Statistics and Computing (2025) 35:154
https://doi.org/10.1007/s11222-025-10688-6

ORIGINAL PAPER

Spatial autoregressive model for interval-valued data and applications

Jinjin Zhang[1] · Kun Li[1] · Aibing Ji[1]

Received: 27 March 2025 / Accepted: 15 July 2025 / Published online: 24 July 2025

Abstract
Interval-valued data, characterized by intrinsic measurement imprecision, uncertainty, and variability, are common in real-world applications. This study introduces a novel spatial autoregressive model tailored for interval-valued data, unifying and generalizing several existing frameworks. To address the limitations of interval representations, we develop a joint quasi-maximum likelihood estimation method that holistically incorporates complete interval information through both center and radius parameters. Crucially, we introduce a novel L_2-type distance metric to quantify interval variance, which systematically captures richer intra-interval information compared to classical Euclidean interval distance metric. The asymptotic properties of the estimators under regularity conditions are established, ensuring statistical robustness. Numerical experiments on synthetic datasets demonstrate the superiority of the proposed method over conventional approaches in prediction accuracy and information retention. Empirical validation on real spatial interval datasets-urban house price domain-confirms the efficiency of the parameter estimation framework and the operational viability of the proposed model.

Keywords Spatial autoregressive model · Interval-valued data · Joint quasi-maximum likelihood estimation

1 Introduction

In statistics and econometrics, it is not uncommon for observations to be recorded as interval-valued data rather than as single point-valued data. One reason is that imprecise observations of quantities result in the measured values being transformed into an interval of possible values, and interval-valued data capture the radius of possible values and uncertainties, providing a more comprehensive description. Another reason is that the resulting classifications of observations invariably involve intervals when observations in large data sets are aggregated into smaller and more manageable data sizes (Billard and Diday 2000). Interval-valued data fully represent the complexity and variability of the real world.

With interval-valued data becoming increasingly significant in statistics and econometrics, Moore (1979) introduces interval operations: interval addition, subtraction, product, scalar multiplication, and division. The most common interval-valued regression methods are established using the centre (CM) (Billard and Diday 2000), the upper and lower bounds (Minmax) (Billard and Diday 2002), and the centre and radius (CRM) (Billard and Diday 2002) of interval-valued variables. The CRM method incorporates the richer utilization of interval data, often outperforming Minmax (Sinova and Aelst 2018; Li et al. 2025). Souza et al. (2017) later proposed a parametric method (PM) to enhance the Minmax method. The linear correlation between the lower and upper bounds of intervals, both serving as predictor variables in PM, may introduce multicollinearity issues. With the advancement of networks, the collection of interval-valued data has become more convenient. Therefore, it is important to develop more methods for analyzing interval-valued data.

Spatial interval-valued data, integrating interval-valued information with spatial dependencies, have become increasingly prevalent in real-world applications. This reflects an essential characteristic: variations in phenomena within specific geographical regions inevitably induce changes in adjacent areas. Such spatial interdependencies are commonly observed across diverse domains including environmental monitoring (air quality), financial markets (stock fluctuations), and real estate (housing price dynamics). Existing research has focused on median estimation (Sinova and Aelst 2018), fuzzy clustering (D'Urso et al. 2023), and auto-

✉ Kun Li
ericehihi@163.com

✉ Aibing Ji
jab@hbu.edu.cn

1 College of Mathematics and Information Science, Hebei University, Baoding 071002, China

correlation index analysis (Freitas et al. 2022) for spatial interval-valued data. Recent methodological advancements include linear autoregressive models: Minmax-based and CRM-based spatial error models (ISER) (Freitas et al. 2024), CRM-based spatial Durbin models with t-distribution assumptions (Huang 2024), and PM-based fixed effects spatial interval-valued panel models (Li et al. 2025). These developments establish a robust analytical framework for handling inherent uncertainty and spatial variability in complex datasets, enhancing capabilities for spatial pattern recognition and decision support. Later, (Li et al. 2025) proposed time-varying spatial panel models.

The increasing availability of spatially correlated interval-valued data underscores the importance of spatial interval analytical methods for informed decision-making. This paper proposes a novel spatial autoregressive model for interval-valued data (ISAR). The CRM-based method are introduced, avoiding collinearity issues as PM-based method (Li et al. 2025). To estimate the coefficients, in the traditional SAR literature, the QMLE has been widely studied since Lee (2004). We follow the tradition QMLE, so that the extreme estimators obey good asymptotic properties. The direct QMLE approach is first proposed, where the spatial interval-valued data for centre and radius are independent of each other is assumed. When the relationship between centre and radius is not independent, we discussed that it might have a poor performance. To address the loss of some interval information, a L_2 type D_k metric (briefly, D_k metric) with respect to the support function is adopted to represent the variance of interval-valued data. The D_k metric proposed by Näther (2006) is a generalized distance measure of d_w metric, measuring distances between two intervals, as discussed in Li et al. (2023); Sun et al. (2018, 2019) and Han et al. (2016). A joint QMLE approach based on the L_2 type D_k metric is then proposed, resulting in estimates with all interval information. One anticipate expect the estimates from the direct QMLE approach are special cases of the joint QMLE approach based on the D_k metric is demonstrated.

Compared with the existing interval-valued models literature, our proposed approach has a number of appealing features. First, we extend conventional constant spatial interval-valued error (ISE) model proposed by Freitas et al. (2024) to spatial autoregressive model, achieving higher predictive accuracy. In particular, we propose a novel ISAR model approach by adding spatial endogenous lags of response variables rather than independence between individuals. The ISAR binary spatial weights selected by our method are allowed to describe the variability of the latitude and longitude coordinates inside of districts, which is consistent with spatial relationships. Second, we also consider the full interval information in terms of the joint QMLE method with weighted matrices and add the relationships between centers and ranges according to the definition of covariances of errors, avoiding a potentially loss of interval information. Third, we propose three ISAR models, fusing the SAR model with the CRM, Minmax and PM methods. As a result, the CRM, Minmax and PM methods can be seen as a special case.

The paper is organized as follows. Section 2 establishes a spatial autoregressive (SAR) model for interval-valued data and develops the joint quasi-maximum likelihood estimation (QMLE) methodology using the D_k metric. Section 3 investigates the asymptotic properties of the proposed D_k-based joint QMLE approach. Section 4 demonstrates that several existing models emerge as special cases of our proposed model. Section 5 presents Monte Carlo simulations evaluating the finite-sample performance of our estimator and compares it with alternatives to direct QMLE and previous methods. Section 6 applies the proposed methodology to analyze Shanghai housing price dynamics. Concluding remarks appear in Section 7. Proofs of theorems are contained in Appendix.

2 Model setting and estimation

2.1 Model

Let $Y_n = (y_{1n}, y_{2n}, \cdots, y_{nn})'$ be the $n \times 1$ vector consisting of one observation on the dependent variable for units in the sample, all y_{in} may potentially be statistically correlated, $X_n = (X_{1n}, X_{2n}, \cdots, X_{nn})'$ be the $n \times p$ matrix of interval exogenous explanatory variables with $X_{in} = (x_{i1n}, x_{i2n}, \cdots, x_{ipn})'$ for $i = 1, 2, \cdots, n$. The classical SAR for single point-valued data is defined by

$$Y_n = \lambda W_n Y_n + X_n \beta + V_n \tag{1}$$

where n is the total number of spatial units, λ is the SAR coefficient, W_n is a nonnegative $n \times n$ spatial weight matrix describing the spatial configuration or arrangement of the units in the samples, $\lambda w_i Y_n$ represents the spillover effect of neighboring spatial units' behavior on spatial unit i, $\beta = (\beta_1, \beta_2, \cdots, \beta_p)'$ is a $p \times 1$ coefficient vector, V_n is an $n \times 1$ vector of disturbance terms, in which each element is assumed to be independently and identically distributed with zero mean and variance σ^2.

For a SAR model for interval-valued data, the dependent vector Y_n and independent matrix X_n are observed in interval form, respectively, i.e., $y_{in}^I = [y_{l,in}, y_{u,in}]$ and $x_{ijn}^I = [x_{l,ijn}, x_{u,ijn}]$ for $i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, p$. The subscripts l and u denote the lower and upper bounds of intervals. Generally, for interval-valued data, the lower bound is smaller than the upper bound. By taking the difference between the lower and upper bounds of Eq.(1), we obtain the following SAR for the radius of Y_n:

$$Y_{r,n} = \lambda W_n Y_{r,n} + X_{r,n}\beta + V_{r,n} \quad (2)$$

where $Y_{r,n} = (Y_{u,n} - Y_{l,n})/2$ and $X_{r,n} = (X_{u,n} - X_{l,n})/2$, $V_{r,n}$ denotes the radius of V_n, where $V_n = [V_{l,n}, V_{u,n}]$, $V_{l,n}$ denotes an $n \times 1$ vector of disturbance terms from $Y_{l,n}$ and $V_{u,n}$ denotes an $n \times 1$ vector of disturbance terms from $Y_{u,n}$, $V_{r,n} = (V_{u,n} - V_{l,n})/2$. Similarly, we can obtain the following SAR for the center of Y_n:

$$Y_{c,n} = \lambda W_n Y_{c,n} + X_{c,n}\beta + V_{c,n} \quad (3)$$

where $Y_{c,n} = (Y_{u,n} + Y_{l,n})/2$ and $X_{c,n} = (X_{u,n} + X_{l,n})/2$. $V_{c,n}$ denotes the center of V_n, $V_{c,n} = (V_{u,n} + V_{l,n})/2$.

The idea of established model (Eqs.(2) and (3)) is found in literature Han et al. (2012) and Han et al. (2016): (i) The full interval information is converted into center and radius; (ii) The established SAR for interval-valued data (Eqs.(2) and (3)) are used same unknown parameters λ, W_n, β, and σ^2. The implication of (ii) is that the endogenous interaction effects among $Y_{r,n}$ is same as the endogenous interaction effects among $Y_{c,n}$, and the effect of $X_{r,n}$ on $Y_{r,n}$ is the same as the effect of $X_{c,n}$ on $Y_{c,n}$. Eqs.(2) and (3) can be considered a reasonable and valid approximation when dealing with the specific issue of linear interval-valued data and spatial correlated interval response variables. From another prospective, ISAR model is a generalization of the interval-valued regression model (CRM) to cross-sectional spatial interval-valued data. However, there exists a drawback for models (2) and (3). For example, the highest house price in a city may influence the highest house price in a neighboring city, but this influence may not be the same as the influence of the lowest house price in that city on the lowest price in the neighboring city.

To relax these restrictions, we impose the coefficients of the center and radius regressions (Eqs.(2) and (3)) to be different, which is similar to the interval-valued regression models, see Lima Neto and de Carvalho (2008); Lima Neto and de Carvalho (2017) and Giordani (2015). The ISAR model takes the following form:

$$Y_{r,n} = \lambda_r W_{r,n} Y_{r,n} + X_{r,n}\beta_r + V_{r,n} \quad (4)$$
$$Y_{c,n} = \lambda_c W_{c,n} Y_{c,n} + X_{c,n}\beta_c + V_{c,n} \quad (5)$$

where the coefficient λ_r (λ_c) measures the strength of dependence between $Y_{r,in}$ and $Y_{r,jn}$ (between $Y_{c,in}$ and $Y_{c,jn}$) for $i, j = 1, 2, \cdots, n$, with a value of zero indicating independence between $Y_{r,in}$ and $Y_{r,jn}$ (between $Y_{c,in}$ and $Y_{c,jn}$). $\beta_r = (\beta_{r,1}, \beta_{r,2}, \cdots, \beta_{r,p})'$ and $\beta_c = (\beta_{c,1}, \beta_{c,2}, \cdots, \beta_{c,p})'$ are two $p \times 1$ coefficient vectors. $V_{r,n}$ and $V_{c,n}$ are two $n \times 1$ vector of disturbance terms. Each elements in $V_{r,n}$ and $V_{c,n}$ are assumed to be independently and identically distributed with zero mean and variance σ_r^2 and σ_c^2, i.e., $V_{r,n} \sim (0, \Sigma_{r,n})$ and $V_{c,n} \sim (0, \Sigma_{c,n})$, with $\Sigma_{r,n} = \sigma_r^2 I_n$ and $\Sigma_{c,n} = \sigma_r^2 I_n$.

Note that a single point-valued data $a \in \mathbb{R}$ can be viewed as a special case of interval-valued data, i.e., $a = [a, a]$. Thus, when the observations of Y_n and X_n are the single point-valued data, the proposed models (4) and (5) degenerate into a classical SAR model (1), where $Y_{r,n}$ and $X_{r,n}$ to be zero, while the center regression model (5) is working.

