Econometric Theory, **0**, 2024, 1–49.
doi:10.1017/S0266466624000173

HETEROSKEDASTICITY ROBUST SPECIFICATION TESTING IN SPATIAL AUTOREGRESSION

JUNGYOON LEE
Royal Holloway, University of London

PETER C. B. PHILLIPS
Yale University, University of Auckland, and Singapore Management University

FRANCESCA ROSSI
University of Verona

Spatial autoregressive (SAR) and related models offer flexible yet parsimonious ways to model spatial and network interactions. 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. Also, 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 a new specification test within the SAR framework that can detect general forms of misspecification including that of the spatial weight matrix, the functional form and the model itself. The test is robust to the presence of heteroskedasticity of unknown form in the disturbances and the approach relates to the conditional moment test framework of Bierens ([1982, *Journal of Econometrics* 20, 105–134], [1990, *Econometrica* 58, 1443–1458]). The Bierens test is shown to be inconsistent in general against spatial alternatives and the new test introduces modifications to achieve test consistency in the spatial setting. A central element is the infinite-dimensional endogeneity induced by spatial linkages. This complexity is addressed by introducing a new component to the omnibus test that captures the effects of potential spatial matrix misspecification. With this modification, the approach leads to a simple pivotal test procedure with standard critical values that is the first test in the literature to have power against misspecifications in the spatial linkages. We derive the asymptotic distribution of the test under the null hypothesis of correct SAR specification and prove consistency. A Monte Carlo study is conducted to study its finite sample performance. An empirical illustration on the performance of the test in modeling tax competition in Finland is included.

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 for sharing his dataset. Address correspondence to Jungyoon Lee, Department of Economics, Royal Holloway, University of London, Egham, United Kingdom; e-mail: Jungyoon.Lee@rhul.ac.uk.

1. INTRODUCTION

The past two decades have witnessed a surge in theoretical and empirical research on the class of spatial econometric models known as spatial autoregressions (SARs). These models were first suggested by Cliff and Ord (1968) and have since been widely extended in directions to suit applied research in many different fields. 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 the theoretical work 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; Lee, 2004) and generalized methods of moments (e.g., Kelejian and Prucha, 1999; 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 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), Anselin (2001), Kelejian and Prucha (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 pre-specified 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 (2017) 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, spatiotemporal, or panel data structures. More recently, Gupta and Qu (2021) 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 (2021) 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 methods do 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 (2016, 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 heteroskedasticity in the error terms. A selection strategy for the correct network structure has also been suggested by Bailey, Holly, and Pesaran (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 (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 (2019) is derived under the assumptions of near-epoch dependence

(NED) (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; Pinkse, Slade, and Brett, 2002, among others).

The goal of the present paper is to complement the above approaches by developing a new omnibus test procedure that can detect general forms of misspecification related to the model, the weight matrix and the functional form of the SAR model. The disturbances are allowed to have heteroskedasticity of unknown form. The approach adopted takes Bierens (1990) conditional moment tests as its starting point, but the formulation of the test is specifically designed to address potential weight matrix misspecification and deliver test consistency in this wider context as well as the regression model formulation. The key feature of general SAR models is the asymptotically infinite-dimensional endogeneities that are induced by spatial matrix linkages. This complexity is addressed by introducing a new component to omnibus testing that is specifically designed to capture these effects and thereby enable the detection of spatial matrix misspecification.

The literature on consistent conditional moment tests has been explored in econometric work from the 1980s (Bierens, 1982; Newey, 1985) and relies on orthogonality condition tests that date back to Ramsey (1969). Under the null hypothesis of correct specification of the regression function, the moment conditions hold 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 nonstationary models in Kasparis (2010). Stinchcombe and White (1998) and Escanciano (2006) have studied the source of test consistency in the Bierens methodology, and positioned the test within a larger class of consistent specification tests that encompasses the tests of Stute (1997) and Escanciano (2006), some of which have been developed further in, for example, Koul and Stute (1999), Stute and Zhu (2005), and Escanciano (2007).

The present paper suggests a test of correct model specification designed particularly for the spatial setting, where outcomes are influenced not only by own individual characteristics but also by the characteristics of neighbors. As indicated briefly above, the design involves the challenge of accounting for the endogenous interactions generated by the spatial weight matrix, which has increasing dimensions by construction and induces neither a natural ordering of observations, nor an obvious notion of decay in correlation strength, unlike the conditions that are conventional in the time series dependence literature. An additional challenge in the spatial setup is the heterogeneity of the regression functions across individuals, a feature that is somewhat similar to the complications entailed by specification testing for nonstationary time series models (e.g., Bierens, 1984).

Similar to Bierens (1990), we define the moment conditions using the moment generating function of individual random variables as weighting functions, rather than their characteristic function, to assist in the derivation of standard limit theory. A distinguishing feature of our test is that while it is constructed using Bierens-type infinite moment conditions driven by a real parameter, it does not rely on such a parameter choice to achieve consistency, unlike the standard Bierens test literature. Instead, the modifications designed to accommodate the spatial setting give rise to a local-to-zero sequence of real parameters complemented by linearization and centering of Bierens-like moment conditions. This in turn leads to a practical test statistic whose properties in the limit no longer depend on the local-to-zero sequence of such a parameter. Thus, our test does not rely on devices such as selection by randomization or supremum/square integral transformation to achieve consistency. Linearization via the local-to-zero sequence, in addition to the choice of the moment generating function over characteristic function in the moment conditions (as in the approach of Bierens (1990) rather than that of Bierens (1982)) allows us to obtain a standard limiting distribution under correct specification which in turn facilitates implementation compared to simulation-based approaches such as a bootstrap method. The advantage of obtaining a standard null distribution comes, therefore, at the cost of choosing the additional local-to-zero tuning sequence. But the use of additional sequences and related assumptions to achieve consistency and standard critical values is in no way new in the literature of conditional moment testing.[1]

In our development, we assume a SAR structure with spatial dependence as a spatial lag and with disturbances that may be heteroskedastic. This framework is a significant base model of interest in the spatial literature and the kernel of more general formulations. Our conditional moment testing approach, with individual outcomes depending on neighbor outcomes and heterogeneous regression functions, should be relevant in other settings. A primary advantage in the approach rests in its applicability to general "spatial" data, where "space" is interpreted more broadly than geographic with no reliance being 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 its limit behavior under general alternatives.

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 (SD) 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

[1] For instance, Bierens (1990) (Theorem 4) and Kasparis (2010) (Lemma 7) rely on a threshold determined by a diverging sequence to derive standard limit theory of their test statistics; and de Jong and Bierens (1994) employ an infinite series of moment conditions whose cardinality needs to increase with the sample size in a specific manner in order to achieve test consistency.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

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 the simple SAR framework to capture dependence structures in the data more satisfactorily.

The paper is organized as follows. The next section presents the model setup, while Section 3 details the limit theory under the null hypothesis of correct specification. Section 4 reports limit theory under a fixed generic alternative, while also showing inconsistency of the Bierens test in the spatial setting. Simulation findings are presented in Section 5. Section 6 provides a heuristic discussion to guide practitioners in the choice of tuning parameters and to interpret test results. Section 7 gives an empirical tax competition illustration of the methods using the model framework and datasets of Lyytikäinen (2012), who dealt with tax competition across Finnish municipalities. Some conclusions and possible extensions are given in Section 8 and proofs are in the Appendix. We also provide a self-contained Supplementary Material with codes for implementing the specification test developed in the paper, along with a small simulated dataset for illustration.

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 SETUP AND MOTIVATION

The so-called mixed regressive SAR model admits a triangular array structure

$$Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n, \tag{2.1}$$

where Y_n is an n-vector of observations, X_n an $n \times k$ matrix of regressors, ϵ_n an n-vector of independent *structural* disturbances, and W_n a sequence of $n \times n$ prespecified spatial weight matrices.[2]

Define $S_n(\lambda) = I_n - \lambda W_n$ and $R_n(\lambda) = W_n S_n^{-1}(\lambda)$. Provided model (2.1) has an equilibrium solution (i.e., under invertibility of $S_n(\lambda)$), the *reduced form* of (2.1) is given by

$$Y_n = S_n^{-1}(\lambda)(X_n \beta + \epsilon_n). \tag{2.2}$$

For individual observations $i = 1, \ldots, n$, the last displayed expression leads to a regression model of the form $Y_{in} = m_{in}(X_n, \lambda, \beta) + u_{in}(\lambda)$, with

[2]To develop inference on the unknown parameters $(\lambda, \beta')'$, some form of normalization of the weight matrix is typically required for identification and stability, inducing a triangular array structure for $W = W_n$. Allowing $Y = Y_n$, $X = X_n$ and $\epsilon = \epsilon_n$ permits relabeling of observations, often required due to the lack of a natural ordering in data that are recorded in space.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

$$m_{in}(X_n,\lambda,\beta) = S_n^{(i)}(\lambda)'X_n\beta = \sum_{j=1}^{n} s_n^{ij}(\lambda)X_{jn}'\beta, \quad \text{and} \quad u_{in}(\lambda) = \sum_{j=1}^{n} s_n^{ij}(\lambda)\epsilon_{jn}, \tag{2.3}$$

where $s_n^{ij}(\lambda)$ denotes the (i,j)th element of $S_n^{-1}(\lambda)$ and $u_{in}(\lambda)$ is the reduced form error of the SAR model.[3]

Model (2.3) is a particular parameterization of a general nonparametric regression of the type

$$Y_{in} = g_{in}(X_n) + \eta_{in},\ \mathbb{E}(\eta_{in}|X_n) = 0,\ \ i = 1,\ldots,n, \tag{2.4}$$

where, as previously defined, $X_n = (X_{1n},\ldots,X_{nn})'$ is an $n\times k$ 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}, namely, $g_{in}(X_n) = \mathbb{E}(Y_{in}|X_n),\ i = 1,\ldots,n$. By defining $g_{in}(\cdot)$ as a function of the entire set of observations X_n, we characterize the above model as a spatial one, whereby individual outcomes are influenced not only by their own characteristics but also by the characteristics of their neighbors. Unlike, for example, Bierens (1990), the general model in (2.4) needs to accommodate a weighting structure such as that in (2.3) and hence we also allow for possible heterogeneity across individuals in the regression function $g_{in}(\cdot)$, as well as dependence on the sample size n. After introducing hypotheses of interest below, we discuss the advantage of basing our analysis on the reduced-form model in (2.2), rather than the structural model (2.1).

The primary concern of the present paper lies in developing an empirical test of whether the regression function $m_{in}(\cdot)$ and error structure $u_{in}(\cdot)$ in (2.3) are a correct parameterization of the unknown $g_{in}(\cdot)$ and error structure η_{in} in (2.4). In what follows we let $\vee$ and $\wedge$ denote the usual *or* and *and* logical operators. Denote $\theta = (\lambda,\beta')'$. For all sufficiently large n, we define the set

$$J_n(\theta) = \{i : \mathbb{P}\big(g_{in} = m_{in}(\theta)\big) < 1\ \vee\ \mathbb{P}(u_{in}(\lambda) = \eta_{in}) < 1\}, \tag{2.5}$$

and let $\text{card}(J_n(\theta))$ denote its cardinality, which measures the extent to which correct specification fails among the observed units. We define the following sets of null and alternative hypotheses.[4]

Null and Alternative Hypotheses
For all sufficiently large n,

$$\mathcal{H}_0:\ \mathbb{P}(m_{in}(\theta_0) = g_{in}) = 1\ \wedge\ \mathbb{P}(u_{in}(\lambda_0) = \eta_{in}) = 1 \tag{2.6}$$

for some $\theta_0 \in \Theta$ and for all $i = 1,\ldots,n$.

$$\mathcal{H}_1:\ \mathbb{P}(m_{in}(\theta) = g_{in}) < 1\ \vee\ \mathbb{P}(u_{in}(\lambda) = \eta_{in}) < 1 \tag{2.7}$$

for all $\theta \in \Theta$, for sufficiently many $i = 1,\ldots,n$ such that $\text{card}(J_n(\theta)) \sim n$ *as $n \to \infty$.*

[3] When $\beta = 0$ ex ante, model (2.3) corresponds to a pure SAR model, in which exogenous regressors are postulated to be jointly irrelevant.

[4] Given the triangular array structure of the quantities involved in the statements below, it would be more precise to index these hypotheses as $\mathcal{H}_{0n}$ and $\mathcal{H}_{1n}$, but the subscript n is omitted for notational simplicity.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

The formulation of the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ is necessarily asymptotic because the SAR model itself is infinite-dimensional. Correspondingly, the statistical test of the null $\mathcal{H}_0$ also relies on asymptotic arguments. More specifically, it is designed to detect an increasing number of potentially misspecified (reduced form) SAR regression functions $m_i(\theta)$ and/or errors $u_i(\lambda)$. Thus, to achieve a test with non-negligible power against violations of $\mathcal{H}_0$ in various directions of departure, the rate condition $\text{card}(J_n(\theta)) \sim n$ under $\mathcal{H}_1$ ensures that the number of units for which misspecification does occur (i.e., the specified functions $m_i(\theta)$ and/or errors $u_i(\lambda)$ are violated in the data) grows as fast as the number of units n. However, it should be noted that the rate assumption $\text{card}(J_n(\theta)) \sim n$ under $\mathcal{H}_1$ is a relevant theory condition arising from the nature of the SAR model rather than a practical limitation.

This is due to a key advantage of considering the reduced-form model in (2.3), rather than the structural model (2.1) arising from the presence of $S_n^{-1}(\cdot)$ in (2.3) that amplifies even sparse/finite deviations from the true W_n to have non-sparse impacts on the system (2.3). The increasing dimensionality of W_n in the structural regression function presents a potentially infinite-dimensional misspecification error and poses a challenge to formulating a straightforward extension of the Bierens test to detect misspecification in the weight matrix. Test consistency against any deviation from the true network structure embodied in W_n may seem, in principle, difficult to achieve due to the increasing dimensionality of W_n as $n \to \infty$. Prima facie, the increasing dimension of W_n might indeed suggest only limited effects from deviations of individual elements W_n from the true weights. Importantly, however, even if there are only sparse deviations from the true weight matrix W_n the effects on the matrix inverse $S_n^{-1}(\cdot)$ of these deviations are not sparse and it is the impact of the specification of W_n on $S_n^{-1}(\cdot)$ that is relevant in the reduced form (2.3) and hence on the data. Indeed, under mild conditions that will be presented later, the following power series expansion of the inverse matrix $S_n^{-1}(\lambda)$ holds

$$S_n^{-1}(\lambda) = (I_n - \lambda W_n)^{-1} = \sum_{j=0}^{\infty} (\lambda W_n)^j. \tag{2.8}$$

From the power series (2.8), it is evident that even a small number of misspecified elements in W_n typically leads to all elements of $S_n^{-1}(\cdot)$ being misspecified. Heuristically, therefore, if a test is able to detect an increasing number of misspecified elements via the role that the inverse matrix $S_n^{-1}(\cdot)$ has in determining $m_{in}(\cdot)$ and/or $u_{in}(\cdot)$, this property will be sufficient to detect the effects of a finite number of misspecified elements in W_n. This is so even when the weight matrix W_n is itself sparse. In this respect, we expect the test developed below to be general enough to identify misspecification in spite of the limitation entailed by W_n having increasing dimensions. We stress that reduced-form spatial autoregressions induce a particular functional form in the error structure that needs to be considered

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

alongside the specification of the regression function, distinguishing it from nonspatial models hitherto considered in the Bierens specification test literature.

For notational simplicity in the sequel we will mostly suppress the subscript n to random and deterministic sequences appearing in our derivations, unless highlighting their dependence on n is important. Similarly, it is convenient to do so in other cases, such as using $S^{-1}(\lambda)$ in place of $S_n^{-1}(\lambda)$.

Before presenting modifications designed to suit spatial model frameworks in the next section, we first demonstrate the inadequacy of a naive implementation of existing Bierens tests in delivering a consistent test in our spatial settings. In nonspatial settings, specification tests in Bierens (1990) rely on a moment condition such as

$$E((Y_i - m_i(\theta))e^{t'X_i}) = 0 \text{ for some } \theta \in \Theta, \tag{2.9}$$

where $t \in \mathbb{R}^k$. Although the exponential in the above moment condition can be replaced by other choices of functions, we retain it as in Bierens (1990) to preserve the tractability of the limit theory, as discussed in the Introduction. The theoretical justification of a consistent test based on (2.9) is given by Lemma 1 in Bierens (1990), which in the setup of (2.3) and (2.4) would amount to proving equivalence between

$$\mathbb{P}(\mathbb{E}((Y_i - m_i(\theta))|X) = 0) = 1 \Leftrightarrow \mathbb{E}\left((Y_i - m_i(\theta))e^{t'X_i}\right) = 0 \tag{2.10}$$

for almost all $t \in \mathbb{R}^k$ and for all $i = 1, \ldots, n$, under correct specification. In our spatial setup defined in (2.3) and (2.4), it is straightforward to conclude the "$\Rightarrow$" implication, but "$\Leftarrow$" in (2.10) does not hold in general. Instead one can only prove the following implication which is not general enough for our purpose.

Claim 1.

$$\mathbb{E}\left((Y_i - m_i(\theta))e^{t'X_i}\right) = 0 \Longrightarrow \mathbb{P}(\mathbb{E}((Y_i - m_i(\theta))|X_i) = 0) = 1 \tag{2.11}$$

for all $t \in \mathbb{R}^k$ up to a zero-measured set.

