CONTINUOUSLY UPDATED INDIRECT INFERENCE IN HETEROSKEDASTIC SPATIAL MODELS

By

Maria Kyriacou, Peter C. B. Phillips, and Francesca Rossi

October 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2208

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

http://cowles.yale.edu/

Continuously Updated Indirect Inference in Heteroskedastic Spatial Models*

Maria Kyriacou† Peter C. B. Phillips‡ Francesca Rossi§

October 5, 2019

Abstract

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, standard 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 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 indirect inference procedure to take account of the parametrization of the variance-covariance matrix of the disturbances. Finite sample performance of the new estimator is assessed in a Monte Carlo study and found to offer advantages over existing methods. 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.

JEL Classification C13; C15; C21

Keywords: Spatial autoregression; Unknown heteroskedasticity; Indirect inference; Robust methods; Weights matrix.

*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.

†University of Southampton

‡Yale University, University of Auckland, University of Southampton, Singapore Management University

§University of Verona.

1 Introduction

In recent years spatial models have stimulated growing interest and application in various areas in economics. Economic data frequently exhibit strong spatial 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 know as spatial autoregressions (SARs hereafter). SAR models, like vector autoregressions, have the great advantages 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 electronic proximity as well as many other socio-economic characteristics.

For SAR models with homoskedastic innovations a wide range of estimation procedures are available. These include maximum likelihood/quasi maximum likelihood (ML/QMLE) estimation (Lee (2004)), two-stage least squares (2SLS) (Kelejian and Prucha (1998)), generalised method of moments (GMM) estimation and indirect inference (Kyriacou, Phillips and Rossi (2017)). Spatial units typically vary over many observed and unobserved characteristics, leading to potentially heterogeneous innovations that introduce bias and invalidate these commonly used estimation procedures.

Although ML/QML methods provide an obvious general approach to parameter estimation (Lee (2004)), in the presence of unobserved heterogeneity these methods produce inconsistencies in spatial parameter and coefficient estimation (e.g. Liu and Lee (2010)). This lack of robustness under heterogeneity is possibly the main shortcoming of ML/QML methods for spatial data, as 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 US states (Bisin, Topa and Verdier (2004)), house selling prices (Harrison and Rubinfeld (1978), LeSage (1999)) and crime rates and social interactions across contiguous US states (Glaeser, Sacerdote and Scheinkman (1996)), where these effects are important.

In contrast to these methods, the simple use of ordinary least squares (OLS) estimation of the parameters of a SAR model with exogenous regressors is consistent under certain restrictive assumptions on the limit behaviour of the spatial design, as discussed in Lee (2002). OLS may also 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 neighbours of a given spatial unit remains fixed as the sample size grows, even in the simpler setting of homoskedastic disturbances . Importantly, the necessary assumptions on the limit behaviour of weights structure which ensure consistency are difficult to verify in practical situations, making OLS estimation a questionable choice for practitioners.

At present, there few techniques that can capably account for heterogeneity of general form as well as general weight matrix structures. Three options are presently available. Lin and Lee (2010) propose a robust generalized method of moments (RGMM) estimator which delivers consistent estimation of the parameters of SAR models with heterokedastic errors. Kelejian and Prucha (2010) consider a GMM-type method which is robust to heteroskedasticity with a particular focus on the SARAR(1,1) model structure [1]. More recently, Liu and Yang (2015) propose a modified QLE/MLE estimator (MQML) that restores consistency by adjusting the score function for the spatial parameter to accommodate general forms of heteroskedasticity.

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 this case, the II procedure converts an inconsistent OLS estimator to a consistent estimator.

We here propose a similar enhancement in the case of heterogeneous spatial errors. The key idea of the approach is to accommodate more realistic error structures by parametrizing the variance-covariance structure in terms of unknown parameters of interest that are incorporated into a suitable binding function within the indirect inference mechanism of estimation. The 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

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

[2] See Durbin (1963, 1988) for an early version of this idea in the context of efficiently estimating structural equation models by iterative instrumental variable methods.

weights schemes that are relevant to empirical work. Within the SAR framework with exogenous regressors (SARX, in the sequel), we show that the CUII estimator is consistent and asymptotically normal. Simulation and empirical results confirm that CUII enjoys excellent finite sample properties under very general spatial design and heteroskedasticity structures and outperforms existing estimation methods under some network structures.

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 behaviour is explored in Section 4. Section 5 presents a Monte Carlo exercise to compare the finite sample behaviour of the CUII estimator to OLS, QML, RGMM and MQML, while Section 6 reports a comparison of estimation methods for inference on the spatial parameter in the context of house price data in the Boston area.

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 ijth element and the transpose of the $i-$th 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 $n \to \infty$.

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, λ_0 is the unknown scalar spatial autoregressive coefficient, and β_0 is 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]

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

Under standard stability conditions, the model in (2.1) can be re-written 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 variance $\sigma_i^2 > 0$. In addition, for some $\delta > 0$,*

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

Let $E(\epsilon\epsilon') = \Omega_0 > 0$. As is common practice in the spatial literature, restrictions on the parameter space and the asymptotic behaviour 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 $i = 1, ..., n$.

(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 $i, j = 1, ..., n$, $W_{ij} = O(1/h)$, 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 non singular. 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 to develop asymptotic theory. These conditions (or similar ones) are required to ensure existence of the reduced form in (2.2) 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 $\text{rank}(X) = k$ *for all sufficiently large* n. *In addition,* $\text{rank}(G'M_XG) \sim n$ *as* $n \to \infty$.

The latter condition rules out the case in which the columns of G and X are perfectly collinear in the limit.

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

$$l(\lambda) = K - \frac{1}{2} ln(y'S(\lambda)'M_X S(\lambda)y) + \frac{1}{n} ln|S(\lambda)| \tag{2.3}$$

and let

$$\hat{\lambda}_{QML} = \underset{\lambda \in \Lambda}{\text{argmax}}\ l(\lambda). \tag{2.4}$$

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

$$p \lim_{n \to \infty} \frac{1}{n} l^{(1)}(\lambda_0) = 0. \tag{2.5}$$

This condition is generally 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 (2.5) and hence a required condition for consistency of $\hat{\lambda}_{QML}$ is

$$\frac{1}{n} \sum_{i=1}^{n} \left(G_{ii} - \frac{1}{n} trG \right) \left(\sigma_i^2 - \frac{1}{n} \sum_{j=1}^{n} \sigma_j^2 \right) \to 0, \tag{2.6}$$

as $n \to \infty$, where G_{ii} is the i'th diagonal element of G. The condition in (2.6) 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, 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.

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

$$\Omega_0(\gamma) = \sigma^2 \begin{pmatrix} e^{z_1\gamma} & 0 & 0 & \dots & 0 \\ 0 & e^{z_2\gamma} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & e^{z_n\gamma} \\ \end{pmatrix} \tag{2.7}$$

for 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 (2.7), the LHS of (2.6) becomes

$$\sum_{t=0}^{\infty} \frac{\gamma^t}{t!} \frac{1}{n} \sum_{i=1}^{n} \left(G_{ii} - \frac{1}{n} trG \right) \left(z_i^t - \frac{1}{n} \sum_{j=1}^{n} z_j^t \right) \tag{2.8}$$

The latter expression confirms that, even in presence of a very simple form of heteroskedasticity such as that in (2.7), the condition displayed in (2.6) is difficult to check for general W and $z_1, ..., z_n$. 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 (2.6) holds. But if instead that sample covariance is constant and non-zero across t (at least for sufficiently large n) the LHS of (2.6) becomes $K(e^\gamma - 1)$, which vanishes only when $\gamma \to 0$. Simple calculations confirm that (2.8) 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

$$\sum_{t=0}^{\infty} \frac{\gamma^t}{t!} \frac{1}{n} \sum_{i=1}^{n} \left(G_{ii} - \frac{1}{n} trG \right) \left(z_i^t - \frac{1}{n} \sum_{j=1}^{n} z_j^t \right) \to_{a.s.} \mathbb{E}\{(G_{ii} - \mu_G) e^{\gamma z_i}\}, \tag{2.9}$$

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

Little has, as yet, been said about the bias of $\hat{\lambda}_{QML}$ under Assumption 1. Starting from the results in Bao (2013), we may compute the bias of $\hat{\lambda}_{QML}$ for $\Omega_0(\gamma)$ given in (2.7). For illustration, we limit our analysis to the Gaussian case, although more general results can be obtained at the expense of extra computation. To assist in the bias calculation we derive the following explicit moment expressions

$$E(l^{(1)}(\lambda_0)) = \frac{tr(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} - \frac{1}{n} tr(G) + o(1), \tag{2.10}$$

$$E(l^{(2)}(\lambda_0)) = -\frac{\beta_0' X' G' M G X \beta_0 + tr(G' G \Omega_0(\gamma))}{tr(\Omega_0(\gamma))} + \frac{2tr^2(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} - \frac{1}{n} tr(G^2) + o(1), \quad (2.11)$$

$$\begin{aligned} E(l^{(2)}(\lambda_0) l^{(1)}(\lambda_0)) = & - \frac{tr(G\Omega_0(\gamma))\left(tr(G'G\Omega_0(\gamma)) + \beta_0 X' G' M G X \beta_0\right)}{tr^2(\Omega_0(\gamma))} - \frac{1}{n} tr(G^2) \frac{tr(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} \\ & + \frac{1}{n} tr(G) \frac{tr(G'G\Omega_0(\gamma)) + \beta_0 X' G' M G X \beta_0}{tr(\Omega_0(\gamma))} - \frac{2}{n} \frac{tr^2(G\Omega_0(\gamma))}{tr^2(\Omega_0(\gamma))} + 2\frac{tr^3(G\Omega_0(\gamma))}{tr^3(\Omega_0(\gamma))} \\ & + \frac{1}{n^2} tr(G) tr(G^2) + o(1), \end{aligned} \quad (2.12)$$

$$\begin{aligned} E(l^{(3)}(\lambda_0)) = & - 6\frac{tr(G\Omega_0(\gamma))\left(\beta_0' X' G' M G X \beta_0 + tr(G'G\Omega_0(\gamma))\right)}{tr^2(\Omega_0(\gamma))} + \frac{8tr^3(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} \\ & - \frac{2}{n} tr(G^3) + o(1) \end{aligned} \quad (2.13)$$

and

$$E(l^{(1)}(\lambda_0)^2) = \frac{tr^2(G\Omega_0(\gamma))}{tr^2(\Omega_0(\gamma))} + \frac{1}{n^2} tr^2(G) - \frac{2}{n} tr(G) \frac{tr(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} + o(1). \quad (2.14)$$

Let $B(\gamma, \lambda_0) = E(\hat{\lambda}_{QML}) - \lambda_0$. From these calculations and Bao (2013), we deduce the following result.