Let $Y_{b,n} = (Y'_{r,n}, Y'_{c,n})'$. $V_{b,n}$ are similarly defined as $Y_{b,n}$, $\beta_b = (\beta'_r, \beta'_c)'$, $X_{b,n}$ and $W_{b,n}$ are the block diagonal matrices, $X_{b,n} = \operatorname{diag}\{X_{r,n}, X_{c,n}\}$, $W_{b,n} = \operatorname{diag}\{W_{r,n}, W_{c,n}\}$ and $\lambda_b = \operatorname{diag}\{\lambda_r, \lambda_c\}$. Then, models (4) and (5) are equivalent to the following bivariate model by converting the interval to a center and radius vector.

$$Y_{b,n} = \lambda_b \otimes I_n W_{b,n} Y_{b,n} + X_{b,n}\beta_b + V_{b,n} \quad (6)$$

where $'\otimes'$ denotes the Kronecker product and I_n is the identity matrix of order n. Define $S_{b,n}(\lambda_b) = I_{2n} - \lambda_b \otimes I_n W_{b,n}$ for any λ_b. At the true parameters $\lambda_{b,0}$, $S_{b,n} = S_{b,n}(\lambda_{b,0})$ Then, presuming $S_{b,n}$ is invertible, the bivariate model (6) can be rewritten as

$$Y_{b,n} = S_{b,n}^{-1} X_{b,n}\beta_b + S_{b,n}^{-1} V_{b,n} \quad (7)$$

where $V_{b,n} \sim (0, \Sigma_b)$ with $\Sigma_b = \sigma_b^2 \otimes I_n$, where σ_b^2 is 2×2 matrix whose diagonal elements are σ_r^2 and σ_c^2 and off-diagonal elements are zero. The goal of the proposed models is to construct consistent estimator for unknown parameters: the spatial coefficient λ_b, the coefficient vector β_b, and the variance σ_b^2.

2.2 Joint quasi-maximum likelihood estimation

To establish the maximum likelihood function of SAR model for interval-valued data, the variance of center and radius regression models, $\Sigma_{b,n}$, can be used. The quasi log-likelihood function of models (4) and (5) is:

$$\begin{aligned} \ln L_{b,n}(\theta_b) = &-n\ln(2\pi) - \frac{1}{2}\ln|\Sigma_{b,n}| \\ &+ \ln|S_{b,n}(\lambda_b)| - \frac{1}{2}V'_{b,n}(\delta_b)\Sigma_{b,n}^{-1}V_{b,n}(\delta_b) \end{aligned} \quad (8)$$

where $\theta_b = (\theta'_r, \theta'_c)'$, $\theta_r = (\beta'_r, \lambda_r, \sigma_r^2)'$ and $\theta_c = (\beta'_c, \lambda_c, \sigma_c^2)'$. $\delta_b = (\delta_r, \delta_c)$ where $\delta_r = (\beta_r, \lambda_r)$ and $\delta_c = (\beta_c, \lambda_c)$. Because the off-diagonal elements of $\Sigma_{b,n}$, $W_{b,n}$, $S_{b,n}$ and are zero matrices, Eq.(8) cen be separated into two parts and each parts includes different unknown parameters θ_c and θ_r, i.e.,

$$\ln L_{b,n}(\theta_r) = \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_r^2 + \ln|S_{r,n}|$$

$$
- \frac{1}{2\sigma_r^2} V'_{r,n}(\delta_r) V_{r,n}(\delta_r) \tag{9}
$$

$$
\ln L_{b,n}(\theta_c) = \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_c^2 + \ln|S_{c,n}| - \frac{1}{2\sigma_c^2} V'_{c,n}(\delta_c) V_{c,n}(\delta_c) \tag{10}
$$

Eqs.(9) and (10) are the single point-valued quasi log-likelihood functions of models (4) and (5), respectively.

However, this treatments not considers the correlation between center and radius. The essence of this result lies in the last term of Eq.(8). Ignoring variance $\Sigma_{b,n}^{-1}$, $V'_{b,n}V_{b,n}$ is the Euclidean distance for point-valued data ($Y_{r,n}$ and $\lambda_r W_{r,n} Y_{r,n} + X_{r,n}\beta_r$, and $Y_{c,n}$ and $\lambda_c W_{c,n} Y_{c,n} + X_{c,n}\beta_c$), which does not reflect the interval-valued nature. Without creating correlation coefficient between center and radius, the D_k distance function for intervals is introduced as a substitute for the Euclidean distance between single point-valued data in quasi log-likelihood function (8).

Denote by $\mathcal{K}_{\mathcal{C}}(\mathbb{R})$ or $\mathcal{K}$ the collection of all non-empty bounded closed intervals in $\mathbb{R}$. The D_k metric

$$
D_k^2(A, B) = \sum_{(u,v)\in S^0\times S^0} (s_A(u) - s_B(u))(s_A(v) - s_B(v))K(u, v) \tag{11}
$$

where K is a symmetric positive definite kernel function on unit space $S^0 = \{u \in \mathbb{R}, |u| = 1\} = \{1, -1\}$, $s_A(u)$ is a support function of the interval A, i.e., $s_A(u) = \sup_{a\in A}\langle u, a\rangle$, $u \in \mathbb{R}$. $s_B(u)$, $s_A(v)$, and $s_B(v)$ are similarly defined as $s_A(u)$. It can be equivalently represented by the lower and upper bounds as

$$
D_k^2(A, B) = K(1, 1)(A_l - B_l)^2 + K(-1, -1)(A_u - B_u)^2 - (A_l - B_l)(A_u - B_u)(K(1,-1) + K(-1, 1))
$$

or equivalently by the center and radius as

$$
D_k^2(A, B) = A_{11}(A_r - B_r)^2 + A_{22}(A_u - B_u)^2 + 2A_{12}(A_l - B_l)(A_u - B_u)
$$

where

$$
\begin{aligned}
A_{11} &= K(1, 1) + K(-1, -1) - (K(1, -1) + K(-1, 1))\\
A_{22} &= K(1, 1) + K(-1, -1) + (K(1, -1) + K(-1, 1))\\
A_{12} &= A_{21} = K(1, 1) - K(-1, -1)
\end{aligned}
$$

Then, the joint quasi log-likelihood function of (6) is

$$
\ln L_{b,n}(\theta_b) = -n\ln(2\pi) - \frac{1}{2}\ln|\Sigma_{b,n}| + \ln|S_{b,n}(\lambda_b)| - \frac{1}{2} V'_{b,n}(\delta_b)\mathbf{L}'_{2n}\Sigma_{b,n}^{-1}\mathbf{L}_{2n} V_{b,n}(\delta_b) \tag{12}
$$

where $\mathbf{L}_{2n}$ is composed of eigenvalues and eigenvectors of $\mathbf{K}_{2n}$, $\mathbf{K}_{2n}$ is a $2n \times 2n$ symmetric kernel matrix in the form as

$$
\mathbf{K}_{2n} = \begin{pmatrix} A_{11}I_n & A_{12}I_n \\ A_{21}I_n & A_{22}I_n \end{pmatrix}
$$

$\theta_b = (\beta'_b, \lambda'_b, \Sigma_{b,n})$, and $\Sigma_{b,n}$ is the variance for bivariate disturbance terms (interval-valued data), i.e., $Var([V_{l,n}, V_{u,n}]) = E(V'_{b,n}\mathbf{K}_{2,n}V_{b,n}) = \Sigma_{b,n}$. However, in general, it is a difficult task that setting up all elements of σ_b^2 are unknown. Thus, similarly to Xu and Qin (2023), we suppose that the covariance matrix is partial unknown, i.e., $\Sigma_{b,n} = \Sigma_{2n}\sigma^2$, where the matrix Σ_{2n} is known but the scale σ is unknown. Thus, the joint quasi log-likelihood function (12) is equivalent to

$$
\ln L_{b,n}(\theta_b) = -n\ln(2\pi) - \frac{1}{2}\ln|\Sigma_{2n}| - n\ln\sigma^2 + \ln|S_{b,n}(\lambda_b)| - \frac{1}{2\sigma^2} V'_{b,n}(\delta_b)\mathcal{W}_{2n} V_{b,n}(\delta_b) \tag{13}
$$

where $\mathcal{W}_{2n} = \mathbf{L}'_{2n}\Sigma_{2n}\mathbf{L}_{2n}$.

Note that, ignoring the variance Σ_{2n}, numerically, although the last term on the right-hand side of Eq.(12)

$$
\begin{aligned}
&V_{b,n}(\delta_b)\mathbf{L}'_{2n}\mathbf{L}_{2n}V_{b,n}(\delta_b) = V'_{b,n}\mathbf{K}_{2n}V_{b,n}\\
&= A_{11}V'_{r,n}(\delta_r)V_{r,n}(\delta_r) + A_{22}V'_{c,n}(\delta_c)V_{c,n}(\delta_c)\\
&+ (A_{12} + A_{21})V'_{r,n}(\delta_r)V_{c,n}(\delta_c)
\end{aligned} \tag{14}
$$

is an alternate representation of the distance between intervals $Y_{b,n}$ and $X_{b,n}\beta_b$ in terms of the center and range information of intervals, it is the D_k distance between intervals Y_n and $X_n\beta$. D_k distance is in essence an integral over the distances between all pairs of points in intervals $Y_{b,n}$ and $X_{b,n}\beta_b$ by the choice of the kernel function K. In other words, the joint quasi log-likelihood function consider the full interval information to fit the SAR models (4) and (5) using the D_k distance between intervals $Y_{b,n}$ and $X_{b,n}\beta_b$. D_k distance proposed by Körner (1997) and Körner and Näther (2002) is measure two sets of arbitrary dimension, which includes intervals as a special case. Recently, it is widely used in regression for fuzzy random data and interval-valued time series, see Näther (2006); Trutschnig et al. (2009); Bertoluzza et al. (1995); Li et al. (2023); Sinova et al. (2014); Han et al. (2012); Sun et al. (2018, 2019); Han et al. (2016), but not in a SAR for interval-valued data. Based on the operation of the interval bounds with the center and radius of the interval, Eq.(14) is equivalent to

$$
\begin{aligned}
&V_{b,n}\mathbf{L}'_{2n}\mathbf{L}_{2n}V_{b,n} = V'_{b,n}\mathbf{K}_{2n}V_{b,n}\\
&= K(1, 1)V'_{l,n}V_{l,n} + K(-1, -1)V'_{u,n}V_{u,n}
\end{aligned}
$$

$$- (K(1,-1) + K(-1,1))V'_{l,n}V_{u,n}$$

where $V_{l,n}$ and $V_{u,n}$ are similarly defined as $V_{r,n}$ and $V_{c,n}$ but using the lower and upper bounds of interval variables. The joint QMLE $\hat{\theta}_b$ is the extremum estimator derived from the maximization of (12). From the log-likelihood function (12), given λ, the joint QMLE of β_b is

$$\hat{\beta}(\lambda) = (X'_{b,n}\mathcal{W}_{2n}X_{b,n})^{-1}X'_{b,n}\mathcal{W}_{2n}S_{b,n}(\lambda_b)Y_{b,n} \tag{15}$$

and the joint QMLE $\hat{\sigma}^2$ of σ^2 is

$$\begin{aligned}\hat{\sigma}^2(\lambda_b) = \frac{1}{2n}&\left[S_{b,n}(\lambda_b)Y_{b,n} - X_{b,n}\hat{\beta}_b(\lambda_b)\right]'\mathcal{W}_{2n}\\ &\left[S_{b,n}(\lambda_b)Y_{b,n} - X_{b,n}\hat{\beta}_b(\lambda_b)\right]\\ &= \frac{1}{2n}Y'_{b,n}S'_{b,n}(\lambda_b)M_{b,n}S_{b,n}(\lambda_b)Y_{b,n}\end{aligned}$$

where $M_{b,n} = (I_{2n} - R_{b,n}))'\mathcal{W}_{2n}(I_{2n} - R_{b,n})$ with $R_{b,n} = X_{b,n}(X'_{b,n}\mathcal{W}_{2n}X_{b,n})^{-1}X'_{b,n}\mathcal{W}_{2n}$. The concentrated log-likelihood function of λ_b is

$$\begin{aligned}&\ln L_{b,n}(\lambda_b) = -n(\ln(2\pi) + 1) - \frac{1}{2}\ln|\Sigma_{2n}|\\ &- n\ln\hat{\sigma}_b^2(\lambda_b) + \ln|S_{b,n}(\lambda_b)|\end{aligned} \tag{16}$$

The joint QMLE $\hat{\lambda}_b$ of λ_b maximizes the concentrated log-likelihood (16). The QMLEs of β_b and σ^2 are, respectively, $\hat{\beta}_{b,n}(\hat{\lambda}_b)$ and $\hat{\sigma}^2(\hat{\lambda}_b)$.

Remark 1 When $K(1,-1) = K(-1,1) = 0$ or $A_{12} = A_{21} = 0$, the joint quasi log-likelihood function (12) degenerates two separated quasi log-likelihood functions (9) and (10) which do not consider the correlation between center and radius,i.e., the direct QMLE approach is a spacial case of joint QMLE approach.

3 Assumptions and asymptotic properties

To provide a rigious analysis of the QMLE, we make some regularity conditions.