The proof of Claim 1 is in the Appendix. For independent $X_j, j = 1, \ldots, n$, from Claim 1, we deduce that a test based on (2.9) offers a necessary and sufficient condition for correct specification in case the difference $m_i(\cdot) - g_i(\cdot)$ only depends on X_i, but this represents only a limited and unlikely case in our spatial setup. We are, in general, not able to conclude $\mathbb{E}\left((Y_i - m_i(\theta))e^{t'X_i}\right) = 0 \Longrightarrow \mathbb{P}(\mathbb{E}((Y_i - m_i(\theta))|X) = 0) = 1$, which is instead required to establish consistency of the simple Bierens test based on (2.9) when a weighting structure that connects all the X_j, for $j = 1, \ldots, n$, enters into the regression function $m_i(\theta_0)$. Thus, in general, a test based on (2.9) in the spatial setup would mimic a necessary, but not a sufficient condition to conclude correct specification of $m_i(\cdot)$ in (2.3). Fundamentally, $X_1, \ldots, X_n$ rather than X_i is the relevant conditioning set for the regression function in our spatial setting and (2.9) is therefore inadequate to provide a consistent test.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

In addition, the argument in the previous paragraph holds for each i. Due to heterogeneity of g_i and m_i across i, instead of considering individual moment condition (2.9), one should in fact consider the following average moment condition:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left((Y_i-m_i(\theta))e^{t'X_i}\right),\quad t\in\mathbb{R}^k, \tag{2.12}$$

and rely on a weak law of large numbers for heterogeneous data when formulating a test statistic such as

$$\frac{1}{n}\sum_{i=1}^{n}\left((Y_i-m_i(\hat{\theta}))e^{t'X_i}\right), \tag{2.13}$$

where $\hat{\theta}$ denotes any estimator that is consistent under $\mathcal{H}_0$ and has a well-defined probability limit to a pseudo-sequence $\theta_n^\sharp=\theta^\sharp$ under $\mathcal{H}_1$. In Claim 2 reported in Section 4, we will provide a counterexample of the failure of a test based on (2.13).

3. TEST STATISTIC AND LIMIT THEORY UNDER $\mathcal{H}_0$

We first discuss estimation before introducing the new test statistic below. Heuristically, under suitable regularity conditions and assumptions on the behavior of W as n increases, the model parameters $\theta=(\lambda,\beta')'$ can be estimated by minimizing a suitable objective function $\mathbb{Q}$ over a compact parameter space Θ, giving

$$\hat{\theta}=\underset{\theta\in\Theta}{\operatorname{argmin}}\ \mathbb{Q}(\theta). \tag{3.1}$$

Limit theory is established by means of standard consistency criteria and suitable central limit theorems for triangular arrays. In this paper, we estimate θ by IV/2SLS since it has the advantage of a relatively simple closed form and is robust to unknown heteroskedasticity. Alternative estimation methods may be considered at the cost of some algebraic modifications.

Define $\mathbb{Z}=\mathbb{Z}_n$ as $\mathbb{Z}=[Z;X]$, with $Z=Z_n$ being an $n\times(p-k)$ matrix of suitable instruments, with $p\geq k+1$ and $\mathbb{X}=[WY;X]$. The IV/2SLS objective function is then

$$\mathbb{Q}=\frac{1}{n}(Y-\mathbb{X}\theta)'\mathbb{P}_{\mathbb{Z}}(Y-\mathbb{X}\theta), \tag{3.2}$$

where $\mathbb{P}_{\mathbb{Z}}=\mathbb{P}_{n,\mathbb{Z}}=\mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'$. From standard arguments (e.g., Kelejian and Prucha, 1998), the IV/2SLS estimate of θ_0 is given by $\hat{\theta}=\left(\mathbb{X}'\mathbb{P}_{\mathbb{Z}}\mathbb{X}\right)^{-1}\mathbb{X}'\mathbb{P}_{\mathbb{Z}}Y$, and

$$\sqrt{n}(\hat{\theta}-\theta_0)=\left(\mathbb{B}'\mathbb{A}^{-1}\mathbb{B}\right)^{-1}\mathbb{B}'\mathbb{A}^{-1}\frac{1}{\sqrt{n}}\mathbb{Z}'\epsilon+O_p\left(\frac{1}{\sqrt{n}}\right), \tag{3.3}$$

where $\mathbb{A}$ and $\mathbb{B}$ are defined as

$$\operatorname{plim}_{n\to\infty}\frac{1}{n}\mathbb{Z}'\mathbb{Z}=\mathbb{A} \tag{3.4}$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

and

$$\text{plim}_{n\to\infty}\frac{1}{n}\mathbb{Z}'\mathbb{X} = \mathbb{B}, \tag{3.5}$$

respectively.

Let $R = R(\theta_0) = WS^{-1}(\lambda_0)$ and define

$$Q = Q(\theta_0) = S^{-1}(\theta_0)\left(I_n - \frac{1}{n}\left(RX\beta_0, \quad X\right)\left(\mathbb{B}'\mathbb{A}^{-1}\mathbb{B}\right)^{-1}\mathbb{B}'\mathbb{A}^{-1}\mathbb{Z}'\right). \tag{3.6}$$

Set $\bar{Y} = \sum_i^n Y_i/n$ and let $\bar{X}$ be the k-vector containing the sample averages of the components of X. Also the $1\times n$ vector of the column averages of S^{-1} is denoted by $\bar{S}^{-1'} = \sum_{i=1}^n S^{(i)'}/n$, and the column-demeaned version of S^{-1} is

$$S^d = S^{-1} - 1\bar{S}^{-1'}. \tag{3.7}$$

Let $\underline{e}(t) = (e^{t'(X_1-\bar{X})},\ldots,e^{t'(X_n-\bar{X})})'$ and $f(t) = (\underline{e}(t), 1_n)'$, with 1_n being an n-vector of ones. Define the $2\times n$ matrix

$$\Psi(t) = \Psi(t,\lambda_0,\beta_0,\mathbb{X}) = f(t)'Q. \tag{3.8}$$

In general, we indicate estimated counterparts (evaluated at $\hat{\theta}$) of previously defined quantities by $\hat{(\cdot)}$. In the Appendix, we prove the following auxiliary result.

LEMMA 1. *Under Assumptions 2–5 and 8, for all θ in Θ and conditionally on X,*

$$\|Q(\theta)\|_\infty + \|Q(\theta)'\|_\infty < K. \tag{3.9}$$

In view of the limitations of (2.13), we introduce a new test statistic and derive its limit properties under $\mathcal{H}_0$. To this extent, we introduce a deterministic, positive sequence p_n satisfying the conditions

$$p_n \to \infty \quad \text{and} \quad \frac{p_n}{n} = o(1) \quad \text{as} \quad n\to\infty, \tag{3.10}$$

and construct the following sample vector:[5]

$$M_n(\hat{\theta},t,t_Y) = M(\hat{\theta},t,t_Y) = \begin{pmatrix} M_1(\hat{\theta},t) \\ M_2(\hat{\theta},t_Y)\end{pmatrix} = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n \left(Y_i - m_i(\hat{\theta})\right)e^{t'(X_i-\bar{X})} \\ \sum_{i=1}^n \left(Y_i - m_i(\hat{\theta})\right)e^{t_Y\frac{Y_i-\bar{Y}}{p_n}} - \frac{t_Y}{p_n}tr(\hat{S}^{d'}\hat{Q}\hat{\Sigma})\end{pmatrix}, \tag{3.11}$$

[5]Following Bierens (1990), in the construction of (3.11), we use a demeaned version of Y_i and X_i to construct the sample analog, without affecting the theoretical properties of the test (see, e.g., Bierens, 2015). Preliminary simulations indeed show superiority of performance when demeaned data are employed.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

where $\hat{\Sigma} = \text{diag}(\hat{\epsilon}_1^2, \ldots, \hat{\epsilon}_n^2)$, $\hat{\epsilon} = Y - \hat{\lambda}WY - X\hat{\beta}$ and $\text{diag}(a)$ returns an $n \times n$ diagonal matrix with the components of the generic $n \times 1$ vector a on its diagonal. As explained below, the vector $M_n(\hat{\theta}, t, t_Y)$ is the key component in our test statistic $\hat{T}(t, t_Y) = nM(\hat{\theta}, t, t_Y)'\hat{\mathbb{V}}(t)^{-1}M(\hat{\theta}, t, t_Y)$ which is defined later in (3.18).

The first component in (3.11) corresponds to the naive extension of Bierens approach given in (2.13). The second component in (3.11) is based on the sample covariance between residuals and a function of Y_i, centered such that the resulting expression has zero mean under $\mathcal{H}_0$ as $n \to \infty$. By construction, the dependent variables Y_i, $i = 1, \ldots, n$, involve an increasing set of covariates weighted by the network structure, according to either (2.3) under correct specification of the regression function, or (2.4) otherwise. Hence, the true network structure plays a direct role in the test statistic formed from (3.11) via Y_i, $i = 1, \ldots, n$. The local-to-zero sequence t_Y/p_n has been introduced to allow a convenient derivation of the centering sequence appearing in the second component of (3.11) following a Taylor expansion of the exponential function. Additional details on the derivation of the centering sequence appearing in the second component of (3.11) will be discussed in the context of Theorem 1, after introducing a relevant set of assumptions to assist in the development of the limit theory of (3.11) under $\mathcal{H}_0$.

Assumption 1. For all n, ϵ_i are independent random variables with zero mean and unknown variances $\sigma_i^2 > 0$ satisfying $sup_{i \geq 1}\sigma_i^2 < K$ and, for some $\delta > 0$,

$$\sup_{0<i\leq n} \mathbb{E}|\epsilon_i|^{2+\delta} \leq K. \tag{3.12}$$

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

Assumption 3. $\lambda_0 \in \Lambda,$ 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 $i,j = 1, \ldots, n$, $W_{ij} = O(1/h)$, where $h = h_n$ is a sequence bounded away from for all n and $h/n \to 0$ as $n \to \infty$.

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

Assumption 1 postulates independent, but not identically distributed disturbances; and the uniform higher-order moment condition assists in establishing central limit theory under $\mathcal{H}_0$. The condition on boundedness of X_i is typical in the spatial literature and it is retained here to avoid introducing further trimming arguments. We stress that boundedness of X_i, $i = 1, \ldots, n$, could be easily relaxed by means of a bounded one-to-one function in case of a simple test based on (2.9). However, the formulation of our augmented test would require additional trimming arguments to accommodate unbounded regressors. Independence across X_i and ϵ_j for all (i,j) could be relaxed to strict exogeneity of X at the cost of some modifications of the derivations in the following sections. Assumptions 3–5 are standard in the SAR literature to ensure that (2.1) and (2.2) are well defined, and for further discussion we refer to, for example, Lee (2004).

Assumption 6 is specific to IV/2SLS (e.g., Kelejian and Prucha, 1998) and it ensures standard non-singularity and relevance of the instruments $\mathbb{Z}$.

Assumption 6. For $i = 1,\ldots,n$ and for all n, $\mathbb{Z}_i$ is a set of i.i.d. bounded random variables in $\mathbb{R}^p$. For $i = 1,\ldots,n$ and all n, the elements of $\mathbb{Z}_i$ are not correlated with ϵ_i. Furthermore, we require, for all sufficiently large n and some constant $c > 0$,

$$\text{eig}_{\min}\left(\frac{1}{n}\mathbb{Z}'\mathbb{Z}\right) \geq c \quad \text{and} \quad \text{rank}\left(\frac{1}{n}\mathbb{Z}'\mathbb{X}\right) = k+1. \tag{3.13}$$

Also, $\mathbb{A}$ defined in (3.4) is positive definite and $\mathbb{B}$ in (3.5) is an $p \times (k+1)$ matrix with full rank $k+1$, where both limits are assumed to exist.

We can in principle relax the boundedness condition on $\mathbb{Z}_i$, $i = 1,\ldots,n$ to the existence of moments of order two. However, in practice $\mathbb{Z}$ contains X and spatially weighted X (e.g., Kelejian and Prucha, 1998; Lee, 2003) and thus we write Assumption 6 in line with the requirements on X imposed in Assumption 2. Finally, we need to impose a standard non-singularity condition for the variance–covariance matrix appearing in Theorem 2, as

Assumption 7. Conditionally on X, the limit $\lim_{n\to\infty} n^{-1}\Psi(t)\Sigma\Psi(t)'$ exists pointwise in t and a.s. as $n \to \infty$, and is positive definite.

Some additional technical conditions on tuning parameters are employed in the aforementioned Taylor expansion of the exponential function in the construction of (3.11) and these also involve the tail slope parameters of the distribution of $(Y_1,\ldots,Y_n)$. These conditions are detailed in Assumption A1 of the Appendix and assist in establishing the following result that gives a convenient asymptotic representation of the key sample statistic $M(\hat{\theta},t,t_Y)$.

THEOREM 1. *Under Assumption 1 with $\delta = 2$, Assumption 2–6 and 8, conditionally on X, under $\mathcal{H}_0$ in (2.6)*

$$\sqrt{n}M(\hat{\theta},t,t_Y) = \frac{1}{\sqrt{n}}\Psi(t)\epsilon + o_p(1), \tag{3.14}$$

with $\Psi(t)$ defined according to (3.8).

Theorem 1 establishes that the test statistic in (3.11) is linear in the disturbances up to negligible terms. The centering sequence in $M_2(\hat{\theta},t_Y)$ in (3.11) delivers a mean-zero statistic under $\mathcal{H}_0$, whereas correct centering is lost under $\mathcal{H}_1$, thereby enabling test consistency. In (A.24) in the proof of Theorem 1, we show that the first negligible term involves a centered quadratic form in the disturbance term, which is formally similar to the Moran I statistic and/or to a residual-based type of statistic designed to assess the presence of spatial correlation (i.e.,

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

Robinson, 2008). Without the correct centering sequence under $\mathcal{H}_0$, that quadratic form diverges and the approximation in Theorem 1 fails. Thus, the first neglected term of the Taylor expansion derived in Theorem 1 is negligible under correct specification of the reduced form weighting structure as the centering sequence is correctly determined, whereas it diverges when the reduced form weighting structure is misspecified as the centering sequence is lost.

The following asymptotic result under the null is established in the Appendix.

THEOREM 2. *Under Assumptions 1–7 and A1, under $\mathcal{H}_0$ in (2.6), as $n \to \infty$*

$$\sqrt{n}\mathbb{V}(t)^{-1/2}M(\hat{\theta},t,t_Y) \to_d \mathcal{N}(0,I_2), \tag{3.15}$$

pointwise in t, conditionally on X, where the standardizing variance–covariance matrix of $\sqrt{n}M(\hat{\theta},t,t_Y)$ is given by $\mathbb{V}(t) = \lim_{n\to\infty} \mathbb{V}_n(t)$, with

$$\mathbb{V}_n(t) = \frac{1}{n}\Psi(t)\Sigma\Psi(t)', \tag{3.16}$$

where $\Sigma = diag(\sigma_1^2,\ldots,\sigma_n^2)$.

The matrix $\mathbb{V}(t)$ exists pointwise in t and almost surely, and it is non-singular under Assumption 7. Since (3.15) holds almost surely for every realization of X, Theorem 2 also holds unconditionally, giving the unconditional distribution of the statistic with $\mathbb{V}(t) = \text{plim}_{n\to\infty}\mathbb{V}_n(t)$. We have the further result.

THEOREM 3. *Under Assumption 1 with $\delta = 2$, Assumptions 2–7, under $\mathcal{H}_0$ in (2.6), as $n \to \infty$*

$$\frac{1}{n}\hat{\Psi}(t)\hat{\Sigma}\hat{\Psi}(t)' - \frac{1}{n}\Psi(t)\Sigma\Psi(t)' = o_p(1), \tag{3.17}$$

where $\hat{\Psi}(t)$ is obtained by replacing unknown parameters by their consistent estimates and $\hat{\Sigma} = diag(\hat{\epsilon}_1^2,\ldots,\hat{\epsilon}_n^2)$.

Let $\hat{\mathbb{V}}(t) = \hat{\Psi}(t)\hat{\Sigma}\hat{\Psi}(t)'/n$ and define the test statistic

$$\hat{T}(t,t_Y) = nM(\hat{\theta},t,t_Y)'\hat{\mathbb{V}}(t)^{-1}M(\hat{\theta},t,t_Y). \tag{3.18}$$

Theorems 2 and 3 lead directly to the following limit theory for the statistic $\hat{T}(t,t_Y)$.

COROLLARY 1. *Under Assumption 1 with $\delta = 2$, Assumptions 2–7 and A1, under $\mathcal{H}_0$ in (2.6), as $n \to \infty$*

$$\hat{T}(t,t_Y) \underset{d}{\to} \chi_2^2. \tag{3.19}$$

Some discussion on the practical implementation of the test using the component (3.11) and statistic $\hat{T}(t,t_Y)$ is reported in Section 6, after establishing consistency properties in the next section.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

4. BEHAVIOR OF $\hat{T}(T, T_Y)$ UNDER MISSPECIFICATION

This section explores the behavior of the test statistic $\hat{T}(t, t_Y)$ under $\mathcal{H}_1$. We stress that we expect test consistency to rely critically upon the derivation of the correct centering sequence appearing in $M_2(\hat{\theta}, t_Y)$ in (3.11), namely,

$$\sum_{i=1}^{n} \left(Y_i - m_i(\hat{\theta})\right) e^{t_Y \frac{Y_i - \bar{Y}}{p_n}} - \frac{t_Y}{p_n} tr(\hat{S}^{d'} \hat{Q} \hat{\Sigma}). \tag{4.1}$$

Thus, the role of the usual specification test component like (2.13), or the component $M_1(\hat{\theta}, t)$ in (3.11), is not necessary to achieve test consistency against $\mathcal{H}_1$. Indeed, as discussed earlier in Section 2, a naive test based on Bierens (1990) moment condition in (2.12) fails in general to detect misspecification when a spatial weighting structure W enters into the reduced form model via $S^{-1}(\cdot)$. Nonetheless, simulation exercises[6] suggest that a specification test based on (2.12) in the spatial setting has good size properties. Hence, including a Bierens-type component in (3.11) such as $M_1(\hat{\theta}, t)$ is likely to improve the finite sample test size since its small-sample behavior is not affected by the approximation error induced by the choice of the tuning parameter p_n discussed above. Importantly, however, $M_1(\hat{\theta}, t)$ in (3.11) does not contribute to test consistency. Further, the linearization and centering elements in the construction of the statistic mean that as $n \to \infty$ the particular choice of t_Y becomes less relevant and test consistency against $\mathcal{H}_1$ is achieved for each $t_Y \neq 0$. We therefore expect the test to deliver consistency regardless of the choice for (t, t_Y), as long as the trivial case $t_Y = 0$ is ruled out.

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 (2.2) with network structure W.