Corollary 1 *Let ϵ be a vector of n independent random variables, normally distributed and such that $E(\epsilon\epsilon') = \Omega_0(\gamma)$, where $\Omega_0(\gamma)$ is defined in (2.7) with $\sigma^2 = 1$. Let Assumptions 2-4 hold. The leading term of $B(\gamma, \lambda_0)$ is given by*

$$\begin{aligned} B(\gamma, \lambda_0) = & - 2\left(E(l^{(2)}(\lambda_0))\right)^{-1} E(l^{(1)}(\lambda_0)) + \left(E(l^{(2)}(\lambda_0))\right)^{-2} E(l^{(2)}(\lambda_0) l^{(1)}(\lambda_0)) \\ & - \frac{1}{2}\left(E(l^{(2)}(\lambda_0))\right)^{-3} E(l^{(3)}(\lambda_0)) E(l^{(1)}(\lambda_0)^2). \end{aligned} \quad (2.15)$$

Under Assumption 1 terms in (2.10), (2.12) and (2.14) do not vanish as n increases, unless $\gamma = 0$ (i.e. the homoskedastic case) and/or some specific structure of W is imposed which ensures that a condition related to (2.6) holds. Given (2.3), the calculation of (2.10)-(2.13) is based on the explicit computation of moments of ratio of quadratic form. Most of the moments of ratios involved are indeed exactly ratio of moments, as ratios of the form $\epsilon' A\epsilon/\epsilon' M_X \epsilon$ for a generic $n \times n$ matrix A are independent of $\epsilon' M_X \epsilon$[4]. However, since we are only interested in the leading terms of (2.15), we can approximate moments of ratios as ratios of moments even when the independence conditions fails. The computation of moments is standard (Bao and Ullah (2007)) and details are omitted here.

[4]See, for example, Conniffe and Spencer (2001), for an analysis and history of this result on ratios of quadratic forms and other moments.

In Figure 1, shown in the Appendix, we report plots of $B(\gamma, \lambda_0)$ for different values of λ_0 and for four different choices[5] of W against $\gamma \in [-10, 10]$ at $n = 200$. The elements of the vector $(z_1, \ldots, z_n)'$ 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 $\lambda_0 = -0.8, 0, 0.4, 0.8$. 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

$$W_n = I_r \otimes B_m, \quad B_m = \frac{1}{m-1}(l_m l_m' - I_m), \tag{2.16}$$

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 (2.16). Similarly, the plot in the top right of Figure 1 reports $B(\gamma, \lambda_0)$ when W is chosen as a circulant with two neighbours 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 bottom left of Figure 1 is the bias function when W is randomly generated as an $n \times n$ 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 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 $i-$th location along the interval $[0, n]$, which is randomly generated from $Unif[0, n]$. 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 "random" and "exponential" are then rescaled by their respective spectral norm, and they tend to be much more relevant to empirical work than other choices such as (2.16) or circulant matrices, as they mimic contiguity-based weight matrices. Under either "random" or "exponential", under Assumption 1, 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.

Figures shown 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). Also, the bias tend to be generally more severe for "exponential" W, as shown in the plot in the bottom right of Figure 1. As expected from (2.10)-(2.14), the leading terms of (2.15) 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 $\hat{\lambda}_{QML}$).

[5] In figure (7) we depict the structure of the four choices of weight matrices used in the sequel, to illustrate the degree of sparseness and/or symmetry.

3 Continuously Updated Indirect Inference based on OLS estimates

As discussed in Section 2, 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 behaviour of W are satisfied. Specifically, as shown in Lee (2002), the OLS estimator of λ_0 is consistent as long as the sequence h defined in Assumption 3(iv) diverges, and it is asymptotically normal if $\sqrt{n}/h = o(1)$ as $n \to \infty$. 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 neighbours 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.

The limit conditions on h that justify consistency of OLS 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 behaviour of h. So, reasons for using OLS are commonly and justifiably ignored. On the other hand, OLS can be used as the simple building block of the new technique developed in the present paper, taking advantage of its computational simplicity. Our methodology 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" as

$$\hat{\beta}(\lambda) = (X'X)^{-1}X'S(\lambda)y, \tag{3.1}$$

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

$$\hat{\lambda} = \frac{y'W'M_X y}{y'W'M_X W y}. \tag{3.2}$$

Similar to the discussion in KPR, we can obtain a formal expansion for the expected value of the latter ratio based on Lieberman's (1992) result as

$$E(\hat{\lambda}) = \frac{E(y'W'M_X y)}{E(y'W'M_X W y)} + O\left(\frac{1}{n}\right). \tag{3.3}$$

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

$$E(\hat{\lambda}) = \frac{tr(P\Omega_0) + \beta_0' X' P X \beta_0}{tr(Q'Q\Omega_0) + \beta_0' X' Q' Q X \beta_0} + O\left(\frac{1}{n}\right). \tag{3.4}$$

Following the formal expansion in (3.4), we define the binding function $\tau_n(\lambda)$ as

$$\tau_n(\lambda, \Omega_\lambda, \hat{\beta}(\lambda)) = \tau_n(\lambda) = \frac{tr(P(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)' X' P(\lambda) X \hat{\beta}(\lambda)}{tr(Q(\lambda)'Q(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)' X' Q(\lambda)' Q(\lambda) X \hat{\beta}(\lambda)} + O_p\left(\frac{1}{n}\right), \tag{3.5}$$

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

$$\tau_n^*(\lambda, \Omega_\lambda, \hat{\beta}(\lambda)) = \tau_n^*(\lambda) = \frac{tr(P(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)' X' P(\lambda) X \hat{\beta}(\lambda)}{tr(Q(\lambda)'Q(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)' X' Q(\lambda)' Q(\lambda) X \hat{\beta}(\lambda)}, \tag{3.6}$$

with $\hat{\beta}(\lambda)$ defined according to (3.1), and

$$\Omega_\lambda = \text{diag}(\epsilon(\lambda)\epsilon(\lambda)'), \quad \epsilon(\lambda) = (y - \lambda W y - X\hat{\beta}(\lambda)), \tag{3.7}$$

where diag(A) for a generic $n \times n$ matrix A returns an $n \times n$ diagonal matrix containing only the main diagonal of A and whose other entries are zero.

Our new continuously updated indirect inference (CUII) estimator of λ, $\hat{\lambda}_{CUII}$ is defined as

$$\hat{\lambda}_{CUII} = \underset{\lambda \, \in \, \Lambda}{\text{argmin}} \{\hat{\lambda} - \tau_n(\lambda, \Omega_\lambda, \hat{\beta}(\lambda))\}^2. \tag{3.8}$$

A detained discussion on the robustness advantages of using $\tau_n^*(.)$ rather than its standard simulated version (e.g. Gouriéroux et al. (2000)) is contained in KPR. In this setting, the standard II approach of simulating pseudo-data to construct the binding function would require even more structure compared to standard estimation problems under homoskedasticity, as not only distributional assumptions, but also the specification of the form of heteroskedasticity would be necessary. From substantial numerical work the objective function in (3.8) 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 imposed to establish the limit theory, we introduce the following conditions.

Assumption 6

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

(ii) $p \lim_{n\to\infty} \tau_n^{(1)}(\lambda_0)$ *exists and 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 Assumption 6. Under Assumption 6, we have the inverse function representation of the CUII estimator

$$\hat{\lambda}_{CUII} = \tau_n^{-1}(\hat{\lambda}). \tag{3.9}$$

4 Limit Theory

This section derives the asymptotic properties of the estimator (3.8) for model (2.1) when the case $\beta_0 = 0$ *a priori* is ruled out. From (3.5) and (3.6) we consider the centring random sequence

$$\begin{aligned} \tau_n(\lambda_0) &= \frac{tr(G'M_X S^{-1}\Omega_{\lambda_0})/n + \hat{\beta}(\lambda_0)'X'S^{-1\prime}W'M_X S^{-1}X\hat{\beta}(\lambda_0)/n}{tr(G'M_X G\Omega_{\lambda_0})/n + \hat{\beta}(\lambda_0)'X'G'M_X GX\hat{\beta}(\lambda_0)/n} + O\left(\frac{1}{n}\right) \\ &= \frac{tr(P\Omega_{\lambda_0})/n + \hat{\beta}(\lambda_0)'X'PX\hat{\beta}(\lambda_0)/n}{tr(Q'Q\Omega_{\lambda_0})/n + \hat{\beta}(\lambda_0)'X'Q'QX\hat{\beta}(\lambda_0)/n} + O\left(\frac{1}{n}\right), \end{aligned} \tag{4.1}$$

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

$$\begin{aligned} V_n = & \frac{4}{n}\begin{pmatrix} \beta_0'X'P'M_X\Omega_0 M_X PX\beta_0 & \beta_0'X'P'M_X\Omega_0 M_X Q'QX\beta_0 \\ \beta_0'X'Q'QM_X\Omega_0 M_X PX\beta_0 & \beta_0'X'Q'QM_X\Omega_0 M_X Q'QX\beta_0 \end{pmatrix} \\ & + \frac{4}{n}\sum_i\sum_{j<i}\sigma_i^2\sigma_j^2 \begin{pmatrix} \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix}. \end{aligned} \tag{4.4}$$

In order to assure the existence of each limit appearing in (4.4) we impose

Assumption 7 *As* $n \to \infty$, $\lim V_n$ *exists.*

Also, define the limits

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

and

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

so that

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

whose existence and non-singularity is assured by Assumption 6.