Assumption 1 The disturbances $\{v_{b,in}\}$ in $V_{b,n} = (v_{b,1n}, v_{b,2n}, \cdots, v_{b,un})$, for all $i = 1, 2, \cdots, n$, are i.i.d. across all i and t. The odd-order moments $E([V_{l,n}, V_{u,n}]^s) = E(v^s_{b,in}|X_{b,in}) = 0$ where s is an arbitrary infinity odd. The second moments $E([V_{l,n}, V_{u,n}]^2) = E(v'_{b,in}k_{2n}v_{b,in}|X_{b,in}, X_{c,in}) = \sigma^2$ and the even-order moment$E([V_{l,n}, V_{u,n}]^t) = E((v'_{b,in}k_{2n}v_{b,in})^t|X_{b,in}, X_{c,in}) = \mu_t$ where t is an arbitrary infinity enve number greater than 2. For some $\gamma_b > 0$, $\mathrm{E}(v^{4+\gamma_r}_{b,in}|X_{b,in})$ exists.

Assumption 2 The elements $\{w_{r,ijn}\}$ of $W_{r,n}$ for $i, j = 1, 2, \cdots, n$ are of most of order $h^{-1}_{r,n}$, denoted by $O(1/h_{r,n})$, uniformly in all i and j, where the rate sequence $\{h_{r,n}\}$ can be bounded or divergent. As a normalization, $w_{r,ijn} = 0$ for all i. The properties of $W_{c,n}$ is assumed as that of $W_{r,n}$.

Assumption 3 The ratios $h_{r,n}/h, h_{c,n}/h \to 0$, as n goes to infinity.

Assumption 4 The matrices $S_{r,n}$ and $S_{c,n}$ are nonsingular.

Assumption 5 The sequences of matrices $W_{r,n}$, $W_{c,n}$, $S^{-1}_{r,n}$, and $S^{-1}_{c,n}$ are uniformly bounded in both row and column sums.

Assumption 6 The regressors $X_{r,in}$ and $X_{c,in}$ for $i = 1, 2, \cdots, n$ are vectors of constants and are uniformly bounded. The limits $\lim_{n\to\infty}\frac{1}{n}X'_{r,n}X_{r,n}$, $\lim_{n\to\infty}\frac{1}{n}X'_{c,n}X_{c,n}$, $\lim_{n\to\infty}\frac{1}{n}X'_{r,n}X_{c,n}$, and $\lim_{n\to\infty}\frac{1}{n}X'_{c,n}X_{r,n}$ exist and all are nonsingular.

Assumption 7 The regressors $X_{r,in}$ and $X_{c,in}$ for $i = 1, 2, \cdots, n$ are vectors of constants and are uniformly bounded. The limits $\lim_{n\to\infty}\frac{1}{n}X'_{r,n}X_{r,n}$, $\lim_{n\to\infty}\frac{1}{n}X'_{c,n}X_{c,n}$, $\lim_{n\to\infty}\frac{1}{n}X'_{r,n}X_{c,n}$, and $\lim_{n\to\infty}\frac{1}{n}X'_{c,n}X_{r,n}$ exist and all are nonsingular.

Assumption 8 The kernel $K(u, v)$ is a symmetric positive function such that for $u, v \in S^0 = \{-1, 1\}$, $K(1, 1) > 0$, $K(1, 1)K(-1, -1) > K(1, -1)^2$, and $K(1, -1) = K(-1, 1)$. $K(1, 1) > 0$, $K(-1, -1)$, and $K(1, -1)$ are uniformly bounded.

Assumption 9 (i) The parameter space Θ is a finite-dimensional compact space of $\mathbb{R}^m$, where $m = 2\times2+2\times2+2p$. (ii) $\theta_{b,n0}$ is an interior point in Θ, where $\theta_{b,n0} = \left\{\beta'_{b,0}, \lambda'_{b,0}, \sigma^2_{b,0}\right\}$ is the true parameter vector value given in bivariate model 6.

Assumption 10 $\lim_{n\to\infty}\frac{1}{2n}(G_{2,n}X_{2,n}\beta_b)'\mathcal{W}_{2n}(G_{2,n}X_{2,n}\beta_b)$ exist and is nonsingular.

Assumption 10 is a condition for the identification of $\lambda_{b,0}$, which is similar to Assumption 8 in Lee (2004); Liang et al. (2021). This assumption is a sufficient condition for global identification of $\theta_{b,n0}$.

Theorem 1 *Under Assumptions 1-10, $\theta_{b,n0}$ is a globally identifiable and $\hat{\theta}_{b,n}$ is consistent estimator of $\theta_{b,n0}$, i.e.,*

$$\hat{\theta}_{b,n} \xrightarrow{p} \theta_{b,n0} \tag{17}$$

Intuitively, the statistics $\frac{1}{2n}\ln L_{b,n}(\theta_{b,n})$ converges in probability $E(\frac{1}{2n}\ln L_{b,n}(\theta_{b,n}))$ uniformly in Θ as $n \to \infty$. Furthermore, the true parameter $\theta_{b,n0}$ is the unique minimizer of $E(\ln L_{b,n}(\theta_{b,n}))$. It then follows from the extremum estimator theorem (see, Anselin (1988)) that $\hat{\theta}_{b,n} \xrightarrow{p} \theta_{b,n0}$ as $n \to \infty$.

Theorem 2 *Under Assumptions 1-10,* $\sqrt{n}(\hat{\theta}_{b,n} - \theta_{b,n0}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\theta_b}^{-1} + \Sigma_{\theta_b}^{-1}\Omega_{\theta_b}\Sigma_{\theta_b}^{-1})$*, where* $\Omega_{\theta_b} = \lim_{n\to\infty}\Omega_{\theta_{b,n}}$ *and*

$$\Sigma_{\theta_b} = \lim_{n\to\infty} -E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\theta_{b,n}\partial\theta'_{b,n}}\right),$$

where

$$\Omega_{\theta_{b,n}} = \begin{pmatrix} 0 & * & * & * \\ \frac{\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^1_{b,iin}X_{b,in} & \Omega_{\theta_{b,22n}} & * & * \\ \frac{\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^2_{b,iin}X_{b,in} & 0 & \Omega_{\theta_{b,33n}} & * \\ \frac{\mu_3}{4n\sigma_0^6} l_{2n}\Sigma^{\frac{1}{2}}\mathbf{L}_{2n}X_{b,n} & \frac{\mu_3}{4n\sigma_0^6} l_{2n}\tilde{G}_{b,in}X_{b,n}\beta_{b,0} + \frac{\mu_3}{4n\sigma_0^6}tr(G_{b,n}) & \frac{\mu_4 - 3\sigma_0^4}{8n\sigma_0^8} & \frac{1}{2\sigma_0^4} \end{pmatrix},$$

and

$$-E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\theta_{b,n}\partial\theta'_{b,n}}\right) = \begin{pmatrix} \frac{1}{2n\sigma_0^2}X'_{b,n}\mathcal{W}_{2n}X_{b,n} & * & * & * \\ \frac{1}{2n\sigma_0^2}(X_{b,n}\beta_{b,0})'G'_{b,n}\odot\mathbf{e}_1\mathcal{W}_{2n}X_{b,n} & \mathbf{T}_{22,1} & * & * \\ \frac{1}{2n\sigma_0^2}(X_{b,n}\beta_{b,0})'G'_{b,n}\mathbf{e}_2\odot\mathcal{W}_{2n}X_{b,n} & 0 & \mathbf{T}_{22,2} & * \\ 0 & \frac{1}{2n\sigma_0^2 tr(G'_{b,n}\odot\mathbf{e}_1)} & \frac{1}{2n\sigma_0^2 tr(G'_{b,n}\odot\mathbf{e}_2)} & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

where

$$\Omega_{\theta_{b,22n}} = \frac{2\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^1_{b,iin}\tilde{G}^1_{b,in}X_{b,n}\beta_b + \frac{\mu_4 - 3\sigma_0^4}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^1_{b,iin}$$

$$\Omega_{\theta_{b,33n}} = \frac{2\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^2_{b,iin}\tilde{G}^2_{b,in}X_{b,n}\beta_b + \frac{\mu_4 - 3\sigma_0^4}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^2_{b,iin}$$

$$\begin{aligned}\mathbf{T}_{22,1} =& \frac{1}{2n\sigma_{b,0}^2}(X_{b,n}\beta_{b,0})'G'_{b,n}\odot\mathbf{e}_1\mathcal{W}_{2n}\mathbf{e}_1\odot G_{b,n}(X_{b,n}\beta_{b,0}) \\ &+ \frac{1}{2n}[tr(\mathbf{e}_1\odot G'_{b,n}\mathbf{e}_1\odot G_{b,n}) \\ &+ tr(\mathbf{e}_1\odot G_{b,n}\mathbf{e}_1\odot G_{b,n})]\end{aligned}$$

$$\begin{aligned}\mathbf{T}_{22,2} =& \frac{1}{2n\sigma_{b,0}^2}(X_{b,n}\beta_{b,0})'G'_{b,n}\odot\mathbf{e}_2\mathcal{W}_{2n}\mathbf{e}_2\odot G_{b,n}(X_{b,n}\beta_{b,0}) \\ &+ \frac{1}{2n}[tr(\mathbf{e}_2\odot G'_{b,n}\mathbf{e}_2\odot G_{b,n}) \\ &+ tr(\mathbf{e}_2\odot G_{b,n}\mathbf{e}_2\odot G_{b,n})]\end{aligned}$$

with $\tilde{G}^1_{b,n} = \Sigma^{\frac{1}{2}}\mathbf{L}_{2n}G_{b,n}\odot\mathbf{e}_1$ *and* $\tilde{G}^2_{b,n} = \Sigma^{\frac{1}{2}}\mathbf{L}_{2n}G_{b,n}\odot\mathbf{e}_2$*. Besides,* $-E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\theta_{b,n}\partial\theta'_{b,n}}\right)$ *is the average Hessian matrix which is nonsingular due to Assumption 10.*

4 Several interval SAR models and binary spatial weight matrices

The above model (Eqs.(4) and (5)) is established by integrating the SAR model and the CRM method. Furthermore, a method of fusing the SAR model with the Minmax method and the PM method can be adopted. The constructing steps of the model (Eqs.(4) and (5)) provides useful reference ideas and methods for the construction of other models.

CASE I. Using the bounds of interval variables, the model fusing SAR model and Minmax method is

$$Y_{l,n} = \lambda_l W_{l,n}Y_{l,n} + X_{l,n}\beta_r + V_{l,n} \tag{18}$$

$$Y_{u,n} = \lambda_u W_{u,n}Y_{u,n} + X_{u,n}\beta_c + V_{u,n} \tag{19}$$

CASE II. Using the bounds of interval variables and considering the lower and upper correlation, the model fusing SAR model and PM method is

$$Y_{l,n} = \lambda_l W_{l,n}Y_{l,n} + X_{l,n}\beta_l + X_{u,n}\gamma_l + V_{l,n} \tag{20}$$

$$Y_{u,n} = \lambda_u W_{u,n}Y_{u,n} + X_{l,n}\beta_u + X_{u,n}\gamma_u + V_{u,n} \tag{21}$$

The parameters of the models (Eqs. 18 and 19, Eqs. 20 and 21) can be solved by the direct QMLE method and the joint QMLE method. The detailed steps for solving are described in in Section 2.2.

The binary spatial weight matrices $(W_{l,n}, W_{u,n})$ and $(W_{c,n}, W_{r,n})$ are computed based on distance criterion. The the corresponding location of the interval-valued house prices in each district D can be identified as a bivariate interval

$([l_a, l_b], \{g_a, g_b]$) that describes the variability of the latitude and longitude coordinates inside of D, where l and g represent the latitude and longitude, respectively. The subscripts a and b indicates the lower and upper house prices, respectively. Let the vectors $v_s = ([l_{as}, l_{bs}], [g_{as}, g_{bs}])$ and $v_f = ([l_{af}, l_{bf}], [g_{af}, g_{bf}])$ represent the districts D_s and D_f, the elements of spatial weight matrices is given by

$$w_{z,n}^{sf} = \begin{cases} 1, & d^z(v_s, v_f) < \mathrm{d}; \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

where $z = l, u, c, r$, d is a critical value, $d^z(v_s, v_f)$ is the distance between D_s and D_f and computes by the metric of symbolic data analysis: the City block distance (CBD) and the squared Euclidean distance (SED).

The CBD for interval-valued data is defined as:

$$\text{CBD:} \quad d^z(v_s, v_f) = |l_{zs} - l_{zf}| + |g_{zs} - g_{zf}|. \tag{23}$$

The SED for interval-valued data is defined as:

$$\text{SED:} \quad d^z(v_s, v_f) = (l_{zs} - l_{zf})^2 + (g_{zs} - g_{zf})^2. \tag{24}$$

5 Simulation

This section presents a study of Monte Carlo on the D_k distance. We present two different data sets in order to estimate the parameters of the spatial regression model for interval-valued data, and evaluate the performance of the model. All response variables are required to be independent and normally distributed. We make the column vector of the explanatory variable matrix X of dimension $p = 5$ and set center and range for the interval-valued data and their coefficients, respectively. Then we have:

$$Y_n^{(r)} = \lambda_r W_n^{(r)} Y_n^{(r)} + X_n^{(r)} \beta_r + V_n^{(r)}$$

$$Y_n^{(c)} = \lambda_c W_n^{(c)} Y_n^{(c)} + X_n^{(c)} \beta_c + V_n^{(c)}$$

where $V_n^{(r)}$ and $V_n^{(c)}$ represents each element of $V_{r,n}$ and $V_{c,n}$, and so do the others.