1. The true weight matrix structure is given by V and the practitioner uses $W \neq V$ in estimation of the model, that is, W is misspecified. Thus, $g(X) = (I - \lambda_0 V)^{-1} X\beta_0$.
2. The weight matrix W is correctly specified, but the exogenous component of the regression is nonlinear in $X_1, \ldots, X_n$ and/or in the parameters $\beta_1, \ldots, \beta_k$, so that $g(X) = (I - \lambda_0 W)^{-1} \rho(X, \beta_0)$, for some function $\rho(\cdot)$.
3. The data generating process is a SD model with weight matrices W_1, W_2, so that
$$g(X) = (I - \lambda_0 W_1)^{-1} X\beta_0 + (I - \lambda_0 W_1)^{-1} W_2 X \gamma_0, \tag{4.2}$$
where γ_0 is a k-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
$$g(X) = X\beta_0 + WX\gamma_0. \tag{4.3}$$

[6]Simulation results supporting this claim are reported in each Monte Carlo table displayed in Section 5. For each table, the number in round (respectively, square) brackets reports empirical size/power of a test based on (2.12) (respectively, (4.1)) only.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

We introduce some additional assumptions to establish the limit properties of our test when the postulated regression function in (2.3) is not correct. Assumption A1 imposes some basic requirements on the true regression function in (2.4): although no specific functional structure is imposed, the true conditional expectation functions $g_i(\cdot)$ are required to satisfy some continuity and dependence conditions. In addition, Assumption 9 establishes some condition on the errors η_i of the true regression function in (2.4). Let $g(\cdot) = (g_1(\cdot), \cdots, g_n(\cdot))'$ be the n-vector of individual $g_i(\cdot) = g_{in}(\cdot)$ functions and set $\Omega_g = \mathbb{V}ar(g)$.

Assumption 8. For $i = 1, \ldots, n$ and all n, $g_i(\cdot)$ are continuous functions of $X_1, \ldots, X_n$ and satisfy $||\Omega_g||_\infty < K$.

Assumption 9. For all n, η_i is independent of X_j for all $i,j = 1, \ldots, n$. For $i = 1, \ldots, n$, $\mathbb{E}(\eta_i|X) = \mathbb{E}(\eta_i) = 0$, $\max_{1\le i\le n} \mathbb{E}(\eta_i^2) < \infty$, and $\max_{1\le i\le n} \sum_{k=1}^{n} |Cov(\eta_i, \eta_k)| = O(1)$.

Assumption A1 accommodates all the special cases of interest that are discussed above. The last statement in Assumption 9 corresponds to a weak dependence condition, which needs to hold under $\mathcal{H}_1$. We point out that, under $\mathcal{H}_0$, this weak dependence condition is guaranteed by Assumption 5 and by finiteness of the second moment of ϵ_i, $i = 1, \ldots, n$.

To establish the behavior of the test based on (3.11) under $\mathcal{H}_1$, we need in turn to prescribe the behavior of the estimator $\hat{\theta}$ under $\mathcal{H}_1$, which is assured by the following high-level condition.

Assumption 10. 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 $\mathcal{H}_1$.

Trivially, under $\mathcal{H}_0$, $\theta^\sharp = \theta_0$. Under $\mathcal{H}_1$

$$\theta^\sharp = \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} \mathbb{X}' \mathbb{P}_Z \mathbb{X} \right)^{-1} \frac{1}{n} \mathbb{X}' \mathbb{P}_Z Y = \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} \mathbb{X}' \mathbb{P}_Z \mathbb{X} \right)^{-1} \frac{1}{n} \mathbb{X}' \mathbb{P}_Z (g + \eta), \tag{4.4}$$

which exists and is finite under Assumptions 2, 6, 8, and 9.

As anticipated at the end of Section 2, we report a counterexample to show explicit failure of a test based solely on (2.13). For simplicity of notation, we drop the reference to θ in the definition of $M_1(\cdot, t)/M_2(\cdot, t_Y)$, that is, we let $\hat{M}_1(t) = M_1(\hat{\theta}, t)$ and $\hat{M}_2(t_Y) = M_2(\hat{\theta}, t_Y)$. By standard arguments, pointwise in t, $\sqrt{n}\hat{M}_1(t) \underset{d}{\to} \mathcal{N}(0, Var(\sqrt{n}\hat{M}_1(t)))$ under $\mathcal{H}_0$. A test based on (2.13) is consistent against any form of misspecification if $\sqrt{n}M_1(\hat{\theta})$ grows without bounds under $\mathcal{H}_1$ as $n \to \infty$ for almost all $t \in \mathbb{R}^k$. On the other hand, consistency of the test cannot be established in general when $\sqrt{n}M_1(\hat{\theta}) = O_p(1)$ for a set of $t \in \mathbb{R}^k$ with positive Lebesgue measure. Without loss of generality, we set $k = 1$, that is, X is an $n \times 1$ vector. Let the true data generating process be

$$Y = \lambda_0 WY + \tau_0 VY + \beta_0 X + \epsilon, \quad \text{with reduced form} \quad Y = \tilde{S}^{-1}(\lambda_0, \tau_0)(\beta_0 X + \epsilon), \tag{4.5}$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

where $\tilde{S}^{-1} = \tilde{S}^{-1}(\lambda_0, \tau_0) = (I - \lambda_0 W - \tau_0 V)^{-1}$, but the practitioner erroneously omits τVY and adopts the incorrect reduced form $Y = S^{-1}(\lambda)(\beta X + \epsilon)$. Thus, in terms of S^{-1} and $\tilde{S}^{-1}$, the true data generating process in reduced form is

$$\begin{aligned} Y &= \beta_0 S^{-1} X + S^{-1}\epsilon + \tau_0 S^{-1} VY \\ &= \beta_0 S^{-1} X + S^{-1}\epsilon + \tau_0 S^{-1} V \tilde{S}^{-1}\epsilon + \tau_0 \beta_0 S^{-1} V \tilde{S}^{-1} X, \end{aligned} \quad \textbf{(4.6)}$$

with $\tau_0 \neq 0$. The Appendix establishes the following claim.

Claim 2. Let the true data generating process be as in (4.6) with $\tau_0 \neq 0$, and let Assumptions 1–6 hold. Let V be a standard $n \times n$ matrix satisfying $||V||_\infty + ||V'||_\infty \leq K$ and $v_{ii} = 0$ for each i. Suppose that the maximum number of nonzero elements of each row of V is given by the sequence $\nu(n)$ such that $\nu(n)/\sqrt{n} = O(1)$ and that $\sup_{\lambda,\tau} ||\tilde{S}^{-1}||_\infty + ||\tilde{S}^{-1\prime}||_\infty \leq K$. Under these conditions, $\sqrt{n} M_1(\hat{\theta}) = O_p(1)$ as $n \to \infty$, for almost all $t \in \mathbb{R}^k$.

Claim 2 shows the general lack of consistency of a test based on $M_1(\hat{\theta})$ when each row and column of V has at most $\sqrt{n}$ nonzero elements. In case, the weight matrix is completely misspecified and W is employed in place of the true matrix W^* the deviation matrix is $V = W^* - W$. We notice that the structural form weighting schemes (and hence V in Claim 2) are typically sparse matrices. Consistency of a standard Bierens test such as that in (2.13) might be achievable in such a case if the number of misspecified elements in some rows increased with the sample size at a rate exceeding $\sqrt{n}$, but this seems an unlikely scenario in practical cases.

We now turn to show consistency of our test based on (3.11). The sample statistic $\hat{M}_2(t_Y)$ in (3.11) employs the average across units of a sample analogue of (centered) expectations. We therefore need to rule out the case in which individual misspecifications in the regression functions offset each other (e.g., in the presence of an unlikely systematic symmetry in the misspecification form and direction), so as to ensure that the average amount of misspecification is non-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 difference between our Assumption 11 and the restriction in Bierens (1984) lies in the fact that, due to the local approximation in $\hat{M}_2(t_Y)$ that is induced by the deterministic, divergent sequence p_n defined in (3.10), test consistency can be achieved for any $t_Y \neq 0$ and hence Assumption 11 does not depend on the choice of t_Y.

Assumption 11. As $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{n} \left| \sum_{i=1}^{n} \mathbb{E}\left(g_i(\mathbb{X}) - m_i(\mathbb{X}, \theta^\sharp)\right) \right| > 0. \quad \textbf{(4.7)}$$

Additionally, we assure non-singularity in the limit of $\hat{\mathbb{V}}_n(t)$ in (3.19) under $\mathcal{H}_1$ by modifying Assumption 7 as follows.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

Assumption 7'. Conditionally on $\mathbb{X}$, $\lim_{n\to\infty} \mathbb{V}_n(t)$ exists and is positive definite uniformly in θ, pointwise in t and a.s., where $\mathbb{V}_n(t)$ is defined in (3.16).

The following result establishes consistency of the test statistic $\hat{T}(t,t_Y)$ and is proved in the Appendix.

THEOREM 4. *Under $\mathcal{H}_1$ in (2.7), for the deterministic sequence p_n in (3.10), and under Assumptions 2–6, 4, and 8–11, for all $c > 0$,*

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

for each $t \in \mathbb{R}^k$ and each $t_Y \neq 0$.

As is clear from the proof of Theorem 4 and, as anticipated in the discussion after Theorem 1 in Section 3, under $\mathcal{H}_1$ the centering sequence in $\hat{M}_2(t_Y)$ fails to deliver a zero expected value for the statistic for each $t_Y \neq 0$ as $n \to \infty$. Hence, an interesting feature of Theorem 4 is that the specific values of (t,t_Y) do not affect test consistency provided $t_Y \neq 0$.

5. 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 (3.19), exploring both size and power. We generate data from the SAR specification in (2.2), with an intercept and two regressors that are uniform random variables $X_{id} \sim_{iid} U(0,4)$, $d = 1,2$, with parameter settings $\beta_0 = (0.7, 2, -1)'$, $\lambda_0 = 0.4$, and sample sizes $n \in \{200, 300, 400, 500, 600, 700\}$. Throughout this section, we generate the matrix of instruments Z as WX (with exclusion of the column of ones).[7] The ϵ_i are generated, for $i = 1, \dots, n$, as

$$\epsilon_i = \sigma_i \zeta_i, \tag{5.1}$$

with $\zeta_i \sim_{iid} \mathcal{N}(0,1)$ and we distinguish between two mechanisms for the scale parameter σ_i:

a) Direct construction using the formula

$$\sigma_i = c \frac{d_i}{\sum_{j=1}^n d_j/n}, \tag{5.2}$$

where the constant c is set to unity and d_i denotes the number of neighbors of unit i, such that, for each generic W, $d_i = \text{card}(j: \ w_{ij} \neq 0, i \neq j)$.

b) σ_i^2s are randomly generated values from a chi-square distribution with 5 degrees of freedom ($\chi^2(5)$).

[7] The pattern of results displayed in this section is virtually unchanged if we include higher-order lags of X, such as W^2X, or if we adopt the so-called optimal instruments as in Lee (2003).

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

With both methods, the σ_i are kept fixed across simulations and across different parameter scenarios. The heteroskedasticity design in (5.2) is in line with earlier simulation work of Kelejian and Prucha (2010) and Arraiz et al. (2010) and is motivated by situations in which heteroskedasticity arises as units across different regions may have different numbers of neighbors.

Two different weight matrices are used:

1) Exponential distance weights, that is, $w_{ij} = exp(-|\ell_i - \ell_j|)1(|\ell_i - \ell_j| < \log n)$, 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 $n \times n$ symmetric 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 1,000 replications and across different experimental scenarios.

The choice of p_n and t_Y drives the trade-off between size and power for small n and it becomes increasingly less important for test performance as n increases. In this simulation exercise, we set $p_n = n^{1/3}$, $t = (1.5, 1.5, 1.5)'$ and $t_Y = 0.4$. An additional analysis for different choices of t_Y and p_n will be reported in Section 6. Also, similar to Bierens (1990), we replace the argument of the exponential function in $\hat{M}_1(t)$ in (3.11) with $t' \tan^{-1}(X_i - \bar{X})$ for each $i = 1, \ldots, n$, where $\tan^{-1}(X_i - \bar{X}) = (\tan^{-1}(X_{i1} - \bar{X}_1), \ldots, \tan^{-1}(X_{ik} - \bar{X}_k))'$ and $\bar{X}_d$ denotes sample mean for $d = 1, \ldots, k$. Given the support of X in this simulation exercise, the $\tan^{-1}(\cdot)$ contribution turns out to be virtually irrelevant. We stress the power inadequacy of a test based only on $\hat{M}_1(t)$ and the oversized performance of a test based on $\hat{M}_2(t_Y)$ only, by reporting for each of the following tables empirical sizes and powers of the test based on $\hat{T}(t, t_Y)$ with both components of $\hat{M}(t, t_Y)$ in (3.11), together with the corresponding results based separately on $\hat{M}_1(t)$ and $\hat{M}_2(t_Y)$ (in round and square brackets, respectively).

We first examine the performance of the test statistic in (3.19) under $\mathcal{H}_0$ in (2.6), and report in Tables 1 and 2 empirical sizes for nominal significance levels $s = 0.1, 0.05, 0.01$ and both weight matrix models (1) and (2) for the heteroskedasticity design in (a) and (b), respectively. Both Tables 1 and 2 show that empirical size converges to the nominal value as the sample size increases, with a noticeable size distortion for the smallest sample sizes especially for scenario (1) with the heteroskedasticity design given in (b). As expected from Claim 1 and in line with our comment at the beginning of Section 4, the size of a test based solely on $\hat{M}_1(t)$ is stable across sample sizes and it is in line (or even superior to) with that based on the full vector $\hat{M}(t, t_Y)$, while a test based on $\hat{M}_2(t_Y)$ is oversized, especially for smaller samples.

Table 1. Empirical size with σ_i as in (a)

W	(1)			(2)		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.147 (0.098)[0.167]	0.074 (0.031)[0.106]	0.021 (0.005)[0.032]	0.109 (0.106)[0.121]	0.047 (0.04)[0.06]	0.010 (0.003)[0.014]
300	0.130 (0.088)[0.145]	0.075 (0.037)[0.089]	0.015 (0.002)[0.032]	0.098 (0.083)[0.126]	0.040 (0.034)[0.061]	0.004 (0.004)[0.011]
400	0.122 (0.086)[0.140]	0.051 (0.036)[0.079]	0.013 (0.004)[0.020]	0.095 (0.082)[0.114]	0.039 (0.030)[0.058]	0.005 (0.004)[0.007]
500	0.086 (0.095)[0.118]	0.052 (0.037)[0.079]	0.007 (0.007)[0.01]	0.097 (0.089)[0.111]	0.046 (0.036)[0.056]	0.010 (0.005)[0.011]
600	0.097 (0.089)[0.101]	0.053 (0.046)[0.055]	0.016 (0.005)[0.013]	0.101 (0.101)[0.113]	0.043 (0.044)[0.051]	0.007 (0.006)[0.005]
700	0.096 (0.108)[0.106]	0.046 (0.052)[0.053]	0.008 (0.006)[0.01]	0.099 (0.093)[0.128]	0.048 (0.045)[0.057]	0.010 (0.006)[0.009]

Note: Test of $\mathcal{H}_0$ in (2.6) based on $\hat{T}(t,t_Y)$ in (3.19) for nominal significance levels $s \in \{10\%, 5\%, 1\%\}$. The empirical size of the $\hat{M}_1(t)$ ($\hat{M}_2(t_Y)$) component-only test is reported in round (square) brackets.

Table 2. Empirical size with σ_i as in (b)

W	(1)			(2)		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.185 (0.105)[0.212]	0.116 (0.045)[0.145]	0.052 (0.003)[0.069]	0.126 (0.101)[0.156]	0.056 (0.037)[0.074]	0.015 (0.004)[0.015]
300	0.152 (0.079)[0.188]	0.101 (0.029)[0.122]	0.035 (0.004)[0.049]	0.116 (0.084)[0.141]	0.056 (0.039)[0.079]	0.007 (0.004)[0.015]
400	0.113 (0.094)[0.139]	0.059 (0.030)[0.075]	0.020 (0.005)[0.022]	0.103 (0.075)[0.145]	0.043 (0.029)[0.069]	0.007 (0.004)[0.009]
500	0.118 (0.091)[0.144]	0.066 (0.038)[0.084]	0.011 (0.004)[0.021]	0.095 (0.088)[0.135]	0.051 (0.033)[0.056]	0.009 (0.002)[0.016]
600	0.108 (0.105)[0.123]	0.062 (0.044)[0.069]	0.023 (0.006)[0.025]	0.106 (0.103)[0.124]	0.047 (0.045)[0.057]	0.003 (0.003)[0.006]
700	0.107 (0.104)[0.122]	0.054 (0.046)[0.065]	0.009 (0.007)[0.016]	0.104 (0.093)[0.127]	0.049 (0.030)[0.062]	0.005 (0.001)[0.008]

Note: Test of $\mathcal{H}_0$ in (2.6) based on $\hat{T}(t,t_Y)$ in (3.19) for nominal significance levels $s \in \{10\%, 5\%, 1\%\}$. The empirical size of the $\hat{M}_1(t)$ ($\hat{M}_2(t_Y)$) component-only test is reported in round (square) brackets.