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

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

We derive

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

which is defined in terms of limits appearing also in $\bar{\tau}^{(1)}$. Thus, its existence and non singularity is assured under Assumption 6.

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

Theorem 1

(a) *Under (2.1) with* $\beta_0 \neq 0$ *and Assumptions 1-7*

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

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

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

where $v_{CUII}^2 = \bar{f}' \lim_{n\rightarrow\infty} V_n \bar{f} / \left(\bar{\tau}^{(1)}\right)^2$ *exists and is non zero 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 replaced by $\hat{\epsilon}_i^2$, for $i = 1, \ldots.., n$.

Theorem 2 *Let Assumption 1 hold, with* $\delta = 2$. *Under Assumptions 2-7, as* $n \rightarrow \infty$

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

The proofs of Theorems 1 and 2 are given in the Appendix.

Estimation of β_0 in 2.1 under Assumption 1 is generally less problematic than estimation of λ_0 as simple OLS produces consistent estimates under general limit behaviour of W. Nonetheless, from (3.1) we can deduce consistency of $\hat{\beta}_{CUII}$ and its asymptotic normality by using the representation

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

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), (4.9) and (4.10)

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

where

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

and

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

Defining

$$\bar{\zeta} = \underset{n\to\infty}{\text{plim}} \zeta_n, \tag{4.20}$$

and

$$T_n = \frac{1}{n} \begin{pmatrix} X'X & X'\Omega_0 M_X (P+P') X\beta_0 & 2X'\Omega_0 M_X Q' Q X\beta_0 \\ \beta_0' X'(P+P') M_X \Omega_0 X & 4\beta_0' X' P' M_X \Omega_0 M_X P X\beta_0 & 4\beta_0' X' P' M_X \Omega_0 M_X Q' Q X\beta_0 \\ 2\beta_0' X' Q' Q M_X \Omega_0 X & 4\beta_0' X' Q' Q M_X \Omega_0 M_X P X\beta_0 & 4\beta_0' X' Q' Q M_X \Omega_0 M_X Q' Q X\beta_0 \end{pmatrix}$$

$$+\frac{4}{n}\sum_i \sum_{j<i} \sigma_i^2\sigma_j^2 \begin{pmatrix} 0_{(k\times k)} & 0_{(k\times 1)} & 0_{(k\times 1)} \\ 0_{(1\times k)} & \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ 0_{(1\times k)} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix}, \tag{4.21}$$

we deduce the following result.

Corollary 2 *Under (2.1) with $\beta_0 \neq 0$, under Assumptions 1-7*

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

as $n \to \infty$.

The variance-covariance matrix $\bar{\zeta}' \lim_{n\to\infty} T_n \bar{\zeta}$ exists and is non-singular under Assumptions 5-7. The proof of Corollary 2 follows in a similar way to that of part (a) of Theorem 1 and is omitted.

5 Simulations

We conduct a set of Monte Carlo experiments to compare the finite sample performance of the CUII estimators with the standard QML and OLS estimators, as well as to the robust GMM (RGMM, henceforth) of Lin and Lee (2010) and to the modified QML (MQML, henceforth) of Liu and Yang (2015).

We consider different scenarios for our simulation exercise, and for each design the number of exogenous regressors is set at $k = 3$, with X_1 being the intercept and the other two drawn randomly from two independent uniform distributions on the support $[0, 1]$. The regressors are then kept fixed across replications. In each scenario, we set $\beta_0 = (0.3, 0.5, -0.5)$, 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 $i = 1, ..., n$, as

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

where $\zeta_i \sim i.i.d.\ \mathcal{N}(0, 1)$ and σ_i is either constructed as

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

where the constant c is set to $c = 1$ and d_i denotes the numbers of neighbours of each unit i, or σ_i is drawn from a χ^2 distribution with 5 degrees of freedom and kept fixed across simulations and across different scenarios. The heteroskedasticity design in (5.2) is in line with the simulation work in Kelejian and Prucha (2007, 2010) and is motivated by spatial situations in which heteroskedasticity arises as units across different regions may have different number of neighbours.

We consider two different choices for W, already introduced in Section 2 and reported here for the reader's convenience. The first one, denoted as "random" in the sequel and in the Tables, is randomly generated as an $n \times n$ matrix of zeros and ones and then re-scaled with its spectral norm, while the number of "ones" is restricted to 20% of the total number of elements in W. This choice is empirically motivated as it mimics a dense contiguity matrix. The second choice, denoted as "exponential" in the sequel, 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. Both choices of W are generated once for each sample size and are kept fixed across different scenarios and across the 1000 Monte Carlo replications. We stress that for both these choices of W the MLE/QML is not expected to return consistent estimators in the presence of unknown heteroskedasticity, as the condition in (2.6) is not met.

In each table we report bias and mean square error (MSE) for the OLS, ML, RGMM[6], CUII and MQML estimators of λ for $n = 30, 50, 100, 200$. Tables 1 and 2 report results for "random" W and σ_i in (5.1) generated as (5.2) and as χ^2_5, respectively, while Tables 3 and 4 report corresponding results for "exponential" W.

[6]The RGMM estimator corresponds to what Liu and Lee (2010) denote as optimal RGMM, and it is constructed using the same algorithm described in Liu and Lee (2010).

[Tables 1-4 about here]

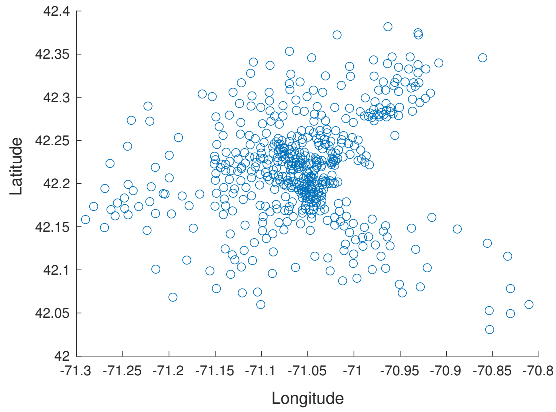
As expected, since our choices of W do not satisfy the limit condition for OLS consistency, the OLS results are severely biased for all sample sizes and across all scenarios. A similar comment holds for the ML estimate of λ, which displays severe bias that does not improve as n increases in the 'random' W case, whereas for the 'exponential' W case, the bias of the ML estimate is severe for small n but appears to decrease with sample size.

The performance of $\hat{\lambda}_{CUII}$ has to be compared to its counterparts that are robust to heteroskedasticity. Across different scenarios, we can see that bias and MSE of $\hat{\lambda}_{CUII}$ is comparable to results obtained by MQML, and outperforms MQML in terms of MSE for most values of λ_0 and n for the 'random' W case. Moreover, from a computational perspective, CUII appears easier to implement and is much less sensitive to the (arbitrary) starting value of the optimization routine. Tables 1 and 2 reveal poor performance of RGMM, which suffers from severe bias even for moderate sample sizes. This behaviour is probably due to the fact that 'random' W is a dense choice of weighting matrix, where standard GMM types of procedures do not perform well in small-moderate samples. The bias appears to decrease with n, confirming the asymptotic properties of RGMM in presence of heteroskedastic errors. In the 'exponential' W case, the performance of RGMM is poor for small sample sizes and improves for $n = 100$ but its finite sample performance remains inferior to that of CUII and MQML.

6 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 therefore more realistic of typical implementations than standard Monte Carlo designs.

Specifically, we use the Boston house price data (Harrison and Rubinfeld (1978)) and its 'corrected' version (Gilley and Pace, 1996), 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 the figure 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 socio-economic variables[7]:

crim	per capita crime rate by town;
zn	proportion of residential land zoned for lots over 25,000 sq.ft;
indus	proportion of non-retail 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 centres;
rad	index of accessibility to radial highways;
tax	full-value property-tax rate per $10,000$$;
ptratio	pupil-teacher ratio by town;
black	$1000 * (Bk - 0.63)^2$, 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

$$log(MEDV)_i = \alpha + \lambda \sum_{i \neq j} w_{ij} ln(MEDV)_j + x_i^{'}\beta + \epsilon_i \quad i = 1, ..., 506, \tag{6.1}$$

where the covariate vector x_i contains rm_i^2, age_i, $log(dis)_i$, $log(rad)_i$, tax_i, $ptratio_i$, $black_i$,

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

$log(stat)_i, crim_i, zn_i, indus_i, chas_i, nox_i^2$. 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 (6.1) and thus on estimation and significance of λ, rather than the covariate coefficient vector β. We recall that estimation of β in the general model (2.1) and in the specific model (6.1) poses fewer consistency and efficiency issues compared to inference on λ. Empirical results for the estimates of the β coefficients in the various weight matrix scenarios can be obtained from the authors.

We conjecture that several measure of proximity might play a role in the house price determination process of $MEDV_i$, so that 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 (6.1) reflect geographical proximity and rely on the geo-distance between tract 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}: $w_{ij} = 1/geo_{ij}$;
- $W^{geo,exp}$: $w_{ij} = exp\,(-|geo_{ij}|)\,\mathbb{1}(|geo_{ij}| < log(n))$;
- $W^{geo,0.9}$: $w_{ij} = \mathbb{1}(|geo_{ij}| < D^*)$, where we set $D^* = 2.5\ km$ to obtain a matrix sparseness 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 neighbouring properties, where 'neighbour' 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 it 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 socio-economic 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.

		QML	CUII	RGMM	MQML
W^{geo}	λ	0.0399	0.0426	0.0442	0.0391
	t-ratio	1.2543	14.6806	1.2595	2.6462
$W^{exp,geo}$	λ	0.0546	0.0665	0.0657	0.0563
	t-ratio	2.7419	7.3380	3.9306	11.4364
$W^{geo,0.9}$	λ	0.0130	0.0131	0.0127	0.0132
	t-ratio	2.4790	44.7070	3.0448	72.5195
W^{tax}	λ	0.0217	0.0230	0.0143	0.0235
	t-ratio	1.2368	12.3797	1.0886	12.7909
W^{school}	λ	-0.0269	-0.0268	-0.0271	-0.0268
	t-ratio	-2.9089	-91.3553	-2.3550	-17.9526

Table 5: Estimates and t-statistic of λ in (6.1) computed by QML, CUII, RGMM and MQML for different choices of weighting structures.