In addition to this, each element of the upper and lower bounds of the error term V is required to follow a normal distribution, i.e. $V_n^{(r)} \sim N(0, \sigma_r^2)$ and $V_n^{(c)} \sim N(0, \sigma_c^2)$. So we set our parameter vector

$$\theta = (\lambda_r, \lambda_c, \beta_{r1}, \beta_{r2}, \beta_{r3}, \beta_{r4}, \beta_{r5}, \beta_{c1}, \beta_{c2}, \\ \times \beta_{c3}, \beta_{c4}, \beta_{c5}, \sigma_r^2, \sigma_c^2)$$

and make them equal to (0.4, 0.2, 10, 0.5, 0.1, 1, 5, 0, −0.5, 5, −1, 10, 0.3, 0.4) respectively. We varied the number of individuals $N = 50, 100, 250, 500$ to see how the parameter estimation performs at different numbers of individuals.

Several simulations with different kernel K are performed. The following gives different kernel settings for the covariance between the error vectors, respectively. In the simulation test, two types of data are respectively proposed for the correlation between error term $V_n^{(r)}$ and $V_n^{(c)}$:
Setting 1: Range error term $V_n^{(r)}$ is not related to center error term $V_n^{(c)}$, i.e., $cov(V_n^{(r)}, V_n^{(c)}) = 0$
Setting 2: Range error term $V_n^{(r)}$ is related to center error term $V_n^{(c)}$, i.e., $cov(V_n^{(r)}, V_n^{(c)}) = 1$

There are two different estimation methods proposed, direct QMLE and joint QMLE, to estimate the coefficient vector θ. The two estimates represent the two ways the nuclear matrix is set up. Direct QMLE means $K(1, 1) = K(-1, -1) = 1$ and $K(1, -1) = K(-1, 1) = 0$ and joint QMLE means $K(1, 1) = 1$, $K(-1, -1) = 4$, $K(1, -1) = 2$, $K(-1, 1) = 3$.

For each experiment we set the middle term of the weight matrix to be equal to 0, and the rest of the elements to belong to the weight values from 0 to 1, which indicates that the observed values account for their own prediction weights ranging from 0 to 1. We will repeat each experiment 1000 times, and therefore the actual mean, standard deviation, bias and RMSE of the estimated parameters of θ will be given in Tables (1)-(2), where the MAE($\hat{\theta}_i$) is calculated as $\frac{1}{1000}\sum_{m=1}^{1000} |\hat{\theta}_i^{(m)} - \hat{\theta}_i^{(0)}|$ and the RMSE($\hat{\theta}_i$) is calculated as $\sqrt{\frac{1}{1000}\sum_{m=1}^{1000} (\hat{\theta}_i^{(m)} - \hat{\theta}_i^{(0)})^2}$. The weight matrix will be repeated 1000 times during the run.

It can be seen from Tables (1)-(2) that after 1000 experiments, the estimated values of each coefficient in setting 1 are very close to the preset values, and the calculated MAE and RMSE are also very small. This shows that both direct QMLE and joint QMLE can estimate the value of each parameter well when there is no correlation between the error terms. In setting 2, when there is correlation between the covariances, it can be seen from Tables (3)-(4) that joint QMLE can give better estimates of various parameters than direct QMLE, which is consistent with our conclusion. By using joint QMLE, our proposed model can effectively solve the problem that the error terms are correlated and the covariance is not equal to 0 according to kernel function. In general, the results show that the model can well analyze and process the center and radius characteristics of the interval value data and make predictions. In addition, it can be found that as N increases, the estimated value is closer to the preset value, and the error term gradually decreases. This shows that our proposed model is more suitable for scenarios with large N. When the number of individuals N in various economic problems is large, this model can be well used to solve such issues with accurate prediction and excellent performance.

Table 1 Indicator estimates for each parameter obtained by direct QMLE in Setting 1

	Mean			Std($10^{(-1)}$)			MAE($10^{(-1)}$)			RMSE($10^{(-2)}$)		
	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500
$\lambda_r = 0.4$	0.3806	0.3812	0.3854	0.0802	0.0632	0.0432	0.1964	0.1983	0.2056	0.1223	0.1698	0.2101
$\lambda_c = 0.2$	0.1947	0.1948	0.1946	0.0555	0.0378	0.0285	0.0637	0.0544	0.0549	0.0443	0.0523	0.0612
$\beta_{r1} = 10$	9.8498	9.8520	9.8430	1.5087	1.2331	0.7992	1.7552	0.1571	1.5931	1.2281	1.5710	1.7602
$\beta_{r2} = 0.5$	0.4702	0.4902	0.4893	1.0941	0.8093	0.6025	0.9078	0.0663	0.4678	0.6537	0.6635	0.6090
$\beta_{r3} = 0.1$	0.0873	0.0897	0.0833	1.1270	0.8435	0.5665	0.9080	0.0692	0.4719	0.6537	0.6915	0.5880
$\beta_{r4} = 1$	0.9759	0.9673	0.9695	1.2568	0.7894	0.5534	1.0054	0.0696	0.5014	0.7377	0.6958	0.6293
$\beta_{r5} = 5$	4.9394	4.9245	4.9138	1.2885	0.9489	0.6552	1.1534	0.0988	0.9205	0.8210	0.9881	1.0812
$\beta_{c1} = 0$	-0.0260	-0.0093	-0.0128	1.3826	0.9548	0.7226	1.1255	0.0781	0.5770	0.8110	0.7806	0.7303
$\beta_{c2} = -0.5$	-0.5107	-0.5043	-0.5062	1.3804	0.9758	0.6241	1.1067	0.0795	0.5127	0.7980	0.7948	0.6240
$\beta_{c3} = 5$	4.9724	4.9521	4.9613	1.5781	1.1607	0.7056	1.2783	0.1022	0.6473	0.9234	1.0222	0.8018
$\beta_{c4} = -1$	-1.0136	-1.0026	-1.0164	1.3586	0.9701	0.6840	1.0965	0.0790	0.5517	0.7870	0.7897	0.7000
$\beta_{c5} = 10$	9.9445	9.9496	9.9459	1.7827	1.1194	0.8975	1.5137	0.0999	0.8919	1.0764	0.9994	1.0441
σ^2	0.3801	0.3939	0.3857	0.4192	0.3068	0.3176	0.3211	0.2444	0.2188	0.2344	0.2496	0.2473

Table 2 Indicator estimates for each parameter obtained by joint QMLE in Setting 1

	Mean			Std($10^{(-1)}$)			MAE($10^{(-1)}$)			RMSE($10^{(-2)}$)		
	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500
$\lambda_r = 0.4$	0.3846	0.3840	0.3831	0.0984	0.0559	0.0620	0.1599	0.1605	0.1687	0.1818	0.1695	0.1796
$\lambda_c = 0.2$	0.1904	0.1916	0.1902	0.0960	0.0551	0.0485	0.1160	0.0877	0.0985	0.1356	0.1000	0.1088
$\beta_{r1} = 10$	9.9046	9.8930	9.8878	1.6708	1.2285	1.0071	1.5424	1.3169	1.2602	1.9167	1.6247	1.5039
$\beta_{r2} = 0.5$	0.5031	0.5093	0.5071	1.3859	0.8264	0.9529	1.1097	0.6385	0.7543	1.3793	0.8275	0.9508
$\beta_{r3} = 0.1$	0.1197	0.1173	0.1137	1.3469	0.7219	0.7737	1.0979	0.5845	0.6182	1.3727	0.7389	0.7821
$\beta_{r4} = 1$	1.0069	1.0064	0.9998	1.2687	0.7083	0.8209	0.9538	0.5471	0.6367	1.2642	0.7076	0.8168
$\beta_{r5} = 5$	4.9681	4.9621	4.9489	1.4768	0.8921	0.8611	1.2036	0.7553	0.7961	1.5037	0.9653	0.9970
$\beta_{c1} = 0$	-0.0370	-0.0537	-0.0689	1.4606	1.3923	1.3456	1.8953	1.1721	1.2494	1.4761	1.4156	1.5061
$\beta_{c2} = -0.5$	-0.5310	-0.5208	-0.5107	1.4966	1.3733	1.3517	1.1339	1.3261	1.1469	1.6128	1.5955	1.4374
$\beta_{c3} = 5$	4.9079	4.9328	4.9062	1.6603	1.6258	1.5087	1.3174	1.4761	1.3866	1.8027	1.7515	1.7699
$\beta_{c4} = -1$	-1.0821	-1.0526	-1.0756	1.4795	1.2780	1.2101	1.9603	1.2434	1.1078	1.2412	1.5619	1.4217
$\beta_{c5} = 10$	9.8406	9.8704	9.8566	1.9723	1.8999	1.5391	1.6423	1.3951	1.1716	1.3359	1.2915	1.0975
σ^2	0.3935	0.3906	0.3802	1.0787	0.7460	0.2297	0.8151	0.5759	0.1908	1.0732	0.7422	0.4286

In order to verify the necessity of using interval-valued data, a comparison is made in this paper. Without considering the radius of interval-valued data, it is considered as single-point data and predicted by the spatial autoregressive model. Fig.1 reports a box plot and plotting the RMSE of between interval and single point data of SAR. It shows the results of comparing different N in the two settings. It can be seen that the RMSE of our proposed model is obviously smaller than that of SVR-SP, which indicates that when the interval-valued data is considered as a scalar for prediction, the feature information of the data will be lost and the prediction results will be inaccurate.

6 Empirical Application

In modern economies, housing comprises a large segment of aggregate demand, as well as a large segment of personal investment. Therefore, housing values play a critical role in the stability of national economies and financial markets. At the same time, housing price is one of the most dynamic and unpredictable variables in the economy. The interaction of housing, financial and economic activities, political interventions and geospatial information all contribute to changes in housing values. Typically, on the one hand, housing prices exhibit spatial pattern (Guo and Qu 2019). An increase of the housing values in one neighborhood in a district may affect the housing values in surrounding neighborhoods. While the

Table 3 Indicator estimates for each parameter obtained by direct QMLE in Setting 2

	Mean			Std($10^{(-1)}$)			MAE($10^{(-1)}$)			RMSE($10^{(-2)}$)		
	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500
$\lambda_r = 0.4$	0.3779	0.3821	0.3830	0.1232	0.0715	0.0503	0.2675	0.1581	0.1692	0.1993	0.1734	0.1765
$\lambda_c = 0.2$	0.1858	0.1906	0.1914	0.0972	0.0561	0.0407	0.1151	0.0876	0.0961	0.1361	0.1010	0.1042
$\beta_{r1} = 10$	10.0769	9.9019	9.8838	2.3876	1.5075	1.0316	2.1575	1.4026	1.2510	2.5879	1.8050	1.5261
$\beta_{r2} = 0.5$	0.6120	0.4060	0.4797	1.6353	1.0781	0.7055	1.7790	1.5550	0.5574	1.6271	1.0752	0.7034
$\beta_{r3} = 0.1$	0.1428	0.1245	0.1226	1.7331	1.0424	0.7354	1.3750	0.8402	0.5954	1.7422	1.0493	0.7521
$\beta_{r4} = 1$	1.0980	1.0277	0.9809	1.8112	0.9655	0.7435	1.4373	0.7468	0.6027	1.8052	0.9608	0.7433
$\beta_{r5} = 5$	4.8698	5.1120	4.9596	2.0364	1.1059	0.7319	1.6637	0.9030	0.6723	2.0650	1.1396	0.8323
$\beta_{c1} = 0$	-0.1284	-0.1182	-0.0840	2.4481	1.3839	1.1789	1.9332	1.5893	0.9963	2.4785	1.4620	1.3032
$\beta_{c2} = -0.5$	-0.4709	-0.6097	-0.5875	2.5571	1.6798	1.1208	2.1740	1.4185	1.1564	2.6538	1.7057	1.4175
$\beta_{c3} = 5$	4.8635	5.0331	4.7215	2.6745	1.6358	1.2081	2.4034	1.4727	1.1559	2.8105	1.7588	1.4905
$\beta_{c4} = -1$	-1.2685	-1.2565	-0.9416	2.3248	1.4849	1.2084	1.9944	1.5011	1.0482	2.4455	1.5755	1.3035
$\beta_{c5} = 10$	9.9005	9.7490	9.8660	2.9340	1.9227	1.3648	2.6923	1.8566	1.5472	3.3071	2.3259	1.9077
σ^2	0.4191	0.4002	0.4007	0.5537	0.3699	0.2667	0.4161	0.2846	0.2189	0.5509	0.3681	0.2653

Table 4 Indicator estimates for each parameter obtained by joint QMLE in Setting 2