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

$$Y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} Y_j + X_i'\beta_0 + \frac{1}{4} X_{i1}^2 + \epsilon_i, \quad i = 1, \ldots, n, \tag{5.3}$$

and the misspecified linear SAR with no quadratic term was estimated. Again, we set $\lambda_0 = 0.4$ and $\beta_0 = (0.7, 2, -1)'$. Test power is reported in Table 3 and is evidently close to unity for all sample sizes. The left panel of Table 3 reports results

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

Table 3. Empirical power under functional form misspecification

σ_i	(a)			(b)		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.962 (0.553)[0.935]	0.924 (0.359)[0.895]	0.832 (0.083)[0.790]	0.874 (0.362)[0.849]	0.808 (0.200)[0.802]	0.691 (0.048)[0.707]
300	0.968 (0.683)[0.939]	0.940 (0.539)[0.901]	0.854 (0.188)[0.795]	0.888 (0.477)[0.865]	0.837 (0.316)[0.796]	0.702 (0.083)[0.679]
400	0.972 (0.813)[0.904]	0.955 (0.691)[0.866]	0.859 (0.357)[0.716]	0.901 (0.584)[0.817]	0.833 (0.414)[0.737]	0.637 (0.148)[0.572]
500	0.980 (0.899)[0.876]	0.960 (0.802)[0.804]	0.854 (0.492)[0.636]	0.893 (0.646)[0.796]	0.824 (0.506)[0.704]	0.645 (0.216)[0.524]
600	0.990 (0.934)[0.866]	0.971 (0.876)[0.791]	0.883 (0.623)[0.574]	0.901 (0.726)[0.743]	0.838 (0.605)[0.638]	0.613 (0.309)[0.415]
700	0.998 (0.973)[0.913]	0.992 (0.929)[0.856]	0.948 (0.720)[0.688]	0.944 (0.790)[0.786]	0.890 (0.663)[0.700]	0.719 (0.375)[0.519]

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) when the true model is (5.3) with nominal significance levels $s \in \{10\%, 5\%, 1\%\}$ and W chosen as in (1). The empirical power of $\hat{M}_1(t)$ ($\hat{M}_2(t_Y)$) is reported in round (square) brackets.

for the weight matrix model (1) and the heteroskedasticity design in (a), whereas the right panel reports results for σ_i^2 generated as in (b). From Table 3, empirical power is very high for all sample sizes and, as expected, the practical performance is substantially superior when the test is based on the full vector $\hat{M}(t, t_Y)$ rather than on $\hat{M}_1(t)$ only. Incidentally, the power of a test based on the second component of $\hat{M}_2(t_Y)$ (in square brackets) is itself lower than the corresponding quantity based on the full vector $\hat{M}(t, t_Y)$, revealing that inclusion of the first component may increase power when misspecification involves individual components of X. Also, the pattern of empirical power when the test is based on both components, rather than just the second one, shows a steadier increasing trend toward unity with the sample size. To this extent, the inclusion of the first component might play a useful role in stabilizing variability of the test statistic.

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

i) Both true and misspecified matrices are generated as in (1) but with two independent sets of locations.
ii) The true matrix is (2) but the practitioner erroneously estimates parameters in (2.2) using W as in (1).

In Tables 4 and 5, we report results for scenarios (i) and (ii) (left and right panels, respectively) for both heteroskedasticity designs (a) and (b). From the figures in round brackets of Tables 4 and 5, we notice that a test based on $\hat{M}_1(t)$ has power that is roughly equal to size for all sample sizes. The performance of our combined test is instead satisfactory for all sample sizes. It is worth noting that power based on $\hat{M}_2(t)$ (in square brackets) is always slightly superior to that obtained with both components of the vector $\hat{M}(t, t_Y)$. This slight loss of power against weight matrix

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

Table 4. Empirical power under W misspecification with σ_i as in (a)

	(i)			(ii)		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.794 (0.089)[0.828]	0.752 (0.035)[0.784]	0.682 (0.001)[0.711]	0.777 (0.098)[0.799]	0.742 (0.043)[0.768]	0.667 (0.004)[0.699]
300	0.807 (0.069)[0.832]	0.761 (0.030)[0.793]	0.683 (0.005)[0.716]	0.779 (0.092)[0.811]	0.748 (0.040)[0.773]	0.673 (0.002)[0.698]
400	0.817 (0.108)[0.837]	0.777 (0.050)[0.805]	0.709 (0.007)[0.736]	0.783 (0.113)[0.798]	0.739 (0.051)[0.764]	0.667 (0.007)[0.689]
500	0.820 (0.104)[0.837]	0.790 (0.048)[0.807]	0.703 (0.005)[0.734]	0.791 (0.101)[0.817]	0.763 (0.038)[0.789]	0.695 (0.006)[0.716]
600	0.831 (0.085)[0.849]	0.796 (0.045)[0.821]	0.724 (0.006)[0.754]	0.790 (0.098)[0.813]	0.756 (0.052)[0.773]	0.680 (0.006)[0.708]
700	0.858 (0.101)[0.877]	0.829 (0.040)[0.851]	0.752 (0.005)[0.788]	0.810 (0.121)[0.847]	0.774 (0.052)[0.807]	0.701 (0.010)[0.744]

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) under scenarios (i) and (ii), with nominal significance levels $s \in \{10\%, 5\%, 1\%\}$. The empirical power of $\hat{M}_1(t)$ ($\hat{M}_2(t_Y)$) is reported in round (square) brackets.

Table 5. Empirical power under W misspecification with σ_i as in (b)

	(i)			(ii)		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.804 (0.106)[0.812]	0.768 (0.043)[0.790]	0.700 (0.006)[0.730]	0.837 (0.098)[0.865]	0.806 (0.036)[0.830]	0.747 (0.004)[0.767]
300	0.811 (0.083)[0.834]	0.767 (0.029)[0.801]	0.692 (0.004)[0.721]	0.856 (0.082)[0.872]	0.832 (0.026)[0.847]	0.776 (0.004)[0.800]
400	0.830 (0.088)[0.850]	0.795 (0.028)[0.822]	0.725 (0.002)[0.752]	0.874 (0.100)[0.887]	0.856 (0.034)[0.871]	0.818 (0.004)[0.832]
500	0.833 (0.095)[0.856]	0.808 (0.034)[0.828]	0.740 (0.050)[0.764]	0.85 (0.095)[0.870]	0.823 (0.032)[0.843]	0.760 (0.001)[0.783]
600	0.834 (0.089)[0.855]	0.796 (0.033)[0.822]	0.737 (0.002)[0.762]	0.843 (0.108)[0.858]	0.808 (0.054)[0.827]	0.755 (0.005)[0.781]
700	0.864 (0.101)[0.881]	0825 (0.051)[0.856]	0.757 (0.010)[0.779]	0.881 (0.073)[0.892]	0.853 (0.033)[0.872]	0.795 (0.003)[0.821]

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) under scenarios (i) and (ii), with nominal significance levels $s \in \{10\%, 5\%, 1\%\}$. The empirical power of $\hat{M}_1(t)$ ($\hat{M}_2(t_Y)$) is reported in round (square) brackets.

misspecification entailed by inclusion of the first component of $\hat{M}(t, t_Y)$ is the cost to obtain size stability (as displayed in Tables 1 and 2) and an increase in power against some forms of misspecification (as displayed in Table 3).

We also consider test power against misspecification of the model itself by generating data based on the SD and SLX models (defined in (4.2) and (4.3), respectively), with parameter values $\beta_0 = (0.7, 2, -1)'$, $\lambda_0 = 0.4$, and $\gamma_0 = (0.5, 1)'$ in (4.2), and $\beta_0 = (0.7, 2, -1)'$, $\lambda_0 = 0.4$, and $\gamma_0 = (1.2, 1)'$ for the parameters in (4.3). The settings for γ_0 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. Tables 6 and 7 report results for the heteroskedasticity design in (a) and (b), respectively. The results in

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

TABLE 6. Empirical power under model misspecification with σ_i as in (a)

Model	SD			SLX		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.916 (0.166)[0.931]	0.872 (0.094)[0.900]	0.755 (0.026)[0.807]	0.740 (0.155)[0.814]	0.630 (0.082)[0.726]	0.412 (0.012)[0.487]
300	0.891 (0.161)[0.924]	0.858 (0.071)[0.893]	0.734 (0.019)[0.778]	0.744 (0.138)[0.808]	0.639 (0.067)[0.725]	0.374 (0.013)[0.486]
400	0.971 (0.147)[0.984]	0.954 (0.093)[0.969]	0.899 (0.021)[0.934]	0.881 (0.118)[0.928]	0.816 (0.061)[0.878]	0.595 (0.013)[0.706]
500	0.936 (0.189)[0.953]	0.895 (0.123)[0.929]	0.751 (0.026)[0.834]	0.747 (0.196)[0.819]	0.633 (0.106)[0.716]	0.367 (0.030)[0.461]
600	0.967 (0.224)[0.969]	0.937 (0.137)[0.954]	0.828 (0.035)[0.874]	0.828 (0.217)[0.858]	0.738 (0.135)[0.791]	0.470 (0.032)[0.542]
700	0.993 (0.148)[0.995]	0.982 (0.087)[0.989]	0.962 (0.018)[0.962]	0.921 (0.133)[0.940]	0.849 (0.067)[0.911]	0.628 (0.014)[0.742]

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) when the true models are SD in (4.2) and SLX in (4.3), with nominal significance level s. The empirical power of the first (second) component-only test is reported in round (square) brackets.

TABLE 7. Empirical power under model misspecification with σ_i as in (b)

Model	SD			SLX		
$n\backslash s$	10%	5%	1%	10%	5%	1%
200	0.772 (0.135)[0.826]	0.666 (0.059)[0.755]	0.461 (0.003)[0.534]	0.625 (0.110)[0.712]	0.477 (0.062)[0.587]	0.225 (0.007)[0.319]
300	0.798 (0.110)[0.858]	0.711 (0.057)[0.787]	0.507 (0.007)[0.588]	0.671 (0.112)[0.749]	0.546 (0.052)[0.638]	0.273 (0.008)[0.388]
400	0.906 (0.122)[0.948]	0.850 (0.063)[0.913]	0.646 (0.007)[0.748]	0.746 (0.113)[0.838]	0.632 (0.050)[0.742]	0.324 (0.004)[0.466]
500	0.914 (0.163)[0.934]	0.881 (0.079)[0.904]	0.713 (0.019)[0.791]	0.661 (0.145)[0.775]	0.531 (0.069)[0.65]	0.264 (0.011)[0.356]
600	0.901 (0.163)[0.937]	0.844 (0.098)[0.892]	0.627 (0.023)[0.752]	0.791 (0.146)[0.847]	0.676 (0.081)[0.772]	0.384 (0.014)[0.512]
700	0.944 (0.206)[0.972]	0.910 (0.132)[0.950]	0.780 (0.041)[0.840]	0.806 (0.108)[0.895]	0.682 (0.048)[0.798]	0.350 (0.008)[0.530]

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) when the true models are SD in (4.2) and SLX in (4.3), with nominal significance level s. The empirical power of the first (second) component-only test is reported in round (square) brackets.

Tables 6 and 7 show good performance of our test in this case, in contrast to the test based on $\hat{M}_1(t)$ for which power is only marginally larger than nominal size. Again, as discussed in the context of Tables 4 and 5, from the figures in square brackets, we notice that the inclusion of the first component entails a small loss in power compared to the full test.

In addition to reporting power against a fixed alternative (i.e., the weights are fully misspecified), as in Tables 4 and 5, we consider a "true" data generating process in which the level of misspecification increases with sample size as $\sqrt{n}$, that is,

$$Y = (I - \lambda_0 W - \tau_0(V - W))^{-1}(X\beta_0 + \epsilon), \qquad \textbf{(5.4)}$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

TABLE 8. Empirical power under W misspecification with $\tau_0 = 0.4\sqrt{n}/\sqrt{700}$, $s = 5\%$

n	(i), (a) $\hat{T}(t, t_Y)$	(i), (a) Moran-I	(i), (b) $\hat{T}(t, t_Y)$	(i), (b) Moran-I	(ii), (a) $\hat{T}(t, t_Y)$	(ii), (a) Moran-I	(ii), (b) $\hat{T}(t, t_Y)$	(ii), (b) Moran-I
200	0.291 (0.038)[0.347]	0.013	0.535 (0.039)[0.562]	0.018	0.232 (0.041)[0.269]	0.027	0.574 (0.034)[0.605]	0.015
300	0.376 (0.030)[0.430]	0.012	0.601 (0.029)[0.634]	0.014	0.366 (0.030)[0.428]	0.017	0.639 (0.033)[0.675]	0.017
400	0.451 (0.049)[0.500]	0.018	0.685 (0.043)[0.732]	0.017	0.411 (0.059)[0.453]	0.023	0.745 (0.052)[0.772]	0.016
500	0.617 (0.028)[0.661]	0.025	0.706 (0.033)[0.735]	0.018	0.554 (0.037)[0.599]	0.021	0.740 (0.034)[0.770]	0.025
600	0.752 (0.037)[0.785]	0.021	0.774 (0.035)[0.793]	0.032	0.669 (0.039)[0.697]	0.028	0.778 (0.041)[0.808]	0.011
700	0.829 (0.040)[0.851]	0.020	0.825 (0.051)[0.856]	0.018	0.774 (0.052)[0.807]	0.015	0.853 (0.033)[0.872]	0.017

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) when the true data generating process is given by (5.4) with $\tau_0 = 0.214, 0.262, 0.302, 0.338, 0.370, 0.4$ for $n = 200, 300, 400, 500, 600, 700$, respectively. The empirical power of $\hat{M}_1(t)$ ($\hat{M}_2(t_Y)$) is reported in round (square) brackets.

where W is, consistently with previous notation, the weighting structure adopted by the practitioner. In the left panel of Table 8, we report results for scenario (i) for both heteroskedasticity structures in (a) and (b), while the right panel reports corresponding results for scenario (ii). We set the parameter $\tau_0 = \tau_{0n}$, which controls the amount of misspecification, as $\tau_{0n} = 0.4\sqrt{n}/\sqrt{700}$. The latter choice is driven by the fact that we need to guarantee invertibility of $(I - \lambda_0 W - \tau_0(V - W))$ and for the largest sample size of $n = 700$, we have $\tau_0 = 0.4$, which in turn corresponds to the fixed alternative reported in Tables 4 and 5. For each scenario and each sample size, we also report the comparison with power obtained by the heteroskedasticity robust version of the Moran-I test with asymptotic critical values, as given in Kelejian and Prucha (2001). As expected, power increases steadily with n when the test statistic is either the combined one in (3.11) or its second component only, $\hat{M}_2(t_Y)$, while a test based on $\hat{M}_1(t)$ has virtually no power. Also, power of the robust version of Moran I is inferior to size for all n.

We conclude this section by reporting a small simulation to assess finite sample support for Claim 2. According to this claim, consistency of a test based on $\hat{M}_1(t)$ is expected to fail when the maximum number of nonzero elements in each row of V does not exceed $\sqrt{n}$. Claim 2 does not indicate that a test based on $\hat{M}_1(t)$ would be consistent if the order of the number of nonzero elements exceeds $\sqrt{n}$ for some number of rows in V, but intuition suggests that power should increase with the number of nonzero elements in the rows of V. To assess this intuition, we set the true data generating process as

$$Y = (I - \lambda_0 W - \tau_0 V)^{-1}(X\beta_0 + \epsilon), \quad \textbf{(5.5)}$$

TABLE 9. Empirical power of $\hat{M}_1(t)$ for varying sparsity of W misspecification

$n\backslash p$	0.001	0.01	0.02	0.08	0.15	0.30
200	0.121	0.124	0.164	0.270	0.309	0.357
300	0.113	0.114	0.156	0.242	0.284	0.341
400	0.082	0.100	0.165	0.256	0.313	0.353
500	0.072	0.114	0.225	0.248	0.302	0.332
600	0.065	0.114	0.143	0.240	0.263	0.312
700	0.063	0.118	0.146	0.225	0.290	0.292

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) based on $\hat{M}_1(t)$, with nominal significance level $s = 5\%$. σ_i generated as in (a).

with $\tau_0 = 0.4$, weight matrix W generated as a Toeplitz circulant structure with two-behind and two-ahead neighbors and each row a cyclic shift of the row above it, X, β_0, and λ_0 as described at the beginning of this section and ϵ_i, $i = 1, \ldots, n$, as in (5.1) with σ_i as in design (a). The deviation matrix V is constructed, once for each scenario, as follows: the elements in each row of V are generated as independent Bernoulli random variables with parameter $p = 0.001, 0.01, 0.02, 0.08, 0.15, 0.30$. The first three choices of p correspond to "sparse" rows of V (i.e., number of nonzero elements smaller than $\sqrt{n}$), while $p = 0.8, 0.15, 0.30$ correspond to "dense" rows (i.e., number of nonzero elements larger than $\sqrt{n}$). Both matrices W and V are rescaled so that each column sums to unity.[8] In Table 9, we report power for $s = 5\%$ in the aforementioned scenarios.

From the results in Table 9, it is clear that power fails to increase with n, so consistency of a Bierens-type test based on $\hat{M}_1(t)$ is not assured, at least in this particular design, even for the empirically unnatural scenario with "dense" rows of V. As expected, power does increase with the probability of nonzero elements p in V, that is, the power of a simple test based on $\hat{M}_1(t)$ improves with the number of nonzero elements in the rows of V.

6. PRACTICAL IMPLEMENTATION OF THE TEST: A HEURISTIC DISCUSSION

This section offers some practical illustrations regarding the choice of the ratio t_Y/p_n, while acknowledging from our simulation work that the choice of t does not materially affect either size or power. As illustrated in Section 4, the test remains consistent for any value of $t_Y \neq 0$ and p_n satisfying (3.10). However, the ratio t_Y/p_n impacts the finite sample properties of the test and requires some practical investigation, even though a formal analysis of local power of our test to establish an optimal ratio t_Y/p_n remains beyond our scope. Also, under $\mathcal{H}_0$,

[8] The pattern of results remains the same even for alternative choices of normalization.

TABLE 10. Empirical size with W chosen as in (1) and σ_i generated as in (a)

$n\backslash(t_Y/p_n)$	$0.3/n^{1/3}$	$0.5/n^{1/3}$	$0.4/\log(n)$	$0.4/n^{2/3}$
200	0.044	0.164	0.103	0.023
300	0.044	0.122	0.101	0.032
400	0.055	0.106	0.096	0.046
500	0.039	0.067	0.069	0.044
600	0.049	0.076	0.082	0.050
700	0.042	0.081	0.078	0.042

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) based on $\hat{T}(t, t_Y)$ in (3.19), with nominal significance level $s = 5\%$ and various combinations of t_Y/p_n.