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 QML, CUII, MQML and RGMM, the main exception being that the estimates obtained by CUII and RGMM are slightly larger than those obtained by QML and MQML when the weight matrix W^{geo} is adopted in (6.1). 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 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 by RGMM are not significant for the weighting structures W^{tax} and W^{geo}, a result that contrasts sharply with the robust estimators.

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 figures in Table 5, and we generate a $n \times B$ matrix of i.i.d random variables from the Rademacher distribution, with typical element indicated by r_{ij} for $i = 1, ...n$ and $j = 1, ...B$. For each estimator N and each choice of W we then generate B sets of pseudo-data vectors as

$$y^*_{N,j} = S^{-1}(\hat{\lambda}_N)\left(X\hat{\beta}_N + u^*_{N,j}\right), \quad j = 1, ..., B, \tag{6.2}$$

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 $j = 1, ..., B$, so that a measure of bias and MSE can be computed as

$$Bias_N = \frac{1}{B}\sum_{j=1}^{B}\hat{\lambda}^*_{N,j} - \hat{\lambda}_N, \quad MSE_N = Bias^2_N + \frac{1}{B}\sum_{j=1}^{B}\left(\hat{\lambda}^*_{N,j} - \frac{1}{B}\sum_{j=1}^{B}\hat{\lambda}^*_{N,j}\right)^2. \tag{6.3}$$

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 $MSE_N/\hat{\lambda}^2_N$.

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

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

Results in Table 6 confirm that $\hat{\lambda}_{CUII}$ has a better performance in terms of bias and/or MSE than its robust counterparts RGMM and MQML for several choices of W. 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 behaviour 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 confirms 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.

7 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 indirect inference transformation of standard OLS estimation that parametrizes 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 converge quickly even when an artificially dense W matrix is designed. 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 $\psi_{ij} = (\ \psi_{1ij} \quad \psi_{2ij}\ )' = (\ (P+P')_{ij}/2 \quad (Q'Q)_{ij}\ )'$ and $\tilde{\psi}_{ij} = (\ \tilde{\psi}_{1ij} \quad \tilde{\psi}_{2ij}\ )' = (\ (M_X P)_{ij} \quad (M_X Q'Q)_{ij}\ )'$, 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

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

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

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

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

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

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 rest of the proof is similar to KPR (2017). In order to avoid repetition we refer to their proof when steps follow in a similar way. Define

$$u_i = (\ u_{1i} \quad u_{2i}\ )' = 2\epsilon_i\sum_j\tilde{\psi}_{ij}X_j'\beta_0 + 2\epsilon_i\sum_{j<i}\psi_{ij}\epsilon_j, \tag{A.8}$$

so that $U_n = \sum_{i=1}^n u_i + o_p(1)$, according to (A.1). The $\{u_i, 1 \leq i \leq n, n = 1, 2, ...\}$ form a triangular array of martingale differences with respect to the filtration formed by the σ-field generated by $\{\epsilon_j; j < i\}$. Let

$$A = Var\left(\sum_{i=1}^n u_i\right) = 4\sum_{i=1}^n\sigma_i^2\sum_{j=1}^n\sum_{t=1}^n\tilde{\psi}_{ij}X_j'\beta_0\beta_0'X_t\tilde{\psi}_{it}' + 4\sum_{i=1}^n\sum_{j<i}\sigma_i^2\sigma_j^2\psi_{ij}\psi_{ij}'. \tag{A.9}$$

Define $z_{in} = \eta' A^{-1/2}u_i$, where η is a 2×1 vector satisfying $\eta'\eta = 1$. By Theorem 2 of Scott (1973) $\sum_i^n z_{in} \to_d \mathcal{N}(0,1)$ if the following stability and Lindeberg conditions hold:

$$\sum_{i=1}^n\mathbb{E}(z_{in}^2|\epsilon_j; j<i) \xrightarrow{p} 1, \tag{A.10}$$

and

$$\sum_{i=1}^{n}\mathbb{E}(z_{in}^2 1(|z_{in} > \xi|)) \to 0 \quad \forall \xi > 0. \tag{A.11}$$

As $n \to \infty$,

$$A/n \to \lim_{n\to\infty} V_n, \tag{A.12}$$

where

$$\begin{aligned}V_n = \; & \frac{4}{n}\begin{pmatrix} \beta_0' X' P' M_X \Omega_0 M_X P X \beta_0 & \beta_0' X' P' M_X \Omega_0 M_X Q' Q X \beta_0 \\ \beta_0' X' Q' Q M_X \Omega_0 M_X P X \beta_0 & \beta_0' X' Q' Q M_X \Omega_0 M_X Q' Q X \beta_0 \end{pmatrix} \\ & + \frac{4}{n}\sum_i \sum_{j<i} \sigma_i^2 \sigma_j^2 \begin{pmatrix} \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix} \\ & = C_1 + C_2, . \end{aligned} \tag{A.13}$$

where C_1 and C_2 contain the first and second terms in (A.13), respectively. All terms in C_1 are $O(1)$, while those in C_2 are bounded by $O(1/h)$ under Assumptions 3 and 4, and by standard algebra. Existence of limits in (A.13) is guaranteed under Assumption 7, and non singularity of C_1 is ensured by Assumptions 2, 3(ii) and 5. The expression in (4.4) is obtained from (A.13) after routine calculations. Thus, we can replace A by n when showing (A.10) and (A.11).

We start by establishing (A.10), which can equivalently be written as

$$\sum_i E\left(z_{in}^2 | \epsilon_j, j < i\right) - \eta' A^{-1/2} A A^{-1/2} \eta \underset{p}{\to} 0. \tag{A.14}$$

The latter, by standard manipulations and (A.12), is equivalent to showing

$$\frac{4}{n}\eta' \left(\sum_i \sigma_i^2 \left(\sum_{j<i}\epsilon_j \psi_{ij}\right)\left(\sum_{j<i}\epsilon_j \psi_{ij}\right)' - \sum_i \sum_{j<i} \sigma_i^2 \sigma_j^2 \psi_{ij}\psi_{ij}'\right)\eta \underset{p}{\to} 0, \tag{A.15}$$

and

$$\frac{4}{n}\eta' \left(\sum_i \sum_j \sum_{t<i} \sigma_i^2 \beta_0' X_j \left(\tilde{\psi}_{ij}\psi_{it}' + \psi_{it}\tilde{\psi}_{ij}'\right)\epsilon_t\right)\eta \underset{p}{\to} 0 \tag{A.16}$$

as $n \to \infty$.

In order to avoid replications, we omit the proof of (A.15), referring to KPR and observing that

$$||P||_\infty + ||P'||_\infty < K, \quad ||Q||_\infty + ||Q'||_\infty < \infty \tag{A.17}$$

and both P_{ij} and Q_{ij}, for $i,j=1,....,n$, are uniformly bounded by $O(1/h)$, so that ψ_{1ij} and ψ_{2ij} have, respectively, similar asymptotic properties to $(G+G')_{ij}/2$ and $(G'G)_{ij}$ appearing in the proof of Theorem 1 in KPR. We verify (A.16) by examining the convergence of each typical element, i.e. by showing

$$\frac{1}{n}\sum_i\sum_j\sum_{t<i}\sigma_i^2\beta_0'X_j\tilde{\psi}_{sij}\psi_{vit}\epsilon_t \underset{p}{\rightarrow} 0 \tag{A.18}$$

for each $s,v=1,2$. Under Assumption 5, i.e. for uniformly bounded X_{ij} for $i,j=1,....,n$, the LHS of (A.18) has mean zero and variance bounded by

$$\frac{1}{n^2}K|\sum_i\sum_j\sum_u\sum_h\sum_{t<i,u}\tilde{\psi}_{sij}\tilde{\psi}_{suh}\psi_{vit}\psi_{vut}| \leq \frac{1}{n^2}K\sum_i\sum_j\sum_u\sum_h\sum_t|\tilde{\psi}_{sij}\tilde{\psi}_{suh}\psi_{vit}\psi_{vut}|$$
$$\frac{1}{n}K\sup_{0<i\leq n}\sum_j|\tilde{\psi}_{sij}|\sup_{0<u\leq n}\sum_h|\tilde{\psi}_{shu}|\sup_{0<t\leq n}\sum_i|\psi_{vit}|\sup_{0<u\leq n}\sum_t|\psi_{vut}| = O\left(\frac{1}{n}\right), \tag{A.19}$$

since (A.17) holds and

$$||M_XP||_\infty + ||P'M_X||_\infty < K, \quad ||M_XQ'Q||_\infty + ||Q'QM_X||_\infty < \infty \tag{A.20}$$

In order to prove (A.11) we verify the sufficient Lyapunov condition

$$\sum_{i=1}^n E|z_{in}|^{2+\delta} \rightarrow 0 \tag{A.21}$$

by considering a typical standardized element of u_i, i.e. $\sum_i E|(1/n)^{1/2}u_{si}|^{2+\delta}$ for $s=1,2$. Under Assumption 1, using $\sum_i E|u_{si}|^{2+\delta}=\sum_i E(E|u_{si}|^{2+\delta}|\epsilon_j,j<i))$ and the c_r inequality,

$$\left(\frac{1}{n}\right)^{1+\delta/2}\sum_i E|u_{si}|^{2+\delta} \leq \left(\frac{1}{n}\right)^{1+\delta/2}K\sum_i E|\sum_{j<i}\psi_{sij}\epsilon_j|^{2+\delta} + \left(\frac{1}{n}\right)^{1+\delta/2}K\sum_i|\sum_j\beta_0'X_j\tilde{\psi}_{sij}|^{2+\delta}. \tag{A.22}$$

Convergence to zero of the first term at the RHS of (A.22) can be shown as in KPR. Convergence of the third term at the RHS of (A.22) can be shown after observing that

$$|\sum_j\beta_0'X_j\tilde{\psi}_{sij}|^{2+\delta} \leq K\sup_{0<j\leq n}|\beta_0'X_j|^{2+\delta}\sum_j|\tilde{\psi}_{sij}|^{2+\delta}, \tag{A.23}$$

where $\beta_0'X_j$ is uniformly bounded under Assumption 5. Thus, the second term at the RHS of (A.22) is bounded by

$$\left(\frac{1}{n}\right)^{1+\delta/2} K\sum_i\sum_j|\tilde{\psi}_{sij}|^{2+\delta} \le \left(\frac{1}{n}\right)^{1+\delta/2} K\sum_i\left(\sum_j\tilde{\psi}_{sij}^2\right)^{1+\delta/2}$$
$$\le \left(\frac{1}{n}\right)^{1+\delta/2} K\left(\sup_i\sum_j\tilde{\psi}_{sij}\right)^{\delta/2}\sum_i\sum_j\tilde{\psi}_{sij}^2 = O\left(\frac{1}{n}\right)^{\delta/2} \tag{A.24}$$

similarly to KPR, under Assumptions 3-5.

