Econometric Theory, **39**, 2023, 107–145.
doi:10.1017/S0266466621000384

CONTINUOUSLY UPDATED INDIRECT INFERENCE IN HETEROSKEDASTIC SPATIAL MODELS

MARIA KYRIACOU
University of Southampton

PETER C.B. PHILLIPS
Yale University
University of Auckland
University of Southampton
Singapore Management University

FRANCESCA ROSSI
University of Verona

Spatial units typically vary over many of their characteristics, introducing potential unobserved heterogeneity which invalidates commonly used homoskedasticity conditions. In the presence of unobserved heteroskedasticity, methods based on the quasi-likelihood function generally produce inconsistent estimates of both the spatial parameter and the coefficients of the exogenous regressors. A robust generalized method of moments estimator as well as a modified likelihood method have been proposed in the literature to address this issue. The present paper constructs an alternative indirect inference (II) approach which relies on a simple ordinary least squares procedure as its starting point. Heteroskedasticity is accommodated by utilizing a new version of continuous updating that is applied within the II procedure to take account of the parameterization of the variance–covariance matrix of the disturbances. Finite-sample performance of the new estimator is assessed in a Monte Carlo study. The approach is implemented in an empirical application to house price data in the Boston area, where it is found that spatial effects in house price determination are much more significant under robustification to heterogeneity in the equation errors.

1. INTRODUCTION

In recent years, spatial models have stimulated growing interest and application in various areas in economics. Economic data frequently exhibit strong spatial

We thank the Co-Editor, Dennis Kristensen, and two anonymous referees for helpful comments and suggestions that have improved the quality of the paper. Phillips acknowledges research support from the NSF under Grant No. SES 18-50860 and a Kelly Fellowship at the University of Auckland. Rossi acknowledges research support from MIUR under the Rita Levi Montalcini scheme. Address correspondence to Maria Kyriacou, Department of Economics, University of Southampton, Southampton, UK; e-mail: m.kyriacou@soton.ac.uk.

patterns that need to be accounted for in applied research. Common examples include real estate pricing data, R&D spillover effects, crime rates, unemployment rates, regional economic growth patterns, and environmental characteristics in urban, suburban, and rural areas. Econometric modeling of such phenomena now makes extensive use of formulations that accommodate spatial dependence through autoregressive specifications known as spatial autoregressions (SARs). SAR models, like vector autoregressions, have the great advantage of simplicity and ready implementation. These models have been found to flexibly describe many different networks of spatial interactions by appropriate ex ante specification of weighting matrices that embody dependencies considered to be of primary relevance in specific empirical applications. Weight matrices may incorporate notions of "economic distance" that include geographic and economic proximity as well as many other socioeconomic characteristics.

For SAR models with homoskedastic innovations, a wide range of estimation procedures are available, ranging from maximum likelihood/quasi maximum likelihood (ML/QML) methods (e.g., Lee, 2004) to two-stage least squares (2SLS; e.g., Kelejian and Prucha, 1998) and generalized method of moments (GMM; e.g., Kelejian and Prucha, 1999). However, data recorded across space are frequently heterogeneously distributed, due to such elements as aggregation of "rate variables," social interactions, preferences, as well as variation in demographic characteristics like income or size across different regions. Examples in the recent empirical literature stress the importance of capturing the inherent heterogeneity in spatial units in modeling and estimation. Inter alia, we cite intermarriage decisions across U.S. states (Bisin, Topa, and Verdier, 2004), house selling prices (Harrison and Rubinfeld, 1978; LeSage, 1999), and crime rates and social interactions across contiguous U.S. states (Glaeser, Sacerdote, and Scheinkman, 1996), where these effects are important. Thus, spatial units typically vary over many observed and unobserved characteristics, leading to potentially heterogeneous innovations in regressions that may introduce bias and invalidate the aforementioned commonly used procedures.

More specifically, although ML/QML methods provide an obvious general approach to parameter estimation (Lee, 2004), in the presence of unobserved heterogeneity, they produce inconsistent estimates (e.g., Lin and Lee, 2010). This lack of robustness to heteroskedasticity is possibly the main shortcoming of ML/QML methods for spatial data. On the other hand, methods based on 2SLS/GMM enjoy robustness toward unknown heteroskedasticity in the disturbances. The simple 2SLS estimator introduced by Kelejian and Prucha (1998) and developed further by Lee (2003, 2007) would produce consistent estimates in case of heteroskedastic disturbances, even though their usual asymptotic distribution would not be correct. Along these lines, Kelejian and Prucha (2007) suggest a heteroskedasticity and autocorrelation consistent (HAC) procedure to consistently estimate the variance matrix of 2SLS-type estimators that is robust to departures from spherical disturbances. The main drawbacks of 2SLS estimators are lack of optimality and the fact that they fail if there are no exogenous regressors

in the model (as well as in cases where the weight matrix is row normalized and the only exogenous regressor in the model is an intercept). More recently, Lin and Lee (2010) propose a robust generalized method of moments estimator which delivers consistent estimation of the parameters of SAR models with heteroskedastic errors, while Kelejian and Prucha (2010) consider a robust GMM with a particular focus on the SARAR(1,1) model structure[1] with heteroskedastic disturbances. The approach developed in Kelejian and Prucha (2010) has been generalized to accommodate higher-order spatial lags in Badinger and Egger (2011). An interesting analysis of the finite-sample performance of 2SLS/GMM estimators in the presence of heteroskedasticity and possibly a spatial lag in the disturbances is reported in Arraiz et al. (2010). More recently, Liu and Yang (2015) propose a modified QLE/MLE estimator modified Quasi Maximum Likelihood (MQML) that restores consistency by adjusting the score function for the spatial parameter to accommodate general forms of heteroskedasticity. Furthermore, Jin and Lee (2019) develop a generalized empirical likelihood method to estimate SARAR models that is robust to heteroskedastic errors.

A separate remark is needed for simple ordinary least squares (OLS) estimation of the parameters of a SAR model with exogenous regressors, which is known to be consistent under certain restrictive assumptions on the limit behavior of the spatial design, as discussed in Lee (2002). OLS may enjoy some robustness to unknown heteroskedasticity in the disturbances, but again, this is only achieved under highly restrictive weight matrix specifications, which may not be pertinent to empirical situations of interest. As a practical example, OLS would not be consistent when the network structure is defined according to a contiguity criterion where the number of neighbors of a given spatial unit remains fixed as the sample size grows, even in the simpler setting of homoskedastic disturbances. In general, the necessary restrictions on the limit behavior of the weight structures that ensure consistency are difficult to verify in practical situations, making OLS estimation a questionable choice for practitioners.

The present paper develops a new method of robust estimation for the SAR model with unknown heteroskedasticity that is based on a continuously updated version of the indirect inference (II) estimator of Kyriacou, Phillips, and Rossi (2017) (KPR henceforth). The II estimator in KPR was designed to modify (inconsistent) OLS estimation of a pure SAR model (that is, SAR without exogenous regressors) with homoskedastic innovations, leading to consistent, asymptotically normal estimates that enjoy good finite-sample properties. In that case, the II procedure converts an inconsistent OLS estimator into a consistent one.

We propose a similar enhancement in the case of a SAR model with exogenous regressors (SARX, in the sequel) and heterogeneous spatial errors. The key idea in this approach is to accommodate more realistic error structures by parameterizing the variance–covariance structure in terms of unknown parameters of interest within the II mechanism via a suitable, feasible, binding function. The

[1] SARAR denotes "spatial autoregression with spatial autoregressive disturbances."

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

idea relates to the "continuous-updating" GMM estimator considered in Hansen, Heaton, and Yaron (1996),[2] where the covariance matrix is continuously altered as the parameter vector in question is updated sequentially in the minimization routine. The proposed continuously updated indirect inference (CUII) estimator is computationally straightforward and can flexibly allow for various forms of unknown heteroskedasticity and realistic spatial weight schemes that are relevant to empirical work. We then show that our new CUII estimator is consistent and asymptotically normal, while simulation and empirical results confirm its satisfactory finite-sample properties under general spatial designs and heteroskedasticity structures. In particular, the simulations show that the new CUII estimator offers a substantial improvement over standard 2SLS and robust GMM as it does not rely on the construction of optimal instruments and, even more importantly, on the joint relevance of such instruments. In independent recent work that appeared after our paper was completed, Bao, Liu, and Yang (2020) extended results in KPR to SARX models with heteroskedastic error terms, and suggested an II-type transformation that produces results comparable to our findings reported in Sections 3 and 4. Bao et al. (2020) adopted a different parameterization of the (nonconstant) variance–covariance matrix of the disturbances, resulting in a different asymptotic variance for their estimator.

The rest of the paper is organized as follows. Section 2 introduces the SARX model and its underlying assumptions. The bias of the QMLE under heteroskedasticity is explored in a working example by using the bias expansion of Bao (2013). We show that the QMLE can be severely biased when the spatial weights deviate from a Toeplitz network structure, such as block diagonal or circulant. The CUII procedure based on OLS is introduced in Section 3, and its limit behavior is explored in Section 4. Section 5 presents some special cases of the general SAR model, while in Section 6, we construct a Moran I test for lack of spatially correlated disturbances. Section 7 reports simulations comparing the finite-sample performance of the CUII estimator to existing methods. An empirical comparison of estimation methods for inference on the spatial parameter in the context of house price data in the Boston area is given in Section 8. Conclusions are in Section 9, proofs are in the Appendix, and the Online Supplementary Material provides additional technical material and simulation results.

In the sequel, λ_0, β_0, and σ_0^2 denote true values of these parameters, while λ, β, and σ^2 denote admissible values. We use A_{ij} and A_i to signify the (i,j)th element and the transpose of the ith row of the generic matrix A. We use $||.||$ and $||.||_\infty$ to denote the spectral norm and uniform absolute row sum norm, respectively, and $K > 0$ represents an arbitrary finite, positive constant. For any function $v(x)$, we define $v^{(r)}(x) : dv^r(x)/dx^r$, and $a_n \sim n$ for any sequence a_n indicates $a_n/n \to K$ as $n \to \infty$. Furthermore, $\eta_i(A)$, $i = 1, \ldots, q$, denote the eigenvalues of a generic real $q \times q$ matrix A, while $\bar{\eta}(A) = \max_{i=1,\ldots q} \{|\eta_i(A)|\}$ and $\underline{\eta}(A) = \min_{i=1,\ldots,q} \{|\eta_i(A)|\}$.

[2]See Durbin (1988) for an early version of the idea of continuous updating in the context of efficiently estimating structural equation models by iterative instrumental variable methods.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

2. THE SARX MODEL WITH UNKNOWN HETEROSKEDASTICITY

Our focus is the linear SARX model

$$y_n = \lambda_0 W_n y_n + X_n \beta_0 + \epsilon_n, \tag{2.1}$$

where n denotes sample size, y_n is an n-vector of observations, X_n is an $n \times k$ matrix of observations of exogenous regressors, and ϵ is a vector of disturbances. We denote by W_n the given $n \times n$ matrix of spatial weights, while λ_0 and β_0 are, respectively, the unknown scalar spatial autoregressive coefficient and a k-vector of coefficients of the exogenous variables. The pure SAR model (with no exogenous regressors) is a special case of (2.1) with $\beta_0 = 0$. In what follows, we assume the presence of exogenous regressors and rule out the possibility of $\beta_0 = 0$. Henceforth, we drop the subscript n even though quantities generally denote triangular arrays, i.e., $y = y_n$, $X = X_n$, $W = W_n$, and $\epsilon = \epsilon_n$.[3]

Under standard stability conditions, the model in (2.1) can be rewritten in reduced form as

$$y = S^{-1}(\lambda_0) X \beta_0 + S^{-1}(\lambda_0)\epsilon, \tag{2.2}$$

where $S = S(\lambda_0) = I_n - \lambda_0 W$. Allowing for unanticipated heteroskedasticity in (2.1), we impose the following condition.

Assumption 1. For all n and for $i = 1, \ldots, n$, the $\{\epsilon_i\}$ are a set of independent random variables, with mean 0 and unknown variances $\sigma_i^2 > 0$. In addition, for some $\delta > 0$,

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

Let $\mathbb{E}(\epsilon\epsilon') = \Omega_0 > 0$. As is common practice in the spatial literature, restrictions on the parameter space and the asymptotic behavior of W are imposed to ensure existence of the reduced form SARX in (2.2) and to establish the limit theory. We therefore impose the following additional conditions.

Assumption 2. $\lambda_0 \in \Lambda$, where Λ is a closed subset in $(-1, 1)$.

Assumption 3. (i) For all n, $W_{ii} = 0$, for $i = 1, \ldots, n$.
(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 bounded away from zero for all n and $h/n \to 0$ as $n \to \infty$.

Assumptions 2 and 3(ii) guarantee that $S^{-1}(\lambda)$ exists for all $\lambda \in \Lambda$ and is nonsingular. It is well documented (e.g., KPR; Kelejian and Prucha, 2010) that the restriction on the parameter space given in Assumption 2 and a condition on the spectral norm, such as the one given in Assumption 3(ii), are not strictly necessary

[3]The subscript n is retained only when we want to particularly stress the importance of some sequential dependence on sample size.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

to develop asymptotic theory. However, these conditions (or similar ones) are required to ensure existence of the reduced form in (2.2), measurability of the dependent variable, and existence of the likelihood function.

Assumption 4. For all sufficiently large n, $\sup_{\lambda\in\Lambda}||S^{-1}(\lambda)||_\infty + ||S^{-1}(\lambda)'||_\infty < K$.

We also impose conditions on existence of limits and no collinearity for large n. Let $M_X = I - X(X'X)^{-1}X'$, and set $G = G(\lambda_0) = WS^{-1}(\lambda_0)$.

Assumption 5. All elements of the $n \times k$ matrix X are uniformly bounded for all n, and, for all sufficiently large n,

$$\underline{\eta}\left(\frac{(X, G'X\beta_0)'(X, G'X\beta_0)}{n}\right) > 0. \tag{2.3}$$

The last condition rules out cases where the columns of G and X are perfectly collinear, for all sufficiently large n.