TABLE 11. Empirical power under scenario (ii) with σ_i generated as in (a)

$n\backslash(t_Y/p_n)$	$0.3/n^{1/3}$	$0.5/n^{1/3}$	$0.4/\log(n)$	$0.4/n^{2/3}$
200	0.570	0.846	0.786	0.044
300	0.578	0.871	0.847	0.019
400	0.535	0.852	0.843	0.022
500	0.538	0.856	0.871	0.017
600	0.497	0.850	0.875	0.015
700	0.528	0.860	0.893	0.012

Note: Test of $\mathcal{H}_0$ in (2.6) against $\mathcal{H}_1$ in (2.7) based on $\hat{T}(t, t_Y)$ in (3.19), with nominal significance level $s = 5\%$ and various combinations of t_Y/p_n.

the rate of the sequence p_n affects the rate of the remainder term in the expansion reported in Theorem 1 and thus the error of the central limit theorem approximation reported in Theorem 2. On one side, a fast-diverging p_n (or, equivalently, a small t_Y/p_n) guarantees size stability, as the error of the approximations in Theorems 1 and 2 would then be small. On the other side, a slow diverging p_n (equivalently, a large t_Y/p_n) increases the power of the test as it assures relevance to the second component of (3.11) via substantial covariation between the reduced form residuals and the exponential term $t_Y(Y_i - \bar{Y})/p_n$. In Tables 10 and 11, we report, respectively, size and power for some combinations of t_Y/p_n. Table 10 is to be compared with the left panel of Table 1; Table 11 is to be compared with the right panel of Table 4. Description of the respective scenarios was reported in Section 5 and is not repeated here. In order to avoid overly large tables, we only report here results corresponding to $s = 5\%$, although a similar pattern was detected in unreported simulations for $s = 10\%$ and $s = 1\%$. We remind the reader that results in Section 5 have been obtained with $t_Y = 0.4$ and $p_n = n^{1/3}$.

From results displayed in Table 10, we infer that for each combination of t_Y/p_n empirical sizes approach the nominal 5%. However, as expected, the finite sample size distortion is more severe for larger t_Y/p_n. On the other hand, from Table 11, we notice that power against network misspecification drops even below size for a fast-diverging p_n, while it achieves its maximum for the smallest ratios t_Y/p_n. Thus, our choice of t_Y/p_n adopted in Section 5 seems to offer the most reasonable compromise between size stability and power against the most problematic sources of misspecification, that is, that concerning the weighting structure.

As discussed in Claims 1 and 2 and further supported in Tables 3–8, our test statistic has the highly unusual special property that its first component-only version (i.e., based on $\hat{M}_1(t)$) can detect some misspecification with respect to X but dramatically fails in presence of misspecification of spatial linkages, while the second component-only version (i.e., based on $\hat{M}_2(t_Y)$) can detect misspecification in both directions. Hence, when a given specification is rejected by our full test statistic $\hat{T}(t, t_Y)$ in (3.11), disparity between conclusions of the two single-component variants could indicate misspecification in the direction of spatial linkages. There is a body of literature on specification search methods for spatial econometric models, (see, e.g., Florax, Folmer, and Rey, 2003; Elhorst, 2010). Development of a multistep specification search method is beyond the scope of the current paper while the unique feature of our test in having two components that are suited for different directions of misspecification could be a promising avenue of future work that may overcome the current limitations of specification search methods, most notably the intractability of significance levels of sequential tests.

7. EMPIRICAL ILLUSTRATION

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 recent decades, which has been typically attributed to competition between neighboring 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 modeling has been widely applied to investigate the presence of such fiscal interaction. Empirical results have frequently found evidence of positive dependence in neighboring government tax rates (see Allers and Elhorst, 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 underprovision 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 to present findings that contrast with earlier literature. Lyytikäinen (2012), in particular, used policy-based instrumental variables (IV) to

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

estimate the spatial autoregressive parameter in SAR models with fixed effects. He found this parameter to be insignificant, in contrast to the preceding literature (e.g., Allers and Elhorst, 2005). Lyytikäinen (2012) additionally reported positive and highly significant spatial parameter estimates based on standard SAR using QMLE and 2SLS without accounting for fixed effects, which is representative of the aforementioned literature's model specifications. The contradiction suggests that caution should be exercised in accepting the findings of previous work showing positive spatial dependence in neighboring government tax rates. He points out that the fitted standard SAR model is unlikely to be correctly specified in practice and that the resulting residual spatial correlation in the errors may result in regressor endogeneity and biased findings. Lyytikäinen (2012) does not consider explicitly the problem of misspecification of the weight matrix W but notes that standard techniques are likely to fail to deliver credible inference if the SAR models are not correctly specified.

This section presents empirical applications of our specification test to data from Lyytikäinen (2012) with the aim of assessing the suitability of SAR specifications in analyzing tax competition data. We find that careful consideration of model specification, similar to that used for policy-based IV estimation, helps to mitigate significantly the noted disparity in the conclusions drawn from the estimates of conventional SAR (without individual fixed effects), and, the refined SAR model estimated by Lyytikäinen (2012). These findings highlight the usefulness of specification testing. The test procedure developed in the present paper may therefore provide a valid starting point toward 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.

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, and μ_i and τ_t are municipality and year fixed effects, respectively. The regressors X_{it} are the municipality's socioeconomic attributes including per capita income, per capita grants, unemployment rate, percent of age 0–16, percent of age 61–75 and percent of age 75+. We refer to Lyytikäinen (2012) for a detailed description of the data and setting.

In order to alleviate possible endogeneity sources arising from unobservable time-invariant characteristics, Lyytikäinen (2012) focused on 1-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 411. This exogenous policy change was used to construct a suitable instrument and estimate parameters of the model

$$\Delta t_i = \lambda \sum_{j=1}^{n} w_{ij} \Delta t_j + \Delta X_i' \beta + \gamma_0 + \gamma_1 P_i + \gamma_2 M_i + \Delta \epsilon_i, \quad i = 1, \ldots, 411, \tag{7.2}$$

where $\Delta\epsilon_i = \epsilon_{i,2000} - \epsilon_{i,1999}$, P_i is a dummy variable indicating whether the 1998 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. Lyytikäinen (2012) found the spatial 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.

For comparability with the policy-based IV estimator in Lyytikäinen (2012), we consider the model based on differenced data, that is, model (7.2) with differences taken between 2000 and 1999, 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 $\Delta\epsilon_i = \epsilon_{i,2001} - \epsilon_{i,2000}$. While Lyytikäinen (2012) uses policy-based IV to estimate the parameters, we use WX as the instrument in 2SLS estimation. 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 normalizing transformation to obtain W. We set $p_n = n^{1/3}$ as in the simulation exercises. Further, in the first component of $M(\hat{\theta}, t, t_Y)$ of (3.11), we use $\exp(t' \tan^{-1}(X_i - \bar{X})$ rather than $\exp(t'X_i)$, for $i = 1, \ldots, n$. The choice of the k-vector $t = t_0(1, 1, \ldots, 1)'$ is calibrated so that $\sum_{i=1}^{n} \exp(t' \tan^{-1}(X_i - \bar{X}))/n = 10$, where 10 is approximately the value of this average in the simulation setup. Once the value of t_0 is established, t_Y is fixed so that $t_0/t_Y = 3.75$, as in Section 5. Results for both general and residential building property tax rates are reported in columns 1–4 of Table 12, 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 building tax rates.

Evidently, all four estimates of λ are insignificant, in agreement with what Lyytikäinen (2012) found with his policy-based IV estimation. This is in contrast to the λ estimates from conventional SAR models without municipality fixed effects, P_i and M_i that Lyytikäinen (2012) used as a comparison. Our specification test strongly rejects the model with differences taken between 2001 and 2000 for general property tax rate, but does not reject in the other three columns. In the rejected case, $\hat{\lambda}$ was above unity with a large standard error, indicating a possible mismatch between the model and the data at hand and it is helpful to discard this case based on our specification test.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

TABLE 12. Estimates and specification tests from first-differenced model (7.2)

	General		Residential	
	1	2	3	4
$\hat{\lambda}$	1.2921 (1.5343)	0.0352 (0.264)	0.4781 (1.1619)	0.1734 (0.6727)
$\hat{T}(t, t_Y)$	766.62***	0.0817	1.0096	0.0759

Note: Left panel: Columns 1 and 2 report 2SLS estimates of λ and their *t*-statistics (in brackets), and the value of the test $\hat{T}(t, t_Y)$, 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 with differences taken between 2001 and 2000 and between 2000 and 1999, respectively, with residential building tax rate as the dependent variable. Row-normalized weight matrices are used. **p*-value < 0.1; ***p*-value < 0.05; ****p*-value < 0.01.

TABLE 13. Estimates and specification tests from (7.1)

	General		Residential	
	1	2	3	4
$\hat{\lambda}$	0.4829 (0.016)	0.3191*** (4.357)	0.5106 (0.029)	0.3115*** (2.823)
$\hat{T}(t, t_Y)$	0.014	11.606***	0.111	8.141**
Municipality FE	Yes		Yes	

Note: Left panel: Columns 1 and 2 report 2SLS estimates of λ and their *t*-statistics (in brackets), and the value of the test $\hat{T}(t, t_Y)$ for model (7.1) with lagged X_{it-1} with and without municipality fixed effects, respectively, with the general property tax rate as dependent variable. Right panel: Columns 3 and 4 report results for model (7.1) with lagged X_{it-1} with and without municipality fixed effects, respectively, with the residential building tax rate as dependent variable. Row-normalized weight matrices are used. **p*-value < 0.1; ***p*-value < 0.05; ****p*-value < 0.01. Data from 1993 to 2001.

We have so far used differenced data for comparability with the policy-based IV estimation results of Lyytikäinen (2012). We now apply our specification test to pooled-level data in typical specifications used in conventional spatial literature, to see if it can be useful in finding suitable SAR specifications. We pool data of all available years 1993–2001 and estimate (7.1) with lagged X_{it-1} which was suggested in Lyytikäinen (2012) as a typical SAR specification used in the tax competition literature with municipality fixed effects omitted (i.e., $\mu_i = 0, i = 1, \ldots, n$). Carrying out a pooled estimation is a plausible scenario if the practitioner is not aware of the aforementioned policy intervention that was implemented in 2000. The danger of such analysis is obtaining spuriously inflated SAR coefficients from the direct impacts of the tax policy: because of the pre-existing positive spatial correlation in property tax rates of neighboring municipalities a municipality whose neighbors are affected by the policy of raising the common statutory lower bound is also likely to experience an imposed increase in tax rate.

We report $\hat{\lambda}$ and our specification test results for (7.1) with lagged X_{it-1} and with and without municipality fixed effects in Table 13. For both general property

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

and residential building tax rates, the models that omit municipality fixed effects report significant SAR coefficients but are rejected by our specification test. In contrast, the models that include municipality fixed effects report insignificant SAR coefficients and are not rejected by the specification test. The implication is that incorporating municipality fixed effects is important for these data, illustrating how our specification test can provide guidance in finding suitable SAR specifications. Although a full replication of the results in Lyytikäinen (2012) is not attempted, the specification test findings and the empirical results in Tables 12 and 13 suggest that 2SLS estimates of appropriate SAR models deliver results that are in line with the policy-based IV estimator of Lyytikäinen (2012).

8. CONCLUDING REMARKS

This paper provides 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. To the best of our knowledge, this is the first test in the literature to have power to detect misspecification in the latter two cases. In view of complications arising from the presence of spatial interactions, the test is specifically constructed to address potential misspecifications in weight matrix components that involve an increasing number of spatial links. A new moment condition is introduced to achieve sensitivity to such misspecification and, after suitable centering and linearization, the resulting test statistic has a standard pivotal limit distribution under the null hypothesis and a set of assumptions that regulate the validity of the linearization process.

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 the study by Lyytikäinen (2012) sheds some light on the much contested suitability of SAR modeling with conventional estimation methods 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 method Lyytikäinen (2012) used with policy-based IV estimation. The present work has focused on specification testing in the basic SAR model with possible heteroskedastic errors. The extension to test for correct specification of a SD model requires some additional work to account for dependence in the exogenous regressors induced by the term WX and is currently under investigation in a separate work. More general applicability requires adaptations of the proposed test to panel data settings.

APPENDIX

We first introduce a technical condition, Assumption A1, which is used in the proof of Theorem 1, to ensure that the linearization remains well defined on the tails of the joint distribution of $(Y_1,\ldots,Y_n)$. We then present Lemma A.1, followed by proofs of Theorems 1–4, proofs of Lemmas 1 and A.1, and lastly proofs of Claims 1 and 2.

Assumption A1. Let p_n and α_n be deterministic, positive sequences satisfying (3.10), $\alpha_n \to \infty$ as $n \to \infty$, and

$$\frac{n^{3/2}}{p_n \alpha_n^{2+\delta}} \to 0 \quad \text{and} \quad \frac{n}{\alpha_n^{2+\delta}} \to 0 \tag{A.1}$$

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

Assumption A1 is a technical requirement on the relative expansion rates among the sequences α_n and p_n, as $n \to \infty$. The sequence α_n is a technical device to ensure that the expansion in Theorem 1 remains well defined on the tails of the distribution of $(Y_1,\ldots,Y_n)$, since the latter does not have bounded support. The practitioner does not have to choose α_n as it does not enter explicitly in the test statistics, but its rate impacts the error of the approximation in Theorem 1 as becomes clear in the proof of Theorem 1, and thereby the central limit theorem approximation in Theorem 2. Indeed, in the tail probability given in (A.17) in that proof, it is evident that the properties of the sequence α_n relate directly to the distribution of ϵ_i via the Markov inequality

$$\mathbb{P}\left(\max_{i\le n}|\epsilon_i| > K\alpha_n\right) = O\left(\frac{n}{\alpha_n^{2+\delta}}\right) \quad \text{with} \quad \frac{n}{\alpha_n^{2+\delta}} \to 0. \tag{A.2}$$

More specifically, the heavier the tails of the distribution of the ϵ_i (i.e., the smaller is the additional moment exponent δ in Assumption 1), the faster α_n needs to diverge to guarantee that $n/\alpha_n^{2+\delta} \to 0$.

Evidently, (A.2) and the properties of the sequence α_n relate directly to the extreme value properties of the innovations ϵ_i. Since Assumption 1 imposes a uniform bound on the variance of ϵ_i, we can write $\epsilon_i = \sigma_i\zeta_i$ with uniformly bounded scale coefficients σ_i and an i.i.d. sequence ζ_i. By extreme value theory, if $\zeta_i \sim_{iid} \mathcal{N}(0,1)$, we have $max_{1\le i\le n}|\zeta_i| = O_p(\sqrt{\log n})$ and thus $\max_{1\le i\le n}|\epsilon_i| = O_p(\sqrt{\log n})$. Thus, a sequence such as $\alpha_n = n^b$ with any $b > 0$ satisfies (A.2). On the other hand, if ζ_i has Pareto tails of the form $\frac{A(z)}{|z|^{1+a}}$ for some $a \ge 1$ and $A(z) = O(1)$ as $|z| \to \infty$, then $\max_{1\le i\le n}|\zeta_i| = O_p(n^{1/a})$ and hence $\max_{1\le i\le n}|\epsilon_i| = O_p(n^{1/a})$. Moreover, by standard extreme value distribution theory, for example, Resnick (2008), for ζ_i, we would have

$$\mathbb{P}(\max_{1\le i\le n}|\zeta_i| \le x) \sim e^{-1/x^a}, \text{ as } x \to \infty. \tag{A.3}$$

Correspondingly, given boundedness of the σ_is, for $\max_{1\le i\le n}|\epsilon_i|$, we also have, as $x \to \infty$,

$$\mathbb{P}(\max_{1\le i\le n}|\epsilon_i| \le x) \sim e^{-1/x^a}, \tag{A.4}$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

and hence $\mathbb{P}(\max_{1\leq i\leq n}|\epsilon_i|\geq x)\sim 1-e^{-1/x^a}\sim \frac{1}{x^a}$. It follows that any sequence α_n for which $n^{1/a}/\alpha_n \to 0$ will satisfy the second part in (A.2) since, under Assumption 1, a must satisfy $a>2+\delta$, with $\delta>0$.

Also, in view of $n/\alpha_n^{2+\delta}\to 0$, a sufficient condition for the first relative rate condition in Assumption A1 to hold is $\sqrt{n}/p_n = O(1)$. In general, a fast diverging p_n implies size stability of our test. However, the slower p_n is, the higher the power of the test against a network misspecification, since a slow-diverging p_n ensures a relevant covariation between the reduced-form residuals and the exponential function e^{Y_i/p_n}, as appearing in (3.11). Thus, choosing p_n such that $\sqrt{n}/p_n = O(1)$ is not necessarily a good empirical strategy, even though it would be compatible with Assumption A1. In Section 6, we discuss the relevant practical implication of the choice of p_n.

We state the following Lemma A.1 which is used in the proofs of theorems. The proof of Lemma A.1 can be found after the proofs of the main theorems.

