend the nonparametric testing procedure of Fan and Li (1996) to spatial data in order to test for correct linear functional form specification in the SAR model. Further, by means of Lagrange Multiplier statistics, Anselin (2001) developed tests to detect misspecification arising from different types of spatial error correlation. A general development of limit theory for this kind of residual-based procedure that includes tests for covariance structures in SAR models as special cases has been developed in Robinson (2008). Also, Delgado and Robinson (2015) offer a testing procedure to discriminate non-nested models for covariance structures that can accommodate spatial, spatio-temporal, or panel data structures. More recently, Gupta and Qu (2020) derive a test of correct specification of the regression functional form while allowing for cross-sectional correlation in the error term by means of series estimation of a nonparametric regression function. The Gupta and Qu (2020) approach includes the work of Su and Qu (2017) on regression specification testing as a special case.

The aforementioned testing approaches enjoy favorable large and small sample properties including good power if the practitioner has prior information about the components of the model structure that are most likely subject to misspecification. But these methods typically do not deliver a general methodology in the absence of such information. In addition, and possibly more importantly, the aforementioned research does not offer a general approach to testing the specific network dependence structure, which limits the scope for practical use in light of the common use of an exogenously chosen weight matrix. To illustrate the possible implications, consider a simple Lagrange Multiplier test to detect a spatial component that might take the form of a spatial lag of the dependent variable or a spatial error structure. In cases where the weight structure of dependence is misspecified, the practitioner might expect the test to retain correct size, but test power is likely to be adversely affected because the focus of the test is not directed at the real source of misspecification.

A more direct approach to tackle the choice of the weight matrix in spatial models has been adopted by Beenstock and Felsenstein (2012), who use the sample covariance matrix of the data to infer the network structure in a panel context. Although promising, this approach is inevitably affected by dimension and suffers from bias when the number of sample units has the same order of magnitude as the number of the time periods. Taking another promising high-dimensional approach, Lam and Souza (2015, 2020) suggest estimating the most effective weighting structure via LASSO procedures, by combining information from multiple specifications. This approach may be employed as a useful implicit test of specific weight structures.

In order to remedy concerns regarding the choice of a network weight matrix while avoiding the challenging task of estimating high-dimensional structures, a relatively narrow branch of the spatial econometric literature has focused on offering model selection procedures between competing models. Along these lines, Kelejian (2008) and Kelejian and Piras (2011, 2016) provide increasingly more general J-type tests which can be used to select among competing choices of weight matrices in SAR models with spatially correlated errors (SARAR). Kelejian's (2008) procedure has been extended in Debarsy and Ertur (2019) to allow for unknown heteroscedasticity in the error terms. A selection strategy for the correct network structure has also been suggested by Bailey et al. (2016), who employ multiple testing to deduce nullity, positivity or negativity of the elements of a weight matrix, while Liu and Prucha (2018) generalize the well-known Moran I statistic to test whether a linear combination of pre-specified weight matrices suitably describes the data within a given spatial autoregression. Even more recently, Liu and Lee (2019) offer a more general method that chooses between two specifications within/between SARAR or matrix exponential spatial specification (MESS), that can be nested or non-nested. That approach relies on a likelihood-ratio test in the spirit of Vuong (1989) and, importantly, allows both of the competing models to be misspecified under the null. The limit theory in Liu and Lee (2019) is derived under the assumptions of Near Epoch Dependence (Jenish and Prucha (2009, 2012)), which limits the scope of application to data that have a geographical interpretation and dependence that can be defined in terms of a decreasing function of distance between observations. Accordingly, it is not directly applicable when 'space' is defined according to a more general notion of economic distance (e.g. Case (1991) and Pinske et al. (2002), among others).

The goal of the present paper is to complement the above approaches by developing an omnibus test procedure that can detect quite general forms of misspecification related to the model, the weight matrix and the functional form for the SAR model. The approach we adopt is in the spirit of the Bierens (1990) conditional moment tests. The literature on consistent conditional moment tests has been widely explored starting in the 1980s (Bierens, 1982; Newey, 1985) and relying on orthogonality condition tests that date back to Ramsey (1969). Under the null hypothesis of correct specification of the regression function, the moment condition(s) holds with probability one, while consistency against general misspecification is achieved by means of a set of weighting functions that depend on some real parameter. The idea of consistent conditional moment tests in Bierens (1982) was originally developed for data that are independent and identically distributed but it has been extended to time series models in Bierens (1984, 1988), de Jong (1996) and, more recently, to non-stationary models in Kasparis (2010). Bierens (1990) suggests a particularly appealing procedure as the resulting test statistic has a standard limiting distribution under the null hypothesis and does not require randomization to achieve consistency, as opposed to Bierens (1982). In this paper we extend the Bierens (1990) test to the spatial setting, characterized by the fact that individual outcomes are influenced not only by their own individual characteristics but also by the characteristics of their neighbours. An extra challenge in the spatial model setup and limit theory is the fact that the regression function is heterogeneous across individuals.

In our development we assume a SAR structure with spatial dependence as a spatial lag since it is a significant base model of interest in the spatial literature and the kernel of many more general formulations. Our conditional moment testing approach, with individual outcomes depending on neighbour outcomes and heterogeneous regression functions, will be relevant in other settings. A primary advantage in the approach is its applicability to general 'spatial' data, where 'space' is interpreted more generally than geographic, as no reliance is placed on NED conditions to limit spatial dependence. We establish the limit distribution of our specification test under the null of correct model specification, including the form of the spatial weight matrix, and establish test consistency against general model misspecification.

Simulations are conducted to explore the finite sample behavior of the test, allowing for cases of geographic distance and random linkages in the weight matrix as well as spatial Durbin and spatial lag X formulations. The results confirm that the test has stable size properties across models and good power performance in distinguishing misspecification in the weight matrix structure and in other aspects of model formulation. The methodology is applied in an empirical study of tax competition among municipalities. The results suggest that the specification test is helpful in guiding refinement of simple SAR models to capture dependence structures in the data more satisfactorily.

The paper is organized as follows. The next section presents the model setup and main assumptions. Section 3 details the extension of the Bierens (1982, 1990) model specification work to the spatial context and discusses the formulation of relevant null and alternative hypotheses. Sections 4 and 5 report the limit theory under the null of correct specification and under a fixed generic alternative. The simulation findings are presented in Section 6. Section 7 provides the tax competition illustration using the model framework and datasets of Lyytikäinen (2012), who dealt with tax competition across Finnish municipalities and Parchet (2019), who applied a spatial analysis to tax rates across Swiss municipalities. Some conclusions and possible extensions are given in Section 8. Proofs and discussion of a case not covered by our assumptions in Section 4 are given in the Appendices.

Throughout the paper, we denote by A_{in} and $A_n^{(i)}$ the vectors formed by taking the transpose of the ith rows of a matrix A_n and its inverse A_n^{-1}, respectively, provided the inverse exists; and a_{ij} and a^{ij} are the (i,j)th elements of A and A^{-1}. The symbol $1 = 1_n$ denotes an $n \times 1$ vector of ones, $||.||$ and $||.||_\infty$ represent spectral and uniform absolute row sum norms, A' is the transpose of A, and $K > 0$ is an arbitrary finite constant whose value may change in each location. The symbol $\approx$ signifies 'approximate equality' and $\sim$ indicates 'asymptotic equivalence'.

2 Model Set-up and Assumptions

We consider a regression model of the following form

$$Y_{in} = g_{in}(\mathbb{X}_n) + \eta_{in}, \ \mathbb{E}(\eta_{in}|\mathbb{X}_n) = 0, \quad i = 1, ..., n, \tag{2.1}$$

where $\mathbb{X}_n = (X_{1n}, \cdots, X_{nn})'$ is $n \times k$ matrix of regressors of *all* sampled units, which may or may not include a column of ones, with the true conditional expectation function for the ith observation denoted by g_{in}, viz., $g_{in}(\mathbb{X}_n) = \mathbb{E}(Y_{in}|\mathbb{X}_n), \ i = 1, ..., n$.

By conditioning on the matrix $\mathbb{X}_n$, instead of on the individual vector $X_i = X_{in}$, we characterise the above model as a spatial one whereby individual outcomes are influenced not only by their own individual characteristics but also by the characteristics of their neighbours.

To allow more flexible modelling and 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-vector observation form by the system

$$Y_n = \lambda W_n Y_n + \mathbb{X}_n \beta + \epsilon_n, \tag{2.2}$$

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

$$Y_n = S_n^{-1}(\lambda)(\mathbb{X}_n \beta + \epsilon_n). \tag{2.3}$$

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

$$\begin{aligned} &Y_{in} = m_{in}(\mathbb{X}_n, \lambda, \beta) + u_{in}(\lambda), \quad \text{where} \\ &m_{in}(\mathbb{X}_n, \lambda, \beta) = S_n^{(i)}(\lambda)' \mathbb{X}_n \beta = \sum_{j=1}^{n} s_n^{ij}(\lambda) X_{jn}' \beta, \quad \text{and} \\ &u_{in}(\lambda) = \sum_{j=1}^{n} s_n^{ij}(\lambda) \epsilon_{jn}, \end{aligned} \tag{2.4}$$

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 unknown parameters of (2.2), denoted by $\theta = (\lambda, \beta')'$, can be estimated by minimizing a suitable objective function over a compact parameter space Θ under general assumptions as

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmin}} \ \mathcal{Q}(\theta). \tag{2.5}$$

The functions $g_{in}(\cdot)$ and $m_{in}(\cdot)$, the quantities in (2.1) and (2.4), as well as most random and deterministic sequences appearing in the sequel, are triangular arrays because of their dependence on n. But it is convenient to suppress the affix n for notational simplicity unless we specifically want to highlight the dependence on n. Similarly, it is convenient to do so in other cases, such as using $R(\lambda)$ in place of $R_n(\lambda)$.

Our concern in the present paper is in testing whether the regression function $m_i(\cdot)$ of (2.4) is a correct characterization of the unknown true regression function $g_i(\cdot)$ of (2.1), i.e. whether $g_i = m_i(\theta)$ for some $\theta \in \Theta$ with probability one. To provide a rigorous development 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_0^2 *and, for some* $\delta > 0$, $\mathbb{E}|\epsilon_i|^{4+\delta} \leq K$.

Assumption 2 *For* $i = 1, ..., n$ *and for all* n, X_i *is a set of iid bounded random variables in* $\mathbb{R}^k$. *For* $i, j = 1, ..., n$ *and all* n, *the elements of* X_i *are independent of* ϵ_j.

Moment existence to order exceeding 4 is required to establish the central limit theorem for quadratic forms, reported in Section 4. The condition on boundedness of X_i is retained for simplicity, but the case of unbounded X_i could be dealt with by introducing a bounded one-to-one function $\phi(X_i)$ (e.g., Bierens (1990)) and an additional trimming argument in the spirit of the discussion in Section 3. Finally, independence across X_i and ϵ_j for all (i, j) could be relaxed to strict exogeneity of $\mathbb{X}$ at expense of some modifications of the derivations in the following sections.

As it is standard in the SAR literature, we impose some conditions on W to ensure that (2.2) and (2.3) are well defined.

Assumption 3 $\lambda_0 \in \Lambda$, *where* Λ *is a closed subset in* $(-1, 1)$.

Assumption 4

(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 a sequence bounded away from zero for all* n *and* $h/n \to 0$ *as* $n \to \infty$.

Assumption 5 *For all sufficiently large* n, $\sup_{\lambda \in \Lambda} \left(||S^{-1}(\lambda)||_\infty + ||S^{-1}(\lambda)'||_\infty\right) \leq K$.

Let $g(\cdot) = (g_1(\cdot), \cdots, g_n(\cdot))'$ be the $n \times 1$ vector of individual $g_i(\cdot) = g_{in}(\cdot)$ functions and $\Omega_g = \mathbb{V}ar(g)$. Although no specific functional structure is imposed, the true conditional expectation functions $g_i(\cdot)$ are required to satisfy some continuity and dependence conditions, as follows

Assumption 6 *For* $i = 1, ..., n$ *and all* n, $g_i(\cdot)$ *are continuous functions of* $X_1, ..., X_n$ *and satisfy* $||\Omega_g||_\infty < K$.

Assumption 6 accommodates all the special cases of interest that are discussed later in Section 5. Additional conditions are imposed on the errors η_i of the true regression function in (2.1).

Assumption 7 *For all* n, η_i *is independent of* X_j *for all* $i, j = 1, ..., n$. *For* $i = 1, ..., n$, $\mathbb{E}(\eta_i|\mathbb{X}) = \mathbb{E}(\eta_i) = 0$, $\sup_{1 \leq i \leq n} \mathbb{E}(\eta_i^2) < \infty$, and $\max_{1 \leq i \leq n} \sum_{k=1}^{n} |Cov(\eta_i, \eta_k)| = O(1)$.

A natural implementation of the Bierens (1990) approach is to construct a test of model specification using a sample equivalent of the moment condition

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left((Y_i - m_i(\theta))e^{t'X_i}\right) = 0, \tag{2.6}$$

which would depend on the average covariance between the residuals $Y_i - m_i(\theta)$ and a function of the corresponding independent variable X_i. A preliminary Monte Carlo exercise[1] shows that such a test would have good size performance and satisfactory power against common sources of model misspecification, apart from those that are linked to W. As an example, if $g(X_1, ..., X_n)$ is the reduced form of the Spatial Durbin model or if the spatial autoregressive component is correct but the linear function form of $X_1, ..., X_n$ is not, a test based on (2.6) will reject the null of correct model specification with probability that increases rapidly with the sample size. However, a test based on (2.6) fails dramatically when the only source of misspecification is the choice of W, with power that is close to size even for large sample sizes.

To explore the reason for this failure a simple illustration using an omitted variable argument is helpful. Suppose that $g = (I - \lambda V)^{-1}\mathbb{X}\beta$, where V is a weight matrix satisfying the standard assumptions, but the practitioner erroneously chooses W when estimating the parameters of the model. The practitioner then effectively estimates the parameters of the augmented model

$$Y = \lambda_1 WY + \lambda_2 VY + \mathbb{X}\beta + \epsilon, \tag{2.7}$$

but with the component VY omitted. If the true network structure of the data is captured by V, then λ_1 is zero and W is irrelevant in describing the spatial process. Thus, if the full model in (2.7) is estimated, we can expect to obtain an estimate of λ_1 close to zero. Further, when VY is omitted from (2.7), we expect to obtain an estimate for λ_1 close to zero whenever the correlation between WY and VY is small, for in that case WY would not mimic the spatial effect of VY. On the other hand, we expect to obtain a non-negligible estimate of λ_1 whenever the components WY and VY display a certain degree of correlation. By contrast, since the choice of the weight structure is strictly exogenous and uncorrelated with the independent variables of the model, the estimates of the coefficients β in (2.7) are expected to be almost unbiased (with the exception of the intercept coefficient when that is present) even when VY is omitted, as the correlation between X_j for $j = 1, ..., k$, and VY is typically small. More specifically, when the true weight structure is V, but the practitioner estimates parameters in (2.7) without including VY, $\hat{\lambda}_1 \approx 0$, and, in vector form the residual vector appearing in (2.6) would be

$$Y - (I - \hat{\lambda}_1 W)^{-1}\mathbb{X}\hat{\beta} \approx \lambda_2 VY - (\hat{\beta} - \beta)\mathbb{X} + \epsilon \approx \lambda_2 VY + \epsilon, \tag{2.8}$$

[1] Details are available from the authors upon request.

which covaries little with functions of X_j, for any $j = 1, ..., n$. Hence, when the network structure is severely misspecified and the correlation between VY and WY is small, a test based on (2.6) will have almost no power. The test failure is alleviated when the misspecification of the network structure is not as severe, and WY is able to partially mimic the true spatial component. The latter may well be a common outcome in empirical work because practitioners frequently obtain evidence of spatial dependence and non-zero estimates of spatial parameters even though the choice of the weight matrix is almost certainly only a crude approximation.

In practical terms, in cases where W and V share some similarities in their structures (such as a circulant and a block diagonal matrix), the estimate of λ_1 in (2.7) may well be nonnegligible and, in turn, the power of a test based on (2.6) may be low but exceed size. On the other hand, if W and V were two independently generated spatial structures, we would expect that $\hat{\lambda}_1 \approx 0$ and test power to be close to size for all n.

These difficulties pose a challenge to formulating a straightforward extension of the Bierens test to detect misspecification in the weight matrix. As argued above, such a specification test is often of crucial interest in practical work where there is only general guidance in the formulation of the weight matrix. This provides a strong incentive to develop a refined test procedure that gives direct attention to the possibility of weight matrix misspecification in spatial autoregression.

3 Hypothesis Formulation

In view of the limitations of the standard Bierens approach, we develop a modified set of moment conditions and a new test statistic to detect general sources of misspecification in spatial models including those associated with the weight matrix. The formulation involves some additional complexity because it is necessary to supplement the Bierens moment condition (2.6) with a condition designed to assess the weight matrix specification.

For all n and $i = 1, ..., n$, let $\mathbb{1}_i(\alpha_n) \equiv \mathbb{1}(|\eta_i| \leq \alpha_n, \max_{1 \leq j \leq n} |\varepsilon_j| \leq \alpha_n)$, where $\mathbb{1}(\cdot)$ is the indicator function and α_n is a deterministic sequence such that $\alpha_n \to \infty$ as n increases. Under Assumptions 2, 5, 6 and the definition of m_i in (2.4), $g_i(\cdot)$ and $\sup_\theta m_i(\theta)$ are bounded, so for $i = 1, ..., n$ and for all sufficiently large n

$$\mathbb{1}_i(\alpha_n) = 1 \implies \mathbb{1}(|g_i + \eta_i| \leq \alpha_n) = 1 \tag{3.1}$$

and

$$\mathbb{1}_i(\alpha_n) = 1 \implies \mathbb{1}(\sup_\theta |m_i(\theta) + u_i(\lambda)| \leq \alpha_n) = 1. \tag{3.2}$$

A brief remark on the definition of $\mathbb{1}_i(\alpha_n)$ is in order here. Even though $u_i(\lambda)$ and η_i are obviously related in case of correct regression specification so that $u_i(\lambda_0) = \eta_i$ and that (for each n) there exists a sequence of values of θ (denoted as $\theta_n^{\#}$ in the sequel) so that $u_i(\lambda_n^{\#}) = g_i - m_i(\theta_n^{\#}) + \eta_i$ under misspecification, this is not necessarily true for any value of $\lambda \in \Lambda$. The indicator function is therefore formulated to include both $|\eta_i| < \alpha_n$ and $\sup_\lambda |u_i(\lambda)| < \alpha_n$ components, so that (3.2) and the argument developed in the sequel holds for any $\theta \in \Theta$.

We now define the augmented vector of moment conditions

$$M_n(\theta, t) = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} \mathbb{E}\left((Y_i - m_i(\theta))e^{t'X_i}\right) \\ \sum_{i=1}^{n} \mathbb{E}\left((Y_i - m_i(\theta))e^{t_Y(g_i + \eta_i \mathbb{1}_i(\alpha_n))} - \mathbb{E}(u_i(\lambda)e^{t_Y(m_i(\theta)+u_i(\lambda)\mathbb{1}_i(\alpha_n))}|\mathbb{X}_n)\right) \end{pmatrix} = 0, \tag{3.3}$$

where $\mathbb{1}_i(\alpha_n)$ guarantees that all moments are well defined for each $\theta \in \Theta$.

In expression (3.3), $M_n(\theta, t)$ augments the Bierens moment condition (2.6) with a condition that directs attention to the weight matrix formulation. The first element of $M_n(\theta, t)$ is the average of the standard moment condition discussed in Section 2. The second element is the average of the (centred) conditional covariances between each unit's reduced-form residual and an exponential function of the unit's dependent variable, subject to the tail trimming condition $\mathbb{1}_i(\alpha_n)$. By construction the dependent variable involves the independent variables weighted by the true networking structure which then plays a direct role in the moment condition. Each term of the second element of (3.3) is centred so that it is zero in the limit when the regression model is correctly specified. But when the weight matrix in the model is misspecified, the centering is lost because the misspecification involves the true reduced form which covaries with the exponential function of the dependent variable, as in the simple illustration leading to (2.8). We therefore expect a test based on a sample analogue of (3.3) to be more powerful against general misspecification that involves use of an inappropriate weight matrix than a simpler statistic of the Bierens type that is based on the first component only.