Standard ML/QML-based estimation methods generally lead to inconsistent estimates unless the ϵ_i's are homoskedastic (e.g., Lin and Lee, 2010). To illustrate, define the concentrated pseudo-log-likelihood function in this case

$$l(\lambda) = K - \frac{1}{2}ln(y'S(\lambda)'M_XS(\lambda)y) + \frac{1}{n}ln|S(\lambda)| \tag{2.4}$$

and let

$$\hat{\lambda}_{QML} = \underset{\lambda\in\Lambda}{\text{argmax}}\ l(\lambda). \tag{2.5}$$

Write $l^{(i)}(\lambda_0) = \frac{\partial^i l(\lambda)}{\partial\lambda^i}|_{\lambda_0}$, for $i > 0$. A necessary condition for consistency of $\hat{\lambda}_{QML}$ is

$$\underset{n\to\infty}{\text{plim}}\frac{1}{n}l^{(1)}(\lambda_0) = 0. \tag{2.6}$$

This condition is satisfied under standard SARX assumptions when disturbances are homoskedastic (Lee, 2004), but it is generally violated under Assumption 1. Specifically, Lin and Lee (2010) show that a sufficient condition for (2.6), and hence for consistency of $\hat{\lambda}_{QML}$, is

$$\frac{1}{n}\sum_{i=1}^{n}\left(G_{ii} - \frac{1}{n}trG\right)\left(\sigma_i^2 - \frac{1}{n}\sum_{j=1}^{n}\sigma_j^2\right) \to 0, \tag{2.7}$$

as $n \to \infty$, where G_{ii} is the ith diagonal element of G. The condition in (2.7) is trivially satisfied for any form of heteroskedasticity when almost all the elements of G are equal. However, unless the weight matrix is restricted to have a circulant or block diagonal structure (such as in Case, 1991) or some other very specific structure which ensures that G_{ii}, for $i = 1, \ldots, n$, are equal, general results about consistency of $\hat{\lambda}_{QML}$ cannot be obtained when σ_i^2 is not constant across i.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

For further illustration, consider the simple and typical (e.g., Harvey, 1976) form of multiplicative heteroskedasticity given by

$$\Omega_0(\gamma) = \sigma^2 \begin{pmatrix} e^{z_1\gamma} & 0 & 0 & \dots & 0 \\ 0 & e^{z_2\gamma} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & e^{z_n\gamma} \end{pmatrix} \tag{2.8}$$

for unknown scalar parameters γ and σ^2 and an n-vector of observables $z = (z_1, \dots, z_n)'$. Set $\sigma^2 = 1$ without loss of generality. For $\Omega_0(\gamma)$ defined as in (2.8), the LHS of (2.7) becomes

$$\sum_{t=0}^{\infty} \frac{\gamma^t}{t!} \frac{1}{n} \sum_{i=1}^{n} \left(G_{ii} - \frac{1}{n} trG \right) \left(z_i^t - \frac{1}{n} \sum_{j=1}^{n} z_j^t \right). \tag{2.9}$$

The latter expression confirms that, even in the presence of a very simple form of heteroskedasticity such as that in (2.8), the condition displayed in (2.7) is difficult to check for general W and $z_1, \dots, z_n$. Of course, under the extreme condition that the sample covariance between the diagonal elements of G and z^t is zero for each t for $n \to \infty$, then condition (2.7) holds. But if instead the sample covariance is constant and nonzero across t (at least for sufficiently large n), the LHS of (2.7) becomes $K(e^{\gamma} - 1)$, which vanishes only when $\gamma \to 0$. Simple calculations confirm that (2.9) is nonzero for other cases. For instance, if $\{G_{ii}, z_i\}$ are stationary and ergodic over i with mean $\{\mu_G, \mu_z\}$ and if z_i has finite moment generating function, then

$$\sum_{t=0}^{\infty} \frac{\gamma^t}{t!} \frac{1}{n} \sum_{i=1}^{n} \left(G_{ii} - \frac{1}{n} trG \right) \left(z_i^t - \frac{1}{n} \sum_{j=1}^{n} z_j^t \right) \to_{a.s.} \mathbb{E}\left((G_{ii} - \mu_G) e^{\gamma z_i} \right), \tag{2.10}$$

which is nonzero whenever the covariance $\mathbb{E}\left((G_{ii} - \mu_G) e^{\gamma z_i} \right)$ is nonzero.

We next examine the bias of $\hat{\lambda}_{QML}$ under error heterogeneity as in Assumption 1. Starting from the results in Bao (2013), we may compute the bias of $\hat{\lambda}_{QML}$ given the error variance matrix $\Omega_0(\gamma)$ in (2.8). For illustration, we limit our analysis to the Gaussian case, although more general results can be obtained at the expense of extra computation. Explicit calculations of the bias terms are reported in Section S.1 of the Online Supplementary Material, and we only report here the plots of the bias functions as γ varies, i.e., of $B(\gamma, \lambda_0)$ for different values of λ_0 and for four different choices[4] of W against $\gamma \in [-10, 10]$ at $n = 200$.

The elements of the vector $(z_1, \dots, z_n)'$ are generated once from a uniform distribution with support $[0, 4]$ and kept fixed across γ as well as across different scenarios. For each choice of W, the spatial parameter ranges from $\lambda_0 = -0.8, 0, 0.4, 0.8$.

[4]Figure S1 of the Online Supplementary Material depicts the structure of the four choices of weight matrices used in this paper, to illustrate the degree of sparseness and/or symmetry in each design.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

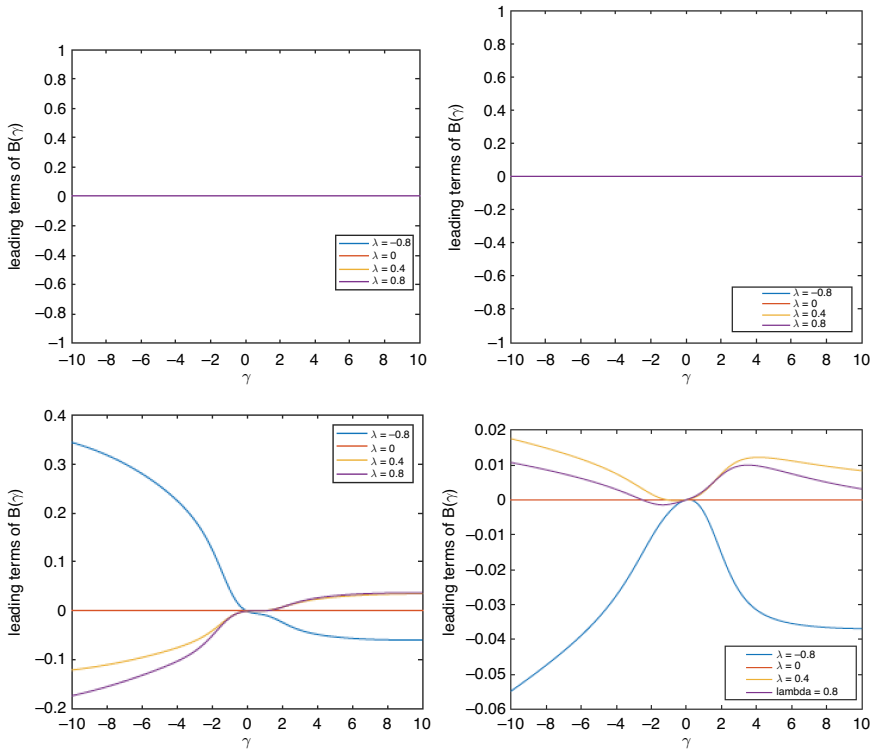

FIGURE 1. $B(\gamma)$ for various weight matrix designs at $n = 200$. Top: (L) block diagonal, (R) circulant, two ahead-two behind; Bottom: (L) "exponential," (R) "random."

The plot depicted on the top left of Figure 1 reports $B(\gamma, \lambda_0)$ when W is chosen as a block diagonal matrix (Case, 1991). Specifically, this first choice of W is defined as

$$W_n = I_r \otimes B_m, \quad B_m = \frac{1}{m-1}(l_m l_m' - I_m), \tag{2.11}$$

where I_s is the $s \times s$ identity matrix, l_m is an m-vector of 1's, and $\otimes$ is the Kronecker product. It is easy to verify that the G_{ii}, for $i = 1, \ldots, n$, are constant across i for W in (2.11). Similarly, the plot in the top right of Figure 1 reports $B(\gamma, \lambda_0)$ when W is chosen as a circulant with two neighbors behind and two ahead. As expected, for both these choices of W, the bias function is zero for all values of λ_0 as γ varies. The plot depicted in the bottom left of Figure 1 is the bias function when W is randomly generated as a symmetric $n \times n$ matrix of zeros and ones, where the number of "ones" is restricted at 20% of the total number of elements in W. This choice of W is generated once for any given n and kept fixed across different γ and λ. Similarly, the plot in the bottom right of Figure 1 displays $B(\gamma, \lambda_0)$ for W based

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

on an exponential distance decay, with $w_{ij} = exp(-|\ell_i - \ell_j|)\mathbb{1}(|\ell_i - \ell_j| < \log n)$, where ℓ_i is the ith location along the interval $[0, n]$, which is randomly generated from a uniform distribution with support $[0, n]$. Again, we generate one W for each sample size and we keep it fixed across scenarios. In the sequel, we refer to these matrices as "random" and "exponential." Both these cases are then rescaled by their respective spectral norm. These examples of W tend to be more relevant to empirical work than other choices such as (2.11) or circulant matrices, as they mimic contiguity-based weight matrices. Under Assumption 1 and with either "random" or "exponential" cases of W, the ML/QML is not expected to return consistent estimators for a general heteroskedastic design as G_{ii}, for $i = 1, \ldots, n$, vary across i.

The plots in the bottom panel of Figure 1 confirm that the finite-sample bias persists even for a moderately sized sample of $n = 200$ and its magnitude varies with λ_0 (e.g., the bias is, in general, larger in absolute value for a large negative λ_0). Furthermore, the bias tends to be generally more severe for "exponential" W, as shown in the plot in the bottom right of Figure 1. As expected, the leading terms of $B(\gamma, \lambda_0)$ (reported explicitly in the Online Supplementary Material) vanish for $\lambda_0 = 0$ and for $\gamma = 0$ (although, even if they are not displayed in Figure 1, terms that vanish as $n \to \infty$ may persist in finite samples and contribute to the overall bias of $\hat{\lambda}_{QML}$).

3. CONTINUOUSLY UPDATED INDIRECT INFERENCE BASED ON OLS ESTIMATES

As discussed above, in the presence of unknown heteroskedasticity, the standard ML/QML methods are, in general, biased and inconsistent. On the other hand, the OLS estimators of the unknown parameters in (2.1) can be consistent even under Assumption 1, as long as some stringent conditions on the asymptotic behavior of W are satisfied. Specifically, as shown in Lee (2002), the OLS estimator of λ_0 is consistent but does not have a standard normal limiting distribution if h defined in Assumption 3(iv) satisfies

$$\frac{1}{h} + \frac{h}{\sqrt{n}} \to 0 \quad \text{as} \quad n \to \infty, \tag{3.1}$$

and it is consistent and asymptotically normal if

$$\frac{1}{h} + \frac{\sqrt{n}}{h} \to 0 \quad \text{as} \quad n \to \infty. \tag{3.2}$$

For instance, OLS estimation of (2.1) will lead to inconsistent estimates in situations whereby the spatial weights are generated via a contiguity criterion (e.g., country borders) and the number of neighbors of a given unit (country in this case) needs to remain constant as the sample size increases, regardless of whether homoskedasticity in the disturbances holds or not.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

The limit conditions (3.1) and (3.2) are hardly verifiable in practical situations as only a finite set of observations is available in most circumstances and we are typically agnostic about the limit behavior of h. Hence, OLS estimation is commonly and justifiably ignored in practice. But OLS can be used as a building block for a modified estimator with improved finite-sample and asymptotic properties as we now show. OLS has the advantage of computational simplicity, but the methodology we describe can, in principle, be extended to QML or other implicitly defined estimators, at the expense of some additional computational and algebraic costs.

Using (2.1), we can "concentrate β out," leading to

$$\hat{\beta}(\lambda) = (X'X)^{-1}X'S(\lambda)y, \tag{3.3}$$

and then focus on estimation of λ_0. The OLS estimator of λ in (2.1), denoted by $\hat{\lambda}$, is defined as

$$\hat{\lambda} = \frac{y'W'M_Xy}{y'W'M_XWy}. \tag{3.4}$$

Similar to the discussion in KPR, we can obtain a formal expansion for the expected value of the latter ratio based on Lieberman's (1994) result as

$$\mathbb{E}(\hat{\lambda}) = \frac{\mathbb{E}(y'W'M_Xy)}{\mathbb{E}(y'W'M_XWy)} + O\left(\frac{1}{n}\right). \tag{3.5}$$

Let $Q(\lambda) = M_XG(\lambda)$, $P(\lambda) = Q(\lambda)'S^{-1}(\lambda)$, $Q = Q(\lambda_0)$, and $P = P(\lambda_0)$. By standard algebra,

$$\mathbb{E}(\hat{\lambda}) = \frac{tr(P\Omega_0) + \beta_0'X'PX\beta_0}{tr(Q'Q\Omega_0) + \beta_0'X'Q'QX\beta_0} + O\left(\frac{1}{n}\right). \tag{3.6}$$

Following the formal expansion in (3.6), we define the binding function $\tau_n(\lambda)$ as

$$\begin{aligned}\tau_n(\lambda, \Omega_\lambda, \hat{\beta}(\lambda)) &= \tau_n(\lambda) \\ &= \frac{tr(P(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)'X'P(\lambda)X\hat{\beta}(\lambda)}{tr(Q(\lambda)'Q(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)'X'Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda)} + O_p\left(\frac{1}{n}\right),\end{aligned} \tag{3.7}$$

and its approximate counterpart (which will be used for practical implementation) as

$$\tau_n^*(\lambda, \Omega_\lambda, \hat{\beta}(\lambda)) = \tau_n^*(\lambda) = \frac{tr(P(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)'X'P(\lambda)X\hat{\beta}(\lambda)}{tr(Q(\lambda)'Q(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)'X'Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda)}, \tag{3.8}$$

with $\hat{\beta}(\lambda)$ defined according to (3.3), and

$$\Omega_\lambda = \text{diag}(\epsilon(\lambda)\epsilon(\lambda)'), \quad \epsilon(\lambda) = (y - \lambda Wy - X\hat{\beta}(\lambda)), \tag{3.9}$$

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

where diag(A) for a generic $n \times n$ matrix A returns the $n \times n$ diagonal matrix containing only the main diagonal of A and with other entries zero.[5]

The methodology we now propose is a continuously updated version of II in which estimates of the error variances that affect bias and the binding function are continuously updated within the II procedure. The resulting estimator is defined as the extremum estimator that satisfies

$$\hat{\lambda}_{CUII} = \underset{\lambda \in \Lambda}{\operatorname{argmin}} \{\hat{\lambda} - \tau_n^*(\lambda, \Omega_\lambda, \hat{\beta}(\lambda))\}^2, \tag{3.10}$$

which we call the CUII estimator, $\hat{\lambda}_{CUII}$, of λ. A detailed discussion on the robustness advantages of using $\tau_n^*(.)$ rather than its standard simulated version (e.g., Gouriéroux, Renault, and Touzi, 1995) is contained in KPR. In this setting, the II approach of simulating pseudodata to construct the binding function would be more restrictive as it would require considerably more structure compared to standard estimation problems under homoskedasticity. In particular, the specific form of heteroskedasticity, in addition to distributional assumptions, would be needed to employ the standard simulation approach.

From substantial numerical work, the objective function in (3.10) is found to be continuous and strictly convex for all values of the parameter space, so that the optimization problem appears to be standard. Since smoothness and monotonicity conditions on $\tau_n(\lambda)$ are required to establish the limit theory, we introduce the following condition.

Assumption 6. (i) For all n, $\tau_n(\lambda)$ is continuously differentiable and strictly increasing for all $\lambda \in \Lambda$ with probability one.

(ii) $\operatorname{plim}_{n\to\infty} \tau_n^{(1)}(\lambda_0)$ exists and it is positive.