	Mean			Std($10^{(-1)}$)			MAE($10^{(-1)}$)			RMSE($10^{(-2)}$)		
	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500	N=100	N=250	N=500
$\lambda_r = 0.4$	0.3804	0.3854	0.3903	0.1061	0.0689	0.0488	0.2003	0.1967	0.1967	0.1285	0.1700	0.2034
$\lambda_c = 0.2$	0.1946	0.1941	0.1951	0.0595	0.0361	0.0250	0.0655	0.0599	0.0593	0.0463	0.0561	0.6405
$\beta_{r1} = 10$	9.8513	9.8471	9.8652	2.0917	1.2025	0.8981	2.0884	1.6714	1.3820	1.4799	1.5862	1.6169
$\beta_{r2} = 0.5$	0.4781	0.4882	0.4836	1.5245	0.9681	0.6030	1.2429	0.7670	0.5011	0.8877	0.7937	0.6221
$\beta_{r3} = 0.1$	0.0898	0.0929	0.0952	1.4483	0.9303	0.6861	1.1666	0.7137	0.5481	0.8368	0.7592	0.6844
$\beta_{r4} = 1$	0.9746	0.9927	0.9731	1.5699	1.0482	0.6880	1.2556	0.8489	0.5449	0.9167	0.8549	0.7353
$\beta_{r5} = 5$	4.9302	4.9119	4.9236	1.7067	1.1472	0.7043	1.4514	0.1171	0.8930	1.5631	1.1785	1.0367
$\beta_{c1} = 0$	-0.0274	-0.0265	-0.0147	1.5153	1.0089	0.7153	1.2378	0.8397	0.5713	1.3877	0.8489	0.7268
$\beta_{c2} = -0.5$	-0.5113	-0.5213	-0.5292	1.4952	0.9928	0.6671	1.1933	0.8165	0.5899	1.4643	0.8264	0.7255
$\beta_{c3} = 5$	4.9704	4.9596	4.9640	1.7447	1.1791	0.7075	1.4132	1.0999	0.5931	1.7201	1.0146	0.7905
$\beta_{c4} = -1$	-1.0132	-1.0148	-1.0152	1.4891	0.9656	0.7448	1.1711	0.7817	0.6239	1.2617	0.7951	0.7565
$\beta_{c5} = 10$	9.9421	9.9425	9.9352	1.8854	1.0996	0.8574	1.5963	0.9788	0.8854	1.5369	1.0103	1.0712
σ^2	0.3804	0.3860	0.3898	0.4263	0.2804	0.2103	0.3334	0.2223	0.1707	0.2561	0.2345	0.2092

building materials of the house itself are not relevant, some unobservable factors may be spatially relevant (e.g., living environment). On the other hand, the house prices in a region are presented as a range of due to the uneven distribution of supporting facilities. Usually, if the surrounding facilities, such as commercial centers and hospitals, are more upscale, the housing price will be on the higher end; conversely, the housing price will be on the lower end. D'Urso et al. (2023) applied the spatial fuzzy clustering method the interval-valued rental values of housing price.

6.1 Data description

In this study, we apply the analysis to examine house price ranges in sixteen districts in Shanghai, China: Qingpu, Yangpu, Minhang, Baoshan, Jiading, Pudong New, Jinshan, Fengxian, Putuo, Jing'an, Changning, Xuhui, Huangpu, Songjiang, Chongming and Hongkou (ordained 1, 2, $\cdots$, 16 in Fig.2, respectively). This study uses two data sources. The first is the housing prices (HP) data of Shanghai provided by the China Real Estate Index System. There are more than 1.7 million rows of records with 16 districts, and the data approximately cover the most urban areas of Shanghai. As show in Fig.2, the spatial coverage of the house prices is quit comprehensive, including 16 districts and reflecting the

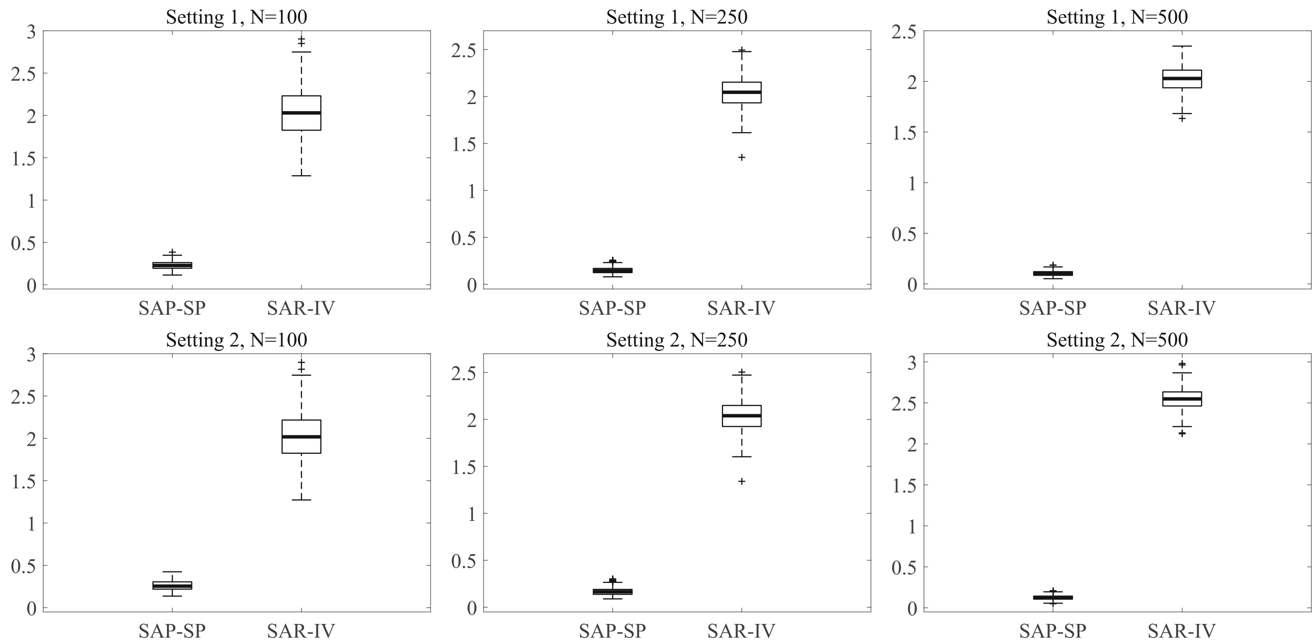


Fig. 1 Compare the RMSE of SAR of single point (SP) and interval-valued (IV) data for $N = 100$, $N = 250$ and $N = 500$ in two settings, respectively

range of house prices for each district. In addition, for each house, the dataset provides floor area ratio (FSR), completion date (CD). The second data source is the 16 districts data set collected from National Bureau of Statistics. For each district, the dataset provides the per capita GDP (PCGDP), and land use data (LUD). It should be emphasized that the land use data here are the sum of high-grade hospitals, high-grade commercial centers (as show in Figs.3 and 4)and transportation hubs. This kind of substitution is common when specific land use data are not precisely known (Ni et al. 2018). In each district, the variables FP, GCR, FSR, CD is the standard interval variables, the variables TP and LUD is the trivial interval variables. The first 12 districts data are applied to model training and the residual 4 districts data are reserved for modeling testing. The two data sources are online and integrated into ArcGis. Part of the dataset is presented in Table 5.

The dataset $S = [X^I_{ijn}, y^I_{in}]$, $i = 1, 2, \cdots, 16$, $y^I_{in} = [y_{l,in}, y_{u,in}]$ is the dependent variable denoting the HP (in order to reduce the absolute value of the data for convenient calculation, let $y^I_{in} = [y_{l,in}, y_{u,in}] = [\log \mathrm{HP}_{l,i}, \log \mathrm{HP}_{u,i}]$); $X^I_{ijn} = [X_{l,ijn}, X_{u,ijn}]$ is considered to be the independent variables, which represents FSR, GCR, CD, TP, and LUD, respectively.

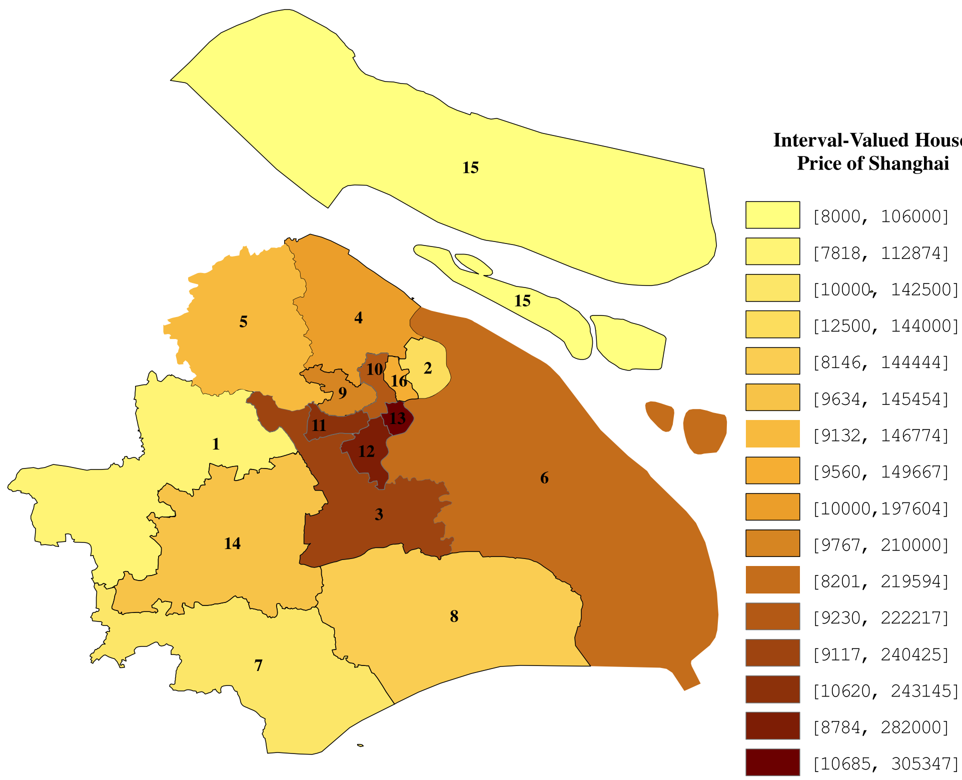
6.2 Model formulation

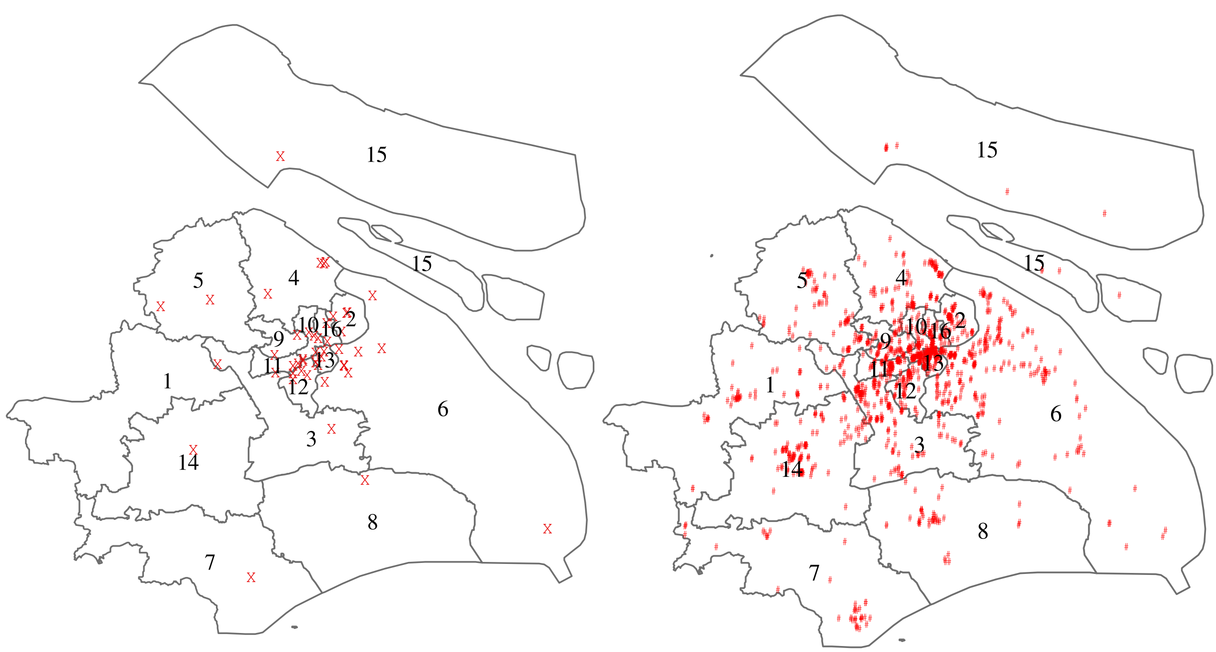
For simplicity, we give abbreviations to the interval-valued regression models and spatial interval-valued autoregressive models:

(1) Minmax: the regression model proposed by Billard and Diday (2002);
(2) D-SAR-Minmax: the proposed spatial autoregressive model based on the bounds of intervals; the parameter estimators is obtained by direct QMLE method;
(3) J-SAR-Minmax: the proposed spatial autoregressive model based on the bounds of intervals; the parameter estimators is obtained by joint QMLE method;
(4) CRM: the regression model proposed by Neto and Carvalho (2008);
(5) D-SAR-CRM: the proposed spatial autoregressive model based on the center and range; and the parameter estimators is obtained by direct QMLE method;
(6) J-SAR-CRM: the proposed spatial autoregressive model based on the center and range, and the parameter estimators is obtained by joint QMLE method;
(7) PM: the regression model proposed by Souza et al. (2017);
(8) D-SAR-PM: the proposed spatial autoregressive model based on the upper and lower bounds; and the parameter estimators is obtained by direct QMLE method.
(9) J-SAR-PM: the proposed spatial autoregressive model based on the upper and lower bounds; and the parameter estimators is obtained by joint QMLE method.