LEMMA A.1. *Let Assumptions 1–6 hold and p_n satisfy (3.10). Let $A=A(\theta)$ be any generic $n\times n$ matrix such that for all sufficiently large n and for all θ, $|a_{ij}(\theta)|=O(1)$ for all $i,j=1,\dots,n$, $\hat{A}=A(\hat{\theta})$, $A=A(\theta_0)$ and $\Sigma=diag(\sigma_1^2,\dots,\sigma_n^2)$. Furthermore, we assume that each component of $A(\theta)$ has continuous derivatives for each $\theta\in\Theta$ and it is such that, for each $p=1,\dots k+1$, $i,j=1,\dots,n$, $\left|\partial a_{ij}(\theta_0)/\partial\theta_p\right|=O(1)$, where $\partial a_{ij}(\theta_0)/\partial\theta_p$ denotes the partial derivative evaluated at $\theta=\theta_0$. We obtain*

$$\frac{t_Y}{\sqrt{n}p_n}\left(tr(\hat{A}\hat{\Sigma})-tr(A\Sigma)\right)=O_p\left(\frac{1}{p_n}\right). \tag{A.5}$$

Proofs of Theorems 1–4

Proof of Theorem 1. We start the proof of Theorem 1 by obtaining a Taylor expansion of the uncentered statistic in (3.11) and thence deducing the correct sequence that ensures (approximately) zero mean under $\mathcal{H}_0$. Let $\underline{e}(t)=(e^{t'(X_1-\bar{X})},\dots,e^{t'(X_n-\bar{X})})'$ and $\underline{e}(t_Y)=(e^{t'(Y_1-\bar{Y})/p_n},\dots,e^{t'(Y_n-\bar{Y})/p_n})'$. In matrix form, the uncentered version of (3.11) is

$$\frac{1}{n}\begin{pmatrix}\underline{e}(t)'(Y-S^{-1}(\hat{\lambda})X\hat{\beta})\\ \underline{e}(t_Y)'(Y-S^{-1}(\hat{\lambda})X\hat{\beta})\end{pmatrix}, \tag{A.6}$$

and by MVT, we obtain

$$\frac{1}{n}\begin{pmatrix}\underline{e}(t)'(Y-S^{-1}(\hat{\lambda})X\hat{\beta})\\ \underline{e}(t_Y)'(Y-S^{-1}(\hat{\lambda})X\hat{\beta})\end{pmatrix}$$
$$=\frac{1}{n}\begin{pmatrix}\underline{e}(t)'\\ \underline{e}(t_Y)'\end{pmatrix}S^{-1}(\lambda_0)\epsilon-\frac{1}{n}\begin{pmatrix}\underline{e}(t)'\\ \underline{e}(t_Y)'\end{pmatrix}S^{-1}(\bar{\lambda})\left(R(\bar{\lambda})X\bar{\beta}\quad X\right)(\hat{\theta}-\theta_0), \tag{A.7}$$

with $||\bar{\beta}_i-\beta_{0i}||<||\hat{\beta}_i-\beta_{0i}||,\ \ i=1,\dots,k$, and $||\bar{\lambda}-\lambda_0||<||\hat{\lambda}-\lambda_0||$. By plugging (3.3) into (A.7), the latter becomes

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

$$\frac{1}{n}\begin{pmatrix}\underline{e}(t)'\\ \underline{e}(t_Y)'\end{pmatrix}S^{-1}(\lambda_0)\epsilon$$
$$-\frac{1}{n}\begin{pmatrix}\underline{e}(t)'\\ \underline{e}(t_Y)'\end{pmatrix}S^{-1}(\bar{\lambda})\left(R(\bar{\lambda})X\bar{\beta}\quad X\right)\left(\left(\mathbb{B}'\mathbb{A}^{-1}\mathbb{B}\right)^{-1}\mathbb{B}'\mathbb{A}^{-1}\frac{1}{n}\mathbb{Z}'\epsilon+O_p\left(\frac{1}{n}\right)\right)$$
$$=\frac{1}{n}\begin{pmatrix}\underline{e}(t)'\\ \underline{e}(t_Y)'\end{pmatrix}S^{-1}(\lambda_0)\epsilon$$
$$-\frac{1}{n}\begin{pmatrix}\underline{e}(t)'\\ \underline{e}(t_Y)'\end{pmatrix}S^{-1}(\bar{\lambda})\left(R(\bar{\lambda})X\bar{\beta}\quad X\right)\left(\mathbb{B}'\mathbb{A}^{-1}\mathbb{B}\right)^{-1}\mathbb{B}'\mathbb{A}^{-1}\frac{1}{n}\mathbb{Z}'\epsilon+O_p\left(\frac{1}{n}\right), \quad \textbf{(A.8)}$$

where the last equality follows after observing that under Assumptions 2–5

$$\frac{1}{n}\begin{pmatrix}\underline{e}(t)'\\ \underline{e}(t_Y)'\end{pmatrix}S^{-1}(\lambda)\left(R(\lambda)X\beta\quad X\right)=O_p(1) \quad \textbf{(A.9)}$$

uniformly in $\theta\in\Theta$. We want to show that the centering sequence $t_Y tr(\hat{S}^{d'}\hat{Q}\hat{\Sigma})/p_n$ allows us to conclude

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\underline{e}(t)'(Y-S^{-1}(\hat{\lambda})X\hat{\beta})\\ \underline{e}(t_Y)'(Y-S^{-1}(\hat{\lambda})X\hat{\beta})-\frac{t_Y}{p_n}tr(S^{d'}Q\Sigma)\end{pmatrix}=\frac{1}{\sqrt{n}}\Psi(t)\epsilon+o_p(1), \quad \textbf{(A.10)}$$

for a suitable definition of $\Psi(t)$. □

We start by a formal Taylor expansion of the exponential function in the second component of (A.6) and we write

$$e^{\frac{t_Y(Y_i-\bar{Y})}{p_n}}=1+\frac{t_Y(Y_i-\bar{Y})}{p_n}+\frac{t_Y^2(Y_i-\bar{Y})^2}{2p_n^2}+O_p\left(\frac{(Y_i-\bar{Y})^3}{p_n^3}\right). \quad \textbf{(A.11)}$$

Since Y_i, $i=1,\ldots,n$, do not have bounded support in general, we ensure that $\mathbb{P}\left((Y_i-\bar{Y})/p_n>K\right)$ remains sufficiently small for some arbitrarily large K as $p_n\to\infty$ by writing

$$e^{\frac{t_Y(Y_i-\bar{Y})}{p_n}}=1+\frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}\left(\max_{i\le n}|\epsilon_i|\le K\alpha_n\right)+\frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i|>K\alpha_n)$$
$$+\frac{t_Y^2(Y_i-\bar{Y})^2}{2p_n^2}\mathbb{P}\left(\max_{i\le n}|\epsilon_i|\le K\alpha_n\right)$$
$$+\frac{t_Y^2(Y_i-\bar{Y})^2}{2p_n^2}\mathbb{P}(\max_{i\le n}|\epsilon_i|>K\alpha_n)+O_p\left(\frac{(Y_i-\bar{Y})^3}{p_n^3}\right), \quad \textbf{(A.12)}$$

where α_n and p_n satisfy Assumption A1.[9]

Note 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 and 5 and the functional form of m_i in (2.3), m_i is bounded and $\mathbb{P}(\max_{i\le n}|\epsilon_i|\le K_1\alpha_n)=\mathbb{P}(\max_{j\le n}|Y_j-\bar{Y}|\le K_2\alpha_n)$ for some suitable constants K_1 and K_2. Therefore, we can focus on the tail probability of ϵ_i, $i=1,\ldots,n$ to suitably bound the tail probability of Y_i. On the other hand,

[9] As discussed after Assumption A1, the tail behavior of ϵ_i plays a key role in determining α_n

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

under Assumptions 1, 2, and 5, the functional form of m_i in (2.3) and the definition of S^d, we have

$$\mathbb{E}(Y_i-\bar{Y})^2 \le K\mathbb{E}\left(\sum_{j=1}^{n} s^{d,ij}\epsilon_j\right)^2 \le K\sum_{j=1}^{n}(s^{d,ij})^2 \le K\max_{i,j\le n}|s^{d,ij}|\max_{i\le n}\sum_{j=1}^{n}|s^{d,ij}| \le K, \tag{A.13}$$

so that, by Markov's inequality, $Y_i-\bar{Y}=O_p(1)$ for each i. Thus, from (A.13), the fourth and fifth terms in (A.12) are $O_p(\mathbb{P}(\max_{i\le n}|\epsilon_i|<K\alpha_n)/p_n^2)$ and $O_p(\mathbb{P}(\max_{i\le n}|\epsilon_i|>K\alpha_n)/p_n^2)$, respectively.

Now,

$$\mathbb{P}\left(\max_{i\le n}|\epsilon_i|\le K\alpha_n\right)=\mathbb{P}\left(|\epsilon_1|\le K\alpha_n\cap|\epsilon_2|\le K\alpha_n\cap\cdots\cap|\epsilon_n|\le K\alpha_n\right)=\prod_{i=1}^{n}\mathbb{P}(|\epsilon_i|<K\alpha_n)$$
$$=\prod_{i=1}^{n}\left(1-\mathbb{P}(|\epsilon_i|>K\alpha_n)\right). \tag{A.14}$$

By Markov's inequality, for each $i=1,\dots,n$,

$$\mathbb{P}\left(|\epsilon_i|>K\alpha_n\right)\le\frac{\mathbb{E}|\epsilon_i|^{2+\delta}}{K^{2+\delta}\alpha_n^{2+\delta}}=O\left(\frac{1}{\alpha_n^{2+\delta}}\right)=o(1), \tag{A.15}$$

where the first equality follows under Assumption 1 for $\delta>0$ and the last equality follows from Assumption A1. Thus

$$\mathbb{P}\left(\max_{i\le n}|\epsilon_i|\le K\alpha_n\right)=1+O\left(\frac{n}{\alpha_n^{2+\delta}}\right), \tag{A.16}$$

where the accuracy of the latter approximation increases for a given α_n when δ is large (i.e., when the distribution of the ϵ_i is thin-tailed). Also, from (A.16),

$$\mathbb{P}\left(\max_{i\le n}|\epsilon_i|>K\alpha_n\right)=1-\mathbb{P}\left(\max_{i\le n}|\epsilon_i|\le K\alpha_n\right)=O\left(\frac{n}{\alpha_n^{2+\delta}}\right), \tag{A.17}$$

and therefore the RHS of (A.12) can be written as

$$1+\frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}\left(\max_{i\le n}|\epsilon_i|\le K\alpha_n\right)$$
$$+\frac{t_Y(Y_i-\bar{Y})}{p_n}\times O\left(\frac{n}{\alpha_n^{2+\delta}}\right)+O_p\left(\max\left(\frac{1}{p_n^2},\frac{n}{p_n^2\alpha_n^{2+\delta}}\right)\right)$$
$$1+\frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}\left(\max_{i\le n}|\epsilon_i|\le K\alpha_n\right)+\frac{t_Y(Y_i-\bar{Y})}{p_n}\times O\left(\frac{n}{\alpha_n^{2+\delta}}\right)+O_p\left(\frac{1}{p_n^2}\right), \tag{A.18}$$

where the last equality follows under Assumption A1.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

By straightforward algebra, using (A.18), (A.8), premultiplied by $\sqrt{n}$, becomes

$$\frac{1}{\sqrt{n}}\left(\begin{pmatrix} \underline{e}(t)' \\ 1_n' + \frac{t_Y}{p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i| \le K\alpha_n)(Y' - \bar{Y}1_n') \end{pmatrix}\right) S^{-1}(\lambda_0)\epsilon$$
$$-\frac{1}{n}\begin{pmatrix} \underline{e}(t)' \\ 1_n' + \frac{t_Y}{p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i| \le K\alpha_n)(Y' - \bar{Y}1_n') \end{pmatrix} S^{-1}(\lambda_0)\left(R(\lambda_o)X\beta_0 \quad X\right)$$
$$\times\left(\mathbb{B}'\mathbb{A}^{-1}\mathbb{B}\right)^{-1}\mathbb{B}'\mathbb{A}^{-1}\frac{1}{\sqrt{n}}\mathbb{Z}\epsilon + O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{p_n^2}, \frac{n^{3/2}}{p_n\alpha_n^{2+\delta}}\right)\right)$$
$$= \frac{1}{\sqrt{n}}\left(\begin{pmatrix} \underline{e}(t)' \\ 1_n' + \frac{t_Y}{p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i| \le K\alpha_n)(Y' - \bar{Y}1_n') \end{pmatrix}\right) Q\epsilon$$
$$+ O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{p_n^2}, \frac{n^{3/2}}{p_n\alpha_n^{2+\delta}}\right)\right), \tag{A.19}$$

where Q is defined in (3.6). Under Assumption A1, $O_p(\max(1/\sqrt{n}, 1/p_n^2, n^{3/2}/(p_n\alpha_n^{2+\delta}))) = o_p(1)$. In addition to the usual $1/\sqrt{n}$ error (arising when replacing $\bar{\lambda}$ and $\bar{\beta}$ with λ_0 and β_0 and from (3.3)), the error of the approximation depends on two extra terms: (i) the error resulting from linearization, as displayed in (A.18), which is bounded by $1/p_n^2$ under Assumption A1 and (ii) the error that is generated by neglecting the (small) probability that $Y_i - \bar{Y}$ (for some $i = 1,\ldots,n$) might assume an extreme value, that is, the error of neglecting the third term in (A.18), which is bounded by $n^{3/2}/(p_n\alpha_n^{2+\delta})$. The latter rate is straightforward to derive after noticing that the dominant neglected term in (A.7) is

$$\frac{t_Y}{\sqrt{n}p_n}(Y' - \bar{Y}1')S^{-1}\epsilon \times O\left(\frac{n}{\alpha_n^{2+\delta}}\right), \tag{A.20}$$

which is overall bounded by $n^{3/2}/(p_n\alpha_n^{2+\delta})$ from standard arguments involving quadratic forms, upon using $Y - \bar{Y}1 = S^d\mathbb{X}\beta + S^d\epsilon$.

With simple manipulations (A.19) becomes

$$\frac{1}{\sqrt{n}}\begin{pmatrix} \underline{e}(t)' \\ 1_n' \end{pmatrix} Q\epsilon + \frac{1}{\sqrt{n}}\begin{pmatrix} 0_{1\times n} \\ \frac{t_Y}{p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i| \le K\alpha_n)\beta_0' X' S^{d'} \end{pmatrix} Q\epsilon$$
$$+ \frac{1}{\sqrt{n}}\begin{pmatrix} 0_{1\times 1} \\ \frac{t_Y}{p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i| \le K\alpha_n)\epsilon' S^{d'} Q\epsilon \end{pmatrix} + O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{p_n^2}, \frac{n^{3/2}}{p_n\alpha_n^{1+\delta}}\right)\right). \tag{A.21}$$

By Lemma 1, $||Q||_\infty + ||Q'||_\infty < K$, and under Assumptions 1–5, we can show by standard arguments that $\mathbb{E}(\epsilon' S^{d'} Q\epsilon) = O(n)$, such that the third term in (A.21) does not have mean zero in the limit, unless we impose $\sqrt{n}/p_n \to 0$ as $n \to \infty$ and it is unbounded in general. Hence, by applying the suitable centering sequence $t_Y tr(\hat{S}^{d'}\hat{Q}\hat{\Sigma})/p_n\sqrt{n}$, we transform (A.21) as

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\begin{pmatrix}\underline{e}(t)'\\ 1_n'\end{pmatrix}Q\epsilon+\frac{1}{\sqrt{n}}\begin{pmatrix}0_{1\times n}\\ \frac{t_Y}{p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i|\le K\alpha_n)\beta_0'X'S^{d'}\end{pmatrix}Q\epsilon\\
&\quad+\frac{1}{\sqrt{n}}\begin{pmatrix}0_{1\times 1}\\ \frac{t_Y}{p_n}\left(\mathbb{P}(\max_{i\le n}|\epsilon_i|\le K\alpha_n)\epsilon'S^{d'}Q\epsilon-tr(S^{d'}Q\Sigma)\right)\end{pmatrix}\\
&\quad+\frac{1}{\sqrt{n}}\begin{pmatrix}0_{1\times 1}\\ \frac{t_Y}{p_n}\left(tr(S^{d'}Q\Sigma)\right)-tr(\hat{S}^{d'}\hat{Q}\hat{\Sigma})\end{pmatrix}\\
&\quad+O_p\left(\max\left(\frac{1}{\sqrt{n}},\frac{1}{p_n^2},\frac{n^{3/2}}{p_n\alpha_n^{2+\delta}}\right)\right)\\
&=\frac{1}{\sqrt{n}}\begin{pmatrix}\underline{e}(t)'\\ 1_n'\end{pmatrix}Q\epsilon+O_p\left(\max\left(\frac{1}{\sqrt{n}},\frac{1}{p_n},\frac{1}{p_n^2},\frac{n}{p_n\alpha_n^{4+\delta}}\frac{n^{3/2}}{p_n\alpha_n^{2+\delta}}\right)\right)+O_p\left(\frac{1}{p_n}\right)\\
&=O_p\left(\max\left(\frac{1}{\sqrt{n}},\frac{1}{p_n},\frac{n^{3/2}}{p_n\alpha_n^{2+\delta}}\right)\right),
\end{aligned}
\tag{A.22}
$$

where the first equality in the last displayed expression follows from Lemma A.1 and by observing that from Lemma 1 and under Assumptions 1–5, we obtain

$$
\begin{aligned}
&\frac{t_Y}{\sqrt{n}p_n}\mathbb{P}(\max_{i\le n}|\epsilon_i|\le K\alpha_n)\beta_0'X'S^{d'}Q\epsilon=\frac{t_Y}{\sqrt{n}p_n}\left(1-\mathbb{P}(\max_{i\le n}|\epsilon_i|>K\alpha_n)\right)\beta_0'X'S^{d'}Q\epsilon\\
&=O_p\left(\max\left(\frac{1}{p_n},\frac{n}{p_n\alpha_n^{2+\delta}}\right)\right)=O_p\left(\frac{1}{p_n}\right),
\end{aligned}
\tag{A.23}
$$

under Assumption A1, and

$$
\begin{aligned}
&\frac{t_Y}{p_n\sqrt{n}}(\epsilon'S^{d'}Q\epsilon-tr(S^{d'}Q\Sigma))-\frac{t_Y}{p_n\sqrt{n}}\left(1-\mathbb{P}(\max_{i\le n}|\epsilon_i|\le K\alpha_n)\right)\epsilon'S^{d'}Q\epsilon\\
&\quad=O\left(\max\left(\frac{1}{p_n},\frac{n^{3/2}}{p_n\alpha_n^{2+\delta}}\right)\right).
\end{aligned}
\tag{A.24}
$$

The second equality in (A.22) follows from Assumption A1.

Thus, from (A.22) and letting $\Psi=\Psi(t)$ as in (3.8),

$$
\sqrt{n}\hat{M}(t)=\frac{1}{\sqrt{n}}\Psi(t)\epsilon+o_p(1). \tag{A.25}
$$

Proof of Theorem 2. 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

$$
u_i=u_{in}(t)=\frac{1}{\sqrt{n}}\sum_{s=1}^{2}b_s\psi_{si}\epsilon_i. \tag{A.26}
$$

We therefore have

$$
v=v_n(t,t_Y)\equiv\mathbb{V}ar(\sqrt{n}b'\hat{M}(t,t_Y))=\sum_{i=1}^{n}\mathbb{V}ar(u_i)=\sum_{i=1}^{n}E(u_i^2)=\frac{1}{n}b'\Psi\Sigma\Psi'b, \tag{A.27}
$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

which is $O(1)$ under Assumptions2–5 and 8 and by Lemma 1, and it is nonzero as $n \to \infty$ under Assumption 9.