As discussed in KPR, the latter is employed as a high-level condition, because the derivation of more primitive assumptions involving general choices of W is not feasible. KPR verified a condition similar to Assumption 6 for a class of W with Toeplitz structures (e.g., circulant and block diagonal structures). However, as is common practice in the simulation-based techniques literature, when W has a more general unspecified structure, practitioners have to rely on numerical methods to verify conditions such as Assumption 6. We stress that $\tau_n(\cdot)$ does not depend on nuisance parameters, and so verification can be achieved numerically in finite samples for any given sample and choice of W in a straightforward manner. In Sections 7 and 8, we report plots of $\tau_n^*(\cdot)$ for $\lambda \in \Lambda$ for the choices of W adopted in the simulations and for a few pseudodatasets $(y_i, x_i')'$, $i = 1, \ldots, n$, and for our empirical setting, respectively.

Under Assumption 6, we have the inverse function representation of the CUII estimator

$$\hat{\lambda}_{CUII} = \tau_n^{*-1}(\hat{\lambda}). \tag{3.11}$$

[5] We outline that our definition of the binding function is different from Bao et al. (2020), as both numerator and denominator are parameterized as a function of λ.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

4. LIMIT THEORY

This section derives the asymptotic properties of the estimator (3.10) for model (2.1) when the case $\beta_0 = 0$ is ruled out a priori. From (3.7) and (3.8), we consider the centering random sequence

$$\tau_n(\lambda_0) = \frac{tr(P\Omega_{\lambda_0})/n + \hat{\beta}(\lambda_0)'X'PX\hat{\beta}(\lambda_0)/n}{tr(Q'Q\Omega_{\lambda_0})/n + \hat{\beta}(\lambda_0)'X'Q'QX\hat{\beta}(\lambda_0)/n} + O\left(\frac{1}{n}\right), \tag{4.1}$$

where

$$\hat{\beta}(\lambda_0) = \beta_0 + (X'X)^{-1}X'\epsilon, \tag{4.2}$$

$$\epsilon(\lambda_0) = M_X\epsilon, \tag{4.3}$$

so that $\Omega_{\lambda_0} = diag(M_X\epsilon\epsilon' M_X)$. Define

$$V_n = \frac{4}{n}\begin{pmatrix} \beta_0'X'PM_X\Omega_0M_XP'X\beta_0 & \beta_0'X'PM_X\Omega_0M_XQ'QX\beta_0 \\ \beta_0'X'Q'QM_X\Omega_0M_XP'X\beta_0 & \beta_0'X'Q'QM_X\Omega_0M_XQ'QX\beta_0 \end{pmatrix} + \frac{4}{n}\sum_i\sum_{j<i}\sigma_i^2\sigma_j^2\begin{pmatrix} \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix}. \tag{4.4}$$

In order to assure the existence of limits of each of the suitably standardized components that appear in (4.4), we impose the following conditions.

Assumption 7. As $n \to \infty$, the following limits exist:

$$\begin{aligned} &\bar{a} = \bar{a}(\lambda_0) = \lim_{n\to\infty}\frac{1}{n}tr(P\Omega_0), \quad \bar{b} = \bar{b}(\lambda_0) = \lim_{n\to\infty}\frac{1}{n}\beta_0'X'PX\beta_0, \\ &\bar{c} = \bar{c}(\lambda_0) = \lim_{n\to\infty}\frac{1}{n}tr(Q'Q\Omega_0), \quad \bar{d} = \bar{d}(\lambda_0) = \lim_{n\to\infty}\frac{1}{n}\beta_0'X'Q'QX\beta_0, \end{aligned} \tag{4.5}$$

$$\begin{aligned} &\bar{a}^{(1)} = \bar{a}^{(1)}(\lambda_0) = \lim_{n\to\infty}\frac{1}{n}\left(tr(G'P\Omega_0 + PG\Omega_0) - 2tr\left(Pdiag(G\Omega_0)\right)\right), \\ &\bar{b}^{(1)} = \bar{b}^{(1)}(\lambda_0) = \lim_{n\to\infty}\frac{1}{n}\left(\beta_0'X'(G'P + PG)X\beta_0 - 2\beta_0'X'P(I - M_X)GX\beta_0\right), \\ &\bar{c}^{(1)} = \bar{c}^{(1)}(\lambda_0) = \lim_{n\to\infty}\frac{2}{n}\left(tr(G'Q'Q\Omega_0) - tr\left(Q'Qdiag(G\Omega_0)\right)\right), \\ &\bar{d}^{(1)} = \bar{d}^{(1)}(\lambda_0) = \lim_{n\to\infty}\frac{2}{n}(\beta_0'X'G'Q'QX\beta_0 - \beta_0'X'Q'Q(I - M_X)G'X\beta_0), \quad \lim_{n\to\infty}V_n. \end{aligned} \tag{4.6}$$

Positive definiteness of V_n in (4.4) in the limit is ensured under Assumption 5, since both terms in (4.4) are positive semidefinite for each n and Assumption 5 establishes strict positive definiteness in the limit for the dominant first term.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

Standard calculations give

$$\bar{\tau}^{(1)} = \bar{\tau}^{(1)}(\lambda_0) = \operatorname*{plim}_{n\to\infty} \tau_n^{(1)}(\lambda_0) = \frac{\bar{a}^{(1)} + \bar{b}^{(1)}}{\bar{c} + \bar{d}} - \frac{(\bar{c}^{(1)} + \bar{d}^{(1)})(\bar{a} + \bar{b})}{(\bar{c} + \bar{d})^2}, \tag{4.7}$$

whose existence and positivity are assured by Assumptions 6 and 7. By virtue of the delta method,

$$\sqrt{n}(\hat{\lambda} - \tau_n(\lambda_0)) = f_n' U_n + o_p(1), \tag{4.8}$$

where

$$U_n = \frac{1}{\sqrt{n}} \begin{pmatrix} \epsilon' P \epsilon - tr(P\Omega_{\lambda_0}) + 2\beta_0' X' P M_X \epsilon \\ \epsilon' Q' Q \epsilon - tr(Q'Q\Omega_{\lambda_0}) + 2\beta_0' X' Q' Q M_X \epsilon \end{pmatrix}, \tag{4.9}$$

and

$$f_n' = \left(\left(\frac{1}{n} y' W' M_X W y \right)^{-1}, \quad \left(\frac{1}{n} y' W' M_X W y \right)^{-2} \left(\frac{1}{n} y' W' M_X y \right) \right). \tag{4.10}$$

We derive

$$\bar{f} = \operatorname*{plim}_{n\to\infty} f_n = \left((\bar{c} + \bar{d})^{-1}, \quad (\bar{c} + \bar{d})^{-2} (\bar{a} + \bar{b}) \right)', \tag{4.11}$$

which is defined in terms of limits that appear in $\bar{\tau}^{(1)}$. In particular, existence and positivity of $\bar{\tau}^{(1)}$ in (4.7) ensure that the common factor $1/(\bar{c} + \bar{d}) \neq 0$. Thus, existence and nonnullity of $\bar{f}$ are assured under Assumptions 6(ii) and 7.

With these results in hand, we obtain the following limit theory.

THEOREM 1. (a) *Under (2.1), with $\beta_0 \neq 0$, and Assumptions 1–5, 6(i), and 7,*

$$\sqrt{n}(\hat{\lambda} - \tau_n(\lambda_0)) \underset{d}{\to} \mathcal{N}(0, \bar{f}' \lim_{n\to\infty} V_n \bar{f}). \tag{4.12}$$

(b) *Under (2.1), with $\beta_0 \neq 0$, and Assumptions 1–7,*

$$\sqrt{n}(\hat{\lambda}_{CUII} - \lambda_0) \underset{d}{\to} \mathcal{N}\left(0, v_{CUII}^2\right), \tag{4.13}$$

where $v_{CUII}^2 = \bar{f}' \lim_{n\to\infty} V_n \bar{f} / \left(\bar{\tau}^{(1)}\right)^2$ exists and is nonzero under Assumptions 2, 3(ii), and 5–7.

Let $\hat{v}_{CUII}^2$ be the estimated version of v_{CUII}^2, obtained by replacing the unknown λ_0 and β_0 by $\hat{\lambda}_{CUII}$ and $\hat{\beta}_{CUII}$, respectively, Ω_0 by $\hat{\Omega} = diag(\hat{\epsilon}\hat{\epsilon}')$, where $\hat{\epsilon} = M_X S(\hat{\lambda}_{CUII}) y$, and σ_i^2 by $\hat{\epsilon}_i^2$, for $i = 1, \ldots, n$.

THEOREM 2. *Let Assumption 1 hold, with $\delta = 2$. Under Assumptions 2–7, as $n \to \infty$,*

$$\hat{v}_{CUII}^2 - v_{CUII}^2 \underset{p}{\to} 0. \tag{4.14}$$

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

Estimation of β_0 in (2.1) under Assumption 1 and exogeneity of X causes less difficulty than estimation of λ_0 as simple OLS regression produces consistent estimates under general limit behavior of W. Nonetheless, from (3.3), we can deduce consistency of $\hat{\beta}_{CUII}$ and its asymptotic normality by using the following representation of the scaled estimation error:

$$\sqrt{n}(\hat{\beta}_{CUII}-\beta_0)=\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{\sqrt{n}}X'\epsilon - \left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'GX\beta_0\sqrt{n}(\hat{\lambda}_{CUII}-\lambda_0)+o_p(1), \tag{4.15}$$

where

$$\sqrt{n}(\hat{\lambda}_{CUII}-\lambda_0)=\frac{1}{\tau_n^{(1)}(\lambda_0)}\sqrt{n}(\hat{\lambda}-\tau_n(\lambda_0))+o_p(1). \tag{4.16}$$

From (4.8)–(4.10),

$$\sqrt{n}(\hat{\beta}_{CUII}-\beta_0)=\zeta_n'R_n, \tag{4.17}$$

where

$$R_n=\frac{1}{\sqrt{n}}\begin{pmatrix} X'\epsilon \\ \epsilon'P\epsilon-tr(P\Omega_{\lambda_0})+2\beta_0'X'PM_X\epsilon \\ \epsilon'Q'Q\epsilon-tr(Q'Q\Omega_{\lambda_0})+2\beta_0'X'Q'QM_X\epsilon \end{pmatrix} \tag{4.18}$$

and

$$\zeta_n=\left(\left(\frac{1}{n}X'X\right)^{-1},\ -\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'GX\beta_0\tau_n^{(1)}(\lambda_0)^{-1}f_n'\right)'. \tag{4.19}$$

Defining

$$\bar{\zeta}=\operatorname*{plim}_{n\to\infty}\zeta_n, \tag{4.20}$$

and

$$T_n=\frac{1}{n}\begin{pmatrix} X'X & X'\Omega_0M_X(P+P')X\beta_0 & 2X'\Omega_0M_XQ'QX\beta_0 \\ \beta_0'X'(P+P')M_X\Omega_0X & 4\beta_0'X'PM_X\Omega_0M_XQ'QX\beta_0 & 4\beta_0'X'PM_X\Omega_0M_XQ'QX\beta_0 \\ 2\beta_0'X'Q'QM_X\Omega_0X & 4\beta_0'X'Q'QM_X\Omega_0M_XP'X\beta_0 & 4\beta_0'X'Q'QM_X\Omega_0M_XQ'QX\beta_0 \end{pmatrix}$$
$$+\frac{4}{n}\sum_i\sum_{j<i}\sigma_i^2\sigma_j^2\begin{pmatrix} 0_{(k\times k)} & 0_{(k\times 1)} & 0_{(k\times 1)} \\ 0_{(1\times k)} & \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ 0_{(1\times k)} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix}, \tag{4.21}$$

we deduce the following result.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

COROLLARY 1. *Under (2.1), with $\beta_0 \neq 0$, and Assumptions 1–7,*

$$\sqrt{n}(\hat{\beta}_{CUII} - \beta_0) \underset{d}{\to} \mathcal{N}(0, \bar{\zeta}' \lim_{n\to\infty} T_n \bar{\zeta}), \tag{4.22}$$

as $n \to \infty$.

The variance matrix $\bar{\zeta}' \lim_{n\to\infty} T_n \bar{\zeta}$ in (4.22) is nonsingular under Assumptions 5–7. The proof of Corollary 1 follows in a similar way to that of part (a) of Theorem 1 and is omitted.

A special remark is needed to clarify the extent of the CUII transformation in case OLS is consistent and/or consistent and asymptotically normal, i.e., under (3.1) and (3.2). The RHS of (3.6) can be equivalently expressed as[6]

$$\mathbb{E}(\hat{\lambda}) = \lambda_0 + \frac{tr(G'M_X\Omega_0)}{\beta_0'X'G'M_XGX\beta_0 + tr(G'M_XG\Omega_0)} + O\left(\frac{1}{n}\right), \tag{4.23}$$

so that, after straightforward manipulation, if $1/h \to 0$,

$$\tau_n(\lambda) = \lambda + \frac{tr(G'(\lambda)M_X\Omega_\lambda)}{\hat{\beta}(\lambda)'X'G'(\lambda)M_XG(\lambda)X\hat{\beta}(\lambda)}\left(1 - \frac{tr(G'(\lambda)M_XG(\lambda)\Omega_\lambda)}{\hat{\beta}(\lambda)'X'G'(\lambda)M_XG(\lambda)X\hat{\beta}(\lambda)} + O_p\left(\frac{1}{h^2}\right)\right) + O_p\left(\frac{1}{n}\right). \tag{4.24}$$

Thus, under (3.1), the CUII transformation removes the finite-sample bias up to order $1/n$ and restores asymptotic normality, while its scope is limited to removing finite-sample bias of order $1/h$ if (3.2) holds. We note that when $h \to \infty$,

$$\begin{aligned}\sqrt{n}(\hat{\lambda} - \tau_n(\lambda_0)) &= \frac{1}{\sqrt{n}} f_n' \begin{pmatrix} \epsilon'G'M_X\epsilon - tr(G'M_X\Omega_{\lambda_0}) + \beta_0'X'G'M_X\epsilon \\ \epsilon'Q'Q\epsilon - tr(Q'Q\Omega_{\lambda_0}) + 2\beta_0'X'Q'QM_X\epsilon \end{pmatrix} \\ &= \frac{1}{\sqrt{n}} f_n' \begin{pmatrix} \beta_0'X'G'M_X\epsilon \\ 2\beta_0'X'Q'QM_X\epsilon \end{pmatrix} + o_p(1), \end{aligned} \tag{4.25}$$

$$f_n = \left(\left(\frac{1}{n}y'W'M_XWy\right)^{-1};\ \left(\frac{1}{n}y'W'M_XWy\right)^{-2}\left(\frac{1}{n}y'W'M_X\epsilon\right)\right)', \tag{4.26}$$

so that

$$\bar{f} = \left(\left(\bar{d}\right)^{-1};\ 0\right)' \tag{4.27}$$

as $y'W'M_X\epsilon/n = o_p(1)$. Therefore, since, from (4.24), $\bar{\tau}(\lambda_0) = 1$ under either (3.1) or (3.2),

$$v_{CUII}^2 = \lim_{n\to\infty} \frac{\beta_0'X'G'M_X\Omega_0M_XGX\beta_0/n}{(\beta_0'X'G'M_XGX\beta_0/n)^2}, \tag{4.28}$$

[6] In the general expression (3.6), λ_0 was not factored out to simplify the algebra in the proof of Theorem 1.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

which is equivalent to the asymptotic variance of $\sqrt{n}(\hat{\lambda}-\lambda_0)$ under (3.2). The latter, in turn, is equivalent to the asymptotic variance of the ML estimator of λ under normality of the errors for $\Omega_0=\sigma^2 I$ (e.g., Lee, 2004).