Let

$$\mathrm{v}_{in}(\theta) = \mathrm{v}_i(\theta) = Y_i - m_i(\theta) = g_i - m_i(\theta) + \eta_i, \tag{3.4}$$

where the second equality follows from (2.1), and

$$\upsilon_{in}(\theta, t_Y) = \upsilon_i(\theta, t_Y) = \mathrm{v}_i(\theta)e^{t_Y(g_i - m_i(\theta) + \eta_i \mathbb{1}_i(\alpha_n))} - \mathbb{E}(u_i(\lambda)e^{t_Y u_i(\lambda)\mathbb{1}_i(\alpha_n)}). \tag{3.5}$$

After standard manipulations, each term of the second component of (3.3) can be written as

$$\mathbb{E}\left(\upsilon_i(\theta, t_Y)e^{t_Y m_i(\theta)}\right), \tag{3.6}$$

outlining that m_i takes on the role of conditioning variable in the new component of our test. Having a non-zero expectation of v_{in} conditional on X_i in the first component when the moment condition (3.3) is violated is readily translated into misspecification of m_i. However, a non-zero expectation of υ_{in} conditioning on m_i calls for clarification regarding which aspects of model misspecification it implies. The exposition below aims to provide precise and intuitive one-to-one correspondences between moment conditions and almost sure equalities given in Corollary 1, which will then be used to formulate hypotheses.

Define the conditional expectations

$$d_{in}(\theta, \mathbb{X}_n) = d_i(\theta) = \mathbb{E}(\mathrm{v}_i(\theta)|\mathbb{X}_n) = g_i - m_i(\theta), \tag{3.7}$$

as $\mathbb{E}(\eta_i|X_1, ..., X_n) = 0$, and

$$\begin{aligned}
&\bar{\upsilon}_{in}(m_i(\theta), t_Y) = \bar{\upsilon}_i(m_i(\theta), t_Y) = \mathbb{E}(\upsilon_i(\theta, t_Y)|m_i(\theta))\\
=&e^{-t_Y m_i(\theta)} \left(\mathbb{E}(e^{t_Y \eta_i \mathbb{1}_i(\alpha_n)}) \left(\mathbb{E}(g_i e^{t_Y g_i}|m_i(\theta)) - m_i(\theta)\mathbb{E}(e^{t_Y g_i}|m_i(\theta))\right)\right)\\
+&e^{-t_Y m_i(\theta)} \left(\left(\mathbb{E}(\eta_i e^{t_Y \eta_i \mathbb{1}_i(\alpha_n)})\mathbb{E}(e^{t_Y g_i}|m_i(\theta)) - \mathbb{E}(u_i(\lambda)e^{t_Y u_i(\lambda)\mathbb{1}_i(\alpha_n)})e^{t_Y m_i(\theta)}\right)\right)\\
=&e^{-t_Y m_i(\theta)} \left(\mathbb{E}(e^{t_Y \eta_i \mathbb{1}_i(\alpha_n)}) \left(\mathbb{E}(g_i e^{t_Y g_i}|m_i(\theta)) - m_i(\theta)\mathbb{E}(e^{t_Y g_i}|m_i(\theta))\right)\right)\\
+&e^{-t_Y m_i(\theta)} \left(\mathbb{E}(\eta_i e^{t_Y \eta_i \mathbb{1}_i(\alpha_n)}) \left(\mathbb{E}(e^{t_Y g_i}|m_i(\theta)) - e^{t_Y m_i}\right)\right)\\
+&\left(\mathbb{E}(\eta_i e^{t_Y \eta_i \mathbb{1}_i(\alpha_n)}) - \mathbb{E}(u_i(\lambda)e^{t_Y u_i(\lambda)\mathbb{1}_i(\alpha_n)})\right).
\end{aligned} \tag{3.8}$$

From (3.8), we deduce that on the support of $(\eta_i, \varepsilon_1, ..., \varepsilon_n)$ such that $\mathbb{1}_i(\alpha_n) = 1$,

$$\mathbb{P}_{m_i}(g_i = m_i(\theta)) = 1 \quad \wedge \quad \mathbb{P}(u_i(\lambda) = \eta_i) = 1 \quad \Longrightarrow \mathbb{P}_{m_i}(\bar{\upsilon}_i(m_i(\theta), t_Y) = 0) = 1, \tag{3.9}$$

where $\mathbb{P}_{m_i}$ denotes the probability induced by $m_i(\cdot)$ only, which is in fact a particular linear combination of the random vectors $\{X_1, ..., X_n\}$, and $\mathbb{P}_{Y_i} = \mathbb{P}$ is the probability induced by Y_i. For the implication in (3.9) to hold we would only need a weaker condition on equality of $u_i(\lambda)$ and η_i in distribution. However, *a.s.* equality is needed for Corollary 2 in the sequel and so it is used here. In order to guarantee the opposite implication, we introduce the following condition.

Assumption 8 *For all sufficiently large n and for all $i = 1, ..., n$ such that*

$$\mathbb{P}_{m_i}(m_i(\theta) = g_i) < 1 \quad \vee \quad \mathbb{P}(u_i(\lambda) = \eta_i) < 1$$

hold, then

$$\mathbb{P}_{m_i}\left(\mathbb{E}\left(e^{tg_i}|m_i(\theta)\right) = e^{tm_i(\theta)}\right) < 1 \quad \vee \quad \mathbb{E}\left(\eta_i e^{t\eta_i}\right) \neq \mathbb{E}\left(u_i(\lambda)e^{tu_i(\lambda)}\right)$$

for all $t \in \mathbb{R}$ apart from a set with Lebesgue measure zero, on the support of $(\eta_i, \varepsilon_1, ..., \varepsilon_n)$ such that $\mathbb{1}_i(\alpha_n) = 1$, for any deterministic divergent sequence α_n.

Assumption 8 is required to rule out the possibility of $\mathbb{P}_{m_i}(\bar{\upsilon}_i(m_i(\theta), t_Y) = 0) = 1$ if either $\mathbb{P}_{m_i}(m_i(\theta) = g_i) < 1$ or $\mathbb{P}(u_i(\lambda) = \eta_i) < 1$. With these conditions we deduce the following proposition which elucidates what the second moment condition of (3.3) implies about model specification.

Proposition 1 *Let Assumptions 1-8 hold. For all sufficiently large n and for all $i = 1, ..., n$ we have the following equivalences*

$$\begin{aligned}
\mathbb{P}_{m_i}(d_i(\theta) = 0) = 1 \quad \wedge \quad \mathbb{P}(u_i(\lambda) = \eta_i) = 1 \quad &\Leftrightarrow \quad \mathbb{P}_{m_i}(\bar{\upsilon}_i(\theta) = 0) = 1 \\
\mathbb{P}_{m_i}(d_i(\theta) = 0) < 1 \quad \vee \quad \mathbb{P}(u_i(\lambda) = \eta_i) < 1 \quad &\Leftrightarrow \quad \mathbb{P}_{m_i}(\bar{\upsilon}_i(\theta) = 0) < 1
\end{aligned}$$

for all $\theta \in \Theta$, on the support of $(\eta_i, \varepsilon_1, ..., \varepsilon_n)$ such that $\mathbb{1}_i(\alpha_n) = 1$, for any deterministic sequence α_n.

With this framework we may now extend Lemma 1 of Bierens (1990).

Theorem 1 *Let Assumptions 1-8 hold.*

a) *For all sufficiently large n and all $i = 1, ..., n$ such that $\mathbb{P}_{X_i}(d_i(\theta) = 0) < 1$, the set of $t \in \mathbb{R}^k$ values for which $\mathrm{v}_i(\theta)$ and $e^{t'X_i}$ are orthogonal, i.e. $S = \{t \in \mathbb{R}^k : \mathbb{E}(\mathrm{v}_i(\theta)e^{t'X_i}) = 0\}$ has Lebesgue measure zero.*

b) *For all sufficiently large n and all $i = 1, ..., n$ so that $\mathbb{P}_{m_i}(\bar{\upsilon}_i(m_i, \theta) = 0) < 1$, the set of $t_Y \in \mathbb{R}$ values for which $\upsilon_i(\theta)$ and $e^{t_Y m_i(\theta)}$ are orthogonal, i.e. $S = \{t_Y \in \mathbb{R} : \mathbb{E}(\upsilon_i e^{t_Y m_i(\theta)}) = 0\}$ has Lebesgue measure zero, for all deterministic sequences α_n such that $\alpha_n \to \infty$ as $n \to \infty$,*

where $d_i(\theta) = d_{in}(\theta, \mathbb{X}_n)$, $\bar{\upsilon}_i(m_i(\theta)) = \bar{\upsilon}_{in}(m_i(\theta))$, $\mathrm{v}_i(\theta) = \mathrm{v}_{in}(\theta)$ and $\upsilon_i(\theta, t_Y) = \upsilon_{in}(\theta, t_Y)$ are defined according to (3.7), (3.8), (3.4) and (3.5).

The proof of Theorem 1 follows with minor modifications to the proof of Lemma 1 in Bierens (1990) and is reported in Appendix 1. From Theorem 1 we deduce the following confirmation of the moment conditions.

Corollary 1 *Let Assumptions 1-8 hold. For all sufficiently large n and for $i = 1, ..., n$*

$$\mathbb{E}\left(\mathrm{v}_i(\theta)e^{t'X_i}\right) = 0 \quad \forall t \in \mathbb{R}^k \text{ up to zero-measured sets} \quad \Leftrightarrow \quad \mathbb{P}_{X_i}(d_i(\theta) = 0) = 1 \qquad (3.10)$$

$$\begin{aligned}&\mathbb{E}\left(\upsilon_i(\theta, t_Y)e^{t_Y m_i(\theta)}\right) = 0 \;\; \forall t_Y \in \mathbb{R} \textit{ up to zero-measured sets}\\ &\quad\Leftrightarrow \;\; \mathbb{P}_{m_i}(d_i(\theta) = 0) = 1 \;\;\wedge\;\; \mathbb{P}(u_i(\lambda) = \eta_i) = 1 \end{aligned} \tag{3.11}$$

for all $\theta \in \Theta$ and all deterministic sequences α_n.

For all sufficiently large n, we define the set

$$J_n(\theta) = J(\theta) = \{i : \mathbb{P}_{X_i}\big(g_i = m_i(\theta)\big) < 1 \;\vee\; \mathbb{P}_{m_i}(g_i = m_i(\theta)) < 1 \;\vee\; \mathbb{P}(u_i(\lambda) = \eta_i) < 1\}, \tag{3.12}$$

and let $\mathrm{card}(J(\theta))$ denote its cardinality, which measures the extent to which the model equivalence $g_i = m_i(\theta)$ fails among the observed units. Correspondingly, in view of the results given in Theorem 1 and Corollary 1, we define the following explicit null and alternative hypotheses.

$$\mathcal{H}_0: \;\mathbb{P}_{X_i}(g_i = m_i(\theta_0)) = 1 \;\wedge\; \mathbb{P}_{m_i}(g_i = m_i(\theta_0)) = 1 \;\wedge\; \mathbb{P}(u_i(\lambda_0) = \eta_i) = 1 \tag{3.13}$$

for some $\theta_0 \in \Theta$ and for sufficiently many i such that $\mathrm{card}(J(\theta_0))/\sqrt{n} = o(1)$ *as $n \to \infty$;*

$$\mathcal{H}_1: \;\mathbb{P}_{X_i}(m_i(\theta^{\#}) = g_i) < 1 \;\vee\; \mathbb{P}_{m_i}(m_i(\theta^{\#}) = g_i) < 1 \;\vee\; \mathbb{P}(u_i(\lambda^{\#}) = \eta_i) < 1 \tag{3.14}$$

for sufficiently many $i = 1, ..., n$ such that $\mathrm{card}(J(\theta^{\#})) \sim n$ *as $n \to \infty$, where $\theta_n^{\#} = \theta^{\#} = (\lambda^{\#}, \beta^{\#\prime})'$ is the sequence of pseudo-true values that maximizes the objective function in (2.5) under $\mathcal{H}_1$.*

These hypotheses consist of multiple statements that arise from the two moment conditions used to construct the test. The first component based on $\mathbb{E}\left(\mathrm{v}_i(\theta)e^{t'X_i}\right) = 0$ from the orthogonality between X_i and η_i corresponds to the first statement in $\mathcal{H}_0$, which is typically given elsewhere in other models as $\mathbb{P}(g_i = m_i(\theta_0)) = 1$ for some θ_0. However our statement in $\mathcal{H}_0$ more precisely involves $\mathbb{P}_{X_i}(\cdot)$ rather than $\mathbb{P}(\cdot)$ and allows for distinction between the roles played by the two components in the statistic. The second component based on $\mathbb{E}\left(\upsilon_i(\theta, t_Y)e^{t_Y m_i(\theta)}\right) = 0$ in the spatial setup gives rise to the second and third statements in $\mathcal{H}_0$, which rely in turn on the equivalences provided in Corollary 1. Thus, we test for almost sure equality of g_i and $m_i(\theta_0)$ conditioning separately on X_i and m_i. In addition, it is necessary to include the almost sure equality of η_i and the SAR reduced-form error $u_i(\lambda_0)$ in $\mathcal{H}_0$. The inclusion of this equality is not surprising given that the reduced form of SAR generates by construction a particular functional structure for the errors and not just for the regression component of the model. Finally, the formulation of $\mathcal{H}_0$ requires that the number ($\mathrm{card}(J(\theta_0))$) of model equivalence failures, i.e., $g_i \neq m_i(\theta_0)$, among the observed units be of smaller order than $\sqrt{n}$ as $n \to \infty$, thereby ensuring that the behavior of the test statistic under the null is dominated by valid specifications with $g_i = m_i(\theta_0)$ rather than failures through misspecification.

As previously discussed, the test cannot detect departures in the direction of weight matrix spatial misspecification of the model when conditioning on X_i alone, i.e. by employing only the first moment condition in (3.3). On the other hand, the obvious choice of multiple conditioning variables, viz. the increasing set $\{X_1, ..., X_n\}$, does not lead to a consistent test in the spirit of Bierens (1990). Instead, Theorem 1 together with Corollary 2 in the sequel show that the new framework delivers a sound basis for testing if we condition specifically on the most relevant linear combination of $\{X_1, ..., X_n\}$ as in (3.11), that is on the known functional form $m_i(\theta)$ – the relevant linear combination of $\{X_1, ..., X_n\}$ under the null $\mathcal{H}_0$.

Our test of $\mathcal{H}_0$ against $\mathcal{H}_1$ relies on asymptotic arguments and is therefore designed to detect an increasing number of potentially misspecified (reduced form) SAR regression functions $m_i(\theta)$. In order to have a well defined limit distribution theory that reflects the null hypothesis $\mathcal{H}_0$, the number of misspecified regression functions must be small enough so as not to influence the limit theory under the null, leading to the requirement that $\mathrm{card}(J(\theta_0)) = o(\sqrt{n})$ in the definition of $\mathcal{H}_0$. On the other hand, to achieve a consistent test against any direction of violation of $\mathcal{H}_0$ the condition $\mathrm{card}(J(\theta^{\#})) \sim n$ under $\mathcal{H}_1$ is used to ensure that the number of units for which misspecification does occur (i.e., the specified function $m_i(\theta)$ is violated in the data) grows as fast as the number of units n. It seems likely that this latter condition might be weakened somewhat and the test may have good practical performance and power for some forms of misspecification, but this possibility is not pursued in the present work.

The following Corollary makes precise what the null hypothesis $\mathcal{H}_0$ in (3.13) implies about the correct generating mechanism.

Corollary 2 *Let Assumptions 1-8 hold. $\mathcal{H}_0$ in (3.13) implies model (2.4) for sufficiently many i such that $card(J(\theta_0))/\sqrt{n} = o(1)$.*

It follows that the model implied by $\mathcal{H}_0$ is SAR in (2.4) up to an error smaller than $K/\sqrt{n}$, which maintains the null limit theory developed in the next section. As is evident from the proof of Corollary 2 in the Appendix, general misspecification in the SAR regression functions would be detected even omitting $\mathbb{P}_{X_i}(g_i = m_i(\theta_0)) = 1$ in (3.13) (and thus omitting the first moment condition in (3.3)). However, inclusion of the first component of (3.3) positively impacts test power and is well suited to detect misspecification in the regressor set as well as their functional form, without having to weight them by the network transformation implied by $S^{-1}(\cdot)$.

4 Test Statistic and Limit Theory Under $\mathcal{H}_0$

In view of the theory presented in Section 3, we construct a statistic for testing the null $\mathcal{H}_0$ in (3.13) against $\mathcal{H}_1$ in (3.14) based on a sample analogue of (3.3) and using the Gaussian quasi-maximum likelihood (QML) estimator $\hat{\theta}$ of θ. Other estimators of θ may be employed with minor algebraic modifications without affecting the following results. To ensure consistency of $\hat{\theta}$ to θ_0 under $\mathcal{H}_0$ we need to impose extra conditions, such as those in Lee (2004). Let $\mathrm{eig}_i(A)$ be the ith eigenvalue of a positive semi-definite matrix A. Let c be an arbitrary, small, positive constant.

Assumption 9 *For all sufficiently large n, uniformly in $(\lambda, \beta')'$*

$$\underline{eig}\left(\frac{1}{n}\left(R(\lambda)\mathbb{X}\beta \quad \mathbb{X}\right)'\left(R(\lambda)\mathbb{X}\beta \quad \mathbb{X}\right)\right) > c > 0, \tag{4.1}$$

where $\underline{eig}(A) = \min_i \,(eig_i(A))$, $\overline{eig(A)} = \max_i \,(eig_i(A))$ for a positive semidefinite matrix A, and c is an arbitrarily small constant.

To establish the limit distribution under $\mathcal{H}_0$ Assumption 9 need only hold at $(\lambda_0, \beta_0')'$. But uniformity of the condition over $(\lambda, \beta')'$ is useful in establishing test consistency in the following section.

Let $\bar{r}_{ij} = (R(\lambda_0) + R(\lambda_0)')_{ij}/2$, $\mu_s = \mathbb{E}(\epsilon^s)$ and

$$\bar{r}^d = \frac{\sigma_0^2}{2} tr\left(\left(R(\lambda_0) + R(\lambda_0)' - \frac{2}{n}tr(R(\lambda_0)I_n)\right)^2\right).$$

To simplify notation in what follows we write $R = R(\lambda_0)$ and $S^{-1} = S^{-1}(\lambda_0)$. Let $\Omega = \Omega(\theta_0)$ be a $(k+1) \times (k+1)$ matrix partitioned as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} = \frac{1}{n\sigma_0^2}\begin{bmatrix} (\bar{r}^d + \beta_0'\mathbb{X}'R'R\mathbb{X}\beta_0) & \beta_0'\mathbb{X}'R'\mathbb{X} \\ \mathbb{X}'R\mathbb{X}\beta_0 & \mathbb{X}'\mathbb{X} \end{bmatrix}. \tag{4.2}$$

Under Assumption 9, $\underline{\mathrm{eig}}(\Omega) > c > 0$, so the inverse Ω^{-1} exists for all sufficiently large n. Let ω^{ij} be the conformable (i, j)-th block of Ω^{-1} for $i, j = 1, 2$. Let $\omega^{(1)}$ be the $(k+1) \times 1$ vector defined as the transpose of the first row of Ω^{-1}. Now define

$$Q = Q(\theta_0) = S^{-1}\left(I_n - \frac{1}{n\sigma_0^2}\left(R\mathbb{X}\beta_0 \quad \mathbb{X}\right)\Omega^{-1}\begin{pmatrix}\beta_0'\mathbb{X}'R' \\ \mathbb{X}'\end{pmatrix}\right). \tag{4.3}$$

Under Assumption 5 and by Lemma 1 (reported in Appendix 2), $||Q||_\infty + ||Q'||_\infty \leq K$. Set $\bar{Y} = \sum_i^n Y_i/n$, the $1 \times n$ vector of the column averages of S^{-1} as $\bar{S}^{-1\prime} = \sum_{i=1}^n S^{(i)\prime}/n$, and the column-demeaned version of S^{-1} as

$$S^d = S^{-1} - 1\bar{S}^{-1'}. \tag{4.4}$$

Let $\underline{e}(t) = (e^{t'X_1}, ..., e^{t'X_n})'$, $f(t) = (\underline{e}(t)', 1')$, the $2 \times n$ matrix

$$\Psi(t, t_Y) = \Psi(t, t_Y, \lambda_0, \beta_0, \mathbb{X}) = f(t)'Q, \tag{4.5}$$

and the 2×1 vector

$$\psi(t, t_Y) = \psi(t, t_Y, \lambda_0, \beta_0, \mathbb{X}) = -\frac{1}{n\sigma_0^2} f(t)' S^{-1} \begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix} \omega^{(1)}. \tag{4.6}$$

Using this notation we indicate the estimated counterparts (evaluated at $\hat{\theta}$) of the previously defined quantities by $\hat{(\cdot)}$. As with previous notation let $(\Psi_i')'$ be the $i-$th row of $\Psi(t, t_Y)'$.