Let $z_i = z_{in}(t, t_Y) = v^{-1/2} u_i = v_n(t, t_Y)^{-1/2} u_i$. We stress that although in the limit the dependence on t_Y is lost, for each finite n, $z_i = z_{in}(t, t_Y)$ and $v = v_n(t, t_Y)$. By the Lindeberg–Feller central limit theorem, if (conditional on $\mathbb{X}$) for each $\zeta > 0$,

$$\sum_{i=1}^{n} \mathbb{E}(z_i^2 1(|z_i > \zeta|)) \underset{p}{\to} 0, \tag{A.28}$$

then $\sum_{i=1}^{n} z_i \to_d \mathcal{N}(0,1)$ pointwise in (t). Thus, the claim (3.15) follows straightforwardly conditional on X, with

$$\mathbb{V}(t) = \lim_{n \to \infty} \frac{1}{n} \Psi \Sigma \Psi'. \tag{A.29}$$

We prove (A.28) by verifying the sufficient Lyapunov condition that pointwise in (t, t_Y) and conditional on X

$$\sum_{i=1}^{n} \mathbb{E}|z_i|^{2+\delta} \to 0. \tag{A.30}$$

Since $v = v_n(t, t_Y) = O(1)$ and is nonzero pointwise in (t, t_Y), we consider equivalently $\sum_i \mathbb{E}|u_i|^{2+\delta}$. Under Assumption 1, we have

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}|u_i|^{2+\delta} &\leq \left(\frac{1}{n}\right)^{1+\delta/2} K \sum_i |\sum_{s=1}^{2} b_s \psi_{si}|^{2+\delta} \\
&\leq K \left(\frac{1}{n}\right)^{1+\delta/2} \max_i |\sum_{s=1}^{2} b_s \psi_{si}|^{\delta} \sum_i \left(\sum_{s=1}^{2} b_s \psi_{si}\right)^2 \\
&\leq K \frac{1}{n^{\delta/2}} \max_i |\sum_{s=1}^{2} b_s \psi_{si}|^{\delta} = o(1)
\end{aligned} \tag{A.31}$$

since $\sum_i (\sum_s^2 b_s \psi_{si})^2 / n = b' \Psi \Psi' b / n = O(1)$ under Assumptions 2–5 and 8 and Lemma 1, and it is nonzero under Assumption 9. Also, from Lemma 1, for each $i = 1, \ldots, n$ and $s = 1, 2$, $|\psi_{si}| = O(1)$. □

Proof of Theorem 3. We prove the claim for the typical (s, v) element, that is,

$$\frac{1}{n} \hat{\Psi}_s' \hat{\Sigma} \hat{\Psi}_v - \frac{1}{n} \Psi_s' \Sigma \Psi_v = o_p(1). \tag{A.32}$$

We write

$$\Psi_s' \Sigma \Psi_v = tr(\Psi_s' \Sigma \Psi_t) = tr(\Sigma \Psi_v \Psi_s'), \tag{A.33}$$

where it is straightforward to verify that the $n \times n$ matrix $\Psi_v \Psi_s'$ has bounded elements since, from Lemma 1, $|\psi_{si}| = O(1)$ for all $i = 1, \ldots, n$ and $s = 1, 2$. Also, from (3.8), by standard algebra each element of $\partial \Psi_v \Psi_s' / \partial \theta_p$ for $p = 1, \ldots, k+1$ is bounded and continuous for each θ. The proof of Theorem 3 follows then from Lemma A.1 with p_n replaced by $\sqrt{n}$. □

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

Proof of Theorem 4. Let α_n be a deterministic sequence such that $\alpha_n \to \infty$ and $n \to \infty$ and p_n as in (3.10). Write $\hat{M}(t,t_Y) = (\hat{M}_1(t), \hat{M}_2(t_Y))'$. In order to prove the claim in Theorem 4, and thus consistency of the test based on (3.11), we show that $\sqrt{n}\hat{M} \to_p \pm\infty$ under Assumption 11. Then, under Assumption 7', $\hat{T}(t,t_Y) = nM(\hat{\theta},t,t_Y)'\hat{A}^{-1}(t)M(\hat{\theta},t,t_Y) \to \infty$.

It is enough to show that under Assumption 11 and $\mathcal{H}_1$ in (2.7), $\text{plim}_{n\to\infty}\hat{M}_2(t_Y) \neq 0$. We can write

$$\begin{aligned}\hat{M}_2(t_Y) &= \frac{1}{n}\sum_{i=1}^{n}\Big(g_i(X) - m_i(X,\theta^\sharp)\Big)\, e^{t_Y(Y_i-\bar{Y})/p_n} + \frac{1}{n}\sum_{i=1}^{n}\eta_i e^{t_Y(Y_i-\bar{Y})/p_n} \\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\big(m_i(X,\hat{\theta}) - m_i(X,\theta^\sharp)\big)e^{t_Y(Y_i-\bar{Y})/p_n} - \frac{t_Y}{np_n}(tr(\hat{S}^d\hat{Q}\hat{\Sigma})). \end{aligned} \quad \textbf{(A.34)}$$

The first term in (A.34) is

$$\frac{1}{n}\sum_{i=1}^{n}\Big(g_i(X) - m_i(X,\theta^\sharp)\Big) + \frac{1}{n}\sum_{i=1}^{n}\Big(g_i(X) - m_i(X,\theta^\sharp)\Big)\Big(e^{t_Y(Y_i-\bar{Y})/p_n} - 1\Big). \quad \textbf{(A.35)}$$

Under Assumption 11,

$$\frac{1}{n}\sum_{i=1}^{n}\Big(g_i(X) - m_i(X,\theta^\sharp)\Big) \underset{p}{\to} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big(g_i(X) - m_i(X,\theta^\sharp)\Big) \neq 0, \quad \textbf{(A.36)}$$

where the limit on the RHS of the last displayed expression exists under Assumptions 2 and 8 and (2.3). The second term can be written as

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\Big(g_i(X) - m_i(X,\theta^\sharp)\Big) \\ &\quad\times\left(\frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}(\eta_i \le K\alpha_n) + \frac{t_Y(Y_i-\bar{Y})}{p_n}\mathbb{P}(\eta_i > K\alpha_n) + O_p\left(\frac{(Y_i-\bar{Y})^2}{p_n^2}\right)\right). \end{aligned} \quad \textbf{(A.37)}$$

Since

$$\mathbb{P}(\eta_i > K\alpha_n) \le \frac{\mathbb{E}(\eta_i)^2}{K^2\alpha_n^2}, \quad \textbf{(A.38)}$$

the expected value of the modulus of (A.37) is bounded by

$$\begin{aligned}&\frac{t_Y}{p_n n}\max_{1\le i\le n}\mathbb{P}(\eta_i \le K\alpha_n)\sum_{i=1}^{n}\mathbb{E}|(g_i(X) - m_i(X,\theta^\sharp))(Y_i-\bar{Y})| \\ &\quad + \max_{1\le i\le n}\frac{\mathbb{E}(\eta_i)^2}{K^2\alpha_n^2}\frac{t_Y}{p_n}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|(g_i(X) - m_i(X,\theta^\sharp))(Y_i-\bar{Y})| + O_p\left(\frac{(Y_i-\bar{Y})^2}{p_n^2}\right) \\ &\le \frac{t_Y}{p_n n}\max_{1\le i\le n}\sum_{i=1}^{n}\mathbb{E}|(g_i(X) - m_i(X,\theta^\sharp))(Y_i-\bar{Y})| \\ &\quad + \max_{1\le i\le n}\frac{K}{\alpha_n^2}\frac{t_Y}{p_n}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|(g_i(X) - m_i(X,\theta^\sharp))(Y_i-\bar{Y})| + O_p\left(\frac{(Y_i-\bar{Y})^2}{p_n^2}\right), \end{aligned} \quad \textbf{(A.39)}$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

under Assumptions 9 and 11. Under Assumptions 2, 8, and 9, by Cauchy–Schwarz for each $i = 1,\ldots,n$,

$$\mathbb{E}|(g_i(\mathbb{X}) - m_i(X,\theta^\sharp)(Y_i - \bar{Y})| \leq \left(\mathbb{E}(g_i(X) - m_i(X,\theta^\sharp))^2\mathbb{E}(Y_i - \bar{Y})^2\right)^{1/2} = O(1), \quad \textbf{(A.40)}$$

so that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|(g_i(X) - m_i(X,\theta^\sharp))(Y_i - \bar{Y})| = O(1), \quad \textbf{(A.41)}$$

and thus the expression in (A.37) is $O_p(1/p_n) = o_p(1)$ from (3.10).

We can show that the second term in (A.34) is $o_p(1)$ under Assumption 9 and (3.10) by using a similar argument to that given above to deal with the linearization on the tails. The details are omitted to avoid repetition. The third term in (A.34) can be written as

$$\frac{1}{n}\sum_{i=1}^{n}(m_i(X,\hat{\theta}) - m_i(X,\theta^\sharp)) + \frac{1}{n}\sum_{i=1}^{n}(m_i(X,\hat{\theta}) - m_i(X,\theta^\sharp))\left(e^{t_Y(Y_i-\bar{Y})/p_n} - 1\right). \quad \textbf{(A.42)}$$

By the mean value theorem, the first term can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{dm_i(X,\bar{\theta}_i)}{d\theta}(\hat{\theta} - \theta^\#) = \frac{1}{n}\sum_{i=1}^{n}S^{(i)'}(\bar{\lambda}_i)\left(R(\bar{\lambda}_i)X\bar{\beta}_i \quad X\right)(\hat{\theta} - \theta^\#), \quad \textbf{(A.43)}$$

where $\bar{\lambda}_i$ and $\bar{\beta}_i$ satisfy, respectively, $|\bar{\lambda}_i - \lambda^\#| < |\hat{\lambda} - \lambda^\#|$ and $||\bar{\beta}_i - \beta^\#|| < ||\hat{\beta} - \beta^\#||$, for $i = 1,\ldots,n$. Under Assumption 10, $\hat{\theta} - \theta^\# = o_p(1)$. Therefore, the last term in (A.42) is $o_p(1)$ as long as we can show that each component of the $1 \times (k+1)$ vector

$$\frac{1}{n}\sum_{i=1}^{n}S^{(i)'}(\bar{\lambda})\left(R(\lambda)X\beta \quad X\right) \quad \textbf{(A.44)}$$

is $O_p(1)$ for each λ and β. For simplicity of notation, in order to assess the rate of (A.44), 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.44) has expectation bounded by

$$K\sup_i\sum_{j=1}^{n}|a_{ij}(\lambda)| = O(1), \quad \textbf{(A.45)}$$

under Assumptions 4 and 5.

By Markov's inequality, the first term in (A.42) is $o_p(1)$. Similarly, the second term in (A.42) can handled by linearization as illustrated above, the details being omitted. Finally, the last term in (A.34) is $O_p\left(\frac{1}{p_n}\right) = o_p(1)$ under Assumptions 2, 4, and 5 and (3.10). □

Proofs of Additional Lemmas and Claims

Proof of Lemma 1. Under Assumption 5, the claim follows as long as

$$\frac{1}{n}||H\Omega^{-1}\mathbb{B}'\mathbb{A}^{-1}Z'||_\infty < K, \quad \textbf{(A.46)}$$

where (limited to the scope of this lemma) we set $H = \left(RX\beta_0 \quad X\right)$and $\Omega = \mathbb{B}'\mathbb{A}^{-1}\mathbb{B}$.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

Let c denote an arbitrarily small constant, which, as usual, is allowed to take different values at each step. Under Assumptions 2–5, $||R||_\infty + ||R'||_\infty < K$, and thus all elements of H are $O(1)$, conditionally on $\mathbb{X}$. Also, under Assumptions 2 and 6, all elements of $\mathbb{B}$ are $O(1)$. Under Assumption 6, $\mathrm{eig}_{\min}(\Omega) > c > 0$ $\mathrm{eig}_{\min}(\mathbb{A}) > c > 0$. Thus,

$$\begin{aligned}\frac{1}{n}||H\Omega^{-1}\mathbb{B}'\mathbb{A}^{-1}Z'||_\infty &= \frac{1}{n}\sup_i \sum_{j=1}^{n}|H_i'\Omega^{-1}\mathbb{B}'\mathbb{A}^{-1}Z_j| \\ &\le \frac{1}{n}\sup_i \sum_{j=1}^{n}||H_i'||||Z_j||||\Omega^{-1}||||\mathbb{A}^{-1}||||\mathbb{B}|| \\ &\le K\sup_{i,j}||H_i'||||Z_j||||\Omega^{-1}||||\mathbb{A}^{-1}||||\mathbb{B}|| \le K,\end{aligned}$$

since under Assumptions 2 and 6 $\sup_i ||H_i|| = \sup_i (H_i'H_i)^{1/2} = O(1)$, $||\mathbb{B}|| = O(1)$ and

$$\begin{aligned}||\Omega^{-1}|| &= \mathrm{eig}_{\max}(\Omega^{-1}) = \frac{1}{eig_{\min}(\Omega)} < \frac{1}{c} < K, \\ ||\mathbb{A}^{-1}|| &= \mathrm{eig}_{\max}(\mathbb{A}^{-1}) = \frac{1}{eig_{\min}(\mathbb{A})} < \frac{1}{c} < K. \end{aligned} \quad \textbf{(A.47)}$$

□

Proof of Lemma A.1. In the sequel, we denote by a_{ij} and $\hat{a}_{ij}$ the (i,j)th elements of $A(\theta_0)$ and $A(\hat{\theta})$, respectively. We prove Lemma A.1 by showing

$$\frac{1}{\sqrt{np_n}}\sum_{i=1}^{n}\left(\epsilon_i^2 - \sigma_i^2\right)a_{ii} = O_p\left(\frac{1}{p_n}\right), \quad \textbf{(A.48)}$$

$$\frac{1}{\sqrt{np_n}}\sum_{i=1}^{n}\left(\hat{\epsilon}_i^2 - \epsilon_i^2\right)a_{ii} = O_p\left(\frac{1}{p_n}\right), \quad \textbf{(A.49)}$$

and

$$\frac{1}{\sqrt{np_n}}\sum_{i=1}^{n}\hat{\epsilon}_i^2\left(\hat{a}_{ii} - a_{ii}\right) = O_p\left(\frac{1}{p_n}\right). \quad \textbf{(A.50)}$$

We show (A.48) by Markov's inequality, after observing that under Assumption 1 and the boundedness of elements of A, the LHS has mean zero and variance bounded by

$$\frac{K}{np_n^2}\sum_{i=1}^{n}a_{ii}^2 = O\left(\frac{1}{p_n^2}\right). \quad \textbf{(A.51)}$$

To demonstrate (A.49), write

$$\hat{\epsilon} = S(\hat{\lambda})y - X\hat{\beta} = \epsilon - X(\hat{\beta} - \beta_0) - (\hat{\lambda} - \lambda_0)RX\beta_0 - (\hat{\lambda} - \lambda_0)R\epsilon, \quad \textbf{(A.52)}$$

and recall $R = R(\lambda_0) = WS^{-1}(\lambda_0)$ and $||R||_\infty + ||R'||_\infty < K$ under Assumptions 4 and 5. Let $\hat{\epsilon}_i = \epsilon_i + \mathsf{b}_i + \mathsf{c}_i$, with

$$\mathsf{b}_i = -(\hat{\lambda} - \lambda_0)R_i'\epsilon \quad \text{and} \quad \mathsf{c}_i = -X_i'(\hat{\beta} - \beta_0) - (\hat{\lambda} - \lambda_0)R_i'X\beta_0, \quad \textbf{(A.53)}$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

and thus

$$\hat{\epsilon}_i^2 - \epsilon_i^2 = \mathsf{b}_i^2 + \mathsf{c}_i^2 + 2\epsilon_i \mathsf{b}_i + 2\epsilon_i \mathsf{c}_i + 2\mathsf{c}_i \mathsf{b}_i. \tag{A.54}$$

The LHS in (A.49) is therefore (where unqualified summations range from 1 to n)

$$\frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\mathsf{b}_i^2 + \frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\mathsf{c}_i^2 + \frac{2}{\sqrt{n}p_n}\sum_i a_{ii}\mathsf{b}_i\epsilon_i + \frac{2}{\sqrt{n}p_n}\sum_i a_{ii}\mathsf{c}_i\epsilon_i + \frac{2}{\sqrt{n}p_n}\sum_i a_{ii}\mathsf{b}_i\mathsf{c}_i. \tag{A.55}$$

Under Assumptions 1–5, $\hat{\lambda} - \lambda_0 = O(1/\sqrt{n})$ (e.g., Kelejian and Prucha, 1998) and the first term in (A.55) is

$$\frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\sum_j\sum_t R_{ij}R_{it}\epsilon_j\epsilon_t(\hat{\lambda}-\lambda)^2, \tag{A.56}$$

where the last factor is of order $O_p(1/n)$. Moreover, we write

$$\frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\sum_j\sum_t R_{ij}R_{it}\epsilon_j\epsilon_t = \frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\sum_j R_{ij}^2\epsilon_j^2 + \frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\sum_j\sum_{t\neq j} R_{ij}R_{it}\epsilon_j\epsilon_t. \tag{A.57}$$

The modulus of the first term in (A.57) has expectation bounded by

$$\frac{K}{\sqrt{n}p_n}\sum_i |a_{ii}|\sum_j R_{ij}^2 \le \frac{K}{\sqrt{n}p_n}\sum_i\sum_j R_{ij}^2 = \frac{K}{\sqrt{n}p_n}tr(RR') = O\left(\frac{\sqrt{n}}{p_n}\right), \tag{A.58}$$

since by standard results $tr(RR') = O(n)$. The second term in (A.57) has mean zero and variance bounded by

$$\begin{aligned}
&\frac{K}{np_n^2}\sum_i\sum_s |a_{ii}||a_{ss}|\sum_j\sum_t |R_{ij}R_{it}R_{sj}R_{st}| \\
&\quad\le \frac{K}{np_n^2}\sum_i\sum_s\sum_j\sum_t |R_{ij}R_{it}|(R_{sj}^2 + R_{st}^2) \\
&\quad\le \frac{K}{np_n^2}\sup_i\sum_t |R_{it}|\sup_j\sum_i |R_{ij}|\sum_s\sum_j R_{sj}^2 + \frac{K}{np_n^2}\sup_i\sum_j |R_{ij}|\sup_t\sum_i |R_{it}|\sum_s\sum_t R_{st}^2 \\
&\quad= O\left(\frac{1}{p_n^2}\right),
\end{aligned} \tag{A.59}$$

where the last equality follows under Assumptions 3–5 and since $tr(RR') = O(n)$. Thus, collecting (A.58), (A.59), and (A.56), the first term in (A.55) is $O_p(1/p_n\sqrt{n})$.