5. TWO SPECIAL CASES: PURE SAR AND LOCAL-TO-PURE SAR

The framework discussed in Sections 3 and 4 is not directly applicable to the pure SAR model

$$y=\lambda_0 Wy+\epsilon, \tag{5.1}$$

which is the special case of (2.1) with $\beta=0$ ex ante and no exogenous inputs in the fitted regression. Limit theory for estimates of λ in (5.1) cannot generally be deduced as a special case of that developed for estimates of λ in (2.1), as the rate of convergence can be slower than the standard $\sqrt{n}$ (e.g., Lee, 2004). Furthermore, OLS estimation of λ in (5.1) is inconsistent unless $\lambda=0$ (Lee, 2002). KPR developed an II transformation of the OLS estimator of λ in (5.1) under homoskedastic errors, to restore its consistency and asymptotic normality. In this section, we extend that work to accommodate unknown heteroskedasticity in the disturbances along the lines of Sections 3 and 4.

The OLS estimator for λ in (5.1) is defined as

$$\hat{\lambda}=\frac{y'W'y}{y'W'Wy}, \tag{5.2}$$

with expected value

$$\mathbb{E}(\hat{\lambda})=\frac{tr(P\Omega_0)}{tr(G'G\Omega_0)}+O\left(\frac{h}{n}\right), \tag{5.3}$$

where $P(\lambda)=G'(\lambda)S^{-1}(\lambda)$ and $P=P(\lambda_0)$. Similar to Section 3, we define

$$\tau_n(\lambda)=\tau(\lambda)=\frac{tr(P(\lambda)\Omega_\lambda)}{tr(G'(\lambda)G(\lambda)\Omega_\lambda)}+O_p\left(\frac{h}{n}\right) \tag{5.4}$$

and its operational version

$$\tau_n^*(\lambda)=\tau(\lambda)=\frac{tr(P(\lambda)\Omega_\lambda)}{tr(G'(\lambda)G(\lambda)\Omega_\lambda)}, \tag{5.5}$$

with $\Omega(\lambda)=diag(\epsilon(\lambda)\epsilon(\lambda)')$ and $\epsilon(\lambda)=S(\lambda)y$. Under Assumption 6 with $\tau_n(\cdot)$ defined according to (5.4), we define again $\hat{\lambda}_{CUII}=\tau_n^{-1}(\hat{\lambda})$.

In the sequel, we adopt the same notation of Section 4 and define

$$\tau_n(\lambda_0)=\frac{\frac{h}{n}tr(P\Omega_{\lambda_0})}{\frac{h}{n}tr(G'G\Omega_{\lambda_0})}+O_p\left(\frac{h}{n}\right) \quad \text{with} \quad \Omega_{\lambda_0}=diag(\epsilon\epsilon'), \tag{5.6}$$

and

$$\bar{\tau}^{(1)}=\bar{\tau}^{(1)}(\lambda_0)=\operatorname*{plim}_{n\to\infty}\tau_n^{(1)}(\lambda_0)=\frac{\bar{a}^{(1)}}{\bar{c}}-\frac{\bar{c}^{(1)}\bar{a}}{\bar{c}^2}, \tag{5.7}$$

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

with $\bar{a}$, $\bar{c}$, $\bar{a}^{(1)}$, $\bar{c}^{(1)}$ defined in (4.5) and (4.6) with $Q = G$ in the present case. Existence and positivity of the limit in (5.7) are guaranteed by Assumptions 6 and 7 with $Q = G$.

As V_n in (4.4) would be singular in the limit in the case of model (5.1), we redefine the main quantities appearing in Section 4 as

$$\sqrt{\frac{n}{h}}(\hat{\lambda} - \tau_n(\lambda_0)) = f_n' U_n + o_p(1), \tag{5.8}$$

where

$$U_n = \sqrt{\frac{h}{n}} \begin{pmatrix} \epsilon' P \epsilon - tr(P\Omega_{\lambda_0}) \\ \epsilon' G' G \epsilon - tr(G'G\Omega_{\lambda_0}) \end{pmatrix}, \tag{5.9}$$

$$f_n = \left(\left(\frac{h}{n} y' W' W y \right)^{-1}, \quad \left(\frac{h}{n} y' W' W y \right)^{-2} \left(\frac{h}{n} y' W' y \right) \right)', \tag{5.10}$$

$$\bar{f} = \operatorname*{plim}_{n\to\infty} f_n = \left((h\bar{c})^{-1}, \quad (h\bar{c})^{-2} (h\bar{a}) \right)', \tag{5.11}$$

and

$$V_n = \frac{4h}{n} \sum_i \sum_{j<i} \sigma_i^2 \sigma_j^2 \begin{pmatrix} \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(G'G)_{ij}}{2} \\ \frac{(P+P')_{ij}(G'G)_{ij}}{2} & (G'G)_{ij}^2 \end{pmatrix}, \tag{5.12}$$

where existence and nonsingularity are again assured under Assumptions 6 and 7. With these reductions for the present case and analogous to Theorem 1, we obtain the following limit theory.

THEOREM 3. (a) *Under (5.1), Assumptions 1–4, 6(i), and 7, with $\tau_n(\cdot)$ and V_n defined according to (5.4) and (5.12), respectively,*

$$\sqrt{\frac{n}{h}}(\hat{\lambda} - \tau_n(\lambda_0)) \underset{d}{\to} \mathcal{N}(0, \bar{f}' \lim_{n\to\infty} V_n \bar{f}). \tag{5.13}$$

(b) *Under (5.1), Assumptions 1–4, 6, and 7, with $\tau_n(\cdot)$ and V_n defined according to (5.4) and (5.12), respectively,*

$$\sqrt{\frac{n}{h}}(\hat{\lambda}_{CUII} - \lambda_0) \underset{d}{\to} \mathcal{N}\left(0, v_{CUII}^2\right), \tag{5.14}$$

where $v_{CUII}^2 = \bar{f}' \lim_{n\to\infty} V_n \bar{f} / \left(\bar{\tau}^{(1)}\right)^2$ exists and is nonzero under Assumptions 2, 3(ii), and 5–7.

A sketch of the proof of Theorem 3 is reported in the Appendix, although it follows with minor modifications from that of Theorem 1.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

Another version of (2.1) is the SARX model with a local-to-zero coefficient vector

$$\beta_0 = \beta_{n,0} = \frac{\delta}{\sqrt{n}}, \tag{5.15}$$

with unknown localized coefficient vector $\delta \neq 0$. This model is a SAR with exogenous regressors of marginal relevance, as measured by $\beta_{n,0}$, which captures decreasing relevance as the sample size increases. As with the notation defined in Section 3, we have $P(\lambda) = Q'(\lambda)S^{-1}(\lambda)$ and $Q(\lambda) = M_X G(\lambda)$. Under Assumptions 2–4, and according to the same argument discussed in (3.6), the expected value of $\hat{\lambda}$ can be written as

$$\begin{aligned}\mathbb{E}(\hat{\lambda}) =& \frac{tr(P\Omega_0) + \beta_0' X'PX\beta_0}{tr(Q'Q\Omega_0) + \beta_0' X'Q'QX\beta_0} + O\left(\frac{1}{n}\right) \\ =& \frac{tr(P\Omega_0)}{tr(Q'Q\Omega_0)}\left(1 + \frac{1}{n}\frac{\delta' X'PX\delta}{tr(P\Omega_0)}\right)\left(1 - \frac{1}{n}\frac{\delta' X'Q'QX\delta}{tr(Q'Q\Omega_0)}\right) + O\left(\frac{1}{n}\right) \\ =& \frac{tr(P\Omega_0)}{tr(Q'Q\Omega_0)} + O\left(\frac{h}{n}\right),\end{aligned} \tag{5.16}$$

so that the binding function $\tau_n(\cdot)$ can be defined as

$$\tau_n(\lambda) = \tau(\lambda) = \frac{tr(P(\lambda)\Omega_\lambda)}{tr(Q'(\lambda)Q(\lambda)\Omega_\lambda)} + O_p\left(\frac{h}{n}\right), \tag{5.17}$$

with $\Omega_\lambda = diag(M_X S(\lambda) yy' S(\lambda) M_X)$.

Proceeding as in Section 4, we obtain

$$\sqrt{\frac{n}{h}}(\hat{\lambda} - \tau_n(\lambda_0)) = f_n' U_n + o_p(1), \tag{5.18}$$

where

$$U_n = \sqrt{\frac{h}{n}}\begin{pmatrix} \epsilon' P\epsilon - tr(P\Omega_{\lambda_0}) \\ \epsilon' Q'Q\epsilon - tr(Q'Q\Omega_{\lambda_0}) \end{pmatrix}, \quad \text{with} \quad \Omega_{\lambda_0} = diag(M_X \epsilon\epsilon' M_X), \tag{5.19}$$

$$f_n = \left(\left(\frac{h}{n} y' W' M_X W y\right)^{-1}, \left(\frac{h}{n} y' W' M_X W y\right)^{-2}\left(\frac{h}{n} y' W' M_X y\right)\right)', \tag{5.20}$$

and

$$\bar{f} = \operatorname*{plim}_{n\to\infty} f_n = \left((h\bar{c})^{-1}, (h\bar{c})^{-2}(h\bar{a})\right)'. \tag{5.21}$$

The equality in (5.18) follows trivially, since

$$\begin{aligned}\sqrt{\frac{h}{n}}(y' W' M_X y - tr(P\Omega_{\lambda_0})) =& \sqrt{\frac{h}{n}}\left(\epsilon' P\epsilon - tr(P\Omega_{\lambda_0}) + \beta_0' X'PX\beta_0 + 2\beta_0' X'P\epsilon\right) \\ =& \sqrt{\frac{h}{n}}(\epsilon' P\epsilon - tr(P\Omega_{\lambda_0})) + o_p(1),\end{aligned} \tag{5.22}$$

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

and

$$\sqrt{\frac{h}{n}}(y'W'M_XWy - tr(Q'Q\Omega_{\lambda_0})) = \sqrt{\frac{h}{n}}(\epsilon'Q'Q\epsilon - tr(Q'Q\Omega_{\lambda_0})) + o_p(1), \tag{5.23}$$

under (5.15). Setting

$$V_n = \frac{4h}{n}\sum_i\sum_{j<i}\sigma_i^2\sigma_j^2 \begin{pmatrix} \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix}, \tag{5.24}$$

we deduce the following limit theory for this local-to-pure SAR case.

THEOREM 4. (a) *Under (2.1), with $\beta_0 = \beta_{n,0}$ as in (5.15), Assumptions 1–4, 6(i), and 7, with $\tau_n(\cdot)$ and V_n defined according to (5.17) and (5.24), respectively,*

$$\sqrt{\frac{n}{h}}(\hat{\lambda} - \tau_n(\lambda_0)) \underset{d}{\to} \mathcal{N}(0, \bar{f}' \lim_{n\to\infty} V_n \bar{f}). \tag{5.25}$$

(b) *Under (2.1), with $\beta_0 = \beta_{n,0}$ as in (5.15), Assumptions 1–4, 6, and 7, with $\tau_n(\cdot)$ and V_n defined according to (5.17) and (5.24), respectively,*

$$\sqrt{\frac{n}{h}}(\hat{\lambda}_{CUII} - \lambda_0) \underset{d}{\to} \mathcal{N}\left(0, v_{CUII}^2\right), \tag{5.26}$$

where $v_{CUII}^2 = \bar{f}' \lim_{n\to\infty} V_n \bar{f} / \left(\bar{\tau}^{(1)}\right)^2$ exists and is nonzero under Assumptions 2, 3(ii), and 5–7.

A sketch of the proof of Theorem 4 is reported in the Appendix. A remark on the specific order $O(n^{-1/2})$ of the shrinking sequence in (5.15) is needed here. Theorem 4 above would hold with essentially no modification in the case $\beta_{0,n} = \delta/\alpha_n$, where α_n represents a generic positive sequence such that $\sqrt{n}/\alpha_n = o(1)$ as $n \to \infty$. In the case

$$\beta_{n,0} = \beta_0 = \frac{\delta}{\sqrt{h}}, \tag{5.27}$$

which corresponds either to a sequence of nonshrinking parameters (in case h is a bounded sequence), or to a shrinking sequence that converges to zero at a rate slower than $\sqrt{n}$, since $h/n = o(1)$ from Assumption 3(iv), the limit theory outlined in Section 4 would apply with virtually no modification after adjusting the normalizing sequences. In particular, Theorem 1 would hold with $\sqrt{n}$ replaced by $\sqrt{n/h}$, and V_n and $\bar{f}$ replaced by hV_n and

$$\bar{f} = \left(\left(h\bar{c} + h\bar{d}\right)^{-1};\ \left(h\bar{c} + h\bar{d}\right)^{-2}\left(h\bar{a} + h\bar{b}\right)\right)', \tag{5.28}$$

respectively.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

6. SPATIALLY CORRELATED DISTURBANCES: A DISCUSSION

The framework derived in Sections 3 and 4 could, in principle, be extended to more general settings such as higher-order SAR models or a SAR model with spatially correlated errors. These extensions would be achieved by adding one (or more) II conditions to match other quantities via suitable binding functions, as opposed to just the first moment as in Section 3. However, such developments come at the cost of additional high-level assumptions that assure the existence of a unique solution. We stress that the present high-level Assumption 6 has the crucial advantage of not depending on nuisance parameters, and thus can be easily checked on a case-by-case basis. This advantage would be lost in more complicated models.

We focus in this section on a heuristic discussion of a CUII-based Moran test for lack of spatially correlated errors by using the approach of Kelejian and Prucha (2001) or Robinson (2008). We consider the following model with SAR errors:

$$y = \lambda_0 Wy + X\beta_0 + u, \quad \text{with} \quad u = \rho_0 Du + \epsilon, \tag{6.1}$$

where D is a weight matrix satisfying Assumptions 3 and 4, and the error vector ϵ satisfies Assumption 1.

The null hypothesis to be tested is $\mathcal{H}_0 : E(uu') = \Omega_0$, i.e., $\rho_0 = 0$, against the alternative hypothesis $\mathcal{H}_1 : E(uu') \neq \Omega_0$. We can construct a test for $\mathcal{H}_0$ based on the vector of residuals $\hat{u}$ under $\mathcal{H}_0$. Thus, from Kelejian and Prucha (2001) or Robinson (2008), we use a suitably studentized quadratic form $\hat{u}'D\hat{u}$ (or its square) to construct a Moran-type of statistic to test $\mathcal{H}_0$. This testing framework in the presence of heteroskedasticity has been discussed by Kelejian and Prucha (2001), using Instrumental Variables IV estimates to construct $\hat{u}$. However, their test cannot accommodate the case $\beta_0 = 0$ ex ante, and it would perform poorly in case the regressors are increasingly irrelevant (e.g., in the case of weak instruments). The derivation of a Central Limit Theorem (CLT) for $\hat{u}'W\hat{u}/\sqrt{n}$ when the residuals $\hat{u}$ are obtained by CUII estimation is much more complex than the analysis of Kelejian and Prucha (2001), as CUII estimates are not simple linear functions of the disturbances.