We now proceed to derive a sample analogue of (3.3) that can be used to form a test statistic. The goal is to develop a statistic that has a standard pivotal limit distribution pointwise in (t, t_Y). The following steps assist by simplifying the limit behavior of the statistic. We introduce a deterministic, positive sequence p_n satisfying the conditions

$$p_n \to \infty, \quad \frac{p_n}{n} = o(1) \ \text{ and } \ \frac{\sqrt{n}}{p_n^2} = o(1) \quad \text{as} \ \ n \to \infty, \tag{4.7}$$

which enable a formal Taylor expansion of the exponential function under some additional technical conditions reported in the proof of the Theorem 2 below. We then derive the sample equivalent of a centered sequence based on the leading terms of this expansion. The resulting sample analogue of the vector of moment conditions in (3.3) has the following explicit form

$$M_n(\hat{\theta}, t, t_Y) = M(\hat{\theta}, t, t_Y) = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} \big(Y_i - m_i(\hat{\theta})\big) e^{t'X_i} \\ \sum_{i=1}^{n} \big(Y_i - m_i(\hat{\theta})\big) e^{t_Y \frac{Y_i - \bar{Y}}{p_n}} - \frac{t_Y \hat{\sigma}^2}{p_n} tr(\hat{S}^{d'} \hat{Q}) \end{pmatrix}, \tag{4.8}$$

which leads to a statistic for testing $\mathcal{H}_0$ in (3.13) against $\mathcal{H}_1$ in (3.14) based on the quadratic form

$$\hat{T}(t, t_Y) = nM(\hat{\theta}, t, t_Y)'\hat{A}^{-1}(t, t_Y) M(\hat{\theta}, t, t_Y), \tag{4.9}$$

where $\hat{A}(t, t_Y)$ is a consistent estimate of $A(t, t_Y) = \lim_{n\to\infty} \mathbb{V}ar(\sqrt{n}M(\hat{\theta}, t, t_Y))$. To ensure pointwise existence and non-singularity of $A(\hat{\theta}, t, t_Y)$, as well as existence of $M(\hat{\theta}, t, t_Y)$, we impose the following conditions.

Assumption 10 *Conditionally on* $\mathbb{X}$, the limit $\lim_{n\to\infty} n^{-1}\Psi(t,t_Y)\Psi(t,t_Y)'$ *exists pointwise in* $(t,t_Y)'$ *and a.s. as* $n\to\infty$*, and is positive definite.*
Assumption 11 *Conditionally on* $\mathbb{X}$*, the limits*

$$\lim_{n\to\infty}\frac{\psi(t,t_Y)\psi(t,t_Y)'}{n}tr\left((R+R')^2\right),\ \lim_{n\to\infty}\frac{1}{n}\psi(t,t_Y)\sum_{i=1}^{n}\bar{r}_{ii}(\Psi_i(t,t_Y)')',\ \lim_{n\to\infty}\frac{\psi(t,t_Y)\psi(t,t_Y)'}{n}\sum_{i=1}^{n}\bar{r}_{ii}^2$$

exist pointwise in $(t,t_Y)'$ *and a.s.* as $n\to\infty$.

In general, Assumption 10 holds as long as $n\to\infty$, rank$(Q)\sim n$ and W does not have constant column sums. More specifically, since S^{-1} is non singular under Assumptions 2 and 3, Assumption 10 requires full rank in the limit of

$$\left(I_n-\frac{1}{n\sigma_0^2}\begin{pmatrix}R\mathbb{X}\beta_0 & \mathbb{X}\end{pmatrix}\Omega^{-1}\begin{pmatrix}\beta_0'\mathbb{X}'R'\\ \mathbb{X}'\end{pmatrix}\right),$$

which holds when the number of regressors k is finite (or grows slower than n). Assumption 10 is violated when W has constant column sums, which amounts to each individual having the same magnitude of influence on others overall. In such a case, the variance matrix $A(t,t_Y)$ of Theorem 2 below suffers from singularity and one cannot carry out inference based on the following theorem. The case where W has constant column sums, although unnatural in practical applications, needs to be studied separately and is analyzed and discussed in Appendix 3.

Assumption R *Let* p_n *and* α_n *be deterministic, positive sequences satisfying (4.7),* $\alpha_n\to\infty$ *as* $n\to\infty$*, and*

$$\frac{\alpha_n}{p_n}\to 0,\qquad \frac{n}{\alpha_n^{4+\delta}}\to 0,\qquad \frac{n^{3/2}}{p_n\alpha_n^{4+\delta}}\to 0 \tag{4.10}$$

as $n\to\infty$*, where* $\delta>0$ *is determined by Assumption 1.*

Assumption R is a technical condition on relative expansion rates among the sequences α_n and p_n, as $n\to\infty$. The relative rates among p_n, α_n and n depend also on the distributional assumption in Assumption 1, i.e. on the positive parameter δ. For instance, if ϵ_i, for $i=1,...,n$, are distributed as either $\mathcal{N}(0,\sigma^2)$ or as t_5 (e.g. as the two extreme cases compatible with Assumption 1), the choice of $p_n=n^{1/3}$ (adopted in the simulation exercise) and $\alpha_n=n^{1/4}$ is acceptable since (4.7) and (4.10) are satisfied as long as $\delta>2/3$. The relative rates of p_n, α_n and n on one hand, and δ implied by Assumption 1 on the other, determine the error of the approximation entailed by the central limit theorem. On the other hand, a slow-diverging p_n typically leads to higher power, since it helps to assure relevance to the second component of (4.8) via substantial co-variation between the residuals and the exponential term $t_Y\frac{Y_i-\bar{Y}}{p_n}$.

The optimal choice of p_n requires an analysis of local power, which exceeds the scope of the present paper and will be addressed in separate work.

The following result provides asymptotics that lead to the null limit distribution of the specification test suggested in (4.9).

Theorem 2 *Let Assumptions 1-5, 9-11 and R hold. Let p_n be a non-negative sequence satisfying (4.7). Under $\mathcal{H}_0$ in (3.13), as $n \to \infty$*

$$\sqrt{n}M(\hat{\theta}, t, t_Y) \to_d \mathcal{N}\left(0, A(t, t_Y)\right), \tag{4.11}$$

pointwise in $(t, t_Y)'$, conditionally on $\mathbb{X}$, where the standardizing variance-covariance matrix of $\sqrt{n}M(\hat{\theta}, t, t_Y)$ is given by $A(t, t_Y) = \lim_{n\to\infty} A_n(t, t_Y)$, with

$$\begin{aligned} A_n(t, t_Y) =& \frac{\sigma_0^2}{n}\Psi(t, t_Y)\Psi(t, t_Y)' + \frac{\sigma_0^4 tr\left((R+R')^2\right)}{2n}\psi(t, t_Y)\psi(t, t_Y)' + \frac{2\mu^{(3)}}{n}\psi(t, t_Y)\sum_{i=1}^{n}\bar{r}_{ii}(\Psi_i(t, t_Y)')' \\ &+ \frac{(\mu^{(4)} - 3\sigma_0^4)}{n}\psi(t, t_Y)\psi(t, t_Y)'\sum_{i=1}^{n}\bar{r}_{ii}^2. \end{aligned} \tag{4.12}$$

The proof of Theorem 2 is reported in Appendix 1.

The matrix $A(t, t_Y)$ exists pointwise in $(t, t_Y)'$ a.s. under Assumptions 10 and 11 and is non singular under Assumptions 3, 4, 9 and 10. Since (A.3.8) holds for every realisation of $\mathbb{X}$, as long as $A(t, t_Y)$ exists pointwise in (t, t_Y) a.s., Theorem 2 also holds unconditionally, giving the unconditional distribution of the statistic with $A(t, t_Y) = \text{plim}_{n\to\infty} A_n(t, t_Y)$. To form the test statistic defined in (4.9), $A(t, t_Y)$ is replaced by the consistent estimate $\hat{A}(t, t_Y)$ obtained by replacing the unknown parameters λ, β, σ^2, μ_3 and μ_4 with their sample versions based on consistent QML estimates $\hat{\theta} = (\hat{\lambda}, \hat{\beta}')'$ and corresponding residuals. From Theorem 2 it follows directly that the test statistic

$$\hat{T}(t, t_Y) \to_d \chi_2^2 \tag{4.13}$$

pointwise in $(t, t_Y)'$ as $n \to \infty$. Finite sample size and power performance of this test are reported in Section 6.

5 Behavior of $\hat{T}(t, t_Y)$ under misspecification

This section explores the behaviour of the test statistic $\hat{T}(t, t_Y)$ under $\mathcal{H}_1$. In order to allow for a general misspecification structure that allows for a generic functional form of the true conditional expectation function $g_{in}(\cdot) = g_i(\cdot)$ for each n we need to impose some high-level Assumptions. These Assumptions can be made more primitive if we are willing to impose more structure on $g_i(\cdot)$, e.g. in case $g_i(\cdot)$ is assumed to display an additive structure in X_1,...,X_n. Let $z_n(\mathbb{X}, t) = z(\mathbb{X}, t) = (z_{1n}(\mathbb{X}, t), \cdots, z_{nn}(\mathbb{X}, t))'$ be the $n \times 1$ vector whose components are the individual $z_{in}(\mathbb{X}, t) = z_i(\mathbb{X}, t) = g_i(\mathbb{X})e^{t'X_i}$ functions. Also, let $\Omega_z(t) = \mathbb{V}ar(z(\mathbb{X}, t))$. We need to integrate the weak dependence condition reported in Assumption 6 with an additional condition.

Assumption 6 (b) *For all $t \in \mathbb{R}^k$ apart from a zero-measured set, $||\Omega_z(t)||_\infty < K$.*

We report below some popular examples of functional structures for $g(\cdot)$, which are often erroneously misspecified and/or simplified by practitioners to the standard SAR in (2.3) with network structure W.

1. The true weight matrix structure is given by V and the practitioner uses W in estimation of the model, i.e. W is misspecified. Thus,

$$g(\mathbb{X}) = (I - \lambda_0 V)^{-1}\mathbb{X}\beta_0.$$

2. The weight matrix W is correctly specified, but the exogenous component of the regression is non-linear in $X_1, ..., X_n$ and/or in the parameters $\beta_1, ..., \beta_k$, so that

$$g(\mathbb{X}) = (I - \lambda_0 W)^{-1}\rho(\mathbb{X}, \beta_0), \text{ for some function } \rho.$$

3. The data generating process is a Spatial Durbin (SD) model with weight matrices W_1, W_2, so that

$$g(\mathbb{X}) = (I - \lambda_0 W_1)^{-1}\mathbb{X}\beta_0 + (I - \lambda_0 W_1)^{-1}W_2\mathbb{X}\gamma_0 \tag{5.1}$$

   where γ_0 is a $k \times 1$ vector of parameters.

4. The endogenous spatial lag is irrelevant, and thus the data generating process is a spatial lagged X (SLX) model, so that

$$g(\mathbb{X}) = \mathbb{X}\beta_0 + W\mathbb{X}\gamma_0. \tag{5.2}$$

All four cases above can be represented by an additive functional form specification $g_i^{add}(\mathbb{X}) = \sum_{j=1}^n (a_{1ij}\rho_1(X_j) + a_{2ij}\rho_2(X_j))$. Assumption 6(b) can be shown to hold for this additive $g^{add}(\mathbb{X})$ if the $n \times n$ matrices $A_1 = A_{1n} = (a_{1ij})$, $A_2 = A_{2n} = (a_{2ij})$ satisfy $||A_1||_\infty + ||A_1'||_\infty + ||A_2||_\infty + ||A_2'||_\infty \leq K$ for all sufficiently large n, and the functions $\rho_1(\cdot), \rho_2(\cdot) : \mathbb{R}^k \Rightarrow \mathbb{R}$ satisfy $\mathbb{E}\rho_1^4(X_1) + \mathbb{E}\rho_2^4(X_1) < \infty$. In the four cases of misspecification given above these conditions are implied by Assumptions 2, 4 and 5.

To establish consistency of the test based on (4.8) we need to prescribe the behavior of the estimator $\hat{\theta}$ under $\mathcal{H}_1$, which is assured by the following high level condition.

Assumption 12 *There exists a sequence of deterministic vectors $\theta^\sharp = \theta_n^\sharp$ of order $O(1)$ such that $\hat{\theta} - \theta^\sharp = o_p(1)$ under $\mathcal{H}_1$.*

In line with the previous section, $\theta^\sharp$ can be interpreted as the (pseudo-true) value that maximises the (misspecified) pseudo log-likelihood function under $\mathcal{H}_1$ in (3.14) and thus $\hat{\theta}$ is the QMLE of $\theta^\sharp$ under $\mathcal{H}_1$. Under $\mathcal{H}_0$, $\theta^\sharp = \theta_0$. Proposition 2 in Appendix 2 shows that, under some standard regularity conditions, $\hat{\lambda} - \lambda^\sharp = o_p(1)$ and therefore $\hat{\beta} - \beta^\# = o_p(1)$ and $\hat{\sigma}^2 - \sigma^{\sharp 2} = o_p(1)$ as $n \to \infty$, where $\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'S(\hat{\lambda})Y$ and $\beta^\sharp = \text{plim}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'S(\hat{\lambda})Y$, while $\hat{\sigma}^2 = Y'S(\hat{\lambda})'S(\hat{\lambda})Y/n$ and $\sigma^{\sharp 2} = \text{plim}Y'S(\hat{\lambda})'S(\hat{\lambda})Y/n$, as $n \to \infty$.

From (3.14) and from Theorem 1, for (almost) all $t \in \mathbb{R}^k$ and $t_Y \in \mathbb{R}$, either

$$\mathbb{E}\left((Y_i - m_i(\theta^\#)e^{t'X_i})\right) \neq 0$$

or

$$\mathbb{E}\left((Y_i - m_i(\theta^\#)e^{t_Y Y_i}) - \mathbb{E}\left(u_i(\lambda^\sharp)e^{t_Y(m_i(\theta^\sharp)+u_i(\lambda^\sharp))}\right)\right) \neq 0$$

for sufficiently many i such that $\text{card}(J(\theta^\#)) \sim n$ as $n \to \infty$. However, the sample statistic in (4.8) considers the average across units of a sample analogue of expectations. Therefore, we need to rule out the case in which individual misspecifications in the regression functions offset each other (e.g. in presence of an unlikely systematic symmetry in the mispecification form and direction), so that the average amount of misspecification is not negligible in the limit. A similar exclusion was used and discussed in Bierens (1984), where nonstationarity in the time series setting may lead to a regression function that varies across time. The following condition achieves this objective in the spatial setting.

Assumption 13 *As $n \to \infty$, in the setting of Theorem 1a)*

$$\lim_{n\to\infty} \frac{1}{n}\left|\sum_{i=1}^{n} \mathbb{E}\left(\left(g_i(\mathbb{X}) - m_i(\mathbb{X}, \theta^\sharp)\right) e^{t'X_i}\right)\right| = \kappa(t) > 0, \tag{5.3}$$

with $\kappa(\cdot)$ being a generic positive function, pointwise in $t \in \mathbb{R}^k$, or in the setting of Theorem 1b)

$$\lim_{n\to\infty} \frac{1}{n}\left|\sum_{i=1}^{n} \mathbb{E}\left(g_i(\mathbb{X}) - m_i(\mathbb{X}, \theta^\sharp)\right)\right| > 0 \tag{5.4}$$

Additionally, we assure non-singularity in the limit of $\hat{A}_n(t, t_Y)$ in (4.9) under $\mathcal{H}_1$ by modifying Assumption 10 as follows.

Assumption 10' *Conditionally on* $\mathbb{X}$*,* $\lim_{n\to\infty} A_n(t, t_Y)$ *exists and is positive definite uniformly in* θ*, pointwise in* (t, t_Y) *and a.s., where* $A_n(t, t_Y)$ *is defined in (4.12).*

Under these conditions we have test consistency.

Theorem 3. *Under* $\mathcal{H}_1$ *in (3.14), and Assumptions 2-5, 6, 6b), 7,8, 9,10' and 12-13, for all* $c > 0$,

$$\mathbb{P}r\Big(\hat{T}(t, t_Y) > c\Big) \to 1 \quad \text{as} \quad n \to \infty,$$

pointwise in (t, t_Y)*.*

6 Simulations

We report the results of a Monte Carlo experiment to examine the finite sample performance of tests for model misspecification based on the $\hat{T}(t, t_Y)$ statistic in (4.9), exploring both size and power. We generate data from the SAR specification in (2.3), with an intercept and two regressors that are *iid* random variables $X_{id} \sim Unif(0, 4)$, $d = 1, 2$, $\epsilon_i \sim N(0, 1)$, for $i = 1, ..., n$, with parameter setting $\beta = (1, 1, 1)'$, $\lambda = 0.4$, and sample sizes $n \in \{100, 200, 300, 400, 500, 600, 700\}$. Two different weight matrices are used:

1) Exponential distance weights, i.e. $w_{ij} = exp(-|\ell_i - \ell_j|)1(|\ell_i - \ell_j| < \log n)$ where ℓ_i is location of i along the interval $[0, n]$ which is generated from $Unif[0, n]$.

2) W is randomly generated as an $n \times n$ matrix of zeros and ones, where the number of "ones" is restricted at 10% of the total number of elements in W.

These weight structures are empirically motivated as they mimic a distance-based matrix generated from real data and a structure based on a contiguity criterion among units. Both matrices are normalized by their respective spectral norm. We generate each matrix once for each n and we keep them fixed across 1000 replications and across different experimental scenarios.

It is straightforward to verify numerically that under both structures 1) and 2) satisfy Assumption 10 for each n. The choice of p_n and t_Y drives the trade-off between size and power for small n but is less important for test performance as n increases. The choice of t does not seem to have an impact on the performance of the test. We set $p_n = n^{1/3}$, $t = (1.5, 1.5, 1.5)'$ and $t_Y = 0.4$. Also, similar to Bierens (1990), we replace the exponential function in the first component of $M(\hat{\theta}, t, t_Y)$ in (4.8) with $t'\text{arctg}(X_i - \bar{X})$ for each $i = 1, ..., n$, where $\text{arctg}(X_i - \bar{X}) = (\text{arctg}(X_{i1} - \bar{X}_1), ..., \text{arctg}(X_{ik} - \bar{X}_k))'$ and $\bar{X}_j$ denotes sample mean for $j = 1, ..., k$. Given the support of $\mathbb{X}$ in this simulation exercise, the $\text{arctg}(\cdot)$ contribution turns out to be virtually irrelevant.

We first examine the performance of the test statistic in (4.9) under $\mathcal{H}_0$ in (3.13), and report in Table 1 empirical size for nominal significance levels $s = 0.1, 0.05, 0.01$ and both weight matrix models 1) and 2). For both matrices there is a slight size distortion for small n, but this quickly improves for $n > 300$. Overall, the size performance is very satisfactory.

Table 1: Empirical size of the test of $\mathcal{H}_0$ in (3.13) based on $\hat{T}(t, t_Y)$ in (4.9) for nominal significance levels $s \in \{10\%, 5\%, 1\%\}$ and with W chosen as in 1) and 2).

	1)			2)		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.149	0.081	0.018	0.124	0.056	0.010
300	0.090	0.037	0.007	0.106	0.049	0.006
400	0.116	0.055	0.009	0.103	0.065	0.015
500	0.080	0.030	0.005	0.097	0.052	0.016
600	0.082	0.039	0.006	0.110	0.052	0.011
700	0.083	0.040	0.006	0.108	0.058	0.006

The empirical power of the test $\hat{T}(t, t_Y)$ was explored in several experiments covering different models, significance levels, and sample sizes. The first scenario aims to show test performance under functional form misspecification. In place of a linear function, the true spatial regression model is assumed to be

$$Y_i = \lambda \sum_{j=1}^{n} w_{ij} Y_j + \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \frac{1}{2} X_{i1}^2 + \epsilon_i, \qquad i = 1, ..., n, \tag{6.1}$$

and the misspecified linear SAR model with no quadratic term was estimated. Again, we set $\lambda = 0.4$ and $(\beta_0, \beta_1, \beta_2)' = (1, 1, 1)'$. Test power is reported in Table 2 and is evidently close to unity for all sample sizes. Table 2 reports results for the weight matrix model 1). Results for W in model 2) were similar and are not reported.