Thus, $A^{-1/2}\sum_i u_i \underset{d}{\to} \mathcal{N}(0,I)$, and the statement in Theorem 1(i) follows by standard delta arguments.

Proof of part (ii). Again, we proceed similarly to KPR and we refer to their proof to avoid repetitions. We rewrite the binding function $\tau_n(\lambda)$ as

$$\tau_n(\lambda,\Omega_\lambda,\hat{\beta}(\lambda)) = \frac{tr(P(\lambda)\Omega_\lambda)+\hat{\beta}(\lambda)'X'P(\lambda)X\hat{\beta}(\lambda)}{tr(Q(\lambda)'Q(\lambda)\Omega_\lambda)+\hat{\beta}(\lambda)'X'Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda)} + O_p\left(\frac{1}{n}\right)$$
$$= \frac{a(\lambda)+b(\lambda)}{c(\lambda)+d(\lambda)} + O_p\left(\frac{1}{n}\right), \tag{A.25}$$

where

$$a(\lambda) = \frac{1}{n}tr(P(\lambda)\Omega_\lambda),\quad b(\lambda) = \frac{1}{n}\hat{\beta}(\lambda)'X'P(\lambda)X\hat{\beta}(\lambda),\quad c(\lambda) = \frac{1}{n}tr(Q(\lambda)'Q(\lambda)\Omega_\lambda),$$
$$d = \frac{1}{n}\hat{\beta}(\lambda)'X'Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda). \tag{A.26}$$

We write

$$\tau_n^{(1)}(\lambda) = \frac{a^{(1)}(\lambda)+b^{(1)}(\lambda)}{c(\lambda)+d(\lambda)} - \frac{(c^{(1)}(\lambda)+d^{(1)}(\lambda))(a(\lambda)+b(\lambda))}{(c(\lambda)+d(\lambda))^2} + O\left(\frac{1}{n}\right), \tag{A.27}$$

where

$$a^{(1)}(\lambda) = \frac{1}{n}tr(G'(\lambda)P(\lambda)\Omega_\lambda) + \frac{1}{n}tr(P(\lambda)G(\lambda)\Omega_\lambda) + \frac{1}{n}tr(P\Omega_\lambda^{(1)}),$$
$$b^{(1)}(\lambda) = -\frac{2}{n}y'W'(I_n-M_X)P(\lambda)X\hat{\beta}(\lambda) + \frac{1}{n}\hat{\beta}(\lambda)'X'G(\lambda)'P(\lambda)X\hat{\beta}(\lambda) + \frac{1}{n}\hat{\beta}(\lambda)'X'P(\lambda)G(\lambda)X\hat{\beta}(\lambda),$$
$$c^{(1)}(\lambda) = \frac{2}{n}tr(G(\lambda)'Q(\lambda)'Q(\lambda)\Omega_\lambda) + \frac{1}{n}tr(Q(\lambda)'M_XQ(\lambda)\Omega_\lambda^{(1)}),$$
$$d^{(1)}(\lambda) = -\frac{2}{n}y'W'(I-M_X)Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda) + \frac{2}{n}\hat{\beta}(\lambda)'X'G(\lambda)'Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda) \tag{A.28}$$

and

$$\Omega_{\lambda}^{(1)} = -2\text{diag}(M_X W y \epsilon(\lambda)'). \tag{A.29}$$

Since

$$\hat{\lambda}_{CUII} - \lambda_0 = \tau_n^{-1}(\hat{\lambda}) - \tau_n^{-1}(\tau_n(\lambda_0)), \tag{A.30}$$

we can derive the limit distribution of $\sqrt{n}(\hat{\lambda}_{CUII} - \lambda_0)$ by the delta method, as long as the asymptotic local relative equicontinuity condition (Phillips, 2012) holds. Thus, similar to KPR, we need to show

$$\left| \frac{\tau_n^{(1)}(\lambda_0) - \tau_n^{(1)}(r)}{\tau_n^{(1)}(r)} \right| \underset{p}{\rightarrow} 0 \tag{A.31}$$

as $n \rightarrow \infty$, uniformly in $\mathcal{N}_\delta = \{r \in \Re : |s(r - \lambda_0)| < \delta, \quad \delta > 0\}$, $s = s_n \rightarrow \infty$ and $s(1/n)^{1/2} \rightarrow 0$. Under Assumption 6(ii), the expression on the LHS of (A.31) is bounded by

$$K \left| \tau_n^{(1)}(\lambda_0) - \tau_n^{(1)}(r) \right|, \tag{A.32}$$

which by the mean value theorem is in turn bounded by

$$K \left| \tau_n^{(2)}(\lambda^*)(\lambda_0 - r) \right|, \tag{A.33}$$

where λ^* is an intermediate point between λ_0 and r. The expression in (A.33) is $O_p(|\lambda_0 - r|) = O_p(s^{-1})$ as long as

$$\tau_n^{(2)}(\lambda^*) = O_p(1), \tag{A.34}$$

which holds under Assumptions 3-5, a derivation of which will be supplied on request.

Therefore, by a delta argument we conclude that

$$\sqrt{n}\tau_n^{(1)}(\hat{\lambda}_{CUII} - \lambda_0) \underset{d}{\rightarrow} \mathcal{N}(0, \bar{f}' \lim_{n\rightarrow\infty} V_n \bar{f}), \tag{A.35}$$

where V_n and $\bar{f}_n$ are defined in (4.4) and (4.11) respectively. The statement in Theorem 1 follows by standard algebra once we write

$$\bar{\tau}^{(1)} = \bar{\tau}^{(1)}(\lambda_0) = \text{p} \lim_{n\rightarrow\infty} \tau_n^{(1)}(\lambda_0), \tag{A.36}$$

in terms of $\bar{a}^{(1)}$, $\bar{b}^{(1)}$, $\bar{c}^{(1)}$ and $\bar{d}^{(1)}$. $\bar{\tau}$ exists and is non singular under Assumption 7(ii).

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

$$\frac{1}{n}\sum_i \sum_{j<i} \left(\hat{\epsilon}_i^2 \hat{\epsilon}_j^2 \hat{\psi}_{sij}\hat{\psi}_{tij} - \sigma_i^2 \sigma_j^2 \psi_{sij}\psi_{tij}\right) = o_p(1), \tag{A.37}$$

$$\frac{1}{n}\left(tr(\hat{\Omega}\hat{A}) - tr(\Omega_0 A)\right) = o_p(1) \tag{A.38}$$

and

$$\frac{1}{n}\left(\hat{\beta}'_{CUII} X' \hat{\tilde{\Psi}}'_s \hat{\Omega} \hat{\tilde{\Psi}}_t X \hat{\beta}_{CUII} - \beta_0' X' \tilde{\Psi}'_s \Omega_0 \tilde{\Psi}_t X \beta_0\right) = o_p(1), \tag{A.39}$$

where, consonant with the notation defined at the beginning of the proof of Theorem 1, $\psi_{ij} = (\ \psi_{1ij} \quad \psi_{2ij}\ )' = (\ (P+P')_{ij}/2 \quad (Q'Q)_{ij}\ )'$, $\tilde{\Psi}_1 = M_X P$, $\tilde{\Psi}_2 = M_X Q'Q$, and $\hat{\psi}_{sij}$, $\hat{\tilde{\Psi}}_t$ for $s, t = 1, 2$ are obtained by replacing the unknown λ_0 by its estimate $\hat{\lambda}_{CUII}$. Also, $\hat{A}$ 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^2_{CUII} is trivial, as it only relies on consistency of $\hat{\lambda}_{CUII}$ and $\hat{\beta}_{CUII}$.
In order to prove (A.37), we need to show

$$\frac{1}{n}\sum_i \sum_{j<i} (\epsilon_i^2 \epsilon_j^2 - \sigma_i^2 \sigma_j^2)\psi_{sij}\psi_{tij} = o_p(1), \tag{A.40}$$

$$\frac{1}{n}\sum_i \sum_{j<i} (\hat{\epsilon}_i^2 \hat{\epsilon}_j^2 - \epsilon_i^2 \epsilon_j^2)\psi_{sij}\psi_{tij} = o_p(1) \tag{A.41}$$

and

$$\frac{1}{n}\sum_i \sum_{j<i} \hat{\epsilon}_i^2 \hat{\epsilon}_j^2 (\hat{\psi}_{sij}\hat{\psi}_{tij} - \psi_{sij}\psi_{tij}) = o_p(1). \tag{A.42}$$