For D-SAR-Minmax, J-SAR-Minmax models, D-SAR-CRM, J-SAR-CRM, D-SAR-PM and J-SAR-PM, the spatial weight matrices is computed by CBD and SED metrics, as described in Section 4. We utilize these models, CRM, Minmanx, D-SAR-Minmax, J-SAR-Minmax, D-SAR-CRM, J-SAR-

Fig. 2 The interval-valued house prices between 16 districts of Shanghai, China





(a) High-grade hospitals

(b) High-grade commercial centers

Fig. 3 Distributions of facilities in Shanghai, China

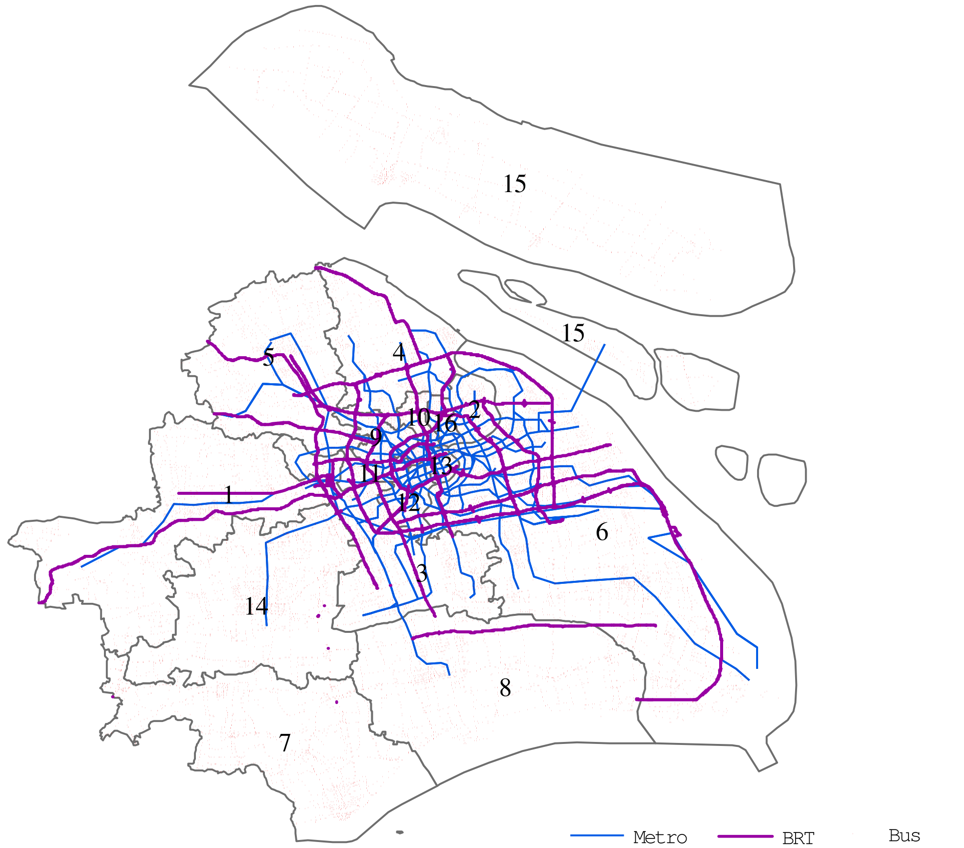


Fig. 4 Urban public transport network in Shanghai, China

Table 5 Housing prices related data

Districts	HP (10^3) [min, max]	FSR ($\% \cdot 10^{-2}$) [min, max]	LUD(10^{-3}) [min, max]	PCGDP(10^{-5}) [min, max]
1	[7.82, 112.87]	[0.27, 4.00]	[1.35, 1.35]	[1.03, 1.13]
2	[12.82, 144.00]	[0.35, 7.20]	[0.39, 0.39]	[1.69, 1.85]
3	[9.12 240.43]	[0.29, 6.53]	[1.56, 1.56]	[1.08, 1.12]
4	[10.00 197.60]	[0.30, 4.30]	[0.86, 0.86]	[0.79, 0.80]
5	[9.13 146.78]	[0.40, 6.70]	[1.11, 1.11]	[1.49, 1.50]
6	[8.20 219.59]	[0.30, 6.80]	[3.91, 3.91]	[2.78, 2.89]
7	[6.67 142.50]	[0.50, 3.40]	[1.48, 1.48]	[1.37, 1.37]
8	[8.15 144.44]	[0.30,4.00]	[1.33, 1.33]	[1.20, 1.26]
...	...	...	...	...

CRM to investigate both intra- and inter-district ranges of housing prices.

6.3 Experimental results

The main objective is to evaluate housing prices assessment from different perspectives. The examination of experimental results is segmented into four facets: the spatial correlation, interval inner correlation, the fitting and predicting performances. The proposed spatial autoregressive models, which includes J-SAR-Minmax, J-SAR-CRM, J-SAR-PM, D-SAR-Minmax, D-SAR-CRM, D-SAR-PM, and the linear regression models, which include Minmax, CRM, PM, are used to analyze the aforementioned housing prices.

The experimental results of these models with weight matrices W^{CBD} and W^{SED} are summarized in Tables 6 and 7. Fig.5 displays the fitting and prediction results of housing prices with weight matrix W CBD.Three main results are listed as follows.

1) Housing prices exhibit spatial correlation across 16 regions. As shown in Tables 6 and 7, the values of all error evaluation indexes show that the fitting and prediction performances of our proposed spatial autoregressive

Table 6 Evaluation indexes of models Minmax, D-SAR-Minmax, J-SAR-Minmax, CRM, D-SAR-CRM, J-SAR-CRM, PM, D-SAR-PM, J-SAR-PM using weight matrix W^{CBD}

Weight Matrix	Models	(a, b, c)	Fitting				Prediction			
			RMSE	MAE	MMER	MSE	RMSE	MAE	MMER	MSE
W^{CBD}	Minmax	–	0.5022	0.5604	0.1564	3.6758	1.1537	1.0456	0.2548	7.1069
	D-SAR-Minmax	–	0.2794	0.3007	0.0991	1.1388	1.0470	1.1144	0.3231	4.3923
	J-SAR-Minmax	(5, 2, 5)	0.2668	0.2948	0.0974	1.0194	0.5568	0.6601	0.2012	1.7602
		(3, 4, 7)	0.2825	0.2959	0.0953	1.3704	0.8407	0.8885	0.2592	3.0883
		(5, 1, 5)	0.2675	0.2931	0.1006	1.0007	0.5304	0.6134	0.1818	1.7185
		(2, −1, 1)	0.2869	0.3582	0.1211	0.5700	0.618	0.6835	0.2047	0.8591
		(10, 6, 10)	0.2701	0.2836	0.0895	1.1219	0.6255	0.711	0.2173	1.9705
		(10, 8, 16)	0.2645	0.2835	0.0907	1.0793	0.6406	0.7347	0.2247	2.0325
		$(a^{opt}, b^{opt}, c^{opt})$	0.2666	0.2936	0.0996	0.9232	0.5699	0.6534	0.2048	1.5973
	CRM	–	0.3325	0.3707	0.1642	0.4379	0.4800	0.5540	0.2975	0.2768
	D-SAR-CRM	–	0.1739	0.1842	0.0863	0.3755	0.3323	0.3984	0.2136	0.2096
	J-SAR-CRM	(5, 2, 5)	0.1653	0.1759	0.0861	0.3678	0.2661	0.3398	0.1759	0.1446
		(3, 4, 7)	0.1684	0.1773	0.0868	0.3026	0.1958	0.2249	0.1117	0.0164
		(5, 1, 5)	0.1671	0.1758	0.0809	0.3868	0.2770	0.3503	0.1797	0.1749
		(2, −1, 1)	0.2172	0.2438	0.1304	0.3255	0.4788	0.5676	0.3454	0.1275
		(10, 6, 10)	0.1613	0.1744	0.0836	0.3365	0.2047	0.2721	0.1364	0.0876
		(10, 8, 16)	0.2317	0.2451	0.1318	0.8391	0.2377	0.2937	0.1468	0.3092
		$(a^{opt}, b^{opt}, c^{opt})$	0.1675	0.1769	0.087	0.3862	0.2865	0.3061	0.1895	0.2141
	PM	–	0.2807	0.2921	0.1606	1.2048	1.4224	0.5540	0.2917	1.7067
	D-SAR-PM	–	0.2586	0.2404	0.0785	0.6785	1.0470	1.1144	0.3231	4.3923
	J-SAR-PM	(5, 2, 5)	0.2786	0.2557	0.0837	0.8777	0.5568	0.6601	0.2012	1.7602
		(3, 4, 7)	0.3146	0.2778	0.0912	1.1268	0.8407	0.8885	0.2592	3.0883
		(5, 1, 5)	0.2846	0.2531	0.0821	0.8491	1.4224	1.3049	0.2917	1.7067
		(2, −1, 1)	0.2520	0.2251	0.0745	0.5225	0.6180	0.6835	0.2047	0.8591
		(10, 6, 10)	0.2678	0.2471	0.0812	0.8031	0.6255	0.7110	0.2173	1.9705
		(10, 8, 16)	0.2619	0.2465	0.0822	0.8128	0.6406	0.7347	0.2247	2.0325
		$(a^{opt}, b^{opt}, c^{opt})$	0.2771	0.26	0.0849	0.9034	0.5699	0.6534	0.2048	1.5973

models for interval-valued data (J-SAR-Minmax, J-SAR-CRM, J-SAR-PM, D-SAR-Minmax, D-SAR-CRM, D-SAR-PM) outperform these of the corresponding previous linear regression models (Minmax, CRM, PM) significantly. Since the error evaluation indexes and these model with different matrices have the familiar performances, we take the RMSE and weight matrix W^{CBD} as an example. For example, when using matrix W^{CBD}, the RMSE of Minmax, D-SAR-Minmax, J-SAR-Minmax are 0.5022, 0.2794 (-0.2228), at around 0.2721 (-0.2300) (The numbers in parentheses represent the gap between the model Minmax and the corresponding proposed spatial autoregressive models for interval-valued data). The RMSE of CRM, D-SAR-CRM, J-SAR-CRM are 0.3325, 0.1739 (-0.1586), at around 0.1826 (-0.1499) (The numbers in parentheses represent the gap between the model CRM and the corresponding proposed spatial autoregressive models). The RMSE of PM, D-SAR-PM, J-SAR-PM are 0.2807, 0.2586 (-0.0221), at around 0.2767 (-0.0040) (The numbers in parentheses represent the gap between the model PM and the corresponding proposed spatial autoregressive models for interval-valued data). These findings underscore the robustness and reliability of spatial autoregressive models for interval-valued data in capturing the nuanced variations across different regions in Shanghai.