Table 2: Empirical power of the test of $\mathcal{H}_0$ in (3.13) against $\mathcal{H}_1$ in (3.14) when the true model is (6.1) with nominal significance levels $s \in \{10\%, 5\%, 1\%\}$ and W chosen as in 1).

$n\backslash s$	10%	5%	1%
200	0.996	0.990	0.967
300	0.998	0.996	0.980
400	0.998	0.995	0.981
500	1.000	0.997	0.986
600	0.999	0.999	0.991
700	1.000	1.000	1.000

To address weight matrix misspecification, the following two models were considered:

a) Both true and misspecified matrices are generated as in 1) but with two independent sets of locations;

b) The true matrix is 2) but the practitioner erroneously estimates parameters in (2.3) using W as in 1).

For both these scenarios the functional specification of the model is (2.3), but the practitioner selects the wrong weight matrix structure. The results are reported in Table 3. For both settings a) and b) the reported empirical power is excellent even for $n = 200$. For both scenarios and all sample sizes the power exceeds 0.90, suggesting a highly satisfactory performance of our test for detecting misspecification of the weight matrix, owing to the second component of (4.8).

Table 3: Empirical power of the test of $\mathcal{H}_0$ in (3.13) against $\mathcal{H}_1$ in (3.14) under scenarios a) and b), with nominal significance levels $s \in \{10\%, 5\%, 1\%\}$.

	a)			b)		
$n \backslash s$	10%	5%	1%	10%	5%	1%
200	0.984	0.979	0.965	0.969	0.964	0.954
300	0.980	0.971	0.951	0.950	0.937	0.915
400	0.998	0.997	0.994	0.962	0.960	0.939
500	0.978	0.967	0.951	0.957	0.949	0.930
600	0.971	0.958	0.925	0.963	0.957	0.936
700	0.985	0.977	0.962	0.961	0.950	0.935

Finally, we consider test power against misspecification of the model itself by generating data based on the SD and SLX models (defined in (5.1) and (5.2)), with parameter values $\beta = (1,1,1)'$, $\lambda = 0.4$ and $\gamma = (1,1)'$ in (5.1), and $\beta = (1,1,1)'$, $\lambda = 0.4$ and $\gamma = (1.5, 1.5)'$ for the parameters in (5.2). The settings for γ are two dimensional vectors as the spatial lag of the intercept is not included. In both cases the same exponential distance weight described in 1) is used for the true and misspecified models. Results reported in Table 4 show that empirical power is close to unity even for n as small as 200 when the true model is SD. When the true model is SLX, empirical power is not so high for smaller values of n, but power improves quickly with sample size.

7 Empirical Application

Investigating the possible existence and nature of interaction between neighboring government tax-setting decisions is a question of much importance at both national and international levels. Many countries have witnessed a common trend of decreasing corporate tax rates over

Table 4: Monte Carlo power of the test of $\mathcal{H}_0$ in (3.13) against $\mathcal{H}_1$ in (3.14) when the true models are SD in (5.1) and SLX in (5.2), with nominal significance level s.

model	SD			SLX		
$n \backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
200	0.998	0.997	0.991	0.583	0.405	0.148
300	0.998	0.995	0.981	0.686	0.528	0.222
400	1.000	1.000	1.000	0.951	0.909	0.698
500	0.999	0.999	0.993	0.969	0.935	0.759
600	1.000	1.000	0.998	0.977	0.949	0.811
700	1.000	1.000	0.999	0.974	0.933	0.781

recent decades, which has been typically attributed to competition between neighbouring governments in their attempts to attract mobile business ventures. This phenomenon has generated policy debates on the desirability of intervention to curb tax competition between local and national governments. Chirinko and Wilson (2017) provide some recent examples in the US and EU. Spatial econometric modelling has been widely applied to investigate the presence of such fiscal interaction. Empirical results have frequently found evidence of positive dependence in neighbouring government tax rates; see Allers and Elhost (2005) and the references therein for an extensive list of empirical papers and results. Findings in these studies broadly support the commonly held view that competition for mobile tax bases has led to a harmful "race to the bottom" in tax rates and subsequent under-provision of public goods.

Some recent empirical papers, concerned by possible endogeneities and model misspecification in previous work, have applied alternative estimation strategies for the spatial interaction in tax rates, aiming to mitigate the effects of endogeneity due to misspecification and present findings that contrast with the earlier literature. Two papers of particular interest are Lyytikäinen (2012) and Parchet (2019), who used policy-based instrumental variables (IV) to estimate the spatial autoregressive parameter in SAR models. They found this parameter to be insignificant and negative significant, respectively, and therefore presented evidence that contrasts with the preceeding iterature (e.g. Allers and Elhorst (2005)). Both Lyytikäinen (2012) and Parchet (2019) additionally report spatial parameter estimates based on conventional methods (such as QMLE) that are positive and highly significant in their data, the contradiction suggesting that caution should be exercised in accepting the findings of previous work showing positive spatial dependence in neighbouring government tax rates. The common ground that casts doubts on the reliability of estimates of the spatial parameters obtained with standard methods is the fact that it is unlikely that the fitted SAR model is correctly specified in practice and that the resulting residual spatial correlation in the errors may result in regressor endogeneity and biased findings. Neither Lyytikäinen (2012) nor Parchet (2019) consider explicitly the problem of misspecification of W. But both articles stress that standard techniques are likely to fail to deliver credible inference if the SAR models are not correctly specified.

The following sub-sections present empirical applications of our specification test to data from these two papers (Lyytikäinen, 2012, and Parchet, 2019) with the aim of assessing the suitability of SAR specifications in analyzing the tax competition data. We find that careful consideration of model specification, similar to that obtained from policy-based IV estimation, helps to mitigate significantly the disparity in the conclusions drawn from the QML and policy-based IV estimators. These findings highlight the usefulness of specification testing. The test procedure developed in the present paper may therefore provide a valid starting point towards developing a suitable SAR specification when alternative models and/or estimation techniques (such as policy-based IV) are not immediately available in practical work to deal with potential endogeneities induced by misspecification.

7.1 Municipality-level tax setting in Finland

Finland's municipalities have autonomy to set their own property tax rates within limits set by the central government. In order to investigate the nature of possible inter-municipality interaction in the determination of Finnish property tax rates, Lyytikäinen (2012) used a SAR model with fixed effects such that

$$t_{it} = \lambda \sum_{j=1}^{n} w_{ij} t_{jt} + X_{it}' \beta + \mu_i + \tau_t + \epsilon_{it} \tag{7.1}$$

where t_{it} denotes either municipality i's general property tax rates or residential building tax rates in year t, the regressors X_{it} include the municipality's socio-economic attributes and μ_i and τ_t are municipality and year fixed effects, respectively. Table 5 reports the variables contained in X_{it}. We refer to Lyytikäinen (2012) for a detailed description of the data and setting.

Table 5: List of variables in X_{it} of model (7.1).

per capita income	per capita grants	unemployment rate
percent of Age 0-16	percent of Age 61-75	percent of Age 75+

In order to alleviate possible endogeneity sources arising from unobservable time-invariant characteristics, Lyytikäinen (2012) focused on one-year differenced data, with $\Delta t_i = t_{i,2000} - t_{i,1999}$ and $\Delta X_i = X_{i,2000} - X_{i,1999}$, where year 2000 coincides with a policy intervention that raised the common statutory lower limit to the property tax rates, and i indexes municipalities that range from 1 to 141. Lyytikäinen used this exogenous policy change to construct a suitable instrument and estimate parameters of the model

$$\Delta t_i = \lambda \sum_{j=1}^{n} w_{ij} \Delta t_j + \Delta X_i' \beta + \gamma_0 + \gamma_1 P_i + \gamma_2 M_i + \Delta \epsilon_i, \quad i = 1, ..., 141, \tag{7.2}$$

where $\Delta\epsilon_i = \epsilon_{i,2000} - \epsilon_{i,1999}$, P_i is dummy variable indicating whether the 1999 tax rate level for municipality i was below the new lower limit imposed in 2000, and M_i indicates the magnitude of the imposed increase for municipality i. P_i and M_i were included to ensure exogeneity of the instrument being used. He found the spatial AR parameter λ to be insignificant for both sets of regressions with either general property tax rate or residential building tax rate, and hence concluded the absence of substantial tax competition between municipalities in Finland.

To complement this analysis, we consider the following two variants of model (7.1) for year 2000:

$$t_i = \lambda \sum_{j=1}^{n} w_{ij} t_j + X_i' \beta + \gamma_0 + \gamma_1 P_i + \gamma_2 M_i + \epsilon_i, \tag{7.3}$$

and

$$t_i = \lambda \sum_{j=1}^{n} w_{ij} t_j + X_i' \beta + \gamma_0 + \gamma_1 P_i + \gamma_2 M_i + \gamma_3' D_i + \epsilon_i, \tag{7.4}$$

where D_i is a vector of county dummies for municipality $i:$ there are 19 counties in our data and model (7.4) includes county-specific controls to partially account for the possibility of omitted variables at the county level. As in Lyytikäinen (2012) we adopt a contiguity matrix with $w_{ij} = 1$ if municipalities i and j share a border and zero otherwise, and apply a row normalising transformation to obtain W. We stress that we need to control for the direct impact of the policy on municipality i (in addition to the socio-economic variables of X_{it}) when using data from post-policy intervention to avoid spuriously inflated estimates of λ as spatial correlation in tax rates means that a municipality whose neighbours are affected by the policy is also likely to experience imposed increase in tax rate.

As in our simulation design, we set $p_n = n^{1/3}$. We calibrate the choice of the vector t and the scalar t_Y so that the modulus of the magnitude of exponentials appearing in (4.8) matches roughly with that of our simulation set up. This strategy results in choosing t as a vector with entries equal to 0.1 and $t_Y = 0.03$. Further, in the first component of $M(\hat{\theta}, t, t_Y)$ of (4.8), we consider $\exp(t' \text{arctg}(X_i - \bar{X}))$, as discussed in the simulation section, rather than $\exp(t'X_i)$, for $i = 1, ..., n$.

Table 6 reports our estimates of λ in models (7.3) and (7.4), for both the general property tax rate (left panel) and the residential building tax rate (right panel). Our test does not reject any of the two specifications in (7.3) and (7.4), for either choice of the dependent variable. The QML estimate for λ is significant for general property tax rate but not for the residential building tax. We observe that including the county dummies somewhat reduces the value of the estimate of λ for both tax rates although it does not affect its statistical significance.

Table 6: Left panel: columns (1) and (2) report QML estimates of λ and their $t-$statistics (in brackets), and the value of the test $\hat{T}(t, t_Y)$ in (4.9) for model (7.3) and (7.4), respectively, with the general property tax rate as dependent variable. Right panel: columns (3) and (4) report results for model (7.3) and (7.4), respectively, with the residential business tax rate as dependent variable. Row-normalized weight matrices are used.
* $p - value < 0.1$; ** $p - value < 0.05$; *** $p - value < 0.01$.

	general property tax		residential building tax	
	(1)	(2)	(3)	(4)
$\hat{\lambda}$	0.2909^{***} (6.0198)	0.2691^{***} (5.2782)	0.0985 (1.5977)	−0.0342 (−0.5151)
$\hat{T}(t, t_Y)$	3.4041	3.1006	0.7912	1.0108

Since municipality fixed effects are not included in models (7.3) and (7.4), for comparability with the policy-based IV estimator in Lyytikäinen (2012), we also consider the model based on differenced data, i.e. model (7.2) with differences taken between 2000 and 1999 (i.e. the year of the policy change and the year before), and model (7.2) with quantities re-defined as $\Delta t_i = t_{i,2001} - t_{i,2000}$ and $\Delta X_i = X_{i,2001} - X_{i,2000}$ and $\Delta\epsilon_i = \epsilon_{i,2001} - \epsilon_{i,2000}$, and the set of X_{it} defined in Table 5. Results for both general and residential building property tax rates are reported in columns 1-4 of Table 7, where columns 1 and 2 contain results with differences calculated between 2001 and 2000 and between 2000 and 1999, respectively, for general property tax rates, and columns 3 and 4 contain corresponding results for residential business tax rates.

Table 7: Left panel: columns (1) and (2) report QML estimates of λ and their $t-$statistics (in brackets), and the value of the test $\hat{T}(t, t_Y)$ in (4.9) for model (7.2) with differences taken between 2001 and 2000 and between 2000 and 1999, respectively, with general property tax rate as the dependent variable. Right panel: columns (3) and (4) report results for model (7.2) with differences taken between 2001 and 2000 and between 2000 and 1999, respectively, with residential business tax rate as the dependent variable. Row-normalized weight matrices are used.
* $p - value < 0.1$; ** $p - value < 0.05$; *** $p - value < 0.01$.

	general property tax		residential building tax	
	(1)	(2)	(3)	(4)
$\hat{\lambda}$	0.0858 (1.1555)	−0.0139 (−0.3193)	0.1462** (2.0222)	0.0610 (1.0910)
$\hat{T}(t, t_Y)$	0.1640	1.1189	2.0383	0.1497

Allowing for municipality fixed effects changes the significance level of the spatial parameter λ, which is now always statistically insignificant for the general property rate regression, as opposed to the results displayed in Table 6, whereas the spatial parameter becomes significant for residential business property tax rates when differences are taken between 2001 and 2000. In both cases, the model is not rejected, suggesting that SAR might be an adequate specification to describe the network dependence in these data. We also observe that the QML estimator for λ in the differenced data remains close to the policy-based IV estimator reported in Lyytikäinen (2012). Although a full replication of the results in Lyytikäinen (2012) is not attempted in this illustration, the specification test findings and the empirical results in Tables 6 and 7 suggest that QML estimates of standard SAR models deliver results that are mostly in line with the policy-based IV estimator of Lyytikäinen (2012), as long as SAR specifications are carefully tailored to account for the policy change (i.e. once P_i, M_i, D_i and municipality fixed effects are accounted for).

We extend the results obtained in Tables 6 and 7 by choosing a different normalization for W. More specifically we impose a weight structure via the same contiguity criterion discussed above but now normalized by its spectral norm rather than by dividing elements in each row by their respective row sums. We report the new results in Tables 8 and 9. Comparing columns 1 and 2 of Table 8 with their counterparts in Table 6, we notice that λ remains significant, although it is much smaller in absolute value. However, when W is chosen as a spectral norm-normalised contiguity structure, our test clearly rejects the SAR specification. A similar pattern is observed when comparing columns 3 and 4 with their counterparts in Table 6, although SAR is rejected for specification (7.4) but not for (7.3). This discrepancy between results obtained by different normalization of the same contiguity structure is not surprising, when one considers that the empirical spatial econometric literature has a long-standing debate on the suitability of row-normalization. In the specific context of spatial analysis with political economy/political science data, row normalization imposes homogenous total exposure to spatial stimulus, without allowing some central municipalities to have a more prominent role compared to more peripheral ones (e.g. Neumayer and Plümper (2016)). Spectral norm normalization, instead, has the advantage of preserving the heterogeneity across different rows as all elements of the contiguity matrix are scaled by the same factor.

For completeness, in Table 9 we report results corresponding to Table 7 for a spectral norm-normalised W. The pattern of results is similar to that reported in Table 7, with the exception of the loss of statistical significance of λ for model (7.2) with differences taken between 2001 and 2000, and residential business property tax rate as the dependent variable. The SAR specification seems adequate for the four cases, similar to Table 7.

Table 8: Left panel: columns (1) and (2) report QML estimates of λ and their $t-$statistics (in brackets), and the value of the test $\hat{T}(t, t_Y)$ in (4.9) for model (7.3) and (7.4), respectively, with the general property tax rate as dependent variable. Right panel: columns (3) and (4) report results for model (7.3) and (7.4), respectively, with the residential business tax rate as dependent variable. Spectral norm normalized weight matrices are used.
* $p - value < 0.1$; ** $p - value < 0.05$; *** $p - value < 0.01$.

	general property tax		residential building tax	
	(1)	(2)	(3)	(4)
$\hat{\lambda}$	-0.0376^{*} (−1.6902)	-0.0446^{**} (−2.0155)	0.0061 (0.2531)	0.0004 (0.0164)
$\hat{T}(t, t_Y)$	13.6353^{***}	6.9837^{**}	3.7663	50.8116^{***}

Table 9: Left panel: columns (1) and (2) report QML estimates of λ and their $t-$statistics (in brackets), and the value of the test $\hat{T}(t, t_Y)$ in (4.9) for model (7.2) with differences taken between 2001 and 2000 and between 2000 and 1999, respectively, with general property tax rate as the dependent variable. Right panel: columns (3) and (4) report results for model (7.2) with differences taken between 2001 and 2000 and between 2000 and 1999, respectively, with residential business tax rate as the dependent variable. Spectral norm normalized weight matrices are used.
* $p - value < 0.1$; ** $p - value < 0.05$; *** $p - value < 0.01$.

	general property tax		residential building tax	
	(1)	(2)	(3)	(4)
$\hat{\lambda}$	0.0930 (1.0225)	−0.0317 (−0.6944)	0.1129 (1.2541)	0.0278 (0.4135)
$\hat{T}(t, t_Y)$	1.3613	2.3674	1.4379	0.1549

7.2 The Swiss case: a multi-tier federal system

The tax system in Switzerland has a special character and is highly decentralized. The total (denoted as "consolidated" in Parchet, 2019) personal income tax rate for a resident in municipality i belonging to canton c, denoted by t_i in the sequel, is composed by $T_i + T_c$, where T_c and T_i are tax rates levied by canton c and municipality i, respectively. For additional details about the Swiss personal income tax system, as well as on the general Swiss federal structure we refer to Parchet (2019) and the references therein. The baseline model considered in Parchet (2019) is a panel version of the following spatial autoregression, where the consolidated tax rate of municipality i is possibly related to a weighted average of consolidated tax rates of neighboring municipalities, with no restriction on whether "neighbors" belong to the same canton as municipality i or not, i.e.

$$t_i = \lambda \sum_{j=1}^{n} w_{ij} t_j + \beta' X_i + \epsilon_i, \quad i = 1, ..., n, \tag{7.5}$$

where X_i is a 37×1 vector that contains a unit constant, various characteristics of municipality i, as well as canton-specific dummies (denoted as D_{ci} in the sequel). In addition to D_{ci}, X_i contains dummies that indicate whether municipality i is an urban area and/or center of an urban area and whether it has a lake shore. A list of the non-binary variables contained in X_i, which capture population, political orientation, economic data, and geographic features, is reported in Table 10, and for additional details we refer the reader to Parchet (2019).

Table 10: List of variables in X_i of model (7.5).

population (in 1,000)	% foreign national
% youth (≤ 20)	% elderly (≥ 80)
% working in primary sector	% working in secondary sector
unemployment rate	total employment per capita
% votes for left-of-center parties	altitude
number of movie theaters within 10 km	

The results reported in Parchet (2019) refer to a panel of observations spanning the period from 1983 to 2012. In this paper we report estimated parameters for model (7.5) using data only from year 2012, although similar results to those presented here were found to hold for the other years. Heuristically, we expect that the theoretical properties of our test hold for the static model (7.5) with data pooled over multiple years. However, robustness of our test in a static panel model has not been formally studied and we therefore use a cross-sectional analysis instead. The 2012 sample used here has $n = 2389$ observations. Following Parchet (2019), the weight matrix is set so that $w_{ij} = 1$ if the road distance between municipalities i

and j does not exceed 10 km, and $w_{ij} = 0$ otherwise, and $w_{ii} = 0$, as is standard.

We report the estimate of λ in (7.5) and the value of our test statistic for two different normalizations of W: column (1) of Table 11 reports results obtained by a row-normalized version of W (i.e. each element w_{ij} is scaled so that elements of each row of the resulting matrix sum to 1), as adopted in Parchet (2019), while column (2) presents the corresponding results obtained when W is normalized by its spectral norm. Henceforth, we denote by W^r and W^s, respectively, the case of row-normalized and spectral-norm-normalized weight matrices.