The third term in (A.55) becomes

$$-\frac{2}{\sqrt{n}p_n}\sum_i a_{ii}\epsilon_i\sum_j R_{ij}\epsilon_j(\hat{\lambda}-\lambda_0), \tag{A.60}$$

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

where, again, $(\hat{\lambda}-\lambda_0)=O_p(1/\sqrt{n})$. We write

$$\frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\epsilon_i\sum_j R_{ij}\epsilon_j=\frac{1}{\sqrt{n}p_n}\sum_i a_{ii}R_{ii}\epsilon_i^2+\frac{1}{\sqrt{n}p_n}\sum_i a_{ii}\sum_{j\neq i}R_{ij}\epsilon_i\epsilon_j. \tag{A.61}$$

By standard arguments, the modulus of the first term in (A.61) has expectation which is $O(\sqrt{n}/p_n)$. The second term in (A.61) has variance bounded by

$$\frac{K}{np_n^2}\sum_i\sum_j a_{ii}^2R_{ij}^2+\frac{K}{np_n^2}\sum_i\sum_j|a_{ii}||a_{jj}||R_{ij}||R_{ji}|\leq\frac{K}{np_n^2}tr(RR')+\frac{K}{p_n^2}\sup_{i,j}|R_{ij}|\sup_j\sum_i|R_{ji}|$$
$$=O\left(\frac{1}{p_n^2}\right), \tag{A.62}$$

under Assumptions 1–5. Thus, by Markov's inequality, the third term in (A.55) is $O_p(1/p_n)$.

The fourth term in (A.55) can be written as

$$-\frac{2}{\sqrt{n}p_n}(\hat{\beta}-\beta_0)'\sum_i a_{ii}\epsilon_iX_i-\frac{2}{\sqrt{n}p_n}(\hat{\lambda}-\lambda_0)\sum_i a_{ii}\epsilon_iR_i'X\beta_0, \tag{A.63}$$

where $\hat{\beta}-\beta_0=O_p(1/\sqrt{n})$ and $\hat{\lambda}-\lambda_0=O_p(1/\sqrt{n})$. The term $\sum_i a_{ii}\epsilon_iX_i/\sqrt{n}p_n$ has mean zero and variance bounded by

$$\frac{K}{np_n^2}\sum_i X_iX_i'=\frac{K}{np_n^2}X'X, \tag{A.64}$$

whose components are $O(1/p_n^2)$ under Assumption 2, such that the first term in (A.63) is $O_p(1/\sqrt{n}p_n)$. Similarly, the term $\sum_i a_{ii}\epsilon_iR_i'X\beta_0/\sqrt{n}p_n$ has mean zero and variance bounded by

$$\frac{K}{np_n^2}\beta_0'X'R'RX\beta_0=O\left(\frac{1}{p_n^2}\right), \tag{A.65}$$

under Assumptions 2, 4, and 5, such that the second term in (A.63) is also $O_p(1/\sqrt{n}p_n)$. Therefore, the fourth term in (A.55) is $O_p(1/\sqrt{n}p_n)$.

The fifth term in (A.55) is

$$\frac{2}{\sqrt{n}p_n}\sum_i a_{ii}R_i'\epsilon\left(X_i'(\hat{\beta}-\beta_0)+(\hat{\lambda}-\lambda_0)R_i'X\beta_0\right). \tag{A.66}$$

Consider the terms

$$\frac{1}{\sqrt{n}p_n}\sum_i\sum_j a_{ii}R_{ij}\epsilon_jX_i' \quad\text{and}\quad \frac{1}{\sqrt{n}p_n}\sum_i\sum_j a_{ii}R_{ij}\epsilon_jR_i'X\beta. \tag{A.67}$$

The first term in (A.67) has mean zero and variance bounded by

$$\frac{K}{np_n^2}\sum_i\sum_s\sum_j|a_{ii}||a_{ss}||R_{ij}||R_{sj}||X_i'X_s|\leq\frac{K}{p_n^2}\sup_{i,s}|X_i'X_s|\sup_j\sum_s|R_{sj}|\sup_i\sum_j|R_{sj}|$$
$$=O\left(\frac{1}{p_n^2}\right), \tag{A.68}$$

and the second term has variance bounded by

$$\frac{K}{p_n^2}\sup_{i,s}\beta_0' X' R_s R_i' X\beta_0 \sup_j \sum_s |R_{sj}| \sup_i \sum_j |R_{ij}| = O\left(\frac{n}{p_n^2}\right). \tag{A.69}$$

Collecting (A.66)–(A.69), we conclude that the fifth term in (A.55) is $O_p(1/p_n)$. By straightforward algebra (noting that it does not contain any term in ϵ_i), we can also show that the second term in (A.55) is $O_p(1/p_n\sqrt{n})$. This concludes the proof of (A.49).

To prove (A.50), write by MVT

$$\hat{a}_{ii} = a_{ii} + \sum_{j=1}^{k+1} \frac{\partial \bar{a}_{ii}}{\partial \theta_j}(\hat{\theta}_j - \theta_{j0}), \tag{A.70}$$

where the partial derivative is evaluated at $\bar{\theta}$, satisfying $|\bar{\theta}_j - \theta_{j0}| < |\hat{\theta}_j - \theta_{j0}|$. Thus, the LHS of (A.50) can be written as

$$\frac{1}{\sqrt{n}p_n}\sum_i \hat{\epsilon}_i^2 \sum_{j=1}^{k+1} \frac{\partial \bar{a}_{ii}}{\partial \theta_j}(\hat{\theta}_j - \theta_{j0}). \tag{A.71}$$

By results in (A.48) and (A.49) and under boundedness and continuity of the components of $\partial A(\theta)/\partial\theta_j$,

$$\frac{1}{n}\sum_i \hat{\epsilon}_i^2 \sum_{j=1}^{k+1} \frac{\partial \bar{a}_{ii}}{\partial \theta_j} \underset{p}{\to} \lim_{n\to\infty}\frac{1}{n}\sum_i \sigma_i^2 \sum_{j=1}^{k+1} \frac{\partial a_{ii}}{\partial \theta_j} = O(1). \tag{A.72}$$

Given that under Assumptions 1–5, for each $j = 1, \dots, k+1$, $\hat{\theta}_i - \theta_{i0} = O_p(1/\sqrt{n})$, (A.71) is $O_p(1/p_n)$, concluding the proof. In the above, we may set $A = S^{d'}Q$ and it is straightforward to verify that it satisfies the conditions imposed on A to prove (A.48)–(A.50). □

Proof of Claim 1. We define

$$d_{in}(\theta, X) = d_i(X) = g_i(X) - m_i(X) = \mathbb{E}(Y_i - m_i(X)|X),$$

since $\mathbb{E}(\eta_i|X) = 0$, where in this context, we drop the dependence on θ for notational simplicity. We define the functions

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

the expected values

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

and the marginal probability measures

$$\nu_{is}(B_i) = \frac{1}{c_{is}}\int_{B_i}\int_{\mathbb{R}^k}\cdots\int_{\mathbb{R}^k} d_{is}(x_1,\dots,x_n)dF(x_1)\dots dF(x_n), \quad s = 1,2, \quad i = 1,\dots,n, \tag{A.73}$$

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 *i.i.d.* random vectors X_j, $j = 1, \dots, n$.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

Define the joint probability measure

$$\nu_{is}(B_1,\ldots,B_n) = \frac{1}{c_{is}}\int_{B_1}\int_{B_2}\cdots\int_{B_n} d_{is}(x_1,\ldots,x_n)dF(x_1)\ldots dF(x_n). \tag{A.74}$$

By the law of iterated expectation,

$$\begin{aligned}\mathbb{E}((Y_i - m_i(X))e^{t'X_i}) = \mathbb{E}(e^{t'X_i}d_i(X)) &= \int_{\mathbb{R}^k}\cdots\int_{\mathbb{R}^k} d_{i1}(x_1,\ldots,x_n)e^{t'x_i}dF(x_1)\ldots dF(x_n)\\ &\quad - \int_{\mathbb{R}^k}\cdots\int_{\mathbb{R}^k} d_{i2}(x_1,\ldots,x_n)e^{t'x_i}dF(x_1)\ldots dF(x_n)\\ &= c_{i1}\int_{\mathbb{R}^k} e^{t'x_i}d\nu_{i1}(x_i) - c_{i2}\int_{\mathbb{R}^k} e^{t'x_i}d\nu_{i2}(x_i), \end{aligned}\tag{A.75}$$

from the definitions of $\nu_{i1}(B_i)$ and $\nu_{i2}(B_i)$. 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 $\nu_{i2}(B_i)$.

If $\mathbb{E}((Y_i - m_i(X))e^{t'X_i}) = 0$ 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

$$c_{i1} = c_{i2}. \tag{A.76}$$

Thus, for each t

$$\int_{\mathbb{R}^k} e^{t'x_i}d\nu_{i1}(x_i) = \int_{\mathbb{R}^k} e^{t'x_i}d\nu_{i2}(x_i), \tag{A.77}$$

implying

$$\nu_{i1}(B_i) = \nu_{i2}(B_i) \quad \forall B_i \in \mathbb{R}^k, \quad i = 1,\ldots,n. \tag{A.78}$$

Therefore, from (A.73),

$$\int_{B_i}\int_{\mathbb{R}^k}\cdots\int_{\mathbb{R}^k} d_i(x_1,\ldots,x_n)dF(x_1)\ldots dF(x_n) = 0 \quad \forall B_i, \tag{A.79}$$

implying

$$\int_{\mathbb{R}^k}\cdots\int_{\mathbb{R}^k} d_i(x_1,\ldots,x_n)dF(x_1)\ldots dF(x_{i-1})dF(x_{i+1})\ldots dF(x_n) = 0, \tag{A.80}$$

and thus $\mathbb{E}(Y_i - m_i(X)|X_i) = 0$, concluding the claim. □

Proof of Claim 2. Since the order of the observations is innocuous in spatial models, for convenience, we rearrange the order so that for each $i = 1,\ldots,n$, $v_{ij} = 0$ for $j = \nu(n)+1,\ldots,n$ (with the exception of $v_{ii} = 0$, by construction). By standard arguments, similar to those used in the proofs of Theorem 2, we can write

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_i (Y_i - m_i(\hat{\theta}))e^{tx_i}\\ &= \frac{1}{\sqrt{n}}\sum_i\sum_j (s^{ij} - s^{ij}(\hat{\lambda}))x_j e^{tx_i} + \frac{\tau\beta}{\sqrt{n}}\sum_i\sum_j\sum_p\sum_q s^{ij}v_{jp}\tilde{s}^{pq}x_q e^{tx_i} + O_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_i\sum_j (s^{ij} - s^{ij}(\hat{\lambda}))x_j e^{tx_i} + \frac{\tau\beta}{\sqrt{n}}\sum_i\sum_j\sum_{p=1}^{\nu(n)}\sum_q s^{ij}v_{jp}\tilde{s}^{pq}x_q e^{tx_i} + O_p(1), \end{aligned}\tag{A.81}$$

where, again, unqualified summations range from 1 to n and $s^{ij} = s^{ij}(\lambda)$, $\tilde{s}^{ij} = \tilde{s}^{ij}(\lambda, \tau)$. Under Assumption 2, the modulus of the second term on the RHS of (A.81) has expectation bounded by

$$\frac{K}{\sqrt{n}} \sum_i \sum_j \sum_{p=1}^{\nu(n)} \sum_q |s^{ij}||v_{jp}||\tilde{s}^{pq}|$$
$$\leq \frac{K}{\sqrt{n}} \sup_p \sum_q |\tilde{s}^{pq}| \sup_j \sum_i |s^{ij}| \sup_p \sum_j |v_{jp}| \nu(n) = O\left(\frac{\nu(n)}{\sqrt{n}}\right), \tag{A.82}$$

which is either bounded or converges to zero under the assumption on the deviation matrix V stated in Claim 2. The first term on the RHS of (A.81), after applying a standard MVT, has expectation bounded by

$$\frac{K}{\sqrt{n}} \sum_i \sum_j |(\bar{S}^{-1} W \bar{S}^{-1})_{ij}||\hat{\lambda} - \lambda|, \tag{A.83}$$

where $\bar{S}^{-1} = S^{-1}(\bar{\lambda})$, with $\bar{\lambda}$ being an intermediate value such that $|\bar{\lambda} - \lambda| < |\hat{\lambda} - \lambda|$. In order to conclude boundedness of the last displayed expression, we observe that under the assumptions of Claim 2

$$\hat{\theta} = \theta + \left(\mathbb{B}'\mathbb{A}^{-1}\mathbb{B}\right)^{-1} \mathbb{B}'\mathbb{A}^{-1} \frac{1}{n} \mathbb{Z}' V Y \tau + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.84}$$

Again under the conditions of Claim 2, $\mathbb{E}|Y_i| \leq K$ for $i = 1, \ldots, n$, and so the modulus of each component $j = 1, \ldots, k+1$ of the second term at the RHS of (A.84) has expectation bounded by

$$\frac{K}{n} \sum_i \sum_{p=1}^{\nu(n)} |\mathbb{Z}_{ij}||v_{ip}| \leq \frac{K}{n} \sup_p \sum_i |v_{ip}| \nu(n), \tag{A.85}$$

since $\mathbb{Z}_{ij} = O_p(1)$ by Assumption 6. Thus, for $j = 1, \ldots, k+1$, collecting (A.84) and (A.85) yields

$$|\hat{\theta}_j - \theta_j| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{\nu(n)}{n}\right)\right) = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.86}$$

where the last equality follows since under the conditions of Claim 2 $\nu(n)/\sqrt{n} = O(1)$. Hence, the expression in (A.83) is bounded by

$$\frac{K}{n} n \sup_j \sum_i |(\bar{S}^{-1} W \bar{S}^{-1})_{ij}| = O(1) \tag{A.87}$$

under Assumptions 4 and 5. Thus $\sqrt{n} M_1(\hat{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - m_i(\hat{\theta})) e^{t x_i} = O_p(1)$ as $n \to \infty$, for almost all $t \in \mathbb{R}^k$, concluding the proof of the claim. □

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Lee, J., Phillips, P. C. B., & Rossi, F. (2024). Supplement to "Heteroskedasticity Robust Specification Testing in Spatial Autoregression," *Econometric Theory* Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466624000173.

REFERENCES

Allers, M. A., & Elhorst, J. P. (2005). Tax mimicking and yardstick competition among local governments in the Netherlands. *International Tax and Public Finance*, 12, 493–513.

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

Arraiz, I., Drukker, D. M., Kelejian, H. H., & Prucha, I. R. (2010). A spatial Cliff–Ord-type-model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science*, 50, 592–614.

Bailey, N., Holly, S., & 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.

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

Beenstock, M., & Felsenstein, D. (2012). Nonparametric estimation of the spatial connectivity matrix using spatial panel data. *Geographical Analysis*, 44, 386–397.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20, 105–134.

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

Bierens, H. J. (1988). ARMA memory index modeling of economic time series. *Econometric Theory*, 4, 35–59.

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

Bierens, H. J. (2015). *Addendum to: Consistent model specification tests (1982)*. Mimeo.

Burridge, P. (1980). On the Cliff–Ord test for spatial correlation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42, 107–108.

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

Chirinko, R. S., & 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.

Cliff, A., & Ord, J. (1968). The problem of spatial autocorrelation. Joint Discussion Paper. University of Bristol: Department of Economics, 26, Department of Geography.

Cliff, A., & Ord, J. (1981). *Spatial processes: Models and applications*. Pion.

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

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

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

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

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, 5, 9–28.

Escanciano, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, 101, 531–541.

Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statistica Sinica*, 17, 115–138.

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

Florax, R. J., Folmer, H., & Rey, S. J. (2003). Specification searches in spatial econometrics: the relevance of Hendry’s methodology. *Regional Science and Urban Economics*, 33, 557–579.

Gupta, A., & X. Qu (2021). *Consistent specification testing under spatial dependence*. Mimeo.

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

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

Kasparis, I. (2010). The Bierens test for certain nonstationary models. *Specification Analysis in Honor of Phoebus J. Dhrymes Journal of Econometrics*, 158, 221–230.

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.

Kelejian, H. H., & Piras, G. (2011). An extension of Kelejian’s J-test for non-nested spatial models. *Regional Science and Urban Economics*, 41, 281–292.

Kelejian, H. H., & Piras, G. (2016). An extension of the J-test to a spatial panel data framework. *Journal of Applied Econometrics*, 31, 387–402.

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

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

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

Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157, 53–67.

Koul, H. L., & Stute, W. (1999). Nonparametric model checks for time series. *Annals of Statistics*, 27, 204–236.

Lam, C., & Souza, P. C. L. (2016). Detection and estimation of block structure in spatial weight matrix. *Econometric Reviews*, 35, 1347–1376.

Lam, C., & Souza, P. C. L. (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, 38, 693–710.

Lee, L.-F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22, 307–335.

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

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

Lee, L.-F., & Yu, J. (2012). The C-type gradient test for spatial dependence in spatial autoregressive models. *Letters in Spatial and Resource Sciences*, 5, 119–135.

Liu, T., & Lung-fei, L. (2019). A likelihood ratio test for spatial model selection. *Journal of Econometrics*, 213, 434–458.

Liu, X., & Prucha, I. R. (2018). A robust test for network generated dependence. *Journal of Econometrics*, 207, 92–113.

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

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

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

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, 120–126.

https://doi.org/10.1017/S0266466624000173 Published online by Cambridge University Press

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

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31, 350–371.

Resnick, S. I. (2008). *Extreme values, regular variation, and point processes* (Vol. 4). Springer Science & Business Media.

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

Stinchcombe, M. B., & White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 14, 295–325.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25, 613–641.

Stute, W., & Zhu, L. (2005). Nonparametric checks for single-index models. *Annals of Statistics*, 33, 1048–1083.

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

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