Under $\mathcal{H}_0$, $\hat{u} = u(\hat{\lambda}_{CUII}) = M_X S(\hat{\lambda}_{CUII})y$, where $\hat{\lambda}_{CUII}$ is the restricted estimate of λ in (6.1), and thus

$$\begin{aligned}
&\frac{1}{\sqrt{n}} u(\hat{\lambda}_{CUII})' Du(\hat{\lambda}_{CUII}) \\
&= \frac{1}{\sqrt{n}} u(\lambda_0)' Du(\lambda_0) - \frac{1}{n} u(\lambda_0)'(D + D')M_X Wy \sqrt{n}(\hat{\lambda}_{CUII} - \lambda_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \epsilon' M_X D M_X \epsilon - \frac{a_n}{\tau_n^{(1)}} f_n' \frac{1}{\sqrt{n}} U_n + o_p(1) \\
&= \left(1 \quad \tfrac{a_n}{\tau_n^{(1)}} f_n'\right) \frac{1}{\sqrt{n}} \begin{pmatrix} \epsilon' M_X D M_X \epsilon \\ \epsilon' P \epsilon - tr(P\Omega_{\lambda_0}) + 2\beta_0' X' P M_X \epsilon \\ \epsilon' Q' Q \epsilon - tr(Q'Q\Omega_{\lambda_0}) + 2\beta_0' X' Q' Q M_X \epsilon \end{pmatrix} + o_p(1),
\end{aligned} \tag{6.2}$$

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

where

$$a_n = \frac{1}{n}\epsilon' M_X(D+D')M_X W S^{-1}\epsilon = \frac{1}{n}\epsilon' M_X(D+D')M_X G\epsilon. \tag{6.3}$$

From the second line on the RHS of (6.2), it is easy to show that the first term is $O_p(1/\sqrt{h})$, while the second is $O_p(1/h)$. Therefore, the argument that follows only holds when h is a bounded sequence (i.e., the setting considered by Kelejian and Prucha (2001, 2010) and Lin and Lee (2010)). A brief remark is made on the case of divergent h at the end of this section, but a thorough analysis in that case is beyond the scope of the present work. We modify Assumption 3 as follows.

Assumption 3'. (i) For all $n, W_{ii} = 0$ and $D_{ii} = 0$, for $i = 1, \ldots, n$.
(ii) For all n, $||W|| \leq 1$ and $||D|| \leq 1$.
(iii) For all sufficiently large n, $||W||_\infty + ||W'||_\infty \leq K$ and $||D||_\infty + ||D'||_\infty \leq K$.
(iv) For all sufficiently large n, uniformly in $i, j = 1, \ldots, n$, $W_{ij} = O(1/h)$, and $D_{ij} = O(1/h)$, where $h = h_n$ is bounded away from zero for all n and it is bounded as $n \to \infty$.

Let $T = M_X D M_X$,

$$\tilde{V}_n = \frac{4}{n}\begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_0' X' P M_X \Omega_0 M_X P' X \beta_0 & \beta_0' X' P M_X \Omega_0 M_X Q' Q X \beta_0 \\ 0 & \beta_0' X' Q' Q M_X \Omega_0 M_X P' X \beta_0 & \beta_0' X' Q' Q M_X \Omega_0 M_X Q' Q X \beta_0 \end{pmatrix}$$
$$+ \frac{4}{n}\sum_i \sum_{j<i} \sigma_i^2 \sigma_j^2 \begin{pmatrix} \frac{(T+T')_{ij}^2}{4} & \frac{(T+T')_{ij}(P+P')_{ij}}{4} & \frac{(T+T')_{ij}(Q'Q)_{ij}}{2} \\ \frac{(T+T')_{ij}(P+P')_{ij}}{4} & \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ \frac{(T+T')_{ij}(Q'Q)_{ij}}{2} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix}, \tag{6.4}$$

and modify Assumption 7 as follows.

Assumption 7'. $\lim_{n\to\infty} \tilde{V}_n$ exists and is nonsingular.

We deduce the following theorem.

THEOREM 5. *Let Assumptions 1, 2, 3', 4–6, and 7' hold. Then, as $n \to \infty$,*

$$\frac{1}{\sqrt{n}} u(\hat{\lambda}_{CUII})' D u(\hat{\lambda}_{CUII}) \xrightarrow{d} \mathcal{N}(0, \tilde{v}^2), \tag{6.5}$$

where

$$\lim_{n\to\infty} \begin{pmatrix} 1 & \frac{a_n}{\tau_n^{(1)}} f_n' \end{pmatrix} \tilde{V}_n \begin{pmatrix} 1 & \frac{a_n}{\tau_n^{(1)}} f_n' \end{pmatrix}'. \tag{6.6}$$

A sketch of the proof of Theorem 5 is given in the Appendix. A consistent estimator, $\hat{\tilde{v}}^2$, of $\tilde{v}^2$ can be obtained as in Theorem 2 by replacing unknowns with suitable estimates, viz., Ω_0 by $\hat{\Omega} = diag(\hat{\epsilon}\hat{\epsilon}')$, where $\hat{\epsilon} = M_X S(\hat{\lambda}_{CUII}) y$, and σ_i^2 by $\hat{\epsilon}_i^2$, for $i = 1, \ldots, n$. Thereupon, a Moran-type test statistic for $\mathcal{H}_0$ can be

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

constructed as

$$\frac{(u(\hat{\lambda}_{CUII})'Du(\hat{\lambda}_{CUII}))^2}{n\hat{\tilde{v}}^2}, \tag{6.7}$$

which has a χ^2 with one degree of freedom limit distribution.

The case of divergent h requires a different normalization rate, as $\tilde{V}$ would be singular for divergent h, and it is not pursued here. In such case,

$$\sqrt{\frac{h}{n}}u(\hat{\lambda}_{CUII})'Du(\hat{\lambda}_{CUII}) = \sqrt{\frac{h}{n}}\epsilon' M_X D M_X \epsilon + O_p\left(\frac{1}{\sqrt{h}}\right). \tag{6.8}$$

7. SIMULATIONS

We report the results of a set of Monte Carlo experiments to compare the finite-sample performance of the CUII estimators with the standard QML (Lee, 2004) and 2SLS estimators (Kelejian and Prucha, 1998), as well as the Robust Generalized Method of Moments (RGMM) procedure of Lin and Lee (2010) and the modified QML (MQML) estimator of Liu and Yang (2015).

We consider different heteroskedastic and spatial weight scenarios. Throughout, the number of exogenous regressors is set at $k = 3$, with the first regressor being an $n \times 1$ column of ones and the other two being randomly drawn from two independent uniform distributions on the support $[0, 1]$, and kept fixed across replications. In each scenario, we set $\beta_0 = (0.2, 0.1, -0.3)$, and consider four different values of λ_0, i.e., $\lambda_0 = -0.5, 0.3, 0.5, 0.8$. We generate ϵ_i, for $i = 1, \ldots, n$, as

$$\epsilon_i = \sigma_i \zeta_i, \tag{7.1}$$

with $\zeta_i \sim i.i.d.t(5)$, i.e., the ζ_i's are generated from a t-distribution with five degrees of freedom. Two mechanisms for the scale parameter σ_i are used: (i) direct construction using the formula

$$\sigma_i = c\frac{d_i}{\sum_{j=1}^{n} d_j/n}, \tag{7.2}$$

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, $d_i = \text{card}(j : \ w_{ij} \neq 0, i \neq j)$ and $\text{card}(\cdot)$ denotes cardinality; or the σ_i are drawn randomly from a χ^2 distribution with five degrees of freedom ($\chi^2(5)$). With both methods, the σ_i are kept fixed across simulations and across different parameter scenarios. The heteroskedasticity design in (7.2) is in line with the simulation work in 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.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

We consider two different choices for W, already partially introduced in Section 2 and reported here for convenience.[7] The first, which is denoted "random" in the tables and in the discussion that follows in this section, is randomly generated as an $n \times n$ symmetric matrix of zeros and ones and then rescaled, so that each row sums to unity, with the number of "ones" restricted to 10% of the total entries in W. This choice is empirically motivated as it mimics a fairly dense contiguity matrix. The second choice, denoted "exponential," is based on an exponential-decay notion of distance, again randomly generated. More specifically, we construct an $n \times 1$ vector of locations by generating n random numbers from a uniform distribution on support $[0, n]$. We then define $w_{ij} = exp\left(-|L_i - L_j|\right) \mathbb{1}(|L_i - L_j| < log(n))$. The resulting matrix is then normalized by its spectral norm.[8] Both choices of W are generated once for each sample size and are kept fixed across different scenarios and across the 1,000 Monte Carlo replications. We stress that for both choices of W, the QML is not expected to return consistent estimators in the presence of unknown heteroskedasticity, as the condition in (2.7) is not met.

Before reporting and discussing the simulation results, we numerically confirm Assumption 6 for a given set of regressors X and a realization of the disturbances for each of the aforementioned scenarios. Figure 2 shows the approximate binding function, $\tau^*(\lambda, \Omega_\lambda, \hat{\beta}(\lambda))$ in (3.8), for one single set of observations X ($k = 3$) and y generated for $\beta_0 = (0.2, 0.1, -0.3)$ and $\lambda = 0.3$. The plots depicted in Figure 2 correspond to all four configurations of heteroskedasticity/weighting structures previously discussed. For each case, Assumption 6 is evidently satisfied as the approximate binding function is strictly monotonic for all $\lambda \in \Lambda$. As expected, Assumption 6 may be violated in the vicinity of $\lambda = 1$, but this "unit root" case is excluded as is customary in the SAR literature.

Tables 1–4 report the bias and mean square error (MSE) for the CUII, QML, MQML, 2SLS, and RGMM procedures for $n = 30,\ 50,\ 100,\ 200$. The RGMM estimator corresponds to what Lin and Lee (2010) denote as optimal RGMM, and is constructed using the same algorithm described in Lin and Lee (2010). The 2SLS estimates are derived by the standard Kelejian and Prucha (1998) approach, i.e., using the linearly independent columns of $[WX, X]$ as instruments for $[Wy, X]$, even though similar results are obtained when using the linearly independent columns of $[WX, W^2X, X]$, or the so-called optimal instruments $[G(\tilde{\lambda})X\tilde{\beta}, X]$ with $(\tilde{\lambda}, \tilde{\beta}')'$ being preliminary estimates (Lee, 2003). Tables 1 and 2 report results for "random" W and σ_i in (7.1) generated as (7.2) and as $\chi^2(5)$, respectively. Tables 3 and 4 report corresponding results for "exponential" W.

Across all scenarios, the QML estimator of λ does not display a severe bias overall, but for the case of "random" W, the bias does not seem to vanish as n increases. On the other hand, the MQML of Liu and Yang (2015) enjoys excellent

[7] Figure S1 in the Online Supplementary Material gives graphical illustrations of these spatial matrix designs.

[8] Here and in the sequel, we define "sparseness" as the ratio between nonzero elements to total number of elements. The sparseness of the exponential distance W matrix decreases with n, and it amounts to 19% for $n = 30$, 14.96% for $n = 50$, 9.02% for $n = 100$, and 5.08% for $n = 200$.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

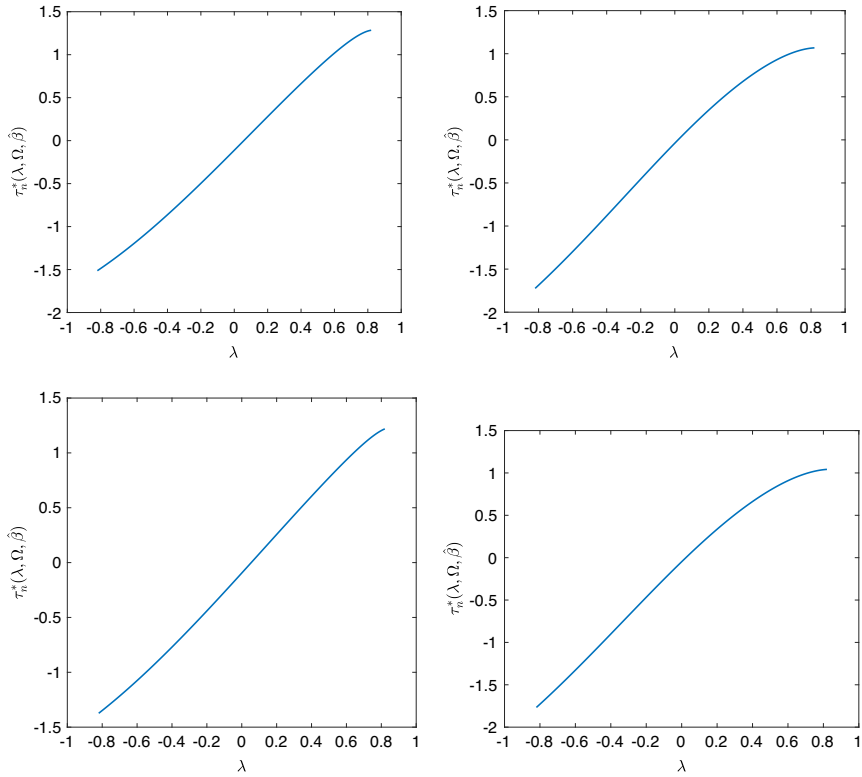

FIGURE 2. Plots of $\tau_n^*(\lambda)$ in (3.8) at $\lambda_0 = 0.3$ with $\beta_0 = (0.2; 0.1; -0.3)$ and $\zeta_i \sim iid\ t(5)$ for $n = 100$. Top: (L) "random" W & $\sigma_i \sim \chi^2(5)$, (R) "exponential" W & $\sigma_i \sim \chi^2(5)$; Bottom: (L) "random" W & σ_i defined as in (7.2), (R) "exponential" W & σ_i defined as in (7.2).

finite-sample performance and is considered our best benchmark. For "random" W, in terms of bias magnitude, MQML seems to slightly outperform CUII in the case where the heteroskedasticity is generated as in (7.2), whereas the opposite holds for σ_is generated as $\chi^2(5)$. For "exponential" W, in most cases, CUII seems to slightly outperform MQML in terms of bias, a feature which is notable especially at $\lambda = 0.5$. In general, the MSE of MQML is slightly smaller than that of CUII, even though the difference becomes progressively smaller as n increases. On the other hand, CUII appears to be nearly insensitive to the starting initialization used in the optimization routine when compared to the MQML.

It has been well documented in the spatial literature that 2SLS is not efficient compared to other methods, and indeed our 2SLS estimates across Tables 1–4 reveal particularly poor performance, having large biases and even larger MSE. Finally, the RGMM performance has a mixed pattern of performance, depending on the value of λ_0 and the choice of σ_i. For both "random" and "exponential" W,

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

TABLE 1. Bias and MSE of CUII, QML, MQML, 2SLS, and RGMM estimators for "random" W using 1,000 Monte Carlo replications. The ϵ_i's are defined as in (7.1) with $\zeta_i \sim iid\ \ t(5)$ and σ_i is defined as in (7.2).