We start by (A.40). We have, for $s, t = 1, 2$

$$\frac{1}{n}\sum_i \sum_{j<i} (\epsilon_i^2 \epsilon_j^2 - \sigma_i^2 \sigma_j^2)\psi_{sij}\psi_{tij} = \frac{1}{n}\sum_i \sum_{j<i} \psi_{sij}\psi_{tij}(\epsilon_i^2 - \sigma_i^2)(\epsilon_j^2 - \sigma_j^2) + \frac{1}{n}\sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\sigma_i^2(\epsilon_j^2 - \sigma_j^2)$$
$$+ \frac{1}{n}\sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\sigma_j^2(\epsilon_i^2 - \sigma_i^2). \tag{A.43}$$

The first term at the RHS of (A.43) has mean zero and variance bounded by

$$\frac{C}{n^2}\sum_i \sum_{j<i} \psi_{sij}^2 \psi_{tij}^2 \le \frac{C}{n^2}\sum_i \sum_j \psi_{sij}^2 \psi_{tij}^2 \le \frac{C}{n^2 h^2}\sum_i \sum_j \psi_{tij}^2 = O\left(\frac{1}{nh^3}\right) \tag{A.44}$$

since

$$\sum_i \sum_j \psi_{tij}^2 = tr(\Psi_t^2) = O\left(\frac{n}{h}\right)$$

for $t = 1, 2$. The second term at the RHS of (A.43) has mean zero and variance bounded by

$$\begin{aligned}
&\frac{C}{n^2} \sum_i \sum_j \sum_u |\psi_{sij}\psi_{tij}\psi_{suj}\psi_{tuj}| \leq \frac{C}{n^2h^2} \sum_i \sum_j \sum_u |\psi_{sij}||\psi_{tuj}| \\
&\leq \frac{C}{nh^2} \sup_j \sum_i |\psi_{sij}| \sup_u \sum_j |\psi_{sij}| = O\left(\frac{1}{nh^2}\right). \qquad \text{(A.45)}
\end{aligned}$$

Similarly, we can show that the third term at the RHS of (A.43) converges to zero in quadratic mean. By Markov's inequality (A.40) follows.

In order to show (A.41) we write

$$\hat{\epsilon}_i = \epsilon_i - \sum_j B_{ij}\epsilon_j - (\hat{\lambda}_{CUII} - \lambda_0)Q_i' X\beta - (\hat{\lambda}_{CUII} - \lambda_0)Q_i'\epsilon, \qquad \text{(A.46)}$$

where Q_i' is the $1 \times n$ vector displaying the $i-$th row of Q and $B_{ij} = X_i'(X'X)^{-1}X_j$, as defined at the beginning of the proof of Theorem 1. By standard arguments, we can show that the last two terms on the RHS of (A.46) are bounded in probability by $1/\sqrt{n}$, uniformly in i. Let

$$\hat{v}_i = \hat{\epsilon}_i - \epsilon_i = -\sum_k B_{ik}\epsilon_k + O_p\left(\frac{1}{\sqrt{n}}\right). \qquad \text{(A.47)}$$

Thus, (A.41) is equivalent to

$$\frac{1}{n} \sum_i \sum_{j<i} \psi_{sij}\psi_{tij}(\hat{v}_i\hat{v}_j + \epsilon_i\hat{v}_j + \epsilon_j\hat{v}_i)(\hat{v}_i\hat{v}_j + \hat{v}_i\epsilon_j + \epsilon_i\hat{v}_j + 2\epsilon_i\epsilon_j) = o_p(1), \qquad \text{(A.48)}$$

as $n \to \infty$. We therefore need to show, as $n \to \infty$, that

$$\frac{1}{n} \sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\hat{v}_i^2\hat{v}_j^2 = o_p(1), \qquad \text{(A.49)}$$

$$\frac{1}{n} \sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\hat{v}_i^2\hat{v}_j\epsilon_j = o_p(1), \qquad \text{(A.50)}$$

$$\frac{1}{n} \sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\hat{v}_i\hat{v}_j\epsilon_i\epsilon_j = o_p(1), \qquad \text{(A.51)}$$

$$\frac{1}{n} \sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\hat{v}_i^2\epsilon_j^2 = o_p(1), \qquad \text{(A.52)}$$

$$\frac{1}{n}\sum_{i}\sum_{j<i}\psi_{sij}\psi_{tij}\hat{v}_j\epsilon_i^2\epsilon_j = o_p(1). \tag{A.53}$$

We only consider the leading term in $\hat{v}_i$ in (A.47) when showing (A.49)- (A.57), but similar routine arguments can be applied to deal with higher order terms.

The modulus of the LHS of (A.49) has expectation bounded by

$$\begin{aligned}
&\frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}||\psi_{tij}|E\left(\hat{v}_i^4\right)^{1/2}E\left(\hat{v}_j^4\right)^{1/2} \leq \frac{C}{n}\sum_{i}\sum_{j}|\psi_{sij}||\psi_{tij}|\left(\sum_{v}B_{iv}^2\right)\left(\sum_{h}B_{jh}^2\right)\\
&\leq \frac{C}{n}\sum_{i}\sum_{j}|\psi_{sij}||\psi_{tij}|B_{ii}B_{jj} \leq \frac{C}{nh^2}\sum_{i}\sum_{j}B_{ii}B_{jj} = O\left(\frac{1}{h^2 n}\right).
\end{aligned} \tag{A.54}$$

Similarly, the modulus of the LHS of (A.50) has expectation bounded by

$$\begin{aligned}
&\frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}||\psi_{tij}|\left(E\hat{v}_j^4\right)^{1/4}\left(E\hat{v}_i^4\right)^{1/2}\left(E\epsilon_j^4\right)^{1/4} \leq \frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}||\psi_{tij}|\left(\sum_{v}B_{jv}^2\right)^{1/2}\left(\sum_{h}B_{ih}^2\right)\\
&\leq \frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}||\psi_{tij}|B_{jj}^{1/2}B_{ii} \leq \frac{C}{nh}\sum_{i}\sum_{j}|\psi_{sij}|B_{ii} \leq \frac{C}{nh}\sup_{i}\sum_{j}|\psi_{sij}|\sum_{i}B_{ii} = O\left(\frac{1}{nh}\right),
\end{aligned} \tag{A.55}$$

as $B_{jj}^{1/2} < 1$. The modulus of the LHS of (A.51) has expectation bounded by

$$\begin{aligned}
&\frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}||\psi_{tij}|\left(E\hat{v}_i^4\right)^{1/4}\left(E\hat{v}_j^4\right)^{1/4}\left(E\epsilon_j^4\right)^{1/4}\left(E\epsilon_i^4\right)^{1/4} \leq \frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}||\psi_{tij}|B_{ii}^{1/2}B_{jj}^{1/2}\\
&\frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}||\psi_{tij}|(B_{ii}+B_{jj}) \leq \frac{C}{nh}\left(\sup_{i}\sum_{j}|\psi_{sij}|\sum_{i}B_{ii} + \sup_{j}\sum_{i}|\psi_{sij}|\sum_{j}B_{jj}\right) = O\left(\frac{1}{nh}\right).
\end{aligned} \tag{A.56}$$

(A.52) can be shown by similar arguments as (A.49)-(A.51), while (A.57) can be written as

$$\frac{1}{n}\sum_{i}\sum_{j<i}\psi_{sij}\psi_{tij}B_{ji}\epsilon_i^3\epsilon_j + \frac{1}{n}\sum_{i}\sum_{j<i}\psi_{sij}\psi_{tij}\epsilon_i^2\epsilon_j^2B_{jj} + \frac{1}{n}\sum_{i}\sum_{j<i}\sum_{u\neq j,i}\psi_{sij}\psi_{tij}\epsilon_i^2\epsilon_j\epsilon_u B_{ju} \tag{A.57}$$

The modulus of the first term in the last displayed expression has expectation bounded by

$$\frac{C}{n}\sum_{i}\sum_{j<i}|\psi_{sij}|\psi_{tij}||B_{ij}| \leq \frac{C}{n}\sum_{i}\sum_{j}|\psi_{sij}|\psi_{tij}|(B_{ii}+B_{jj}) = O\left(\frac{1}{hn}\right), \tag{A.58}$$

as in previous calculations. Similarly, the second term in (A.57) is $O(1/nh)$, while the third term has mean zero and variance bounded by

$$\frac{C}{n^2}\sum_i\sum_j\sum_u\sum_l|\psi_{sij}\psi_{tij}\psi_{sil}\psi_{til}|B_{uj}^2 + \frac{C}{n^2}\sum_i\sum_j\sum_k\sum_l|\psi_{sij}\psi_{tij}\psi_{skl}\psi_{tkl}|B_{lj}^2$$
$$\frac{C}{n^2}\sum_i\sum_j\sum_l|\psi_{sij}\psi_{tij}\psi_{sil}\psi_{til}|B_{jj} + \frac{C}{n^2}\sum_i\sum_j\sum_k\sum_l|\psi_{sij}\psi_{tij}\psi_{skl}\psi_{tkl}|B_{jl}^2. \tag{A.59}$$

Proceeding as before, the first term in the last displayed expression is bounded by $O(1/n^2h^2)$, while the second one is bounded by $O(1/nh^2)$. By Markov's inequality, this conclude the proof of (A.41).

In order to show (A.42) we apply a standard mean value theorem argument, such as

$$\frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2(\hat{\psi}_{sij}\hat{\psi}_{tij}-\psi_{sij}\psi_{tij}) = \frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2\left(\bar{\psi}_{sij}(\hat{\psi}_{tij}-\psi_{tij})+\bar{\psi}_{tij}(\hat{\psi}_{sij}-\psi_{sij})\right), \tag{A.60}$$

where $\bar{\psi}_{sij}$ (or $\bar{\psi}_{tij}$) is an intermediate point between $\hat{\psi}_{sij}$ and ψ_{sij}. From Theorem 1, $\hat{\psi}_{sij} - \psi_{sij} = O_p(1/\sqrt{n})$ and thus $\bar{\psi}_{sij} - \psi_{sij} = o_p(1)$. Therefore, (A.60) is bounded by

$$\sup_{i,j}|\hat{\psi}_{sij}-\psi_{sij}|\frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2|\psi_{tij}|. \tag{A.61}$$

By similar arguments to those applied to prove (A.40) and (A.41), we conclude that as $n\to\infty$

$$\frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2|\psi_{tij}| \underset{p}{\to} \lim\frac{1}{n}\sum_i\sum_{j<i}\sigma_i^2\sigma_j^2|\psi_{tij}|, \tag{A.62}$$

which is $O(1)$ in the limit. Thus, (A.61) is $O_p(1/\sqrt{n})$, concluding the proof of (A.37).