2) The spatial autoregressive models for interval-valued data using the spatial weight matrix W^{SED} are suitable in the current research on the degree of influence of explanatory variables on house prices. Comparing Tables 6 and 7, it can be seen that the values of RMSE, MAE, MMER and MSE of the spatial autoregressive models for interval-valued data using the spatial weight matrix W^{SED} are smaller compared to these of using the spatial W^{CBD} in most fitting cases. It means that the spatial autoregressive models for interval-valued data using the spatial

Table 7 Evaluation indexes of models Minmax, D-SAR-Minmax, J-SAR-Minmax, CRM, D-SAR-CRM, J-SAR-CRM, PM, D-SAR-PM, J-SAR-PM using weight matrix W^{SED}

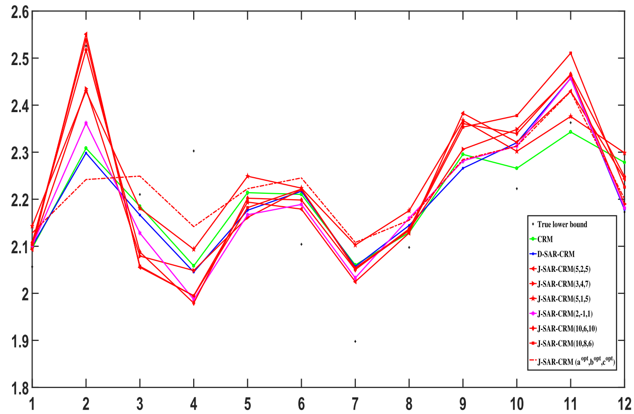
Weight Matrix	Models	(a, b, c)	Fitting				Prediction			
			RMSE	MAE	MMER	MSE	RMSE	MAE	MMER	MSE
W^{SED}	Minmax	–	0.5022	0.5604	0.1564	3.6758	1.1537	1.0456	0.2548	7.1069
	D-SAR-Minmax	–	0.2965	0.3455	0.1229	1.0831	0.797	0.9313	0.2540	4.525
	J-SAR-Minmax	(5, 2, 5)	0.2660	0.3082	0.1064	1.0583	0.9256	1.1070	0.2992	6.3107
		(3, 4, 7)	0.2680	0.3161	0.1079	0.9239	0.8663	1.0749	0.2942	5.7319
		(5, 1, 5)	0.2343	0.2668	0.0889	0.9427	1.0089	1.3193	0.3725	7.6275
		(2, −1, 1)	0.3190	0.3918	0.1312	0.7136	0.5587	0.5263	0.1656	1.2214
		(10, 6, 10)	0.2492	0.2900	0.0964	0.9607	1.0661	1.4154	0.3952	8.8360
		(10, 8, 16)	0.2429	0.2823	0.0947	0.8870	0.9323	1.2149	0.3401	6.6888
		$(a^{opt}, b^{opt}, c^{opt})$	0.2311	0.2664	0.0870	0.8783	1.1746	1.568	0.4331	10.355
	CRM	–	0.3325	0.3707	0.1642	0.4379	0.48	0.554	0.2975	0.2768
	D-SAR-CRM	–	0.1528	0.1646	0.0869	0.2748	0.468	0.4278	0.2897	0.1394
	J-SAR-CRM	(5, 2, 5)	0.1584	0.1740	0.0861	0.3788	0.3782	0.3418	0.2353	0.2241
		(3, 4, 7)	0.1527	0.1748	0.0865	0.3374	0.4009	0.3746	0.2561	0.2415
		(5, 1, 5)	0.1694	0.2018	0.1049	0.2775	0.6010	0.6033	0.3426	0.0977
		(2, −1, 1)	0.1826	0.1953	0.1032	0.2676	0.8693	1.0044	0.5215	1.5960
		(10, 6, 10)	0.1549	0.1741	0.0877	0.3632	0.3850	0.3619	0.2474	0.1975
		(10, 8, 16)	0.1490	0.1681	0.0892	0.2911	0.4649	0.4610	0.2873	0.1689
		$(a^{opt}, b^{opt}, c^{opt})$	0.1803	0.1916	0.1029	0.5178	0.3849	0.4309	0.2933	0.2396
	PM	–	0.2807	0.2921	0.1606	1.2048	1.4224	0.5540	0.2917	1.7067
	D-SAR-PM	–	0.2112	0.2066	0.0659	0.5398	0.797	0.9313	0.2540	2.5250
	J-SAR-PM	(5, 2, 5)	0.1874	0.1939	0.0646	0.3775	0.9256	1.1070	0.2992	2.3107
		(3, 4, 7)	0.2043	0.2042	0.0656	0.4273	0.8663	1.0749	0.2942	2.7319
		(5, 1, 5)	0.2175	0.2087	0.0663	0.6207	1.0089	1.3193	0.3725	2.6275
		(2, −1, 1)	0.1964	0.2155	0.0709	0.3841	0.5587	0.5263	0.1656	1.2214
		(10, 6, 10)	0.2223	0.2200	0.0708	0.6819	1.0661	1.4154	0.3952	2.836
		(10, 8, 16)	0.2284	0.2108	0.0683	0.5468	0.9323	1.2149	0.3401	2.6888
		$(a^{opt}, b^{opt}, c^{opt})$	0.2311	0.2339	0.0786	0.5110	1.1746	1.568	0.4331	2.355

weight matrix W^{SED} are suitable in the current research on the degree of influence of explanatory variables on house prices. It cen be utilized for deeply analyzing the complex internal and spatial relationships between numerous explanatory variables such as FSR, GCR, TP, LUD, PCGOP and house price fluctuations. For example, by using models J-SAR-Minmax, J-SAR-CRM and J-SAR-PM with spatial weight matrix W^{SED}, the coefficients of various factors influencing house prices can be accurately quantified. As shown in Table 8, the factor LUD that plays a positive and significant effects in driving up house prices. This finding is consistent with the conclusions from numerous previous studies,which all indicated the positive connection between the LUD and housing prices (Yii et al. 2022; Li et al. 2025). As for the government departments, these results provide strong support for more targeted real estate regulatory policies to be introduced, so that house prices can be more effectively stabilized, and the healthy and stable development of the real estate market can be effectively ensured. Furthermore, the spatial autoregressive models for interval-valued data using the spatial weight matrix W^{CBD} are suitable for forecasting the future housing price fluctuations based on current urban design. Comparing Tables 6 and 7, it can be seen that the prediction performances of J-SAR-Minmax, J-SAR-CRM and J-SAR-CRM using spatial weight matrix W^{CBD} is better than these of using spatial weight matrix W^{SED}.

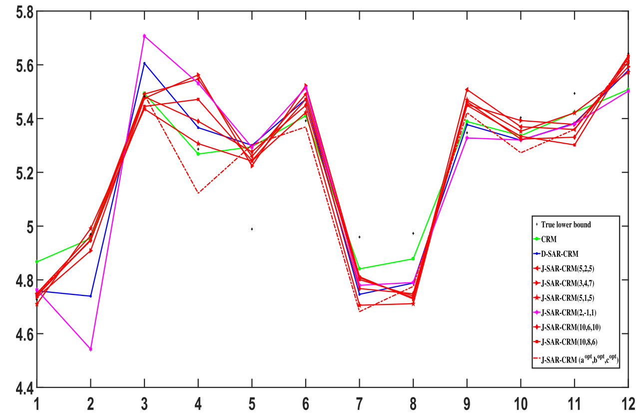
3) The values of all error evaluation indexes show that the interval housing prices of Shanghai have interval inner correlation. In the linear regression models, the model PM has the best fitting and prediction experiments compared with Minmax and CRM. In the spatial autoregressive models, the models J-SAR-Minmax, J-SAR-CRM and J-SAR-PM are better than the models D-SAR-Minmax, D-SAR-CRM and D-SAR-PM, regardless of

Table 8 The coefficients of models J-SAR-CRM using weight matrix W^{SED}

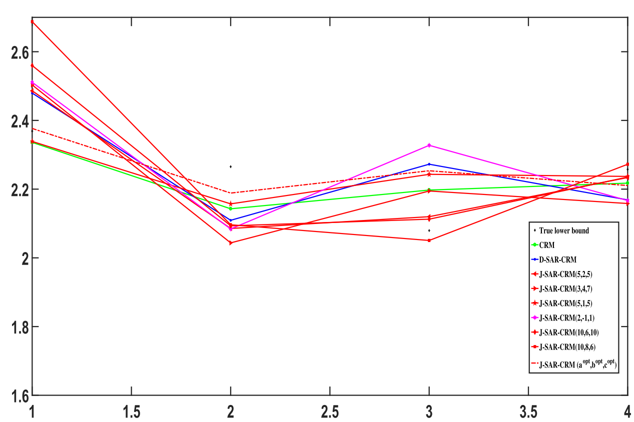
Models	(a, b, c)	Lower					Upper				
		λ_l	Constant	FSR_l	LUD_l	PCGDP_l	λ_u	Constant	FSR_u	LUD_u	PCGDP_u
J-SAR-CRM	(5, 2, 5)	-0.1841	2.6875	-0.0659	2.9980	-1.0005	-0.1731	7.0858	0.0618	2.5345	-0.0061
	(3, 4, 7)	0.1935	0.3868	0.0264	2.3276	-0.3474	0.2299	1.7649	0.1621	1.7445	-0.0202
	(5, 1, 5)	-0.1876	0.9416	-0.0162	2.9252	-0.1979	-0.1785	2.8932	0.0645	3.1518	-0.0142
	(2, −1, 1)	0.3775	-0.1037	-0.0792	4.1829	-1.1951	0.3921	0.8725	0.1051	2.6683	-0.0183
	(10, 6, 10)	-1.0085	2.4166	0.0611	1.3099	-0.6551	-0.6204	5.9584	0.0350	3.4099	-0.0277
	(10, 8, 16)	-0.5728	1.8350	0.0946	0.7793	-0.7392	-0.3149	4.3140	0.0215	3.3234	0.0245
	$(a^{opt}, b^{opt}, c^{opt})$	-2.1057	4.1610	0.0393	1.1307	-0.7505	-1.5326	10.4523	0.0615	2.2323	-0.0303



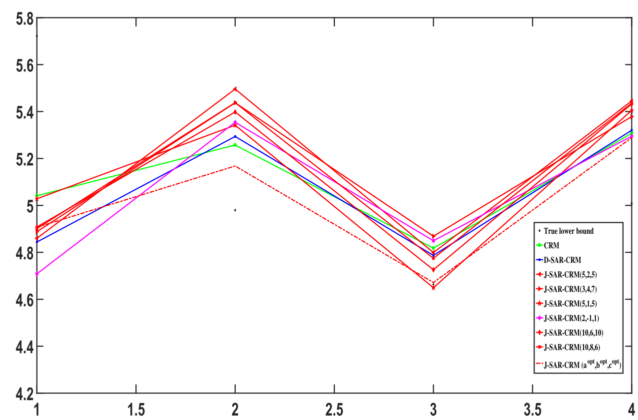
(a) Lower bound fitting of training data



(b) Upper bound fitting of training data



(c) Lower bound prediction of test data



(d) Upper bound prediction of test data

Fig. 5 The fitting and prediction results of housing prices with weight matrix W^{CBD}. Note: In each figure, '+' indicates the bounds of the true interval-valued data; the red lines show the experimental results of response variables of the model J-SAR-CRM with different (a, b, c); the green and blue lines show the experimental results of response variables of the models CRM and D-SAR-CRM, respectively. The magenta line is used to represent the model corresponding to the optimal fitting or prediction results

the weight matrices W^{CBD} or W^{SED}. Actually, from the perspective of model construction, the interval inner correlation of interval-valued data is considered in these models PM, J-SAR-Minmax, J-SAR-CRM and J-SAR-PM. In PM, the explanatory variables corresponding to the upper and lower bounds of the predictor variables are consistent. In J-SAR-Minmax and J-SAR-CRM, the interval inner correlation of the predictor variable is adjusted by the valuesa, b and c. The J-SAR-PM model not only contains the adjustment form of J-SAR-Minmax and J-SAR-CRM, but also the model itself incorporates interaction linear terms between $X_{r,n}$ and $X_{c,n}$. However, as shown in Tables 6 and 7, the error evaluation values of J-SAR-CRM are smaller than those of D-SAR-PM and J-SAR-PM. The reason may be as mentioned in Xu and Qin (2023), the application of the PM expansion models may be affected by collinearity.

4) The values of all error evaluation indexes show that the models J-SAR-CRM and D-SAR-CRM with W^{SED} have the best fitting and prediction performances, in comparison with all rest other models. For example, as shown in Tables 6 and 7, J-SAR-CRM with W^{SED} and $(a, b, c) = (10, 8, 16)$ has the lowest RMSE, at around 1.490, followed by J-SAR-CRM with W^{SED} and $(a, b, c) = (3, 4, 7)$, D-SAR-CRM, J-SAR-CRM with W^{SED} and rest cases of (a, b, c). CRM has a relatively higher error,at around 0.3325.

7 Conclusions and prospects

This paper attempts to model spatial interval-valued data regression, considering all inner information of intervals. We introduce a new SAR model for interval-valued data. To estimate the estimators, We have derived a QMLE approach based on L_2 type D_k metric. Note that it is appealing that the L_2 type D_k metric as a generalization of the direct QMLE approach preserving the asymptotic properties of classical QMLE approach (Lee 2004). The simple expression of the D_k metric in QMLE approach than these of Euclidean metric lies in the fact that D_k metric measures the distance between each pair of points in intervals in terms of the support function with respect to kernel functions. Through Monte Carlo simulations, we present the finite sample properties of the proposed estimation method. The results report that the QMLE approach based on L_2 type D_k metric fit the interval-valued data more adeptly compared to the QMLE approach for point-valued data. When applied to real datasets related to house prices of Shanghai, China for fitting and forecasting, the proposed models demonstrate best results, highlighting their excellent performance. Thus, the proposed models provide effective solutions to practical challenges associated with interval-valued data that exhibit spatial correlation.