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0745	0.1950	−0.0376	0.1264	−0.0144	0.0634	−0.0080	0.0560
	0.3	0.0100	0.1581	0.0260	0.1173	0.0224	0.0800	0.0030	0.0618
	0.5	0.0527	0.1653	0.0692	0.1191	0.0180	0.0751	0.0190	0.0609
	0.8	0.0793	0.1663	0.0738	0.1058	0.0777	0.0847	0.0329	0.0608
QML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0577	0.0803	−0.0031	0.0728	−0.0356	0.0506	−0.0393	0.0504
	0.3	−0.0858	0.0971	−0.0936	0.0805	−0.0797	0.0635	−0.0813	0.0539
	0.5	−0.1374	0.1013	−0.1049	0.0763	−0.1078	0.0571	−0.0950	0.0484
	0.8	−0.1267	0.0816	−0.1410	0.0702	−0.1326	0.0480	−0.1545	0.0498
MQML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	0.0358	0.0868	0.0181	0.0834	0.0060	0.0537	0.0002	0.0521
	0.3	−0.0331	0.0957	−0.0124	0.0817	−0.0015	0.0652	−0.0068	0.0564
	0.5	−0.0407	0.0872	−0.0025	0.0756	−0.0230	0.0557	−0.0015	0.0516
	0.8	−0.0427	0.0751	−0.0231	0.0759	−0.0031	0.0513	−0.0083	0.0486
2SLS		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.7459	2.3226	−0.4963	3.3981	−0.5898	4.4274	−0.5110	7.5457
	0.3	−0.0329	2.5238	0.0149	4.3781	0.0645	4.5001	0.1618	8.4440
	0.5	0.1809	3.1665	0.2530	2.9833	0.1343	2.5633	0.1618	5.0755
	0.8	0.3151	2.6836	0.1274	2.1983	0.2108	1.6882	0.2041	2.4499
RGMM		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.3074	0.6961	−0.1230	0.4128	−0.0629	0.0760	−0.0536	0.0645
	0.3	−0.0291	0.4907	−0.0289	0.2695	−0.0416	0.1224	−0.0557	0.0816
	0.5	0.0149	0.4958	0.0141	0.2493	−0.0319	0.2660	−0.0668	0.1370
	0.8	0.1151	0.5264	0.0721	0.4906	0.0339	0.5678	−0.0326	0.5065

when the σ_i's are generated as (7.2), the finite-sample performance of RGMM is satisfactory for larger sample sizes, although both bias and MSE are particularly poor for $\lambda_0 = -0.5$ in the "random" case. Furthermore, for "random" W, RGMM displays a large MSE for all sample sizes when $\lambda_0 = 0.8$, even though the bias is not particularly severe. The performance of the RGMM in terms of bias and MSE improves for both "random" and "exponential" cases when the σ_i's are generated from $\chi^2(5)$, even though it still remains substantially inferior to that of both CUII and MQML.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

Table 2. Bias and MSE of CUII, QML, MQML, 2SLS, and RGMM estimators for "random" W using 1,000 Monte Carlo replications. The ϵ_i's are defined as in (7.1) with $\zeta_i \sim iid\ \ t(5)$ and $\sigma_i \sim \chi^2(5)$.

		$n=30$		$n=50$		$n=100$		$n=200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0360	0.1262	−0.0158	0.0590	0.0111	0.0589	−0.0006	0.0521
	0.3	0.0132	0.1204	0.0066	0.0733	0.0036	0.0596	−0.0098	0.0523
	0.5	0.0003	0.1128	0.0089	0.0649	0.0040	0.0691	−0.0037	0.0580
	0.8	0.0691	0.1203	0.0136	0.0766	0.0357	0.0742	0.0197	0.0588
QML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0169	0.0733	−0.0301	0.0440	−0.0110	0.0487	−0.0278	0.0480
	0.3	−0.0728	0.0833	−0.0810	0.0606	−0.0751	0.0500	−0.0827	0.0491
	0.5	−0.1527	0.0947	−0.0801	0.0534	−0.1069	0.0555	−0.1051	0.0501
	0.8	−0.0985	0.0626	−0.0950	0.0445	−0.1418	0.0498	−0.1542	0.0511
MQML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	0.0374	0.0813	0.0216	0.0488	0.0295	0.0528	0.0072	0.0499
	0.3	−0.0175	0.0837	−0.0020	0.0588	−0.0047	0.0513	−0.0121	0.0547
	0.5	−0.0300	0.0772	−0.0145	0.0521	−0.0204	0.0538	−0.0143	0.0520
	0.8	−0.0316	0.0640	0.0006	0.0530	−0.0191	0.0466	−0.0036	0.0490
2SLS		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.5405	2.8543	−0.4837	1.8426	−0.6045	3.5927	−0.7222	11.3494
	0.3	−0.2197	1.8963	−0.2851	2.0994	−0.2312	5.9163	−0.1206	11.8204
	0.5	−0.2943	1.5970	−0.2300	1.6587	−0.0086	5.1936	0.0365	7.5799
	0.8	0.1399	1.4416	0.0665	0.7330	0.1751	1.6407	0.2028	3.4770
RGMM		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.1668	0.4255	−0.0858	0.1015	−0.0485	0.0813	−0.0423	0.0582
	0.3	−0.0492	0.2873	−0.0647	0.1140	−0.0705	0.0872	−0.0703	0.0652
	0.5	−0.0662	0.2928	−0.0464	0.1381	−0.0720	0.2267	−0.1203	0.1795
	0.8	0.0492	0.3546	−0.0134	0.1806	−0.0658	0.3921	−0.0725	0.7230

The performance of 2SLS and particularly that of RGMM seems to be heavily influenced by sparsity in W. The results in Tables 1 and 2 correspond to a choice of W with 10% "ones" for each sample size. In the Online Supplementary Material, we report the equivalent of Tables 1 and 2 for a choice of W with 20% "ones," which leads to a denser contiguity-like matrix. As expected, the MSEs reported in Tables S1 and S2 are systematically higher than their counterparts in Tables 1 and 2, and the performance of 2SLS and, especially, of RGMM, deteriorates even more clearly in terms of MSE as sparseness in W declines.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

TABLE 3. Bias and MSE of CUII, QML, MQML, 2SLS, and RGMM estimators for "exponential" W using 1,000 Monte Carlo replications. The ϵ_i's are defined as in (7.1) with $\zeta_i \sim iid\ \ t(5)$ and σ_i is defined as in (7.2).

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.1506	0.3090	−0.0951	0.1605	−0.0358	0.0656	−0.0137	0.0661
	0.3	−0.0838	0.1811	−0.0529	0.1022	−0.0179	0.0424	−0.0142	0.0450
	0.5	−0.0404	0.1688	−0.0268	0.0980	−0.0719	0.0394	−0.0043	0.0437
	0.8	0.0122	0.0741	0.0237	0.0532	0.0111	0.0268	−0.0119	0.0448
QML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.1088	0.1617	−0.0859	0.1155	−0.0251	0.0561	−0.0382	0.0678
	0.3	−0.2065	0.1516	−0.1254	0.0868	−0.0660	0.0378	−0.0470	0.0406
	0.5	−0.1772	0.1282	−0.1228	0.0769	−0.0719	0.0309	−0.0520	0.0340
	0.8	−0.1279	0.0636	−0.1033	0.0394	−0.0694	0.0187	−0.0564	0.0229
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0531	0.1884	−0.0553	0.1261	−0.0238	0.0616	−0.0082	0.0646
	0.3	−0.1020	0.1276	−0.0734	0.0810	−0.0287	0.0374	−0.0243	0.0390
	0.5	−0.1010	0.1097	−0.0750	0.0677	−0.0245	0.0277	−0.0321	0.0319
	0.8	−0.0401	0.0977	−0.0300	0.0462	−0.0187	0.0155	−0.0400	0.0216
2SLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	0.7895	3.0953	1.1721	4.9663	1.3151	6.2323	1.5788	6.2859
	0.3	0.6586	1.4838	0.7798	1.2917	0.9263	1.8396	1.0740	2.2559
	0.5	0.5258	1.2483	0.6827	1.6504	0.7489	1.1029	0.8554	1.5950
	0.8	0.3057	0.2669	0.2789	0.2419	0.3758	0.3419	0.4689	0.6066
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0967	0.5008	−0.0479	0.1979	−0.0075	0.0868	0.0239	0.0813
	0.3	0.0034	0.3632	0.0713	0.2917	0.0343	0.1076	0.0495	0.0985
	0.5	0.1273	0.4544	0.0778	0.2578	0.0849	0.1758	0.0723	0.1372
	0.8	0.1046	0.1596	0.1164	0.1352	0.0903	0.0982	0.0761	0.1027

In addition to the degree of sparsity of W, the poor performance of 2SLS and RGMM observed in Tables 1–4 might also depend on the particular simulation design, which corresponds to rather weak relevance in the instruments. In Tables S3 and S4 in the Online Supplementary Material, we report results derived for $\beta_0 = (2, 1.5, -1)$ and the two columns of X, other than that corresponding to the intercept, being randomly drawn from two independent uniform distributions on the support $[0, 4]$. We only focus on the choice of "exponential" W, but the same pattern holds for "random" W. In Tables S3 and S4, we also report bias

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

Table 4. Bias and MSE of CUII, QML, MQML, IV, and RGMM estimators for 'eExponential" W using 1,000 Monte Carlo replications. The ϵ_i's are defined as in (7.1) with $\zeta_i \sim iid\ t(5)$ and $\sigma_i \sim \chi^2(5)$.

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0990	0.2383	−0.0666	0.1576	−0.0388	0.0696	−0.0374	0.0703
	0.3	−0.0406	0.1812	−0.0459	0.0958	−0.0209	0.0381	−0.0184	0.0530
	0.5	−0.0364	0.1810	−0.0014	0.0891	−0.0094	0.0395	−0.0051	0.0454
	0.8	0.0251	0.0667	0.0337	0.0616	−0.0002	0.0258	−0.0121	0.0421
QML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.1076	0.1653	−0.0686	0.1134	−0.0618	0.696	−0.0589	0.0717
	0.3	0.1540	0.1300	−0.1210	0.0892	−0.0566	0.0381	−0.0594	0.0473
	0.5	−0.1808	0.1365	−0.0956	0.0631	−0.0551	0.0345	−0.0487	0.0376
	0.8	−0.1032	0.0531	−0.1072	0.0394	−0.0508	0.0165	−0.0625	0.0225
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.0385	0.1705	−0.0285	0.1219	−0.0269	0.0649	−0.0299	0.0670
	0.3	−0.0854	0.1182	−0.0665	0.0737	−0.0293	0.0351	−0.0288	0.0467
	0.5	−0.1048	0.1136	−0.0579	0.0551	−0.0319	0.0306	−0.0314	0.0356
	0.8	−0.0473	0.0759	−0.0511	0.0315	−0.0265	0.0154	−0.0227	0.0210
2SLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	0.6186	5.2417	1.3742	5.0555	1.1728	4.5728	1.6285	5.6988
	0.3	0.8710	1.9714	0.7492	1.0895	0.8836	1.4099	1.0015	2.8367
	0.5	0.6006	0.9648	0.7298	0.9990	0.6812	0.9991	0.8861	1.7479
	0.8	0.2798	0.2063	0.3322	0.3881	0.3538	0.2985	0.4931	0.7864
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	−0.5	−0.1083	0.3466	−0.0179	0.1787	−0.0152	0.0952	0.0078	0.0871
	0.3	0.0617	0.4151	0.0801	0.2614	0.0382	0.0787	0.0528	0.1089
	0.5	0.0935	0.3829	0.1226	0.2545	0.1044	0.1706	0.0653	0.1084
	0.8	0.1404	0.1590	0.1238	0.1797	0.1127	0.1202	0.0946	0.0970

and MSE for a continuously updated version of RGMM (denoted by CUGMM henceforth), obtained by using the same set of optimal instruments of Lin and Lee (2010), but by parameterizing the optimal choice of the weight matrix in terms of unknown parameters in the same spirit of Hansen et al. (1996).[9] Results displayed in Table S3 show that all estimators enjoy a very satisfactory performance, with

[9]We omit the comparison with CUGMM for the simulation setup adopted in Tables 1–4, as its performance appears similar to, if not worse than, that of RGMM.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

CUII and CUGMM often outperforming RGMM and MQML in terms of both bias and MSE. In Table S4, instead, CUII often tends to outperform both RGMM and CUGMM in terms of MSE, even though occasionally it has slightly larger bias. The performance of MQML is excellent overall and sometimes even superior to that of CUII. In sum, the finite-sample properties of all the estimators remain very satisfactory for the design adopted in Tables S3 and S4. In particular, we notice that 2SLS performs very well, unlike in Tables 1–4, S1, and S2. This might have been expected, given the strong instrument relevance and the simple closed form of 2SLS.

We also explore robustness of CUII versus competitor estimators in the case of model misspecification. Tables S5 and S6 given in the Online Supplementary Material report results for CUII, QML, MQML, and RGMM when the true data generating process is a pure SAR, while the estimated model is a SARX with intercept and one exogenous regressor which is drawn from a uniform distribution on the support $[0, 1]$. Tables S5 and S6 display results for "exponential" W, with σ_i generated as in (7.2) and from $\chi^2(5)$, respectively. Comparison with 2SLS is not possible here, as the standard instruments would be irrelevant. From Tables S5 and S6, as expected, RGMM has an erratic pattern, which becomes extremely poor for positive λ. This is due to irrelevance of instruments used to construct RGMM. The pattern of CUII, with respect to MQML, across sample sizes and different values of λ, follows that of Tables 3 and 4, confirming robustness of CUII in cases where the practitioner is agnostic about the relevance of regressors.

8. EMPIRICAL ILLUSTRATION

In this section, we report an empirical application of the CUII estimator and compare results to those obtained by the competitor methods QML, RGMM, and MQML. This application provides an illustration of the new method in a practical setting. The application is complemented by a further simulation that is matched to the empirical data and is therefore more realistic than standard Monte Carlo designs.

Specifically, we use Boston house price data (Harrison and Rubinfeld, 1978) and its "corrected" version (Gilley and Pace, 1996),[10] which also includes information on LON (tract point longitudes in decimal degrees) and LAT (tract point latitudes in decimal degrees) for the 506 census tracts in the Boston Standard Metropolitan Area during the early 1970s. The locations of the 506 census tracts are depicted in Figure 3 below.

The main variable of interest is $log(MEDV)$, which is the logarithm of the median price (in thousands of dollars) for owner-occupied houses. The dataset contains additional information about the following environmental and socioeconomic variables:[11]

[10] Data and codes are available from the authors upon request.

[11] For additional information about the dataset, we refer to Simlai (2014) and Harrison and Rubinfeld (1978).