The proofs of (A.38) and (A.39) are omitted as they follow very similar arguments to those applied to show (A.37) and (A.2) at the beginning of the proof of Theorem 1.

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
OLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.2528	0.3940	-0.2149	0.3080	-0.2467	0.2909	-0.2278	-.2821
	0.3	-0.2883	0.4693	-0.2420	0.3128	-0.1987	0.2763	-0.2236	0.0521
	0.5	-0.2070	0.3750	-0.2005	0.2838	-0.2049	0.2579	-0.1991	0.2291
	0.8	-0.2384	0.3907	-0.1614	0.1944	-0.1060	0.1281	-0.1028	0.1216
ML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1931	0.3121	-0.1480	0.2462	-0.1730	0.2266	-0.1485	0.2153
	0.3	-0.3529	0.3595	-0.2855	0.2411	-0.2504	0.2048	-0.2645	0.1933
	0.5	-0.2997	0.2810	-0.2975	0.2193	-0.2978	0.2037	-0.2886	0.1863
	0.8	-0.3972	0.3290	-0.3137	0.1815	-0.2526	0.1215	-0.2554	0.1183
RGMM		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.2024	0.5687	-0.1525	0.3751	-0.1690	0.2897	-0.1353	0.2587
	0.3	-0.3946	0.6164	-0.2915	0.3471	-0.2491	0.2699	-0.2747	0.2706
	0.5	-0.3374	0.6602	-0.3041	0.3265	-0.2864	0.2887	-0.2970	0.2683
	0.8	-0.3912	0.5127	-0.2786	0.2453	-0.1976	0.1375	-0.2010	0.1419
CUII		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.0112	0.3925	0.0076	0.2889	-0.0099	0.2420	0.0083	0.2236
	0.3	-0.0548	0.3694	-0.0234	0.2447	0.0062	0.2327	-0.0120	0.1960
	0.5	-0.0253	0.31229	-0.0097	0.2173	-0.0149	0.2063	-0.0220	0.1843
	0.8	-0.0628	0.2782	-0.0557	0.1462	-0.0141	0.1040	-0.0120	0.1032
MQML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.0270	0.3900	0.01385	0.2857	-0.0095	0.2388	0.0101	0.2327
	0.3	-0.0177	0.3940	-0.0119	0.2517	0.0149	0.2409	-0.0092	0.2001
	0.5	-0.0124	0.3310	-0.0002	0.2363	-0.0049	0.2132	-0.0199	0.1863
	0.8	-0.0567	0.2935	-0.0520	0.1468	-0.0061	0.1098	-0.0069	0.1074

Table 3: Bias & MSE of OLS, ML, RGMM, CUII and MQML estimators for 'random' W. The ϵ_is are defined as in (5.1) with $\zeta_i \sim iidN(0,1)$ and σ_i is defined as in (5.2) (based on 1000 replications).

		$n=30$		$n=50$		$n=100$		$n=200$	
OLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1914	0.3734	-0.2153	0.3095	-0.2158	0.2902	-0.2494	0.3082
	0.3	-0.3044	0.4905	-0.2556	0.3542	-0.2890	0.3242	-0.2387	0.3198
	0.5	-0.3039	0.4640	-0.2852	0.3787	-0.2472	0.3217	-0.2557	0.3175
	0.8	-0.3011	0.3986	-0.2227	0.2966	-0.2287	0.2467	-0.1785	0.2374
ML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1339	0.2918	-0.1452	0.2433	-0.2158	0.2902	-0.1658	0.2315
	0.3	-0.3467	0.3586	-0.3042	0.2568	-0.2890	0.3242	-0.3001	0.2396
	0.5	-0.3780	0.3635	-0.3773	0.2992	-0.2472	0.3217	-0.3522	0.2520
	0.8	-0.4487	0.3551	-0.4131	0.2704	-0.2287	0.2467	-0.3815	0.2230
RGMM		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1792	0.5280	-0.1380	0.3009	-0.1153	0.2730	-0.1622	0.2606
	0.3	-0.3948	0.6056	-0.3313	0.3518	-0.3452	0.3185	-0.3507	0.3622
	0.5	-0.4586	0.6691	-0.4007	0.4150	-0.3906	0.3952	-0.4262	0.4270
	0.8	-0.4505	0.4887	-0.4321	0.4388	-0.3947	0.3868	-0.4334	0.3984
CUII		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0005	0.3945	0.0198	0.2938	0.0213	0.2566	0.0014	0.2478
	0.3	-0.0232	0.3816	-0.0555	0.2806	-0.0454	0.2365	-0.0098	0.2540
	0.5	-0.1000	0.3452	-0.0656	0.2774	-0.0350	0.2483	-0.0266	0.2462
	0.8	-0.1462	0.2789	-0.0752	0.2233	-0.0774	0.1725	-0.0159	0.1916
MQML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.0281	0.4081	-0.0382	0.2904	0.0289	0.2612	0.0069	0.2481
	0.3	-0.0065	0.4158	-0.0208	0.2878	-0.0307	0.2461	0.0007	0.2614
	0.5	-0.0409	0.3721	-0.0008	0.3018	-0.0104	0.2590	-0.0074	0.2533
	0.8	-0.0494	0.2740	-0.0213	0.2245	-0.0144	0.3296	0.0135	0.2019

Table 4: Bias & MSE of OLS, ML, RGMM, CUII and MQML estimators for 'random' W. The ϵ_is are defined as in (5.1) with $\zeta_i \sim iidN(0,1)$ and $\sigma_i \sim \chi^2(5)$ (based on 1000 replications).

		$n=30$		$n=50$		$n=100$		$n=200$	
OLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.7251	1.0790	-0.7724	1.0887	-0.6001	0.6419	-0.6557	0.7643
	0.3	-0.1313	0.4406	-0.0500	0.2654	0.1123	0.1303	0.0956	0.1415
	0.5	0.0460	0.2222	0.0836	0.1599	0.2155	0.1071	0.2372	0.1336
	0.8	0.1066	0.0555	0.1412	0.0536	0.2098	0.0585	0.2180	0.0720
ML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.2040	0.2393	-0.1887	0.1895	-0.0982	0.0934	-0.1370	0.1154
	0.3	-0.1691	0.1818	-0.1275	0.1124	-0.0355	0.0477	-0.0439	0.0556
	0.5	-0.1143	0.1096	-0.0966	0.0783	-0.0174	0.0330	-0.0041	0.0410
	0.8	-0.0738	0.0397	0.0602	0.0310	-0.0145	0.0129	-0.0154	0.0167
RGMM		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1339	0.1968	-0.1320	0.1505	-0.0402	0.0761	-0.0725	0.0841
	0.3	-0.1726	0.1755	-0.1434	0.1051	-0.0646	0.0466	-0.0775	0.0539
	0.5	-0.1473	0.1133	-0.1344	0.0895	-0.0633	0.0384	-0.0618	0.0438
	0.8	-0.0920	0.0662	-0.0955	0.0483	-0.0600	0.0221	-0.0708	0.0382
CUII		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1164	0.2587	-0.1037	0.1638	-0.0210	0.0754	-0.0446	0.0815
	0.3	-0.0749	0.1725	-0.0672	0.1031	-0.0205	0.0431	-0.0406	0.0487
	0.5	-0.0493	0.1221	-0.0407	0.0824	-0.0075	0.0391	-0.0055	0.0493
	0.8	0.0207	0.0617	0.0058	0.0488	0.0239	0.0305	0.0530	0.0351
MQML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0618	0.1781	-0.0724	0.1392	-0.0095	-0.0702	-0.0330	0.0715
	0.3	-0.0892	0.1303	-0.0807	0.0827	-0.0282	0.0391	-0.0487	0.0443
	0.5	-0.0958	0.0900	-0.0762	0.0622	-0.0321	-0.0294	-0.0390	0.0350
	0.8	-0.0743	0.0355	-0.0602	0.o0299	-0.0391	0.0146	-0.0606	0.0211

Table 5: Bias & MSE of OLS, ML, RGMM, CUII and MQML estimators for 'exponential' W. The ϵ_is are defined as in (5.1) with $\zeta_i \sim iidN(0,1)$ and σ_i is defined as in (5.2) (based on 1000 replications).

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
OLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.8157	1.1707	-0.7638	1.0488	-0.6501	0.7101	-0.6185	0.6505
	0.3	-0.0456	0.3867	0.0277	0.2661	0.1165	0.1430	0.1540	0.1768
	0.5	0.0210	0.2823	0.1534	0.1760	0.2548	0.1376	0.2997	0.1807
	0.8	0.0956	0.1034	0.1775	0.0628	0.2572	0.0844	0.3172	0.1374
ML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1530	0.1639	-0.0865	0.1112	-0.0525	0.0706	-0.0282	0.0590
	0.3	-0.1397	0.1195	- 0.1044	0.0877	-0.0717	0.0435	-0.0484	0.0478
	0.5	-0.1641	0.1273	-0.1171	0.0654	-0.0507	0.0338	-0.0562	0.0328
	0.8	-0.1526	0.0742	-0.0571	0.0287	-0.0466	0.0164	-0.0618	0.0208
RGMM		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1866	0.1934	-0.0975	0.1210	-0.0574	0.0727	-0.0344	0.0616
	0.3	-0.1537	0.1366	-0.1160	0.0957	-0.0777	0.0466	-0.0573	0.0514
	0.5	-0.1639	0.1342	-0.1070	0.0713	-0.0496	0.0363	-04435	0.0395
	0.8	-0.1078	0.0969	-0.0468	0.0516	-0.0243	0.0584	-0.0325	0.0278
CUII		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1252	0.2551	-0.0882	0.1459	-0.0383	0.0747	-0.0193	0.0619
	0.3	-0.0561	0.1604	-0.0305	0.0993	-0.0231	0.0466	-0.0192	0.0534
	0.5	-0.0202	0.1453	-0.0077	0.0876	-0.0049	0.0410	0.0049	0.0478
	0.8	0.0265	0.0879	0.0750	0.0311	0.1139	0.0183	0.0899	0.0143
MQML		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0422	0.1630	-0.0508	0.1169	-0.062	0.0698	-0.0133	0.0600
	0.3	-0.0807	0.1110	-0.0514	0.0751	-0.0347	0.0402	-0.0314	0.0327
	0.5	-0.0810	0.0923	-0.0601	0.0548	-0.0276	0.0316	-0.0249	0.0327
	0.8	-0.0831	0.0472	-0.0529	0.0260	-0.0239	0.0144	-0.0344	0.0178

Table 6: Bias & MSE of OLS, ML, RGMM, CUII and MQML estimators for 'exponential' W. The ϵ_is are defined as in (5.1) with $\zeta_i \sim iidN(0,1)$ and and $\sigma_i \sim \chi^2(5)$ (based on 1000 replications).