The QML estimate of λ reported in column (1) is positive and statistically significant. This is in line with Table 2 in Parchet (2019), as well as with Allers and Elhost (2005). Nonetheless, the SAR specification is rejected by our $\hat{T}(t, t_Y)$ test in (4.9) at the 5% level, confirming that some alternative specification is needed in order to perform reliable inference. The inadequacy of the SAR specification supports the policy-based IV strategy discussed in Parchet (2019), at least so far as mitigating possible endogeneity effects arising from misspecification. Interestingly, the estimate of λ becomes negative and significant when W^s, as opposed to W^r, is adopted. The negative, significant λ is consonant with the policy-based IV estimate of λ displayed in Table 3 in Parchet (2019). However, our test strongly rejects SAR also for W^s. These results suggest that, although SAR is inadequate in both cases, spectral normalization instead of the common row normalization is enough to account for the negative sign in Parchet (2019). This discrepancy in sign across the two different normalization is not surprising, and we refer the reader to the discussion about the effects of different normalizations reported in Section 7.1.

Table 11: Columns (1) and (2) report estimates of λ and their t-statistics (in brackets), and the value of the test statistic $\hat{T}(t, t_Y)$ in (4.9) for model (7.5), with W^r and W^s, respectively, for data pertaining to 2012. Columns (3) and (4) report corresponding results for model (7.6), with W^r and W^s, respectively.
* $p - value < 0.1$; ** $p - value < 0.05$; *** $p - value < 0.01$.

	Model (7.5)		Model (7.6)	
	(1): W^r	(2): W^s	(3): W^r	(4): W^s
$\hat{\lambda}$	0.2271 *** (11.1581)	−0.0214*** (−4.4371)	0.1371*** (5.0666)	0.0380 (1.0177)
$\hat{T}(t, t_Y)$	8.1890**	126.7315***	3.6558	5.6807*

To assess whether some of the issues with SAR are linked to omitting relevant fixed-effects in the one-period model in (7.5), we re-estimate a SAR model with data differenced across two periods. We define $\Delta t_i = t_{i,2012} - t_{i,2007}$, where the same notation applies to the other quantities appearing in (7.5)[2]. Controlling explicitly for canton-dummies and including an

[2]We take differences over a 5-year time span to ensure sufficient variation in the X_i, which contain several observables with little change over consecutive years.

intercept, we have the model

$$\Delta t_i = \lambda \sum_{j=1, j\neq i}^{n} w_{ij} \Delta t_j + \gamma + \beta' \Delta X_i + \delta' D_{ci} + \Delta \epsilon_i, \quad i = 1, ..., n, \tag{7.6}$$

where ΔX_i is a 4×1 vector containing differenced data on size of the population, unemployment rate, total employment per capita and percent of votes for left-of-center parties, since all other variables reported in Table 10 do not display significant time variation and are omitted from the model.

Estimates of λ in (7.6) and their associated $t-$ratios as well as the value of the specification test $\hat{T}(t, t_Y)$ in (4.9) for W^r and W^s are reported in columns (3) and (4) of Table 11. Under row normalization, λ appears to be positive and significant at the 1% level, a result that is consistent with Allers and Elhorst (2005) and with Table 2 in Parchet (2019). More importantly, the specification test fails to reject SAR even at the 10% level, suggesting that mispecification issues detected in model (7.5) and reported in column (1) of Table 11 are resolved by differencing the data and thereby removing potential endogeneity due to fixed-effects. However, under spectral norm normalization, λ is no longer significant, and the adequacy of SAR is weakly rejected at 10% level. This finding reveals that although differencing alleviates some of the model misspecification issues, the SAR specification with spectral norm normalization W^s seems much less well suited than row normalization for modeling the spatial structure of these data.

8 Concluding remarks

This paper develops a substantial modification of the Bierens conditional moment test designed to suit the needs of spatial modeling. The test statistic has a convenient standard chi square limit theory and is consistent against general alternatives including those that involve functional form, the spatial/network specification, and weight matrix formulation. In view of the complications arising from the presence of both spatial interactions and systematic regressor components in spatial models, the test framework is formulated with careful attention to the multiple component nature of the null and alternative hypotheses. In particular, the framework elucidates precisely the different forms of misspecification in the model for which the test has discriminatory power and for which the test statistic is explicitly constructed to address.

Since the test has a standard pivotal limit distribution under $\mathcal{H}_0$ it is straightforward to implement using asymptotic critical values and simulations reveal that its practical performance is highly satisfactory with stable size and good power against multiple sources of misspecification. The application of our test to the municipality-level tax competition data from two recent studies by Lyytikäinen (2012) and Parchet (2019) sheds some light on the much-contested suitability of SAR modeling with conventional estimation methods (such as QML) in the tax competition literature. In particular, the specification tests conducted here corroborate the need for careful refinement of the specification or methods designed to address induced endogeneity from misspecification similar to the methods Lyytikäinen (2012) and Parchet (2019) have used with their policy-based IV estimation.

This paper has focused on specification testing in the basic SAR model with homoskedastic errors. More general applicability requires adaptations in both cross-sectional and panel data settings to accommodate broader maintained conditions. We expect that similar methods and results will apply in more sophisticated models, such as those with error heterogeneity and endogeneities, by modifications that address heterogeneities and robust estimation methods such as IV that address endogeneities. For instance, Theorem 2 continues to hold with minor modifications for IV estimation as the basic structure of the existing proof involves linear and quadratic forms in the errors that are preserved with only minor adjustments to account for the relevant projection operators. A heteroskedasticity-robust version of the test and its practical application is currently under investigation in separate work.

Appendix 1

Proof of Theorem 1. For each i satisfying $\mathbb{P}_{m_i}(\bar{\upsilon}_i(m_i,\theta)=0)<1$ and for all sufficiently large n, the proof of part b) of Theorem 1 follows directly from that of Lemma 1 in Bierens (1990). The proof of part a) follows again from Bierens (1990), once we can show that *for every $i=1,...,n$ and for sufficiently large n*

$$\mathbb{P}_{X_i}(\mathbb{E}(\mathrm{v}_i(\theta)|X_1,...,X_n)=0)<1 \to \mathbb{E}(\mathrm{v}_i(\theta)e^{t'X_i})\neq 0 \quad \textit{for at least one } t\in\mathbb{R}^k. \tag{A.1.1}$$

Recall that

$$d_{in}(\theta,X_1,...,X_n)=d_i(X_1,...,X_n)=\mathbb{E}(\mathrm{v}_i(\theta)|X_1,...,X_n),$$

where in this context we drop the dependence from θ for the sake of notational simplicity, with a similar convention for $\mathrm{v}_{in}(\theta)=\mathrm{v}_i$. We define the functions

$$d_{1i}(\cdot)=max\{d_i(\cdot),0\},\quad d_{2i}(\cdot)=max\{-d_i(\cdot),0\},$$

the expected values

$$c_{is}=\mathbb{E}(d_{is}(X_1,...,X_n)),\quad s=1,2,$$

and probability measures

$$\nu_{is}(B_i) = \frac{1}{c_{is}} \int_{B_i} \int_{\mathbb{R}^k} ... \int_{\mathbb{R}^k} d_{is}(x_1, ..., x_n) dF(x_1)...dF(x_n), \quad s = 1, 2, \quad i = 1, ..., n, \quad (A.1.2)$$

where B_i is a Borel set in $\mathbb{R}^k$ and is the range of integration of the variable X_i, and $F(x)$ is the cumulative distribution function of each of the *iid* random vectors X_j, $j = 1, ..., n$. Define the joint probability measure

$$\nu_{is}(B_1, ..., B_n) = \frac{1}{c_{is}} \int_{B_1} \int_{B_2} ... \int_{B_n} d_{is}(x_1, ..., x_n) dF(x_1)...dF(x_n). \quad (A.1.3)$$

We have

$$\begin{aligned}
&\mathbb{E}(\mathrm{v}_i e^{t'X_i}) = \mathbb{E}(e^{t'X_i} d_i(X_1, ..., X_n)) = \\
&= \int_{\mathbb{R}^k} ... \int_{\mathbb{R}^k} d_{i1}(x_1, ..., x_n) e^{t'x_i} dF(x_1)...dF(x_n) - \int_{\mathbb{R}^k} ... \int_{\mathbb{R}^k} d_{i2}(x_1, ..., x_n) e^{t'x_i} dF(x_1)...dF(x_n) \\
&= c_{i1} \int_{\mathbb{R}^k} e^{t'x_i} d\nu_{i1}(x_i) - c_{i2} \int_{\mathbb{R}^k} e^{t'x_i} d\nu_{i2}(x_i),
\end{aligned}$$

from the definitions of $\nu_{1j}(B_j)$ and $\nu_{2j}(B_j)$. Note that $\int_{\mathbb{R}^k} e^{t'x_i} d\nu_{1i}(x_i)$ and $\int_{\mathbb{R}^k} e^{t'x_i} d\nu_{2i}(x_i)$ are the moment generating functions of the probability measures $\nu_{i1}(B_i)$ and $\nu_{i2}(B_i)$.

We proceed by contradiction. If $\mathbb{E}(\mathrm{v}_i e^{t'X_i}) = 0$ for all $t \in \mathbb{R}^k$, substituting $t = 0$ in the equation $c_{i1} \int_{\mathbb{R}^k} e^{t'x_i} d\nu_{i1}(x_j) - c_{i2} \int_{\mathbb{R}^k} e^{t'x_i} d\nu_{i2}(x_i) = 0$ yields

$$c_{i1} = c_{i2}. \quad (A.1.4)$$

Thus, for each t

$$\int_{\mathbb{R}^k} e^{t'x_i} d\nu_{i1}(x_i) = \int_{\mathbb{R}^k} e^{t'x_i} d\nu_{i2}(x_i), \quad (A.1.5)$$

implying

$$\nu_{i1}(B_i) = \nu_{i2}(B_i) \quad \forall B_i \in \mathbb{R}^k, \quad i = 1, ..., n. \quad (A.1.6)$$

Therefore

$$\int_{B_i} \int_{\mathbb{R}^k} ... \int_{\mathbb{R}^k} d_i(x_1, ..., x_n) dF(x_1)...dF(x_n) = 0 \quad \forall B_i. \quad (A.1.7)$$

implying $d_i(x_1, ..., x_n) = 0$ almost surely with respect to $\mathbb{P}_{X_i}$. ■

Proof of Corollary 1. The "⇒" implication in ii) follows directly from part b) of Theorem 1. The "⇐" part in ii) trivially follows from Proposition 1 and the law of iterated expectations. Similarly, the "⇒" implication in i) follows from part a) of Theorem 1. To establish the "⇐"

implication in i) we observe that

$$\mathbb{P}_{X_i}(\mathbb{E}(\mathrm{v}_i(\theta)|X_1,...,X_n)=0)=1 \quad \Longrightarrow \quad \mathbb{P}_{X_i}(\mathbb{E}(\mathrm{v}_i(\theta)|X_i)=0)=1 \tag{A.1.8}$$

since

$$\mathbb{E}(\mathrm{v}_i(\theta)|X_1,...,X_n)=0 \quad \Longrightarrow \quad \mathbb{E}(\mathrm{v}_i(\theta)|X_i)=0 \tag{A.1.9}$$

by the law of iterated expectations. Thus

$$1=\mathbb{P}_{X_i}(E(\mathrm{v}_i(\theta)|X_1,...,X_n)=0)\leq \mathbb{P}_{X_i}(\mathbb{E}(\mathrm{v}_i(\theta)|X_1)=0), \tag{A.1.10}$$

implying that $\mathbb{P}_{X_i}(\mathbb{E}(\mathrm{v}_i(\theta)|X_1)=0)=1$ and the result follows by the law of iterated expectations. ■

Proof of Corollary 2. Recall that, for all $i=1,...,n$ and for all n, $Y_i=g_i(X_1,...,X_n)+\eta_i$, where $g_i=\mathbb{E}(Y_i|X_1,...,X_n)$ and so $\eta_i=Y_i-\mathbb{E}(Y_i|X_1,...,X_n)$. On the other hand we can always write, for all $i=1,...,n$ and for all n,

$$Y_i=\mathbb{E}(Y_i|m_i(\theta_0))+\xi_i(\theta_0)=\mathbb{E}(g_i|m_i(\theta_0))+\xi_i(\theta_0), \quad \xi_i=Y_i-\mathbb{E}(Y_i|m_i(\theta_0)), \tag{A.1.11}$$

since $\mathbb{E}(\eta_i|m_i(\theta_0))=0$. Then, by independence of η_i and m_i,

$$\mathbb{P}_{m_i}(g_i=m_i(\theta_0))=1 \Longrightarrow \mathbb{P}_{Y_i}(g_i=m_i(\theta_0))=1, \tag{A.1.12}$$

and so, under $\mathcal{H}_0$,

$$1=\mathbb{P}_{Y_i}(g_i=m_i(\theta_0))=\mathbb{P}_{Y_i}(\eta_i=\xi_i(\theta_0)). \tag{A.1.13}$$

In view of (3.13), we have $\mathbb{P}_{Y_i}(u_i(\lambda_0)=\eta_i)=1$ and thus $\mathbb{P}_{Y_i}(u_i(\lambda_0)=\xi_i(\theta_0))=1$. Collecting these results under $\mathcal{H}_0$, we obtain

$$Y_i=E(g_i|m_i(\theta_0))+\xi_i(\theta_0)=m_i(\theta_0)+u_i(\lambda_0). \tag{A.1.14}$$

■

Proof of Theorem 2 For ease of notation we frequently omit the subscript n and when there is no risk of confusion use $R=R(\lambda_0)$, $S^{-1}=S^{-1}(\lambda_0)$, $\hat{M}(t,t_Y)=M(\hat{\theta},t,t_Y)$, and $M(t,t_Y)=M(\theta_0,t,t_Y)$. By the mean value theorem (MVT),

$$\begin{aligned}\sqrt{n}\hat{M}(t,t_Y)=&\sqrt{n}M(t,t_Y)+\frac{dM(t,t_Y)}{d\theta}|_{\theta=\bar{\theta}}\sqrt{n}(\hat{\theta}-\theta_0)\\ =&\sqrt{n}M(t,t_Y)+\frac{dM(t,t_Y)}{d\theta}|_{\theta=\theta_0}\sqrt{n}(\hat{\theta}-\theta_0)+O_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{A.1.15}$$

where

$$||\bar{\theta}-\theta_0|| < ||\hat{\theta}-\theta_0||, \tag{A.1.16}$$

and $\hat{\theta}$ is a consistent estimator of θ_0, converging at rate $\sqrt{n}$ to the true value under the null. We have

$$\sqrt{n}(\hat{\theta}-\theta_0) = \Omega^{-1}\begin{pmatrix} \frac{1}{\sigma^2\sqrt{n}}\beta_0'\mathbb{X}'R'\epsilon \\ \frac{1}{\sigma^2\sqrt{n}}\mathbb{X}'\epsilon \end{pmatrix} + \Omega^{-1}\begin{pmatrix} \frac{1}{\sigma^2\sqrt{n}}(\epsilon' R\epsilon - \sigma^2 tr(R)) \\ 0 \end{pmatrix} + O_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.1.17}$$

where Ω is defined according to (4.2).

Formally, we approximate the exponential component in $M(\hat{\theta}, t, t_Y)$ as follows

$$\begin{aligned} e^{\frac{t_Y(Y_i-\bar{Y})}{p_n}} =& 1 + \frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}\left(\sup_i|\epsilon_i| \le K\alpha_n\right) + \frac{t_Y(Y_i-\bar{Y})}{p_n}\left(1-\mathbb{P}(\sup_i|\epsilon_i| \le K\alpha_n)\right) \\ &+ \frac{t_Y^2(Y_i-\bar{Y})^2}{2p_n^2}\mathbb{P}\left(\sup_i|\epsilon_i| \le K\alpha_n\right) + \frac{t_Y^2(Y_i-\bar{Y})^2}{2p_n^2}\left(1-\mathbb{P}(\sup_i|\epsilon_i| \le K\alpha_n)\right) + ..., \end{aligned} \tag{A.1.18}$$

where α_n satisfies Assumption R.

We stress that under $\mathcal{H}_0$, $Y_i - \bar{Y} = \sum_t\sum_j s^{d,it}x_{tj}\beta_j + \sum_j s^{d,ij}\epsilon_j$, where $s^{d,ij}$ denote the $(i-j)$th element of $S^d = S^{-1} - 1\bar{S}^{-1'}$, with $\bar{S}^{-1'} = \sum_{i=1}^n S^{i'}/n$. Thus, given Assumptions 2, 5 and the functional form of m_i in (2.4), $\mathbb{P}\left(\sup_i|\epsilon_i| \le K\alpha_n\right) \le \mathbb{P}\left(\sup_i|Y_i-\bar{Y}| \le K\alpha_n\right)$.

Now, under Assumption 1 and by Markov's inequality,

$$\begin{aligned} \mathbb{P}\left(\sup_i|\epsilon_i| \le K\alpha_n\right) =& \left(\mathbb{P}\left(|\epsilon_i| \le K\alpha_n\right)\right)^n = \left(1-\mathbb{P}(|\epsilon_i| > K\alpha_n)\right)^n \\ =& 1 + O\left(\frac{n}{\alpha_n^{4+\delta}}\mathbb{E}|\epsilon_i|^{4+\delta}\right) = 1 + O\left(\frac{n}{\alpha_n^{4+\delta}}\right) \end{aligned} \tag{A.1.19}$$

for $\delta > 0$, under Assumption 0. Therefore (A.1.18) can be written as

$$\begin{aligned} &1 + \frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}\left(\sup_i|\epsilon_i| \le K\alpha_n\right) + \frac{t_Y(Y_i-\bar{Y})}{p_n}O\left(\frac{n}{\alpha_n^{4+\delta}}\right) + \frac{t_Y^2(Y_i-\bar{Y})^2}{2p_n^2}\mathbb{P}\left(\sup_i|\epsilon_i| \le K\alpha_n\right) \\ &+ \frac{t_Y^2(Y_i-\bar{Y})^2}{2p_n^2}O\left(\frac{n}{\alpha_n^{4+\delta}}\right) + \dots \end{aligned} \tag{A.1.20}$$

By simple algebra using (A.1.17) and (A.1.20), (A.1.15) becomes

$$\frac{1}{\sqrt{n}}\left(\begin{matrix} \underline{e}(t)'S^{-1}\epsilon \\ \left(1' + \frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i| \le K\alpha_n)(Y' - \bar{Y}1')\right)S^{-1}\epsilon - \frac{t_Y}{p_n}\sigma_0^2 tr(S^{d'}Q) \end{matrix} \right)$$

$$-\frac{1}{n}\begin{pmatrix} \underline{e}(t)' \\ 1' + \frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i| \leq K\alpha_n)(Y' - \bar{Y}1') \end{pmatrix} S^{-1}\begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix}\Omega^{-1}\begin{pmatrix} \frac{1}{\sigma^2\sqrt{n}}\beta_0'\mathbb{X}'R'\epsilon \\ \frac{1}{\sigma^2\sqrt{n}}\mathbb{X}'\epsilon \end{pmatrix}$$
$$-\frac{1}{n}\begin{pmatrix} \underline{e}(t)' \\ 1' + \frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i| \leq K\alpha_n)(Y' - \bar{Y}1') \end{pmatrix} S^{-1}\begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix}\Omega^{-1}\begin{pmatrix} \frac{1}{\sigma^2\sqrt{n}}(\epsilon'R\epsilon - \sigma^2 tr(R)) \\ 0 \end{pmatrix}$$
$$+O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{\sqrt{n}}{p_n^2}, \frac{n^{3/2}}{p_n\alpha_n^{4+\delta}}\right)\right), \tag{A.1.21}$$

where Q is defined in (4.3). Under Assumption R, $O_p\left(\max\left(1/\sqrt{n}, \sqrt{n}/p_n^2, n^{3/2}/(p_n\alpha_n^{4+\delta})\right)\right) = o_p(1)$. In addition to the usual $1/\sqrt{n}$ error (as displayed in (A.1.15)), the error of the approximation depends on two extra terms: (i) the error resulting from linearization, i.e. that obtained when fourth (and fifth) terms in (A.1.20) are dropped, is bounded by $\sqrt{n}/p_n^2$; and (ii) the error that is generated by neglecting the (small) probability that $Y_i - \bar{Y}$ (for some $i = 1, ..., n$) might assume an extreme value is bounded by $n^{3/2}/(p_n\alpha_n^{4+\delta})$. The stated rates are straightforward to derive after locating the dominant terms upon using $Y - \bar{Y}1 = S^d\mathbb{X}\beta + S^d\epsilon$ in the expressions.