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

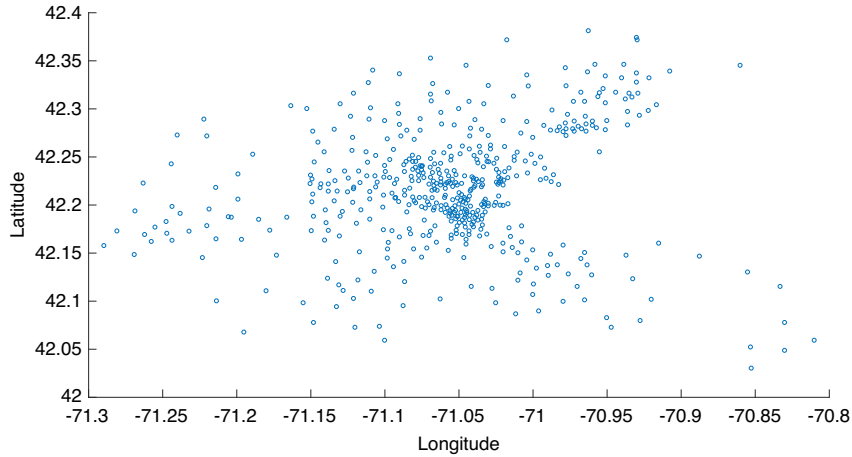

FIGURE 3. Locations of the 506 census tracts.

crim	per capita crime rate by town;
zn	proportion of residential land zoned for lots over 25,000 sq.ft;
indus	proportion of nonretail business acres per town;
chas	Charles River dummy variable (= 1 if tract bounds river; 0 otherwise);
nox	nitrogen oxides concentration (parts per 10 million);
rm	average number of rooms per dwelling;
age	proportion of owner-occupied units built prior to 1940;
dis	weighted mean of distances to five Boston employment centers;
rad	index of accessibility to radial highways;
tax	full-value property-tax rate per 10,000;
ptratio	pupil–teacher ratio by town;
black	$1,000 * (Bk - 0.63)^2$, where Bk is the proportion of blacks by town;
lstat	lower status of the population (percent).

In the spirit of Simlai (2014), we estimate parameters of the model

$$log(MEDV)_i = \alpha + \lambda \sum_{i \neq j} w_{ij} ln(MEDV)_j + x_i^{'}\beta + \epsilon_i \quad i = 1, \ldots, 506, \tag{8.1}$$

where the covariate vector x_i contains rm_i^2, age_i, $log(dis)_i$, $log(rad)_i$, tax_i, $ptratio_i$, $black_i$, $log(stat)_i, crim_i, zn_i, indus_i, chas_i, nox_i^2$. When $\lambda = 0$ ex ante, the model simplifies to the hedonic price model. Since the main scope of this paper is robust estimation and inference on λ, this illustration focuses on the spatial network effect in model (8.1) and thus on estimation and significance of λ, rather than the

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

covariate coefficient vector β. We recall that estimation of β in the general model (2.1) and in the specific model (8.1) poses fewer consistency and efficiency issues compared to inference on λ.

We conjecture that several measures of proximity might play a role in the house price determination process of $MEDV_i$, as both economic distance and geographical distance seem relevant. Accordingly, we design five different choices of weight matrix W, which are denoted, respectively, as W^{geo}, $W^{exp,geo}$, $W^{geo,0.9}$, W^{tax}, and W^{school}. The first three choices for W in (8.1) reflect geographical proximity and rely on the geo-distance between tracts i and j (denoted as geo_{ij} in the sequel) computed using the Haversine formula. The matrices W^{geo}, $W^{exp,geo}$, and $W^{geo,0.9}$ are then constructed as

- W^{geo}: $w_{ij} = 1/geo_{ij}$;
- $W^{geo,exp}$: $w_{ij} = exp\left(-|geo_{ij}|\right)\mathbb{1}(|geo_{ij}| < log(n))$;
- $W^{geo,0.9}$: $w_{ij} = \mathbb{1}(|geo_{ij}| < D^*)$, where we set $D^* = 2.5\ km$ to obtain a matrix sparseness[12] of approximately 9%.

The remaining two choices of W are defined in terms of various economic distances. W^{tax} contains the inverse of pairwise distances between census tracts, where proximity is defined according to how similar their respective full-property tax rates are. Specifically $w_{ij}^{tax} = 1/|tax_i - tax_j|$ if $tax_i \neq tax_j$, and $w_{ij}^{tax} = 1$ if $tax_i = tax_j$. Heuristically, we expect house prices to be affected more from the house prices of neighboring properties, where "neighbor" is now defined as being of similar status, which, in turn, is proxied by the property tax rate. Similarly, we define W^{school} based on the observable *ptratio*, which is known to reflect the quality of schools in each census tract. Again, two census tracts with similar *ptratio* are expected to be similar in terms of their socioeconomic status. We define $w_{ij}^{school} = 1/|ptratio_i - ptratio_j|$, as long as $ptratio_i \neq ptratio_j$, and $w_{ij}^{school} = 1$ in case $ptratio_i = ptratio_j$. For all choices of W, we set $w_{ii} = 0$, and we normalize the matrices, so that elements of each row sum to 1. Figures S2–S6 in the Online Supplementary Material depict the 3D plots for W^{geo}, $W^{geo,exp}$, $W^{geo,0.9}$, W^{tax}, and W^{school}, respectively. These plots offer an intuitive illustration on the spatial structure, the respective level of sparsity, and the respective decay rates of the weights, especially in view of the three nonsparse structures W^{geo}, W^{tax}, and W^{school}, for which the sparseness amounts to approximately 100%, but have decaying weights.

Figure S7 in the Online Supplementary Material displays the approximate binding functions for all the choices of the weight matrices for λ in $(-1, 1)$. Thus, Assumption 6 is satisfied for the present setup, confirming that CUII is well defined.

In Table 5, we report estimates and t-ratios for the parameter λ obtained by QML, CUII, RGMM, and MQML. The estimates of λ do not vary much across

[12]We remind the reader that we are calculating "sparseness" as the ratio of number of nonzero elements over the total number of elements.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

TABLE 5. Estimates and t-statistic of λ in (8.1) computed by QML, CUII, RGMM, and MQML for different choices of weighting structures. CPU times measured in seconds.

		QML	CUII	RGMM	MQML
W^{geo}	λ	0.0367	0.0381	0.0380	0.0379
	t-ratio	1.1990	16.3998	1.1213	2.7596
	CPU times	4.32	2.52	452.22	1.24
$W^{exp,geo}$	λ	0.0241	0.0246	0.0247	0.0247
	t-ratio	1.8542	19.8468	2.1385	17.6448
	CPU times	3.71	4.3	342.64	1.36
$W^{geo,0.9}$	λ	0.0130	0.0131	0.0127	0.0132
	t-ratio	2.4790	44.7070	3.0448	72.5195
	CPU times	4.87	4.15	150.7	1.27
W^{tax}	λ	0.0217	0.0230	0.0143	0.0235
	t-ratio	1.2368	12.3797	1.0886	12.7909
	CPU times	4.48	3.60	378.06	12.79
W^{school}	λ	−0.0269	−0.0268	−0.0271	−0.0268
	t-ratio	−2.9089	−91.3553	−2.3550	−17.9526
	CPU times	3.75	2.89	52.29	1.29
	average CPU times	4.226	3.492	275.18	3.59

QML, CUII, MQML, and RGMM. But a major difference between the methods shows up in the significance of the spatial effects. The methods CUII and MQML return similar results with highly significant t-statistics in all scenarios, with t-statistics substantially larger by an order of magnitude than those of the other methods. The effects of robustification to heterogeneity in the equation errors is therefore materially important in hypothesis testing. The CPU times (in seconds) are also displayed for each estimation method.[13] These values reveal that the CUII is computationally inexpensive, a feature which is robust across all choices of weight matrices as shown by the average CPU time (across all W choices), while the RGMM appears to be the most computationally expensive method. While the MQML has a further computational advantage over CUII for the majority of W choices, its CPU time appears to be sensitive to the choice of W.

The t-statistics obtained by QML are unreliable, because the QML standard errors are not robust to heteroskedasticity of the errors, but the corresponding figures are reported in the tables for completeness. Finally, the t-ratios obtained

[13]This has been calculated using the *cputime* command in Matlab.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

TABLE 6. Bias and MSE for $N = QML, CUII, RGMM, MQML$ computed from $B = 100$ bootstrap samples for various choices of weighting structures.

		QML	CUII	RGMM	MQML
W^{geo}	$Bias_N/\hat{\lambda}_N$	−0.0007	0.0646	0.0546	0.0691
	$MSE_N/\hat{\lambda}_N^2$	0.9204	0.8136	0.7645	0.9678
$W^{exp,geo}$	$Bias_N/\hat{\lambda}_N$	−0.2344	−0.0095	−0.0220	−0.0386
	$MSE_N/\hat{\lambda}_N^2$	0.1452	0.0618	0.0642	0.1647
$W^{geo,0.9}$	$Bias_N/\hat{\lambda}_N$	0.0465	0.0525	0.0554	0.0523
	$MSE_N/\hat{\lambda}_N^2$	0.1277	0.1268	0.1313	0.1259
W^{tax}	$Bias_N/\hat{\lambda}_N$	−0.0967	−0.0382	−0.0610	−0.0359
	$MSE_N/\hat{\lambda}_N^2$	0.4461	0.3658	0.9324	0.3888
W^{school}	$Bias_N/\hat{\lambda}_N$	−0.0154	−0.0199	−0.0176	−0.0201
	$MSE_N/\hat{\lambda}_N^2$	0.1817	0.1732	0.1816	0.1840

by RGMM are not significant for the weighting structures W^{tax} and W^{geo}, a result that contrasts sharply with those obtained by CUII and MQML.

In order to assess the accuracy, and hence the reliability of point estimates and t-statistics displayed in Table 5, we mimic the empirical illustration in a Monte Carlo exercise by means of B bootstrap samples constructed from the point estimates in Table 5 and using a wild bootstrap variant to generate simulation errors from the actual residuals. Specifically, we denote by $\hat{\epsilon}_N$, for $N = QML, CUII, RGMM, MQML$, the vector of residuals based on entries in Table 5, and we generate an $n \times B$ matrix of i.i.d. random variables from the Rademacher distribution, with typical element indicated by r_{ij}, for $i = 1, \dots n$ and $j = 1, \dots, B$. For each estimator N and each choice of W, we then generate B sets of pseudodata vectors as

$$y_{N,j}^* = S^{-1}(\hat{\lambda}_N)\left(X\hat{\beta}_N + u_{N,j}^*\right), \quad j = 1, \dots, B, \tag{8.2}$$

with each component of $u_{N,j}^*$ being constructed as $u_{N,ij}^* = (n/(n-k))^{1/2}\hat{\epsilon}_{N,i}r_{ij}$ ($k = 14$), and we calculate the corresponding B estimators $\hat{\lambda}_{N,j}^*$, for $j = 1, \dots, B$, so that a measure of bias and MSE can be computed as

$$Bias_N = \frac{1}{B}\sum_{j=1}^{B}\hat{\lambda}_{N,j}^* - \hat{\lambda}_N, \quad MSE_N = Bias_N^2 + \frac{1}{B}\sum_{j=1}^{B}\left(\hat{\lambda}_{N,j}^* - \frac{1}{B}\sum_{j=1}^{B}\hat{\lambda}_{N,j}^*\right)^2. \tag{8.3}$$

Results of this bootstrap exercise are reported in Table 6. For ease of interpretation, given the small magnitude of $\hat{\lambda}_N$, for $N = QML, CUII, RGMM, MQML$, in Table 6, we report scaled quantities, i.e., $Bias_N/\hat{\lambda}_N$ and $MSE_N/\hat{\lambda}_N^2$.

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

Results in Table 6 show that $\hat{\lambda}_{CUII}$ generally has similar performance in terms of bias and MSE as its robust counterparts RGMM and MQML and improved performance in some cases, notably for $W^{exp,geo}$. Again, *QML* is not expected to be consistent in these scenarios, although its bias appears to be substantially larger than that of other estimators only for $W^{exp,geo}$. More importantly, the clear advantage in terms of efficiency of $\hat{\lambda}_{CUII}$ and $\hat{\lambda}_{MQML}$ over $\hat{\lambda}_{RGMM}$ outlined in Table 5 for W^{tax} is confirmed by the simulation exercise reported in Table 6. However, this small simulation exercise reveals that results obtained when proximity is defined as W^{geo} are very erratic, and thus practitioners should treat estimates and tests reported in the first line of Table 5 as not particularly reliable. This behavior is probably due to the very dense structure of this choice of weighting structure. Results for $W^{geo,0.9}$ and W^{school} reported in Table 6 partially confirm those in Table 5, as MSE_{CUII} and MSE_{MQML} are lower than MSE_{RGMM}, but the comparative advantage in terms of *MSE* does not fully match the considerable difference in their *t*-statistic values reported in Table 5.

While the main goal of this empirical exercise is to illustrate the implementation of the CUII method in relation to other spatial econometric methods, the results do reveal some interesting features concerning the various channels of spatial correlation in the context of house price determination. In particular, it seems worthy of mention that spatial effects that are present when the network structure is defined in terms of W^{tax} or W^{school} continue to persist when the individual levels of *tax* and *ptratio* are included among regressors, revealing a genuinely significant impact. It is also worth pointing out that the spatial effects induced by W^{tax} and W^{school} differ in sign and thus their interpretation as positive or negative spatial spillovers differ, an empirical feature that might usefully be explored in subsequent research.

9. CONCLUDING REMARKS

Unobserved heteroskedasticity in the disturbances is a frequent occurrence in spatial models due to sample unit heterogeneity across their many individual features, including respective unit size. The new estimation method introduced in this paper directly addresses such heterogeneity, relying on an II transformation of standard OLS estimation that parameterizes the error covariance matrix in terms of the unknown spatial parameter. The procedure follows in the spirit of continuously updated estimators in the broader econometric literature such as GMM. The resulting CUII estimator is consistent and asymptotically normal under some standard model and regularity conditions combined with an additional binding function condition that can be numerically verified in practical work.

The finite-sample performance of the CUII estimator is found in simulations to be very satisfactory when compared to other robust methods such as the GMM robust procedures of Lin and Lee (2010) and Kelejian and Prucha (2010) or the modified QMLE procedure of Liu and Yang (2015). Implementation of CUII is straightforward, and the optimization routine to derive the estimator appears to

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

converge quickly even when an artificially dense W matrix is designed. The finite-sample performance of CUII appears to retain some degree of robustness even in the presence of model misspecification. A simple empirical illustration based on Boston house price data reveals that a major advantage of accommodating heterogeneity in system disturbances lies in hypothesis testing, where significance tests are found to differ considerably across estimation methods, with CUII giving much higher levels of significance to spatial effects across many different choices of the house price network structure.

APPENDIX

Proof of Theorem 1.