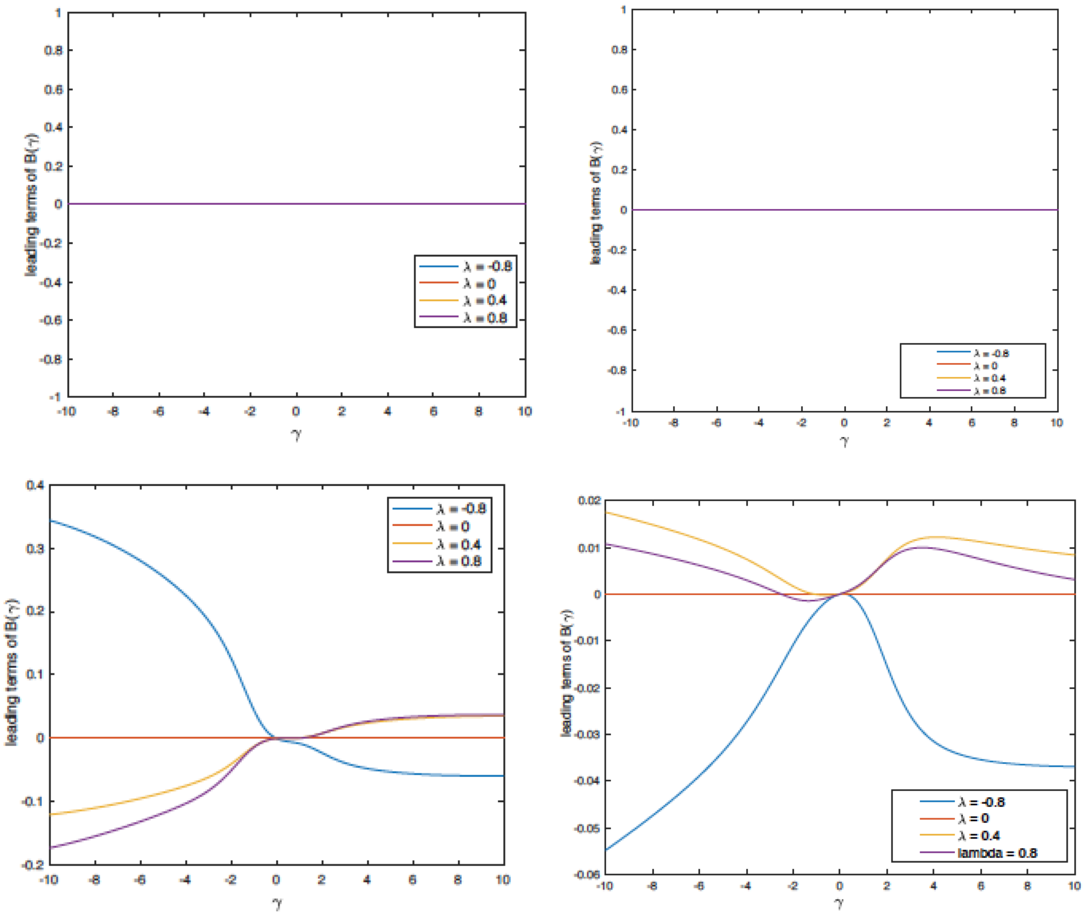


Figure 1: $B(\gamma)$ for various weight matrix designs at $n = 200$. Top: (L) block diagonal, (R) circulant, two ahead-two behind; bottom: (L) 'exponential', (R) 'random'.

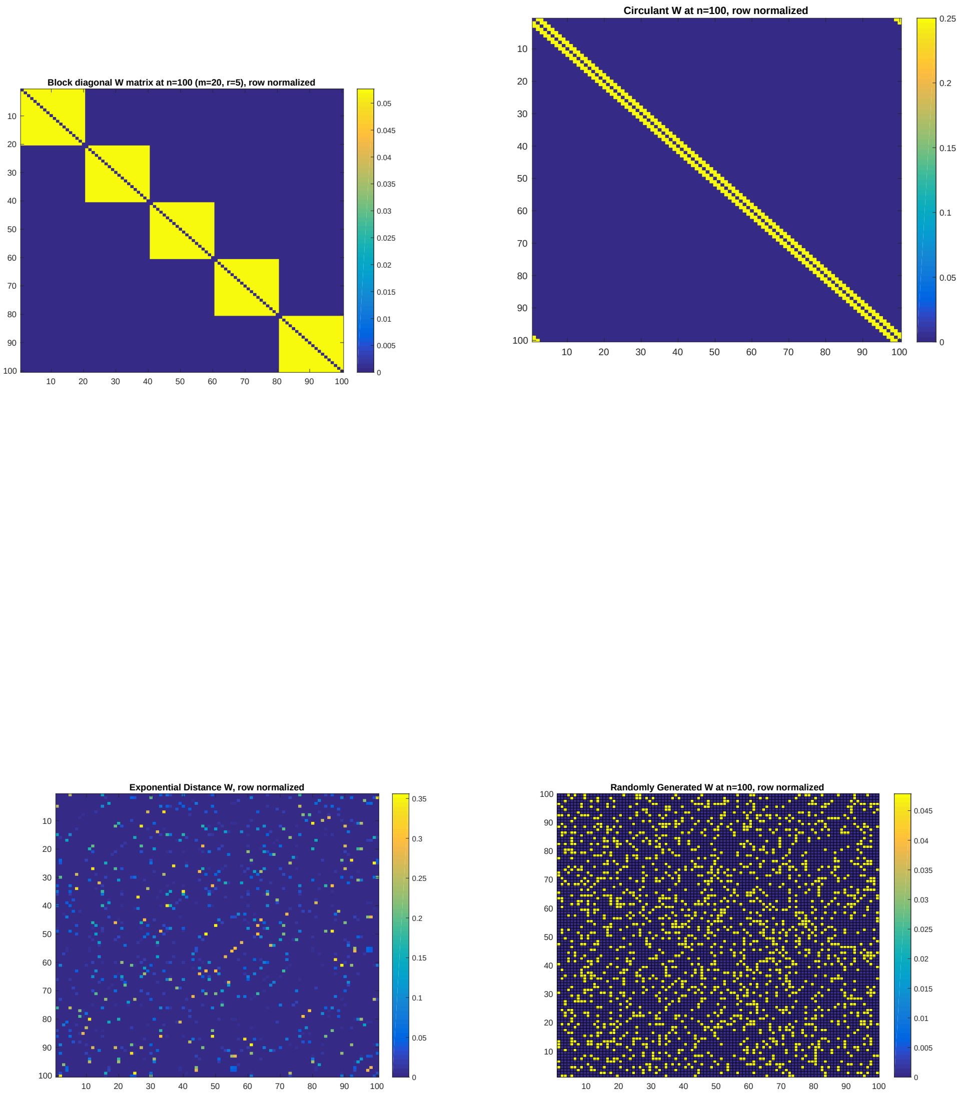


Figure 2: Weight Matrix structures. Top: (L) block diagonal W; (R) circulant, two ahead-two behind ; Bottom: (L) ‘exponential’, (R) ‘random’. $n = 100$

References

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

Bao, Y. and Ullah, A., 2007. Finite sample moments of maximum likelihood estimator in spatial models. Journal of Econometrics 137, 396-413.

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

Case, T., 1991. Spatial patterns in household demand. Econometrica. 59, 953-965.

Conniffe, D. and Spencer J.E., 2001. When moments of ratios are ratios of moments. The Statistician 50(2), 161-168.

Durbin, J. 1988. Maximum likelihood estimation of the parameters of a system of simultaneous regression equations. Econometric Theory 4, 159-170. (Paper presented at the Econometric Society Meetings in Copenhagen, 1963)

Gilley, O. W. and Pace, R. K., 1996. On the Harrison and Rubinfeld data. Journal of Environmental and Economic Management 31, 403-405.

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

Gouriéroux, C., Renault, E. andTouzi, N., 2000. Calibration by simulation for small sample bias correction. In: Mariano, R.S., Schuermann, T., Weeks, M. (Eds.), *Simulation-Based Inference in Econometrics: Methods and Applications.* Cambridge University Press, 328-58.

Hansen, L. P., Heaton, J. and Yaron A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics 14.3, 262-280.

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

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

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

Kelejian H. H. and Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53-67.

Kyriacou, M, Phillips P.C.B. and Rossi F., 2017. Indirect inference in spatial autoregressions. The Econometrics Journal 20, 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., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899-1925.

LeSage, J. P., 1999. The theory and practice of spatial econometrics. University of Toledo. Toledo, Ohio 28, 33.

Lieberman, O., 1994. A Laplace approximation to the moments of a ratio of quadratic forms. Biometrika 81, 681-90.

Lin, X. and Lee, L. F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157, 34-52.

Liu, S. F. and Yang, Z., 2015 Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. Regional Science and Urban Economics 52, 50-70.

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, 17-30.