With simple manipulations (A.1.21) becomes

$$\frac{1}{\sqrt{n}}\begin{pmatrix} \underline{e}(t)' \\ 1' \end{pmatrix} Q\epsilon + \frac{1}{\sqrt{n}}\begin{pmatrix} 0_{1\times n} \\ \frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i| \leq K\alpha_n)\beta_0'\mathbb{X}'^{d'} \end{pmatrix} Q\epsilon$$
$$+\frac{1}{\sqrt{n}}\begin{pmatrix} 0_{1\times 1} \\ \frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i| \leq K\alpha_n)\epsilon'^{d'}Q\epsilon - \frac{t_Y}{p_n}\sigma_0^2 tr(S^{d'}Q) \end{pmatrix}$$
$$-\frac{1}{n}\begin{pmatrix} \underline{e}(t)' \\ 1' \end{pmatrix} S^{-1}\begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix}\Omega^{-1}\begin{pmatrix} \frac{1}{\sigma^2\sqrt{n}}(\epsilon'R\epsilon - \sigma^2 tr(R)) \\ 0_{k\times 1} \end{pmatrix}$$
$$-\frac{1}{n}\begin{pmatrix} 0_{1\times n} \\ \frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i| \leq K\alpha_n)\beta_0'\mathbb{X}'^{d'} \end{pmatrix} S^{-1}\begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix}\Omega^{-1}\begin{pmatrix} \frac{1}{\sigma^2\sqrt{n}}(\epsilon'R\epsilon - \sigma^2 tr(R)) \\ 0_{k\times 1} \end{pmatrix}$$
$$-\frac{1}{n}\begin{pmatrix} 0_{1\times n} \\ \frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i| \leq K\alpha_n)\epsilon'^{d'} \end{pmatrix} S^{-1}\begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix}\Omega^{-1}\begin{pmatrix} \frac{1}{\sigma^2\sqrt{n}}(\epsilon'R\epsilon - \sigma^2 tr(R)) \\ 0_{k\times 1} \end{pmatrix}$$
$$+ O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{\sqrt{n}}{p_n^2}, \frac{n^{3/2}}{p_n\alpha_n^{4+\delta}}\right)\right). \tag{A.1.22}$$

Under Assumptions 1-5 and 9, we can show by standard arguments that the second and fifth terms in (A.1.22) are bounded by $1/p_n$, the third is bounded by $\max(1/p_n, n^{3/2}/(p_n\alpha_n^{4+\delta}))$ and the sixth term is bounded by $1/(\sqrt{n}p_n)$.

Let $\Psi(t, t_Y)$ and $\psi(t, t_Y)$ be the $2 \times n$ matrix and 2×1 vector defined in (4.5) and (4.6).

Henceforth, we drop the dependence on (t, t_Y) in these expressions to simplify notation, i.e. $\Psi = \Psi(t, t_Y)$ and $\psi = \psi(t, t_Y)$. From (A.1.22),

$$
\begin{aligned}
\sqrt{n}\hat{M}(t) =& \frac{1}{\sqrt{n}}\Psi\epsilon + \frac{1}{\sqrt{n}}\psi(\epsilon' R\epsilon - \sigma_0^2 tr(R)) + O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{p_n}, \frac{\sqrt{n}}{p_n^2}, \frac{n^{3/2}}{p_n\alpha_n^{4+\delta}}\right)\right) \\
=& \frac{1}{\sqrt{n}}\Psi\epsilon + \frac{1}{\sqrt{n}}\psi(\epsilon' R\epsilon - \sigma_0^2 tr(R)) + o_p(1) \qquad \text{(A.1.23)}
\end{aligned}
$$

under Assumption R. Now let b be any deterministic 2×1 vector such that $b'b = 1$ and write $\sqrt{n}b'\hat{M}(t, t_Y) = \sum_{i=1}^{n} u_i$, where

$$
u_i = u_{in}(t, t_Y) = \frac{1}{\sqrt{n}}\sum_{s=1}^{2} b_s\Psi_{si}\epsilon_i + \frac{1}{\sqrt{n}} b'\psi\bar{r}_{ii}(\epsilon_i^2 - \sigma_0^2) + \frac{2b'\psi}{\sqrt{n}}\epsilon_i\sum_{j<i}\bar{r}_{ij}\epsilon_j, \qquad \text{(A.1.24)}
$$

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

We therefore have

$$
\begin{aligned}
&\mathbb{V}ar(\sqrt{n}b'\hat{M}(t, t_Y)) \equiv a_n(t, t_Y) = a = \sum_{i=1}^{n}\mathbb{V}ar(u_i) = \sum_{i=1}^{n} E(u_i^2) \\
=& \frac{\sigma_0^2}{n} b'\Psi\Psi' b + \frac{\sigma_0^4 tr\left((R + R')^2\right)}{2n} b'\psi\psi' b + \frac{2\mu^{(3)}}{n} b'\psi\sum_{i=1}^{n}\bar{r}_{ii}(\Psi_i')' b \\
+& \frac{(\mu^{(4)} - 3\sigma_0^4)}{n} b'\psi\psi' b\sum_{i=1}^{n}\bar{r}_{ii}^2, \qquad \text{(A.1.25)}
\end{aligned}
$$

where Ψ_i' is the transpose of the $i-$th row of Ψ'. The leading term of (A.1.25) is the first and by Lemma 1 is non zero as $n \to \infty$.

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

$$
\sum_{i=1}^{n}\mathbb{E}(z_i^2|\epsilon_j; j < i) \underset{p}{\to} 1, \qquad \text{(A.1.26)}
$$

and for each $\zeta > 0$

$$
\sum_{i=1}^{n}\mathbb{E}(z_i^2 1(|z_i > \zeta|)) \underset{p}{\to} 0, \qquad \text{(A.1.27)}
$$

then $\sum_{i=1}^{n} z_i \to_d \mathcal{N}(0, 1)$ pointwise in (t, t_Y). Thus, the claim in Theorem 2 follows straightforwardly, with

$$A = \lim_{n\to\infty}\left\{\frac{\sigma_0^2}{n}\Psi\Psi' + \frac{\sigma_0^4 tr\left((R+R'^2\right)}{2n}\psi\psi' + \frac{2\mu^{(3)}}{n}\psi\sum_{i=1}^{n}\bar{r}_{ii}(\Psi_i')' + \frac{(\mu^{(4)}-3\sigma_0^4)}{n}\psi\psi'\sum_{i=1}^{n}\bar{r}_{ii}^2\right\}. \tag{A.1.28}$$

We start by showing (A.1.26). We can equivalently show that, conditional on $\mathbb{X}$ and pointwise in (t, t_Y),

$$\sum_{i=1}^{n}\left(\mathbb{E}(z_i^2|\epsilon_j; j<i) - \mathbb{E}(z_i^2)\right) \underset{p}{\to} 0. \tag{A.1.29}$$

By standard algebra,

$$\begin{aligned}\sum_{i=1}^{n}\left(\mathbb{E}(z_i^2|\epsilon_j; j<i) - E(z_i^2)\right) =& \frac{4b'\psi}{na}\Bigg(\eta'\psi\Bigg(\sigma_0^2\sum_{i=1}^{n}\sum_{j<i}\sum_{s<i}\bar{r}_{ij}\bar{r}_{is}\epsilon_j\epsilon_s - \sigma_0^4\sum_{i=1}^{n}\sum_{j<i}\bar{r}_{ij}^2\Bigg) \\ &+\sigma_0^2\sum_{i=1}^{n}\sum_{p=1}^{2}b_p\Psi_{pi}\sum_{j<i}\bar{r}_{ij}\epsilon_j + b'\psi\mu_0^{(3)}\sum_{i=1}^{n}\bar{r}_{ii}\sum_{j<i}\bar{r}_{ij}\epsilon_j\Bigg).\end{aligned} \tag{A.1.30}$$

Since $a = O(1)$ and is non-zero for each (t, t_Y) by Lemma 1 and trivially each component of $\psi = O(1)$, result (A.1.26) holds if, pointwise in (t, t_Y) and conditional on $\mathbb{X}$,

$$\frac{1}{n}\left(\sigma^2\sum_{i=1}^{n}\sum_{j<i}\sum_{s<i}\bar{r}_{ij}\bar{r}_{is}\epsilon_j\epsilon_s - \sigma^4\sum_{i=1}^{n}\sum_{j<i}\bar{r}_{ij}^2\right) \underset{p}{\to} 0, \tag{A.1.31}$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{p=1}^{2}\eta_p\Psi_{pi}\sum_{j<i}\bar{r}_{ij}\epsilon_j \underset{p}{\to} 0, \tag{A.1.32}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\bar{r}_{ii}\sum_{j<i}\bar{r}_{ij}\epsilon_j \underset{p}{\to} 0. \tag{A.1.33}$$

The LHS of (A.1.31) can be written as

$$\frac{\sigma^2}{n}\Big(\sum_{i=1}^{n}\sum_{j<i}\bar{r}_{ij}^2(\epsilon_j^2-\sigma^2) + \sum_{i=1}^{n}\sum_{\substack{j,s<i\\ j\neq s}}\bar{r}_{ij}\bar{r}_{is}\epsilon_j\epsilon_s\Big). \tag{A.1.34}$$

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

$$\frac{K}{n^2}\sum_{i,k}\sum_{j<i,k}\bar{r}_{ij}^2\bar{r}_{kj}^2 \le \frac{K}{n^2}\sum_{i,j,k}\bar{r}_{ij}^2\bar{r}_{kj}^2 \le \frac{K}{n^2}\left(\max_j\sum_i\bar{r}_{ij}^2\right)\sum_{k,j}\bar{r}_{kj}^2 = O\left(\frac{1}{nh}\right), \tag{A.1.35}$$

since

$$\sum_{k,j} \bar{r}_{kj}^2 = \frac{1}{4} tr((R + R'^2) = O\left(\frac{n}{h}\right), \tag{A.1.36}$$

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

$$\max_j \sum_i \bar{r}_{ij}^2 = \max_j \frac{1}{4} e_j'(R + R'^2) e_j \leq K \| R + R'^2 \| \leq K, \tag{A.1.37}$$

under Assumptions 3, 4 and 5. The second term in (A.1.34) has again mean zero and variance bounded by

$$\begin{aligned}
&\frac{K}{n^2} \Big| \sum_{i,k} \sum_{j<i,k} \sum_{s<i,k} \bar{r}_{ij} \bar{r}_{is} \bar{r}_{kj} \bar{r}_{ks} \Big| \leq \frac{K}{n^2} \sum_{i,k,j,s} |\bar{r}_{ij} \bar{r}_{is} \bar{r}_{kj} \bar{r}_{ks}| \leq \frac{K}{n^2} \sum_{i,k,j,s} |\bar{r}_{ij} \bar{r}_{is}| (\bar{r}_{kj}^2 + \bar{r}_{ks}^2) \\
&\leq \frac{K}{n^2} \left(\max_j \sum_i |\bar{r}_{ij}| \right) \left(\max_i \sum_s |\bar{r}_{is}| \right) \sum_{k,j} \bar{r}_{kj}^2 + \frac{K}{n^2} \left(\max_i \sum_j |\bar{r}_{ij}| \right) \left(\max_s \sum_i |\bar{r}_{is}| \right) \sum_{k,s} \bar{r}_{ks}^2 \\
&= O\left(\frac{1}{nh} \right),
\end{aligned} \tag{A.1.38}$$

under Assumptions 3, 4 and 5 and by (A.1.36). Thus, collecting (A.1.35) and (A.1.38) we deduce (A.1.31) by the Markov inequality.

The LHS of (A.1.32) has mean zero and variance bounded by

$$\begin{aligned}
&\frac{K}{n^2} \sum_i \sum_s \sum_{j<i,s} \Big| \left(\sum_{p=1}^2 b_p \Psi_{pi} \right) \left(\sum_{q=1}^2 b_q \Psi_{qs} \right) \bar{r}_{ij} \bar{r}_{sj} \Big| \\
&\leq \frac{K}{n^2} \sum_i \sum_s \sum_j |\bar{r}_{ij}| |\bar{r}_{sj}| \left(\left(\sum_{p=1}^2 b_p \Psi_{pi} \right)^2 + \left(\sum_{q=1}^2 b_q \Psi_{qs} \right)^2 \right) \\
&\leq \frac{K}{n^2} \big(\max_i \sum_j |\bar{r}_{ij}| \big) \big(\max_j \sum_s |\bar{r}_{sj}| \big) \sum_i \left(\sum_{p=1}^2 b_p \Psi_{pi} \right)^2 + \frac{K}{n^2} \big(\max_j \sum_i |\bar{r}_{ij}| \big) \big(\max_s \sum_j |\bar{r}_{sj}| \big) \sum_s \left(\sum_{q=1}^2 b_q \Psi_{qs} \right)^2 \\
&= O\left(\frac{1}{n} \right)
\end{aligned} \tag{A.1.39}$$

pointwise in (t, t_Y) under Assumptions 3,4 and 5 and by Lemma 1. Then (A.1.32) follows by Markov inequality.

Similarly, (A.1.33) follows by Markov inequality after observing that the LHS has mean zero and variance bounded by

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

$$\leq \frac{K}{n^2}\big(\max_i \sum_j |\bar{r}_{ij}|\big)\big(\max_j \sum_k |\bar{r}_{kj}|\big)\sum_i \bar{r}_{ii}^2 + \frac{K}{n^2}\big(\max_i \sum_j |\bar{r}_{ij}|\big)\big(\max_k \sum_j |\bar{r}_{kj}|\big)\sum_k \bar{r}_{kk}^2$$
$$= O\left(\frac{1}{nh}\right), \tag{A.1.40}$$

since

$$\sum_i \bar{r}_{ii}^2 \leq \sum_{i,j} \bar{r}_{ij}^2 = O\left(\frac{n}{h}\right). \tag{A.1.41}$$

We prove (A.1.27) by verifying the sufficient Lyapunov condition that pointwise in (t, t_Y) and conditional on $\mathbb{X}$

$$\sum_{i=1}^n \mathbb{E}|z_i|^{2+\delta} \to 0. \tag{A.1.42}$$

Since $a = O(1)$ and is non-zero pointwise in (t, t_Y), we consider equivalently $\sum_i \mathbb{E}|u_i|^{2+\delta}$. We use $\sum_i \mathbb{E}|u_i|^{2+\delta} = \sum_i \mathbb{E}(\mathbb{E}|u_i|^{2+\delta}|\epsilon_j, j < i)$. By the c_r inequality and since $\psi = O(1)$ for each (t, t_Y), we have

$$\sum_{i=1}^n \mathbb{E}|u_i|^{2+\delta} \leq \left(\frac{1}{n}\right)^{1+\delta/2} K\sum_i |\sum_{s=1}^2 b_s \Psi_{si}|^{2+\delta} + \left(\frac{1}{n}\right)^{1+\delta/2} K\sum_i |\bar{r}_{ii}|^{2+\delta}$$
$$+ \left(\frac{1}{n}\right)^{1+\delta/2} K\sum_i \mathbb{E}|\sum_{j<i} \bar{r}_{ij}\epsilon_j|^{2+\delta}. \tag{A.1.43}$$

The first term on the RHS of (A.1.43) is bounded by

$$K\left(\frac{1}{n}\right)^{1+\delta/2} \max_i |\sum_{s=1}^2 b_s \Psi_{si}|^\delta \sum_i \left(\sum_{s=1}^2 b_s \Psi_{si}\right)^2 \leq K\frac{1}{n^{\delta/2}} \max_i |\sum_{s=1}^2 b_s \Psi_{si}|^\delta = o(1), \tag{A.1.44}$$

since $\sum_i (\sum_s^2 b_s \Psi_{si})^2/n = b'\Psi\Psi' b/n = O(1)$ and non-zero by Lemma 1, and for each i, s $|\Psi_{si}| = O(1)$. Similarly, the second term on the RHS of (A.1.43) is bounded by

$$K\frac{1}{n^{\delta/2}} \max_i |\bar{r}_{ii}|^\delta \frac{1}{n}\sum_i \bar{r}_{ii}^2 = o(1), \tag{A.1.45}$$

since

$$|\bar{r}_{ii}| \leq K||R + R'||_\infty \leq K, \tag{A.1.46}$$

and by virtue of (A.1.41). By the Burkholder and von Bahr/Esseen inequalities the last term at the RHS of (A.1.43) is bounded by

$$K\left(\frac{1}{n}\right)^{1+\delta/2}\sum_{i}\mathbb{E}|\sum_{j<i}\bar{r}_{ij}^{2}\epsilon_{j}^{2}|^{1+\delta/2}\leq K\left(\frac{1}{n}\right)^{1+\delta/2}\sum_{i}\sum_{j<i}|\bar{r}_{ij}|^{2+\delta}$$
$$=K\left(\frac{1}{n}\right)^{1+\delta/2}\sum_{i}(\sum_{j<i}\bar{r}_{ij}^{2})^{1+\delta/2}\leq K\left(\frac{1}{n}\right)^{1+\delta/2}(\max_{i}\sum_{j}\bar{r}_{ij}^{2})^{\delta/2}\sum_{i,j}\bar{r}_{ij}^{2}=o(1) \quad \text{(A.1.47)}$$

by (A.1.36) and (A.1.37). Thus, collecting (A.1.44), (A.1.45) and (A.1.47) we conclude that (A.1.27) holds pointwise in t. ■

Proof of Theorem 3

In order to prove the claim in Theorem 3, and thus consistency of the test based on (4.8), we show that $\sqrt{n}\hat{M}\to_{p}\pm\infty$ under Assumption 13. Then, under Assumption 10', $\hat{T}(t,t_{Y})=nM(\hat{\theta},t,t_{Y})'\hat{A}^{-1}(t,t_{Y})M(\hat{\theta},t,t_{Y})\to\infty$.

Write $\hat{M}(t,t_{Y})=(\hat{M}_{1}(t),\hat{M}_{2}(t_{Y}))'$. We aim to show that under $\mathcal{H}_{1}$ in (3.14), either $\text{plim}_{n\to\infty}\hat{M}_{1}(t)\neq 0$ or/and $\text{plim}_{n\to\infty}\hat{M}_{2}(t_{Y})\neq 0$. We have

$$\hat{M}_{1}(t)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\left(g_{i}(\mathbb{X})-m_{i}(\mathbb{X},\theta^{\sharp})\right)e^{t'X_{i}}\right)+\frac{1}{n}\sum_{i=1}^{n}\eta_{i}e^{t'X_{i}}+\frac{1}{n}\sum_{i=1}^{n}\left(g_{i}(\mathbb{X})e^{t'X_{i}}-\mathbb{E}\big(g_{i}(\mathbb{X})e^{t'X_{i}}\big)\right)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\left(m_{i}(\mathbb{X},\theta^{\sharp})e^{t'X_{i}}-\mathbb{E}\big(m_{i}(\mathbb{X},\theta^{\sharp})e^{t'X_{i}}\big)\right)-\frac{1}{n}\sum_{i=1}^{n}\big(m_{i}(\mathbb{X},\hat{\theta})-m_{i}(\mathbb{X},\theta^{\sharp})\big)e^{t'X_{i}}. \quad \text{(A.1.48)}$$

The first term in (A.1.48) is strictly non-zero as n increases under Assumption 13. In order to ensure $\sqrt{n}\hat{M}_{1}(t)\to_{p}\pm\infty$, we show that the remaining terms in (A.1.48) are $o_{p}(1)$.