Proof of part (i) Let ψ_{ij} and $\tilde{\psi}_{ij}$ be the 2×1 vectors defined as $\psi_{ij} = (\psi_{1ij} \;\; \psi_{2ij})' = ((P+P')_{ij}/2 \;\; (Q'Q)_{ij})'$ and $\tilde{\psi}_{ij} = (\tilde{\psi}_{1ij} \;\; \tilde{\psi}_{2ij})' = ((M_X P')_{ij} \;\; (M_X Q'Q)_{ij})'$, respectively. Let $\tilde{\Omega} = diag(\epsilon\epsilon')$ and $\Omega_{\lambda_0} = diag(M_X \epsilon\epsilon' M_X)$, consonant with the notation of Section 4. We first show

$$\begin{aligned} U_n &= \frac{1}{\sqrt{n}} \begin{pmatrix} \epsilon' P\epsilon - tr(P\Omega_{\lambda_0}) + 2\beta_0' X' P M_X \epsilon \\ \epsilon' Q'Q\epsilon - tr(Q'Q\Omega_{\lambda_0}) + 2\beta_0' X' Q'Q M_X \epsilon \end{pmatrix} \\ &= \frac{1}{\sqrt{n}} \begin{pmatrix} \epsilon' P\epsilon - tr(P\tilde{\Omega}) + 2\beta_0' X' P M_X \epsilon \\ \epsilon' Q'Q\epsilon - tr(Q'Q\tilde{\Omega}) + 2\beta_0' X' Q'Q M_X \epsilon \end{pmatrix} + o_p(1). \end{aligned} \tag{A.1}$$

Thus, we need to show

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{ii,s} (\epsilon_i(\lambda_0)^2 - \epsilon_i^2) = o_p(1), \quad s = 1, 2, \tag{A.2}$$

where $\epsilon_i(\lambda_0) = \epsilon_i - \sum_i^n B_{ij}\epsilon_j$, $B_{ij} = X_i'(X'X)^{-1}X_j$.

We have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_i \psi_{ii,s} (\epsilon_i(\lambda_0)^2 - \epsilon_i^2) &= \frac{1}{\sqrt{n}} \sum_i \psi_{ii,s} \sum_j \sum_t \epsilon_j \epsilon_t B_{ij} B_{it} - \frac{2}{\sqrt{n}} \sum_i \psi_{ii,s} \epsilon_i \sum_j B_{ij} \epsilon_j \\ &= \frac{1}{\sqrt{n}} \sum_i \psi_{ii,s} \sum_j \epsilon_j^2 B_{ij}^2 + \frac{1}{\sqrt{n}} \sum_i \psi_{ii,s} \sum_j \sum_{t \neq j} \epsilon_j \epsilon_t B_{ij} B_{it} \\ &\quad - \frac{2}{\sqrt{n}} \sum_i \psi_{ii,s} \epsilon_i^2 B_{ii} - \frac{2}{\sqrt{n}} \sum_i \psi_{ii,s} \epsilon_i \sum_{j \neq i} B_{ij} \epsilon_j. \end{aligned} \tag{A.3}$$

The modulus of the first term in (A.3) has expectation bounded by

$$\frac{C}{\sqrt{n}} \sum_i |\psi_{ii,s}| \sum_j B_{ij}^2 \leq \frac{C}{\sqrt{n}h} \sum_i \sum_j B_{ij}^2 = \frac{C}{\sqrt{n}h} tr(X(X'X)^{-1}X') = o(1), \tag{A.4}$$

as $\psi_{ii,s} = O(1/h)$ under Assumptions 3 and 4. Similarly, the modulus of the third term has expectation bounded by

$$\frac{C}{\sqrt{n}} \sum_i |\psi_{ii,s}| B_{ii} \leq \frac{C}{\sqrt{n}h} \sum_i B_{ii} = o(1). \tag{A.5}$$

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

The second term in (A.3) has mean zero and variance bounded by

$$\frac{C}{n}\sum_i\sum_v\sum_j\sum_{t\neq j}|\psi_{ii,s}||\psi_{vv,s}||B_{ij}B_{it}B_{vt}B_{vj}| \leq \frac{C}{n}\sum_i\sum_v\sum_j\sum_t|\psi_{ii,s}||\psi_{vv,s}||B_{ij}B_{it}B_{vt}B_{vj}|$$
$$\leq \frac{C}{nh^2}\sum_i\sum_v\sum_j\sum_t|B_{ij}B_{it}|(B_{vt}^2+B_{vj}^2) \leq \frac{C}{nh^2}\sup_i\sum_j|B_{ij}|\sup_t\sum_i|B_{it}|\sum_v\sum_t B_{vt}^2$$
$$+\frac{C}{nh^2}\sup_j\sum_i|B_{ij}|\sup_i\sum_t|B_{it}|\sum_v\sum_j B_{vj}^2 \leq \frac{C}{nh^2}, \tag{A.6}$$

under Assumptions 3–5. Similarly, the fourth term in (A.3) has mean zero and variance bounded by

$$\frac{C}{n}\sum_i\sum_{j\neq i}\psi_{ii,s}^2B_{ij}^2+\frac{C}{n}\sum_i\sum_{j\neq i}\psi_{ii,s}\psi_{jj,s}B_{ij}B_{ji} \leq \frac{C}{nh^2}\sum_i\sum_j B_{ij}^2 \leq \frac{C}{nh^2}. \tag{A.7}$$

Then, (A.2) holds by the Markov inequality.

The remainder of the proof of part (i) and the proof of part (ii) are similar to the proofs in KPR and are reported in the Online Supplementary Material. □

Proof of Theorem 2. Let $\hat{\epsilon} = M_X S(\hat{\lambda}_{CUII})y$ and $\hat{\Omega} = diag(\hat{\epsilon}\hat{\epsilon}')$. We need to show, as $n \to \infty$ and $t,s = 1,2$, that

$$\frac{1}{n}\sum_i\sum_{j<i}\left(\hat{\epsilon}_i^2\hat{\epsilon}_j^2\hat{\psi}_{sij}\hat{\psi}_{tij}-\sigma_i^2\sigma_j^2\psi_{sij}\psi_{tij}\right)=o_p(1), \tag{A.8}$$

$$\frac{1}{n}\left(tr(\hat{\Omega}\hat{A})-tr(\Omega_0 A)\right)=o_p(1), \tag{A.9}$$

and

$$\frac{1}{n}\left(\hat{\beta}'_{CUII}X'\hat{\tilde{\Psi}}'_s\hat{\Omega}\hat{\tilde{\Psi}}_tX\hat{\beta}_{CUII}-\beta_0'X'\tilde{\Psi}'_s\Omega_0\tilde{\Psi}_tX\beta_0\right)=o_p(1), \tag{A.10}$$

where, consonant with the notation defined at the beginning of the proof of Theorem 1, $\psi_{ij} = \begin{pmatrix}\psi_{1ij} & \psi_{2ij}\end{pmatrix}' = \begin{pmatrix}(P+P')_{ij}/2 & (Q'Q)_{ij}\end{pmatrix}'$, $\tilde{\Psi}_1 = M_X P$, $\tilde{\Psi}_2 = M_X Q'Q$, and $\hat{\psi}_{sij}$, $\hat{\tilde{\Psi}}_t$, for $s,t = 1,2$, are obtained by replacing the unknown λ_0 by its estimate $\hat{\lambda}_{CUII}$. Furthermore, $\hat{A}$ is the estimated version of a generic matrix $A = A(\lambda_0)$ whose elements are uniformly bounded by $1/h$ and such that $||A(\lambda)||_\infty + ||A(\lambda)'||_\infty < C$ uniformly over λ. Convergence of the other terms appearing in v_{CUII}^2 is trivial, as it only relies on consistency of $\hat{\lambda}_{CUII}$ and $\hat{\beta}_{CUII}$.

We provide a proof of (A.8), while the proofs of (A.9) and (A.10) are omitted, as they follow very similar arguments to those applied to show (A.8) and (A.2) at the beginning of the proof of Theorem 1. In order to prove (A.8), we need to show

$$\frac{1}{n}\sum_i\sum_{j<i}\left(\epsilon_i^2\epsilon_j^2-\sigma_i^2\sigma_j^2\right)\psi_{sij}\psi_{tij}=o_p(1), \tag{A.11}$$

$$\frac{1}{n}\sum_i\sum_{j<i}\left(\hat{\epsilon}_i^2\hat{\epsilon}_j^2-\epsilon_i^2\epsilon_j^2\right)\psi_{sij}\psi_{tij}=o_p(1), \tag{A.12}$$

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

and

$$\frac{1}{n}\sum_{i}\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2(\hat{\psi}_{sij}\hat{\psi}_{tij}-\psi_{sij}\psi_{tij})=o_p(1). \tag{A.13}$$

The rest of the proof is technical, and the details are reported in the Online Supplementary Material. □

Proof of Theorem 3. The details of the proof are virtually identical to that of Theorem 1 once we adjust for the correct normalization rate. Following the same notation adopted in the proof of Theorem 1, let ψ_{ij} be the 2×1 vectors defined as $\psi_{ij}=\begin{pmatrix}\psi_{1ij} & \psi_{2ij}\end{pmatrix}'=\begin{pmatrix}(P+P')_{ij}/2 & (G'G)_{ij}\end{pmatrix}'$, and

$$u_i=\begin{pmatrix}u_{1i} & u_{2i}\end{pmatrix}'=2\epsilon_i\sum_{j<i}\psi_{ij}\epsilon_j, \tag{A.14}$$

so that $\sqrt{n/h}U_n=\sum_{i=1}^n u_i+o_p(1)$, according to (5.9). Furthermore, let

$$A=Var\left(\sum_{i=1}^n u_i\right)=4\sum_{i=1}^n\sum_{j<i}\sigma_i^2\sigma_j^2\psi_{ij}\psi_{ij}', \tag{A.15}$$

such that $hA/n\to\lim_{n\to\infty}V_n$ as $n\to\infty$, where V_n is defined in (5.12). The rest of the proofs follows in a similar way to the proof of Theorem 1 with minor modifications. □

Proof of Theorem 4. The details of the proof are again identical to that of Theorem 1 once we adjust for the correct normalization rate. Unlike the proof of Theorem 3, before proceeding, we need to show

$$\begin{aligned}U_n=&\sqrt{\frac{h}{n}}\begin{pmatrix}\epsilon'P\epsilon-tr(P\Omega_{\lambda_0})\\ \epsilon'Q'Q\epsilon-tr(Q'Q\Omega_{\lambda_0})\end{pmatrix}\\ =&\sqrt{\frac{h}{n}}\begin{pmatrix}\epsilon'P\epsilon-tr(P\tilde{\Omega})+\\ \epsilon'Q'Q\epsilon-tr(Q'Q\tilde{\Omega})\end{pmatrix}+o_p(1),\end{aligned} \tag{A.16}$$

where, consonant with the notation defined in the proof of Theorem 1, $\tilde{\Omega}=diag(\epsilon\epsilon')$ and $\Omega_{\lambda_0}=diag(M_X\epsilon\epsilon'M_X)$. The proof is identical to that reported in (A.2)–(A.7) after replacing the rate $1/\sqrt{n}$ with $\sqrt{h/n}$. The rest of the proof then follows as in Theorem 3, with V_n and $\bar{f}$ defined in (5.24) and (5.21). □

Proof of Theorem 5. The proof again follows that of Theorem 1 after redefining ψ_{ij} and $\tilde{\psi}_{ij}$ to be the 3×1 vectors $\psi_{ij}=(\psi_{1ij}\ \ \psi_{2ij}\ \ \psi_{3ij})'=((T+T')_{ij}/2\ \ (P+P')_{ij}/2\ \ (Q'Q)_{ij})'$ and $\tilde{\psi}_{ij}=\begin{pmatrix}\tilde{\psi}_{1ij} & \tilde{\psi}_{2ij} & \tilde{\psi}_{3ij}\end{pmatrix}'=\begin{pmatrix}0 & (M_XP')_{ij} & (M_XQ'Q)_{ij}\end{pmatrix}'$. □

SUPPLEMENTARY MATERIAL

To view the supplementary material for this article, please visit: http://doi.org/10.1017/S0266466621000384

https://doi.org/10.1017/S0266466621000384 Published online by Cambridge University Press

REFERENCES

Arraiz, I., D. M. Drukker, H. H. Kelejian & I. R. Prucha (2010) A spatial Cliff–Ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science* 50(2), 592–614.

Badinger, H. & P. Egger (2011) Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. *Papers in Regional Science* 90(1), 213–235.

Bao, Y. (2013) Finite sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* 29(1), 68–88.

Bao, Y., X. Liu & L. Yang (2020) Indirect inference estimation of spatial autoregressions. *Econometrics* 8(3), 34.

Bisin, A., G. Topa & T. Verdier (2004) Religious intermarriage and socialization in the United States. *Journal of Political Economy* 112(3), 615–664.

Case, A. C. (1991) Spatial patterns in household demand. *Econometrica: Journal of the Econometric Society* 59, 953–965.

Durbin, J. (1988) Maximum likelihood estimation of the parameters of a system of simultaneous regression equations. *Econometric Theory* 4(1), 159–170.

Gilley, O. W. & R. Pace (1996) On the Harrison and Rubinfeld data. *Journal of Environmental Economics and Management* 31(3), 403–405.

Glaeser, E. L., B. Sacerdote & J. A. Scheinkman (1996) Crime and social interactions. *The Quarterly Journal of Economics* 111(2), 507–548.

Gouriéroux, C., E. Renault & N. Touzi (1995) *Calibration by Simulation for Small Sample Bias Correction*. Institut National De La Statistique et des Etudes Economiques.

Hansen, L. P., J. Heaton & A. Yaron (1996) Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* 14(3), 262–280.

Harrison, D. & D. L. Rubinfeld (1978) Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 5(1), 81–102.

Harvey, A. C. (1976) Estimating regression models with multiplicative heteroscedasticity. *Econometrica* 44(3), 461–465.

Jin, F. & L. F. Lee (2019) GEL estimation and tests of spatial autoregressive models. *Journal of Econometrics* 208(2), 585–612.

Kelejian, H. H. & I. R. Prucha (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics* 17(1), 99–121.

Kelejian, H. H. & I. R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40(2), 509–533.

Kelejian, H. H. & I. R. Prucha (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104(2), 219–257.

Kelejian, H.H. & I.R. Prucha (2007) HAC estimation in a spatial framework. *Journal of Econometrics* 140(1), 131–154. Analysis of spatially dependent data.

Kelejian, H. H. & I. R. Prucha (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157(1), 53–67.

Kyriacou, M., P. Phillips & F. Rossi (2017) Indirect inference in spatial autoregression. *Econometrics Journal* 20(2), 168–189.

Lee, L. F. (2002) Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory* 18(2), 252–277.

Lee, L. F. (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22(4), 307–335.

Lee, L. F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72(6), 1899–1925.

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

LeSage, J. P. (1999) *The Theory and Practice of Spatial Econometrics*, vol. 28. University of Toledo.
Lieberman, O. (1994) A Laplace approximation to the moments of a ratio of quadratic forms. *Biometrika* 81(4), 681–690.
Lin, X. & L.F. Lee (2010) GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157(1), 34–52. Nonlinear and Nonparametric Methods in Econometrics.
Liu, S. F. & Z. Yang (2015) Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. *Regional Science and Urban Economics* 52, 50–70.
Robinson, P. M. (2008) Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics* 147(1), 5–16.
Simlai, P. (2014) Estimation of variance of housing prices using spatial conditional heteroskedasticity (SARCH) model with an application to Boston housing price data. *The Quarterly Review of Economics and Finance* 54(1), 17–30.