Under Assumptions 2 and 7, the second term in (A.1.48) has mean zero and variance equal to

$$\frac{1}{n^{2}}|\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(\eta_{i}\eta_{j})\mathbb{E}\left(e^{t'X_{i}}e^{t'X_{j}}\right)|\leq\frac{K}{n}\sup_{i}\sum_{j=1}^{n}|Cov(\eta_{i},\eta_{j})|=O\left(\frac{1}{n}\right). \quad \text{(A.1.49)}$$

The third term in (A.1.48) has again mean zero and variance bounded by

$$\frac{1}{n^{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}|Cov\left(g_{i}e^{t'X_{i}},g_{j}e^{t'X_{j}}\right)|=O\left(\frac{1}{n}\right), \quad \text{(A.1.50)}$$

under Assumption 6b). The fourth term in (A.1.48) can be decomposed as

$$\frac{1}{n}\sum_{i=1}^{n}s^{ii}\left(X_{i}'\beta^{\sharp}e^{t'X_{i}}-\mathbb{E}\left(X_{i}'\beta^{\sharp}e^{t'X_{i}}\right)\right)+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}s^{ij}\mathbb{E}\left(X_{j}'\beta^{\#}\right)\left(e^{t'X_{i}}-\mathbb{E}\left(e^{t'X_{i}}\right)\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} s^{ij}\mathbb{E}\left(e^{t'X_i}\right)\left(X_j'\beta^{\#}-\mathbb{E}\left(X_j'\beta^{\#}\right)\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} s^{ij}\left(e^{t'X_i}-\mathbb{E}\left(e^{t'X_i}\right)\right)\left(X_j'\beta^{\#}-\mathbb{E}\left(X_j'\beta^{\#}\right)\right). \tag{A.1.51}$$

Under Assumption 2, all terms in (A.1.51) have mean zero and involve sums of independent, zero-mean random quantities. Under Assumption 2, the variance of the first term in (A.1.51) is bounded by

$$\frac{K}{n^2}\sum_{i=1}^{n}(s^{ii})^2 = O\left(\frac{1}{n}\right), \tag{A.1.52}$$

under Assumptions 4-5. Similarly, under Assumption 2 the variance of the last term in (A.1.51) is bounded by

$$\frac{K}{n^2}|\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left((s^{ij})^2+s^{ij}s^{ji}\right)| \leq \frac{K}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left((s^{ij})^2+|s^{ij}||s^{ji}|\right)$$

$$\leq \frac{K}{n}\sup_{i,j}|s^{ij}|\sup_{j}\sum_{i=1}^{n}(|s^{ij}|+|s^{ji}|) = O\left(\frac{1}{n}\right), \tag{A.1.53}$$

under Assumptions 4-5. The second and third terms in (A.1.51) under Assumption 2 have variance bounded by

$$\frac{K}{n^2}|\sum_{i}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{q=1\\ q\neq i}}^{n}s^{ij}s^{iq}| \leq \frac{K}{n^2}\sum_{i}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{q=1\\ q\neq i}}^{n}|s^{ij}||s^{iq}| \leq \frac{K}{n}\sup_{i}\sum_{j=1}^{n}|s^{ij}|\sup_{i}\sum_{q=1}^{n}|s^{iq}| = O\left(\frac{1}{n}\right), \tag{A.1.54}$$

under Assumption 5. By Markov's inequality, we conclude that the second, third and fourth terms on the RHS of (A.1.48) are all $O_p(1/\sqrt{n})$.

The last term in (A.1.48), by the mean value theorem, can be written as

$$\frac{1}{n}\sum_{i=1}^{n}e^{t'X_i}\frac{dm_i(\mathbb{X},\bar{\theta})}{d\theta}(\hat{\theta}-\theta^{\#}) = \frac{1}{n}\sum_{i=1}^{n}e^{t'X_i}S^{(i)\prime}(\bar{\lambda})\left(R(\bar{\lambda})\mathbb{X}\bar{\beta}\quad\mathbb{X}\right)(\hat{\theta}-\theta^{\#}), \tag{A.1.55}$$

where $\bar{\lambda}$ and $\bar{\beta}$ satisfy, respectively, $|\bar{\lambda}-\lambda^{\#}| < |\hat{\lambda}-\lambda^{\#}|$ and $||\bar{\beta}-\beta^{\#}|| < ||\hat{\beta}-\beta^{\#}||$. Under Assumption 12, $\hat{\theta}-\theta^{\#} = o_p(1)$. Therefore the last term in (A.1.48) is $o_p(1)$ as long as we can show that each component of the $(k+1)\times 1$ vector

$$\frac{1}{n}\sum_{i=1}^{n} e^{t'X_i} S^{(i)\prime}(\bar{\lambda})\left(R(\bar{\lambda})\mathbb{X}\bar{\beta} \quad \mathbb{X}\right) \tag{A.1.56}$$

is $O_p(1)$. For simplicity of notation, in order to assess the rate of (A.1.56), let $A(\lambda)$ be equal to either $S^{-1}(\lambda)$ or $S^{-1}(\lambda)R(\lambda)$, its $(i,j)-$th element being $a_{ij}(\lambda)$. Under Assumption 2, the modulus of the typical element of (A.1.56) has expectation bounded by

$$K \sup_i \sum_{j=1}^{n} |a_{ij}(\bar{\lambda})| = O(1), \tag{A.1.57}$$

under Assumptions 4 and 5. By Markov's inequality, the last term in (A.1.48) is $o_p(1)$.

We now deal with the second component of the test statistic $\hat{M}_2(t_Y)$. We can write

$$\hat{M}_2(t_Y) = \frac{1}{n}\sum_{i=1}^{n}\left(g_i(\mathbb{X}) - m_i(\mathbb{X},\theta^\sharp)\right) e^{t_Y(Y_i-\bar{Y})/p_n} + \frac{1}{n}\sum_{i=1}^{n}\eta_i e^{t_Y(Y_i-\bar{Y})/p_n}$$
$$-\frac{1}{n}\sum_{i=1}^{n}\left(m_i(\mathbb{X},\hat{\theta}) - m_i(\mathbb{X},\theta^\sharp)\right)e^{t_Y(Y_i-\bar{Y})/p_n} - \frac{t_Y}{np_n}(\hat{\sigma}^2 tr(\hat{S}^d\hat{Q})) \tag{A.1.58}$$

Under Assumptions 6, and 7, under $\mathcal{H}_1$, we have $\sup_i (Y_i - \bar{Y}) = O_p(1)$. The first term in (A.1.58) is

$$\frac{1}{n}\sum_{i=1}^{n}\left(g_i(\mathbb{X}) - m_i(\mathbb{X},\theta^\sharp)\right) + \frac{1}{n}\sum_{i=1}^{n}\left(g_i(\mathbb{X}) - m_i(\mathbb{X},\theta^\sharp)\right)\left(e^{t_Y(Y_i-\bar{Y})/p_n} - 1\right) \tag{A.1.59}$$

Under Assumption 13,

$$\frac{1}{n}\sum_{i=1}^{n}\left(g_i(\mathbb{X}) - m_i(\mathbb{X},\theta^\sharp)\right) \underset{p}{\rightarrow} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(g_i(\mathbb{X}) - m_i(\mathbb{X},\theta^\sharp)\right) \neq 0, \tag{A.1.60}$$

where the limit on the RHS of the last displayed expression exists under Assumption 2, 6 and (2.4). The modulus of the second term in (A.1.59) is bounded by

$$\sup_i \left| e^{t_Y(Y_i-\bar{Y})/p_n} - 1 \right| \frac{1}{n}\sum_{i=1}^{n}|g_i(\mathbb{X}) - m_i(\mathbb{X},\theta^\sharp)|,$$

which is $o_p(1)$ as $n\to\infty$ since $\sum_{i=1}^{n}|g_i(\mathbb{X}) - m_i(\theta^\sharp)|/n = O_p(1)$ under $\mathcal{H}_1$ in (3.14). Similarly, we can show that the second and the third term in (A.1.58) are $o_p(1)$. The details are omitted to avoid repetition. The last term in (A.1.58) is $O_p\left(\frac{1}{p_n}\right) = o_p(1)$ under Assumptions 2, 4, 5 and 9 and by (4.7). ■

Appendix 2: Additional Lemma and Propositions

Lemma 1 *Conditional on $\mathbb{X}$ and under Assumptions 2, 4, 5 and 9,*

$$||Q||_\infty + ||Q'||_\infty < K \tag{A.2.1}$$

for all sufficiently large n.

Proof of Lemma 1. Under Assumption 4, the claim follows as long as

$$\frac{1}{n}||Z\Omega^{-1}Z'||_\infty < K, \tag{A.2.2}$$

where (limited to the scope of this Lemma) we set $Z = \begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix}$.

Under Assumptions 2-5, $||R||_\infty + ||R'||_\infty < K$, and thus all elements of Z are $O(1)$, conditionally on $\mathbb{X}$. Under Assumption 9, $\underline{\mathrm{eig}}(\Omega) > c > 0$ and then

$$\begin{aligned}\frac{1}{n}||Z\Omega^{-1}Z'||_\infty =& \frac{1}{n}\sup_i \sum_{j=1}^{n}|Z_i'\Omega^{-1}Z_j| \leq \frac{1}{n}\sup_i \sum_{j=1}^{n}||Z_i'||||Z_j||||\Omega^{-1}|| \\ \leq& K\sup_{i,j} ||Z_i'||||Z_j||||\Omega^{-1}|| \leq K,\end{aligned}$$

since $\sup_i ||Z_i|| = \sup_i (Z_i'Z_i)^{1/2} = O(1)$ and

$$||\Omega^{-1}|| = \overline{\mathrm{eig}}(\Omega^{-1}) = \frac{1}{\underline{eig}(\Omega)} < \frac{1}{c} < K. \tag{A.2.3}$$

■

We introduce some additional technical assumptions that are used in Proposition 2. Let $\mathcal{N}_\delta = \{\lambda : |\lambda - \lambda^\sharp| < \delta\}$ and $\bar{\mathcal{N}_\delta} = \Lambda/\mathcal{N}_\delta$ for some $\delta > 0$. Define $\hat{\sigma}^2(\lambda) = \frac{1}{n}y'S(\lambda)'M_X S(\lambda)y$, $M_X = I - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$ and $\tilde{\sigma}^2 := \mathbb{E}_1(\hat{\sigma}^2(\lambda))$, with $\mathbb{E}_1(\cdot)$ denoting expectation under $\mathcal{H}_1$.

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

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

Assumption A is an identification condition on $\lambda^\sharp$ under $\mathcal{H}_1$, akin to Assumption 5 of Delgado and Robinson (2015).

Assumption B Setting $g = (g_1(X_1,...,X_n),\cdots,g_n(X_1,...,X_n))'$ we have

$$\underline{\lim}_{n\to\infty}\inf_{\Lambda}\tilde{\sigma}^2(\lambda) = \underline{\lim}_{n\to\infty}\inf_{\Lambda}\frac{1}{n}\left(\mathbb{E}(g'S(\lambda)'M_XS(\lambda)g) + \mathbb{E}(\eta'S(\lambda)'M_XS(\lambda)\eta)\right) > 0.$$

Assumption C $\sup_{1\le i\le n}\mathbb{E}(\eta_i^4) < K.$

We mention that existence of the fourth moment of η_i, uniformly over i, was not required in Assumption 7, which only imposed uniform boundedness of the row sums of $\mathbb{E}(\eta\eta')$.

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

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

where $N(\lambda) := S(\lambda)'M_XS(\lambda)$.

Sufficient conditions for Assumption D are $\|W\| < K$, $\sup_\lambda\|S(\lambda)\| < K$ which follow from Assumption 4 (ii), and $\|M_X\| = O_p(1)$, which follows from Assumptions 2 and 9. To see this, note that

$$\frac{dN(\lambda)}{d\lambda} = -WM_XS(\lambda)' - S(\lambda)M_XW', \tag{A.2.4}$$

and by the mean value theorem with $\bar{\lambda}$ such that $|\bar{\lambda} - \lambda^\dagger| < |\lambda - \lambda^\dagger|$ we obtain

$$\|N(\lambda) - N(\lambda^\dagger)\| \le |\lambda - \lambda^\dagger|\sup_{\bar{\lambda}:|\bar{\lambda}-\lambda^\dagger|<|\lambda-\lambda^\dagger|}\|\frac{dN(\lambda)}{d\lambda}|_{\lambda=\bar{\lambda}}\|,$$

where uniformly over $\bar{\lambda}$

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

Hence $\|N(\lambda) - N(\lambda^\dagger)\| \le K|\lambda - \lambda^\dagger|$ and setting $\delta = \varepsilon/K$ with $\varepsilon > 0$ satisfies Assumption D.

Let $\mu_g = \mathbb{E}(g)$ and $\Omega_g = \mathbb{V}ar(g)$, as defined in Assumption 6. Both μ_g and Ω_g exist under Assumption 6, and higher order moments of g exist as well under the assumption that $g(\cdot)$ is a continuous function of bounded random variables. Let $\zeta(\lambda) = M_XS(\lambda)g$ and denote by $\mu_\zeta(\lambda)$ and $\Omega_\zeta(\lambda)$ the mean and variance matrix of $\zeta(\lambda)$, both of which exist for each $\lambda\in\Lambda$ under Assumptions 2, 6 and 9. We make the following condition on $\Omega_\zeta(\lambda)$.

Assumption E $\sup_{\lambda\in\Lambda}||\Omega_\zeta(\lambda)||_\infty < K.$

Since $\zeta(\lambda)$ is a null vector under $\mathcal{H}_0$, i.e. when $g = S^{-1}(\lambda)\mathbb{X}\beta$, $\zeta(\lambda)$ may be interpreted as a measure of misspecification in the regression function, conditional on $X_1,...,X_n$. Thus, Assumption E imposes a condition that the degree of misspecification of the function g is bounded in probability.

Proposition 2 *Define a sequence of pseudo true values as* $\lambda_n^\sharp = \lambda^\sharp := \arg\min_{\lambda \in \Lambda} \tilde{\mathcal{L}}(\lambda)$ *where*

$$\tilde{\mathcal{L}}(\lambda) = \log(\tilde{\sigma}^2(\lambda)) + \frac{1}{n} \log |S^{-1}(\lambda)' S^{-1}(\lambda)|. \tag{A.2.5}$$

Under Assumptions A-E, 2-7 and 9, $\hat{\lambda} - \lambda^\sharp = o_p(1)$ *under* $\mathcal{H}_1$*.*

Proof of Proposition 2.

We follow the arguments in the proof of Theorem 1 of Delgado and Robinson (2015). We proceed with the concentrated likelihood $\hat{\lambda} = \arg\min_{\lambda \in \Lambda} (\mathcal{L}(\lambda))$, where

$$\mathcal{L}(\lambda) = \log(\hat{\sigma}^2(\lambda)) + \frac{1}{n} \log |S^{-1}(\lambda)' S^{-1}(\lambda)|. \tag{A.2.6}$$

Using independence between X_i and η_j for each $i, j = 1, ..., n$, and since $Y = g + \eta$ under $\mathcal{H}_1$, we have

$$\tilde{\sigma}^2(\lambda) = \mathbb{E}_1(\hat{\sigma}^2(\lambda)) = \frac{1}{n}\mathbb{E}(g' S(\lambda)' M_X S(\lambda) g) + \frac{1}{n}\mathbb{E}(\eta' S(\lambda)' M_X S(\lambda)\eta),$$

giving

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

Let $\mathbb{P}_1$ denote probability under $\mathcal{H}_1$ and define the neighbourhood $\mathcal{N}_\delta = \{\lambda : |\lambda - \lambda^\sharp| \leq \delta\}$ and $\bar{\mathcal{N}_\delta} = \Lambda \backslash \mathcal{N}_\delta$. The following chain of inequalities holds

$$\mathbb{P}_1(\hat{\lambda} \in \bar{\mathcal{N}_\delta}) \leq \mathbb{P}_1\big(\inf_{\bar{\mathcal{N}_\delta}} \mathcal{L}(\lambda) < \mathcal{L}(\lambda^\sharp)\big) \leq \mathbb{P}_1\big(\sup_{\Lambda} |\mathcal{L}(\lambda) - \tilde{\mathcal{L}}(\lambda)| \geq \inf_{\bar{\mathcal{N}_\delta}} |\tilde{\mathcal{L}}(\lambda) - \tilde{\mathcal{L}}(\lambda^\sharp)|\big).$$

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

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

which in turn implies

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

To complete the proof of Proposition 2 it suffices to verify the following two statements:

$$\inf_{\bar{\mathcal{N}}_\delta} \left(\tilde{\mathcal{L}}(\lambda) - \tilde{\mathcal{L}}(\lambda^\sharp)\right) > \epsilon, \text{ for all sufficiently large } n \text{ and for some } \epsilon > 0, \quad \text{(A.2.7)}$$

$$\sup_{\Lambda} |\mathcal{L}(\lambda) - \tilde{\mathcal{L}}(\lambda)| \to 0, \quad \text{as} \quad n \to \infty. \quad \text{(A.2.8)}$$

(A.2.7) follows from Assumption A. The LHS of (A.2.8) is bounded by

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

and so (A.2.8) follows as long as

$$\sup_{\Lambda} |\hat{\sigma}^2(\lambda) - \tilde{\sigma}^2(\lambda)| \to 0, \quad \text{(A.2.9)}$$

$$\underline{\lim}_{n\to\infty} \inf_{\Lambda} \tilde{\sigma}^2(\lambda) > 0. \quad \text{(A.2.10)}$$

Assumption B implies (A.2.10). To establish (A.2.9), we first verify pointwise convergence in probability of $\hat{\sigma}^2(\lambda) - \tilde{\sigma}^2(\lambda)$. Under $\mathcal{H}_1$,

$$\begin{aligned}
\hat{\sigma}^2(\lambda) - \tilde{\sigma}^2(\lambda) &= \frac{1}{n}\left(\eta' S(\lambda)' S(\lambda)\eta - \mathbb{E}(\eta' S(\lambda)' S(\lambda)\eta)\right) \\
&- \frac{1}{n^2}\left(\eta' S(\lambda)' \mathbb{X}\left(\frac{\mathbb{X}'\mathbb{X}}{n}\right)^{-1} \mathbb{X}' S(\lambda)\eta - \mathbb{E}\left(\eta' S(\lambda)' \mathbb{X}\left(\frac{\mathbb{X}'\mathbb{X}}{n}\right)^{-1} \mathbb{X}' S(\lambda)\eta\right)\right) \\
&+ \frac{1}{n}\left(g' N(\lambda) g - \mathbb{E}(g' N(\lambda) g)\right) + \frac{2}{n}\eta' N(\lambda) g \\
&= \frac{1}{n}\left(\eta' S(\lambda)' S(\lambda)\eta - \mathbb{E}(\eta' S(\lambda)' S(\lambda)\eta)\right) + \frac{1}{n}\left(g' N(\lambda) g - \mathbb{E}(g' N(\lambda) g)\right) \\
&+ \frac{2}{n}\eta' N(\lambda) g + o_p(1), \qquad \text{(A.2.11)}
\end{aligned}$$

where the last step follows by observing that $\mathbb{X}'\mathbb{X}/n$ converges to a non-singular $k \times k$ non-random matrix under Assumptions 2 and 9, $\mathbb{X}' S(\lambda)\eta/n = o_p(1)$ under Assumptions 2, 4 and 7 and $\mathbb{X}' S(\lambda) g/n = O_p(1)$ under Assumptions 2, 4 and 6. Thus, $\hat{\sigma}^2(\lambda) - \tilde{\sigma}^2(\lambda)$ has zero mean and variance bounded by

$$\begin{aligned}
&K\mathbb{E}\left(\frac{1}{n}\eta' S(\lambda)' S(\lambda)\eta - \mathbb{E}\left(\frac{1}{n}\eta' S(\lambda)' S(\lambda)\eta\right)\right)^2 + K\mathbb{E}\left(\frac{1}{n} g' N(\lambda) g - \mathbb{E}\left(\frac{1}{n} g' N(\lambda) g\right)\right)^2 \\
&+ K\mathbb{E}\left(\frac{2}{n} g' N(\lambda)\eta\right)^2, \qquad \text{(A.2.12)}
\end{aligned}$$

by the c_r inequality. Let $\mathbb{E}(\eta\eta'|\mathbb{X}) = \mathbb{E}(\eta\eta') = \Omega_\eta$, where $||\Omega_\eta||_\infty < K$ under Assumption 7. The first term in (A.2.12) is bounded by

$$\frac{K}{n^2} tr\left((S(\lambda)'S(\lambda)\Omega_\eta)^2\right) \leq \frac{K}{n}, \tag{A.2.13}$$

under Assumptions 4, 7 and C. Recall $\zeta(\lambda) = M_X S(\lambda) g$ and let $\bar{\zeta}(\lambda) = \Omega_\zeta^{-1/2}\zeta(\lambda)$. The second term in (A.2.12) is thus

$$\frac{1}{n^2}\left(E(\bar{\zeta}(\lambda)'\Omega_\zeta(\lambda)\bar{\zeta}(\lambda))^2 - (E(\bar{\zeta}(\lambda)'\Omega_\zeta(\lambda)\bar{\zeta}(\lambda)))^2\right) \leq \frac{K}{n^2}\left(\mu_\zeta(\lambda)'\Omega_\zeta\mu_\zeta(\lambda) + 2tr(\Omega_\zeta(\lambda)^2)\right) \leq \frac{K}{n} \tag{A.2.14}$$

under Assumption E. By the law of iterated expectations, the third term in (A.2.12) can be written as

$$\begin{aligned}
&\frac{4}{n^2}E\left(g'S(\lambda)'M_X S(\lambda)\eta\eta' S(\lambda)'M_X S(\lambda)g\right) = \frac{4}{n^2}E\left(g'S(\lambda)'M_X S(\lambda)\Omega_\eta S(\lambda)'M_X S(\lambda)g\right) \\
&= \frac{4}{n^2}E\left(\zeta(\lambda)'S(\lambda)\Omega_\eta S(\lambda)'\zeta(\lambda)\right) = \frac{4}{n^2}E(\bar{\zeta}(\lambda)'\Omega_\zeta(\lambda)^{1/2}\Omega_\eta\Omega_\zeta(\lambda)^{1/2}\bar{\zeta}(\lambda)) \\
&= \frac{4}{n^2}\mu_\zeta(\lambda)'\Omega_\eta\mu_\zeta(\lambda) + \frac{4}{n^2}tr(\Omega_\eta\Omega_\zeta(\lambda)) \leq \frac{K}{n}
\end{aligned} \tag{A.2.15}$$

under Assumptions 7 and E. Collecting (A.2.13), (A.2.14) and (A.2.15), pointwise convergence to zero of $\hat{\sigma}^2 - \tilde{\sigma}^2$ follows by Markov's inequality.

The uniform convergence in (A.2.9) 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 $\mathcal{N}_{\dagger\delta} = \{\lambda : |\lambda - \lambda^\dagger| \leq \delta\}$

$$\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} |(\hat{\sigma}^2(\lambda) - \tilde{\sigma}^2(\lambda)) - (\hat{\sigma}^2(\lambda^\dagger) - \tilde{\sigma}^2(\lambda^\dagger))| = O(\varepsilon) \tag{A.2.16}$$

The LHS of (A.2.16) is bounded by

$$\begin{aligned}
&K\left(\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} |\hat{\sigma}^2(\lambda) - \hat{\sigma}^2(\lambda^\dagger)| + \sup_{\mathcal{N}_{\dagger\delta}} |\mathbb{E}_1(\hat{\sigma}^2(\lambda) - \hat{\sigma}^2(\lambda^\dagger))|\right) \\
&\leq K\left(\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} |\hat{\sigma}^2(\lambda) - \hat{\sigma}^2(\lambda^\dagger)| + \sup_{\mathcal{N}_{\dagger\delta}} \mathbb{E}_1|\hat{\sigma}^2(\lambda) - \hat{\sigma}^2(\lambda^\dagger)|\right) \\
&\leq K\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} |\hat{\sigma}^2(\lambda) - \hat{\sigma}^2(\lambda^\dagger)|.
\end{aligned}$$

The last displayed term is

$$\frac{K}{n}\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} |y'(N(\lambda) - N(\lambda^\dagger))y| = \frac{K}{n}\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} |tr((N(\lambda) - N(\lambda^\dagger))yy')|, \tag{A.2.17}$$

which in turn is bounded by

$$\frac{K}{n}\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} \Big(||N(\lambda) - N(\lambda^\dagger)|| tr(yy') \Big) = \frac{K}{n}\mathbb{E}_1 \left(y'y \sup_{\mathcal{N}_{\dagger\delta}} ||N(\lambda) - N(\lambda^\dagger)|| \right)$$
$$\leq \frac{K}{n} \left(\mathbb{E}_1(y'y)^2\right)^{1/2} \left(\mathbb{E}_1(\sup_{\mathcal{N}_{\dagger\delta}} ||N(\lambda) - N(\lambda^\dagger)||)^2 \right)^{1/2},$$

where the second factor is $O(\varepsilon)$ for $\delta = \varepsilon/K$ as illustrated in the argument reported after Assumption D, and observing that $\mathbb{E}_1 \sup_{\mathcal{N}_{\dagger\delta}} ||N(\lambda) - N(\lambda^\dagger)||$ exists for each $\lambda \in \Lambda$ under Assumptions 2, 4 and 9. On the other hand

$$\frac{1}{n^2}\mathbb{E}_1(y'y)^2 = O(1) \tag{A.2.18}$$

under Assumptions 2, 6, 7, C and E.

Appendix 3: Extension to a Singular Case

This section analyzes a special case that cannot be accommodated under the Assumptions and framework of Section 4. In particular, we outlined in Section 4 that Assumption 10 is violated when the sum of the elements in each column of Q in (4.3) is zero. This happens, for instance, when elements in each column of W sum to the same constant, such as in a block diagonal or in a circulant structure. Incidentally, in case W has constant column sums, the second row of $\Psi(t, t_Y)$ in (4.5) is a $1 \times n$ vector of zeros, and the second component of $\psi(t, t_Y)$ in (4.6) equals zero. In order to accommodate the case of W having constant column sums, which, although unpopular among practitioners, might be of some interest, we re-define the building block of our test statistic $M(\hat{\theta}, t, t_Y)$ to allow the two components to have different normalization rates as

$$M_n^s(\hat{\theta}, t, t_Y) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \big(Y_i - m_i(\hat{\theta})\big) e^{t'X_i} \\ \frac{p_n}{n}\sum_{i=1}^{n} \big(Y_i - m_i(\hat{\theta})\big) e^{t_Y \frac{Y_i - \bar{Y}}{p_n}} - \frac{t_Y \hat{\sigma}^2}{n} tr(\hat{S}^{d'}\hat{Q}), \end{pmatrix} \tag{A.3.1}$$

where Q and S^d are again defined according to (4.3) and (4.4), and with the sequence p_n satisfying

$$p_n \to \infty, \quad \frac{p_n}{n} = o(1) \text{ and } \frac{\sqrt{n}}{p_n} = o(1). \tag{A.3.2}$$

The statistic in (A.3.1) with p_n in (A.3.2) and the limit theory in the sequel, can in principle be applied even when Assumption 10 with Ψ defined as in (4.5) is not violated. But a fast divergent p_n together with a slow convergence rate of the second component in (A.3.1) may compromise the power of the specification test unnecessarily. Some preliminary numerical work could be done to assess whether the model under $\mathcal{H}_0$ falls within the scope of Assumption 10, and hence of Theorem 2, or whether the statistic needs modification as in (A.3.1).

Similar to what defined in Section 4, we set the $2 \times n$ matrix

$$\Psi^s(t, t_Y) = \Psi^s(t, t_Y, \lambda_0, \beta_0, \mathbb{X}) = \begin{pmatrix} \underline{e}(t)' \\ t_Y \beta_0' \mathbb{X}' S^{d\prime} \end{pmatrix} Q, \tag{A.3.3}$$

the 2×1 vectors

$$\psi_1^s(t, t_Y) = \psi_1^s(t, t_Y, \lambda_0, \beta_0, \mathbb{X}) = -\frac{1}{n\sigma_0^2} \begin{pmatrix} \underline{e}(t)' \\ t_Y \beta_0' \mathbb{X}' S^{d\prime} \end{pmatrix} S^{-1} \begin{pmatrix} R\mathbb{X}\beta_0 & \mathbb{X} \end{pmatrix} \omega^{(1)} \tag{A.3.4}$$

and

$$\psi_2^s(t, t_Y) = \begin{pmatrix} 0_{1\times 1} \\ t_Y \end{pmatrix}. \tag{A.3.5}$$

For notational convenience, we define $\bar{sq} = (S^{d\prime}Q + Q'S^{d\prime})/2$. Also, we modify Assumptions 10 and 11 as follows.

Assumption 10$^{\mathbf{s}}$ *Conditionally on $\mathbb{X}$, $\lim_{n\to\infty} \frac{1}{n} \Psi^s(t, t_Y)\Psi^s(t, t_Y)'$ exists pointwise in $(t, t_Y)'$ and a.s. as $n \to \infty$, and is positive definite.*

Assumption 11$^{\mathbf{s}}$ *Conditionally on $\mathbb{X}$, the limits*

$$\lim_{n\to\infty} \frac{\psi_1^s(t,t_Y)\psi_1^s(t,t_Y)'}{n} tr\left((R+R')^2\right), \ \lim_{n\to\infty} \frac{1}{n} tr\left((S^{d\prime}Q + Q'S^d)^2\right),$$
$$\lim_{n\to\infty} \frac{\psi_1^s(t,t_Y)\psi_2^s(t,t_Y)'}{n} tr\left((S^{d\prime}Q + Q'S^d)(R+R')\right), \ \lim_{n\to\infty} \frac{\psi_1^s}{n} \sum_{i=1}^n \bar{r}_{ii}(\Psi_i^s(t,t_Y)')', \ \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \bar{sq}_{ii}(\Psi_i^s(t,t_Y)')',$$
$$\lim_{n\to\infty} \frac{\psi_1^s \psi_1^{s\prime}}{n} \sum_{i=1}^n \bar{r}_{ii}^2, \ \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \bar{sq}_{ii}^2, \ \lim_{n\to\infty} \frac{\psi_1^s \psi_2^{s\prime}}{n} \sum_{i=1}^n \bar{r}_{ii}\bar{sq}_{ii}$$

exist pointwise in $(t, t_Y)'$ and a.s. as $n \to \infty$.

In order to prove the following Theorem 4, we modify Assumption R to

Assumption R' *Let p_n and α_n be deterministic, positive sequences satisfying (A.3.2), $\alpha_n \to \infty$ as $n \to \infty$, and*

$$\frac{\alpha_n}{p_n} \to 0, \qquad \frac{n^{3/2}}{\alpha_n^{4+\delta}} \to 0 \tag{A.3.6}$$

as $n \to \infty$, where $\delta > 0$ is determined by Assumption 1.

Assumption R' is satisfied, for instance, with $p_n = n^{2/3}$ and $\alpha_n = n^{1/3}$, as long as $\delta > 1/2$. Now let

$$\begin{aligned}
A_n^s(t,t_Y) &= \frac{\sigma_0^2}{n}\Psi^s(t,t_Y)\Psi^s(t,t_Y)' + \frac{\sigma_0^4 tr\left((R+R')^2\right)}{2n}\psi_1^s(t,t_Y)\psi_1^s(t,t_Y)' \\
&+\frac{\sigma_0^4 tr\left((S^{d\prime}Q+Q'S^d)^2\right)}{2n}\psi_2^s(t,t_Y)\psi_2^s(t,t_Y)' + \frac{\sigma_0^4 tr\left((S^{d\prime}Q+Q'S^d)(R+R')\right)}{2n}\psi_1^s(t,t_Y)\psi_2^s(t,t_Y)' \\
&+\frac{2\mu^{(3)}}{n}\left(\psi_1^s(t,t_Y)\sum_{i=1}^n \bar{r}_{ii}(\Psi_i^s(t,t_Y)')' + \psi_2^s(t,t_Y)\sum_{i=1}^n \bar{s}q_{ii}(\Psi_i^s(t,t_Y)')'\right) \\
&+\frac{(\mu^{(4)}-3\sigma_0^4)}{n}\left(\psi_1^s(t,t_Y)\psi_1^s(t,t_Y)'\sum_{i=1}^n \bar{r}_{ii}^2 + \psi_2^s(t,t_Y)\psi_2^s(t,t_Y)'\sum_{i=1}^n \bar{s}q_{ii}^2 + 2\psi_1^s(t,t_Y)\psi_2^s(t,t_Y)'\sum_{i=1}^n \bar{r}_{ii}\bar{s}q_{ii}\right).
\end{aligned} \tag{A.3.7}$$

With these modifications the limit behavior of $M_n^s(\hat{\theta},t,t_Y)$ can now be obtained.

Theorem 4 *Let Assumptions 1-5, 9, 10^s, 11^s and R' hold. Let p_n be a non-negative sequence satisfying (A.3.2). Under $\mathcal{H}_0$ in (3.13), as $n \to \infty$*

$$\sqrt{n}M^s(\hat{\theta},t,t_Y) \to_d \mathcal{N}\left(0, A^s(t,t_Y)\right), \tag{A.3.8}$$

pointwise in $(t,t_Y)'$, conditionally on $\mathbb{X}$, where the standardizing variance-covariance matrix of $\sqrt{n}M^s(\hat{\theta},t,t_Y)$ is given by $A^s(t,t_Y) = \lim_{n\to\infty} A_n^s(t,t_Y)$, where $A_n^s(t,t_Y)$ is defined in (A.3.7).

The corresponding limit theory for the test statistic $\hat{T}^s(t,t_Y)$ under $\mathcal{H}_0$ is

$$\hat{T}^s(t,t_Y) = nM^s(\hat{\theta},t,t_Y)'\hat{A}^{s\ -1}(t,t_Y)M^s(\hat{\theta},t,t_Y) \to_d \chi_2^2. \tag{A.3.9}$$

In addition, Theorem 3 in Section 5 goes through with $\hat{T}(t,t_Y)$ replaced by $\hat{T}^s(t,t_Y)$ under the same set of Assumptions as those in Theorem 3. In particular, the second part of the proof of Theorem 3 (in the Appendix) holds with minor modifications to the rates of the component terms in the limit. The dominance of the relevant term concerning the second component of $M^s(\hat{\theta},t,t_Y)$, which is needed for the consistency of the test in (A.3.9), is assured by the second part of Assumption 13.

Proof of Theorem 4

The proof follows closely that of Theorem 2, and we only report the few necessary modifications to avoid repetition. By a Taylor expansion as in (A.1.18) - (A.1.20) and by a similar argument to that adopted in the proof of Theorem 1, (A.1.21) becomes

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\underline{e}(t)'S^{-1}\epsilon \\ \frac{p_n}{\sqrt{n}}\left(1'+\frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i|\leq K\alpha_n)(Y'-\bar{Y}1')\right)S^{-1}\epsilon-\frac{t_Y}{n}\sigma_0^2tr(S^{d'}Q)\end{pmatrix}$$
$$-\begin{pmatrix} \frac{1}{n}\underline{e}(t)' \\ \frac{p_n}{n}\left(1'+\frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i|\leq K\alpha_n)(Y'-\bar{Y}1')\right)\end{pmatrix}S^{-1}\begin{pmatrix}R\mathbb{X}\beta_0 & \mathbb{X}\end{pmatrix}\Omega^{-1}\begin{pmatrix}\frac{1}{\sigma^2\sqrt{n}}\beta_0'\mathbb{X}'R'\epsilon \\ \frac{1}{\sigma^2\sqrt{n}}\mathbb{X}'\epsilon\end{pmatrix}$$
$$-\begin{pmatrix} \frac{1}{n}\underline{e}(t)' \\ \frac{p_n}{n}\left(1'+\frac{t_Y}{p_n}\mathbb{P}(\sup_i|\epsilon_i|\leq K\alpha_n)(Y'-\bar{Y}1')\right)\end{pmatrix}S^{-1}\begin{pmatrix}R\mathbb{X}\beta_0 & \mathbb{X}\end{pmatrix}\Omega^{-1}\begin{pmatrix}\frac{1}{\sigma^2\sqrt{n}}(\epsilon'R\epsilon-\sigma^2tr(R)) \\ 0\end{pmatrix}$$
$$+O_p\left(\max\left(\frac{1}{\sqrt{n}},\frac{\sqrt{n}}{p_n},\frac{n^{3/2}}{\alpha_n^{4+\delta}}\right)\right)=O_p\left(\max\left(\frac{\sqrt{n}}{p_n},\frac{n^{3/2}}{\alpha_n^{4+\delta}}\right)\right). \tag{A.3.10}$$

By substituting the expression for $Y'-\bar{Y}1'$, using $1'Q=0_{1\times n}$ and

$$1'S^{-1}\begin{pmatrix}R\mathbb{X}\beta_0 & \mathbb{X}\end{pmatrix}\omega^{(1)}=0,$$

(A.1.22) becomes

$$\frac{1}{\sqrt{n}}\Psi^s\epsilon+\frac{1}{\sqrt{n}}\psi_1^s(\epsilon'R\epsilon-\sigma_0^2tr(R))+\frac{1}{\sqrt{n}}\psi_2^s(\epsilon'^{d\prime}Q\epsilon-\sigma_0^2tr(S^{d\prime}Q))+O_p\left(\max\left(\frac{\sqrt{n}}{p_n},\frac{n^{3/2}}{\alpha_n^{4+\delta}}\right)\right). \tag{A.3.11}$$

The rest of the proof follows that of Theorem 2 and is omitted. ■

References

[1] Allers, M.A. and P. Elhorst (2005), Tax Mimicking and Yardstick Competition Among Local Governments in the Netherlands, *International Tax and Public Finance*, 12, 493-513.

[2] Anselin, L. (2001), Rao's score test in spatial econometrics, *Journal of Statistical Planning and Inference* 97, 113–139.

[3] Baltagi, B.H. and, Li, D. (2001), LM tests for functional form and spatial error correlation, *International Regional Science Review* 24, 194–225.

[4] Beenstock M. and D. Felsenstein (2012), Nonparametric estimation of the spatial connectivity matrix using spatial panel data, *Geographical Analysis*, 44, 386-397.

[5] Bailey, N., Holly, S., and Pesaran, M. H. (2016), A two stage approach to spatio-temporal analysis with strong and weak cross sectional dependence, *Journal of Applied Econometrics*, 31, 249-280.

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

[7] Bierens, H.J. (1984), Model specification testing of time series regressions, *Journal of Econometrics*, 26, 323-353.

[8] Bierens, H.J. (1988), ARMA memory index modeling of economic time series (with discussion), *Econometric Theory*, 4, 35-59.

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

[10] Burridge, P. (1980), On the Cliff-Ord test for spatial correlation, *Journal of the Royal Statistical Society, Series B* , 42, 107-8.

[11] Cliff, A. and J.K. Ord (1981), *Spatial Processes: Models & Applications*, Pion.

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

[13] Chirinko, R. S. and Wilson, D. J (2017), Tax competition among U.S. states: Racing to the bottom or riding on a seesaw?, *Journal of Public Economics*, 155, 147-163.

[14] 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.

[15] Debarsy, N. and C. Ertur (2019), Interaction matrix selection in spatial autoregressive models with an application to growth theory, *Regional Science and Urban Economics*, 75, 49-69.

[16] de Jong, R.M. (1996), The Bierens test under data dependence, *Journal of Econometrics*, 72, 1-32.

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

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

[19] Gupta, A. and X. Qu (2020), Consistent specification testing under spatial dependence. Mimeo.

[20] 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.

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

[22] Kasparis, Y (2010), The Bierens test for certain nonstationary models, *Journal of Econometrics*, 158, 221-230 .

[23] Kelejian H. H. (2008), A spatial J-test for model specification against a single or a set of non-nested alternatives, *Letters in Spatial and Resource Sciences* 1, 3–11.

[24] Kelejian, H. H. and G. Piras (2011), An Extension of Kelejian's J-test for non-nested spatial models, *Regional Science and Urban Economics*, 41, 281-292.

[25] Kelejian, H. H. and G. Piras (2016), An Extension of the J-test to a spatial panel data framework, *Journal of Applied Econometrics*, 31, 387-402.

[26] 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.

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

[28] 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.

[29] Lam C. and P. Souza (2016), Detection and Estimation of Block Structure in Spatial Weight Matrix, *Econometric Reviews*, 35, 1347-1376.

[30] Lam C. and P. Souza (2020), Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model, *Journal of Business and Economics Statistics*, 38, 693-710.

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

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

[33] Lee, L.F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137, 489-514.

[34] Lee, L.F. and J. Yu (2012). The $C(\alpha)$ -type gradient test for spatial dependence in spatial autoregressive models. *Letters in Spatial and Resource Sciences*, 5, 119-135.

[35] Liu, T. and L.F. Lee (2019), A likelihood ratio test for spatial model selection, *Journal of Econometrics*, 213, 434-458.

[36] Liu, X. and L.F. Lee (2018), A robust test for network generated dependence, *Journal of Econometrics*, 207, 92-113.

[37] Lyytikäinen, T. (2012),Tax competition among local governments: Evidence from a property tax reform in Finland, *Journal of Public Economics*, 96, 584-595.

[38] Martellosio, F. (2012) Testing for spatial autocorrelation: The regressors that make the power disappear. *Econometric Reviews*, 31, 215-240.

[39] Newey, W. K. (1985), Maximum likelihood specification testing and conditional moment tests, *Econometrica*, 53, 1047-1070.

[40] Ord, J.K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, 120-6.

[41] Parchet, R. (2019), Are Local Tax Rates Strategic Complements or Strategic Substitutes?, *American Economic Journal: Economic Policy*, 11(2), 189-224.

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

[43] Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least Squares Regression Analysis, *Journal of the Royal Statistical Society, Series B*, 31, 350-371.

[44] Robinson, P.M. (2008). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics*, 147, 5-16.

[45] Scott, D.J. (1973). Central limit theorems for martingales using a Skorokhod representation approach. *Advances in Applied Probability*, 5, 119-137.

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

[47] Vuong, Q. (1989), Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, 57, 307-333.