This paper focuses exclusively on ISAR models with linear assumption, which is the trade-off between simplicity and interpretability. Here we highlight the potential limitations of the proposed model and several future research directions as follows. Firstly, dealing with the nonlinear or more complex interval-valued variation is essential; however, the ISAR model occasionally seems too rigid to do this during computation. Future work could adopt an integrated perspective, incorporating time-dimensional dynamics and time-varying coefficients (as referenced by Liang et al. (2021) and Zhao et al. (2025)) to directly process time-varying interval-valued data. Alternatively, graph neural networks (Dawn and Bandyopadhyay 2023) could be adapted for spatial interval-valued analysis. Secondly, developing estimators resistant to anomalous data is critical. This requires designing appropriate weight functions during estimation to mitigate outlier influence. As the current model lacks such mechanisms. Therefore, future research could focus on developing robust ISAR models. Thirdly, to address real-world financial issues with spatial correlations, such as cross-market contagion effects, geographical risk spillovers, and spatial dependence in high-frequency order flows, it is crucial to partition the state space into discrete regimes and capture nonlinear dynamics. Specifically, it involves studying exogenous and endogenous thresholds, but the proposed ISAR model is inapplicable to the datasets exhibiting discrete regimes and nonlinear dynamic features. Therefore, future research could focus on developing threshold ISAR models. Our current ISAR-based approach is a solid foundation and baseline for future researchers.

Appendix

A Proofs of Theorem 1.

In the following, the idea for prove the consistency and identification of $\hat{\theta}_{b,n}$ is similar to Lee (2004) and Liang et al. (2021). Define $Q_{b,n}(\lambda_b) = \max_{\beta_{b,n},\sigma^2} E(\ln(\theta_{b,n}))$. The consistency of $\hat{\theta}_{b,n}$ will follow from

$$\frac{1}{2n}[\ln L_{b,n}(\lambda_b) - Q_{b,n}(\lambda_b)] \xrightarrow{p} 0 \ \text{ uniformly on } \ \Lambda_b, \tag{A1}$$

and the uniqueness identification condition is

$$\lim_{n\to\infty} \sup \left\{ \max_{\lambda_b \in N_\epsilon^c(\lambda_{b,0})} \frac{1}{2n}(Q_{b,n}(\lambda_b) - Q_{b,n}(\lambda_{b,0})) \right\} < 0 \ \text{ for any } \ \epsilon, \tag{A2}$$

where $N_\epsilon^c(\lambda_{b,0})$ is the complement of an open neighbourhood of Λ_b of diameter ϵ.

(1) Proof of (A1). Since

$$\begin{aligned} Q_{b,n}(\lambda_b) &= \max_{\beta_b,\sigma^2} E(\ln L_{b,n}(\theta_{b,n})) \\ &= -n(\ln(2\pi)+1) - \frac{1}{2}\ln|\Sigma_{2n}| - n\ln\sigma_b^{2*}(\lambda_b) \\ &\quad + \ln(\mathbf{K}_{2n} S_{b,n}(\lambda_b)). \end{aligned}$$

Thus, the optimal solutions of $Q_{b,n}(\lambda_b)$ are

$$\begin{aligned} &\beta_b^*(\lambda_b) = (X'_{b,n}\mathcal{W}_{2n}X_{b,n})X'_{b,n}\mathcal{W}_{2n}S_{b,n}(\lambda_b)S_{b,n}^{-1}X_{b,n}\beta_{b,0}, \\ &\sigma_b^{2*}(\lambda_b) = \frac{1}{2n}E\left\{[S_{b,n}(\lambda_b)Y_{b,n}\right. \\ &\left.-X_{b,n}\beta_b^*(\lambda_b)]'\mathcal{W}_{2n}[S_{b,n}(\lambda_b)Y_{b,n} - X_{b,n}\beta_b^*(\lambda_b)]\right\} \\ &= \frac{1}{2n}E[(G_{b,n}X_{b,n}\beta_b)'(\lambda_{b,0}-\lambda_b)' \\ &\otimes l_n M_{b,n}(\lambda_{b,0}-\lambda_b)\otimes l_n(G_{b,n}X_{b,n}\beta_b)] \\ &+ \frac{1}{n}E[(X_{b,n}\beta_b)'M_{b,n}(\lambda_{b,0}-\lambda_b)\otimes l_n(G_{b,n}X_{b,n}\beta_b)] \\ &+ \frac{1}{2n}E[(X_{b,n}\beta_b)'M_{b,n}(X_{b,n}\beta_b)] \\ &+ \frac{\sigma^2}{2n}E[S_{b,n}'^{-1}S'_{b,n}(\lambda_b)S_{b,n}(\lambda_b)S_{b,n}^{-1}]. \end{aligned}$$

Then, $\frac{1}{2n}[\ln L_{b,n}(\lambda_b) - Q_{b,n}(\lambda_b)] = -\frac{1}{2}[\ln \hat{\sigma}^2(\lambda_b) - \ln \sigma_b^{2*}(\lambda_b)]$, where

$$\begin{aligned}
\hat{\sigma}^2(\lambda_b) = &\frac{1}{2n} Y'_{b,n} S'_{b,n}(\lambda_b) M_{b,n} S_{b,n}(\lambda_b) Y_{b,n} \\
= &\frac{1}{2n}(G_{b,n} X_{b,n}\beta_b)'(\lambda_{b,0} - \lambda_b)' \otimes l_n M_{b,n} \\
&\times (\lambda_{b,0} - \lambda_b) \otimes l_n (G_{b,n} X_{b,n}\beta_b) \\
&+ \frac{1}{n}(X_{b,n}\beta_b)' M_{b,n}(\lambda_{b,0} - \lambda_b) \otimes l_n (G_{b,n} X_{b,n}\beta_b) \\
&+ \frac{1}{2n}(X_{b,n}\beta_b)' M_{b,n}(X_{b,n}\beta_b) \\
&+ \frac{1}{2n} V'_{b,n} S'^{-1}_{b,n} S'_{b,n} \mathcal{W}_{2,n} S_{b,n}(\lambda_b) S^{-1}_{b,n} V_{b,n} \\
&+ \frac{1}{n}(X_{b,n}\beta_b)' M_{b,n} S_{b,n}(\lambda_b) S^{-1}_{b,n} V_{b,n} \\
&+ \frac{1}{n}(G_{b,n} X_{b,n}\beta_b)'(\lambda_{b,0} - \lambda_b)' \\
&\times \otimes l_n M_{b,n} S_{b,n}(\lambda_b) S^{-1}_{b,n} V_{b,n}.
\end{aligned}$$

According to Lee (2004),

$$\begin{aligned}
&\frac{1}{n}(G_{b,n} X_{b,n}\beta_b)' M_{b,n} S_{b,n}(\lambda_b) S^{-1}_{b,n} V_{b,n} \\
&= o_p(1); \quad \frac{1}{n}(X_{b,n}\beta_b)' M_{b,n}(G_{b,n} X_{b,n}\beta_b) = o_p(1), \\
&\frac{1}{2n} V'_{b,n} S'^{-1}_{b,n} S'_{b,n} \mathcal{W}_{2,n} S_{b,n}(\lambda_b) S^{-1}_{b,n} V_{b,n} \\
&= \frac{\sigma^2}{2n} E[S'^{-1}_{b,n} S'_{b,n}(\lambda_b) S_{b,n}(\lambda_b) S^{-1}_{b,n}] + o_p(1).
\end{aligned}$$

Hence,

$$\hat{\sigma}^2(\lambda_b) - \sigma_b^*(\lambda_b) = o_p(1) \text{ uniformly on } \Lambda_b, \tag{A3}$$

and so that (A1) holds.

(2) Proof of (A2). Consider the pure spatial autoregressive process, i.e., let $\beta_b = 0$ in Eq.(3), $Y_{b,n} = \lambda_b \otimes l_n W_{b,n} + V_{b,n}$, and $V_{b,n} \sim \mathcal{N}(0, \Sigma_{2n}\sigma^2)$, where the matrix Σ_{2n} is known but the scale σ^2 is unknown. Denote the log-likelihood function of this process as $\ln L^p_{b,n}(\lambda_b, \sigma^2)$, it follows that

$$\begin{aligned}
\ln L^p_{b,n}(\lambda_b, \sigma^2) = &-n\ln(2\pi) - \frac{1}{2}\ln|\Sigma_{2n}| \\
&- n\ln\sigma^2 + \ln|S_{b,n}(\lambda_b)| \\
&- \frac{1}{2\sigma^2} Y'_{b,n} S'_{b,n}(\lambda_b) \mathcal{W}_{2n} S'_{b,n}(\lambda_b) Y_{b,n}
\end{aligned}$$

The optimal solution of $\max_{\sigma^2} \ln L_{b,n}(\lambda_b, \sigma^2)$ is

$$\tilde{\sigma}^2 = \frac{\sigma_0^2}{2n} S'^{-1}_{b,n} S'_{b,n}(\lambda_b) \mathbf{K}_{2n} S'_{b,n}(\lambda_b) S^{-1}_{b,n} \tag{A4}$$

Let $\tilde{Q}^p_{b,n}(\lambda_b) := \max_{\sigma^2} E(\ln L^p_{b,n}(\lambda_b, \sigma^2)) = -n\ln(2\pi + 1) - \frac{1}{2}\ln|\Sigma_{2n}| - n\ln\tilde{\sigma}^2 + \ln|S_{b,n}(\lambda_b)|$, it follows that $\tilde{Q}^p_{b,n}(\lambda_b) \le \max_{\lambda_b,\sigma^2} E(\ln L^p_{b,n}(\lambda_b, \sigma^2)) = E(\ln L^p_{b,n}(\lambda_{b,0}, \sigma_0^2)) = \tilde{Q}^p_{b,n}(\lambda_{b,0})$, and, therefore,

$$\frac{1}{2n}(\tilde{Q}^p_{b,n}(\lambda_b) - \tilde{Q}^p_{b,n}(\lambda_{b,0})) \le 0 \text{ uniformly on } \Lambda_b, \tag{A5}$$

As $Y_{b,n} = S^{-1}_{b,n} X_{b,n}\beta_{b,0} + S^{-1}_{b,n} V_{b,n}$, $S_{b,n}(\lambda_b) = S_{b,n} + (\lambda_{b,0} - \lambda_b) \otimes l_n W_{b,n}$, it follows that

$$\begin{aligned}
V_{b,n}(\delta_b) = &S_{b,n}(\lambda_b) Y_{b,n} - X_{b,n}\beta_b \\
= &X_{b,n}(\beta_{b,0} - \beta_b) + (\lambda_{b,0} - \lambda_b) \\
&\otimes l_n W_{b,n} Y_{b,n} + V_{b,n},
\end{aligned}$$

and

$$\begin{aligned}
&V'_{b,n}(\delta_b) \mathcal{W}_{2n} V_{b,n}(\delta_b) \\
&= (\beta_{b,0} - \beta_b)' X'_{b,n} \mathcal{W}_{2n} X_{b,n}(\beta_{b,0} - \beta_b) \\
&\quad + Y'_{b,n} W'_{b,n} \tilde{\mathcal{W}} W_{b,n} Y_{b,n} + V'_{b,n} \mathcal{W}_{2n} V_{b,n} \\
&\quad + 2(\beta_{b,0} - \beta_b)' X'_{b,n} \mathcal{W}_{2n}(\lambda_{b,0} - \lambda_b) \otimes l_n W_{b,n} Y_{b,n} \\
&\quad + 2(\beta_{b,0} - \beta_b)' X'_{b,n} \mathcal{W}_{2n} V_{b,n} \\
&\quad + 2Y'_{b,n} W'_{b,n}(\lambda_{b,0} - \lambda_b)' \otimes l_n \mathcal{W}_{2n} V_{b,n}
\end{aligned}$$

where $\tilde{\mathcal{W}} = (\lambda_{b,0} - \lambda_b)' \otimes l_n \mathcal{W}_{2n}(\lambda_{b,0} - \lambda_b) \otimes l_n$. The term in (A2) can be rewritten as

$$\begin{aligned}
\frac{1}{2n}(Q_{b,n}(\lambda_b) - Q_{b,n}(\lambda_{b,0})) = &\frac{1}{2n}(\tilde{Q}^p_{b,n}(\lambda_b) - \tilde{Q}^p_{b,n}(\lambda_{b,0})) \\
&- \frac{1}{2}[\ln(\sigma_b^*(\lambda_b)) - \ln(\hat{\sigma}^2(\lambda_b))]
\end{aligned} \tag{A6}$$

From (1) Proof of (A1) and Eq.(A3), $\sigma_b^*(\lambda_b) - \hat{\sigma}^2(\lambda_b) \le 0$, so that the last term in above equality $[\ln(\sigma_b^*(\lambda_b)) - \ln(\hat{\sigma}^2(\lambda_b))] \le 0$. Combined with A5, (A2) holds.

If the identification uniqueness condition was not satisfied, without loss generality, there would exist a sequence $\lambda_{b,n}$ converging to $\lambda_{b,+} \ne 0$ such that $\lim_{n\to\infty} \frac{1}{2n}(Q_{b,n}(\lambda_b) - Q_{b,n}(\lambda_{b,0})) = 0$. This would be possible be only if $\lim_{n\to\infty} [\ln(\sigma_b^*(\lambda_{b,n})) - \ln(\hat{\sigma}^2(\lambda_{b,n}))] = 0$ and $\lim_{n\to\infty} \frac{1}{2n}(\tilde{Q}^p_{b,n}(\lambda_{b,n}) - \tilde{Q}^p_{b,n}(\lambda_{b,0})) = 0$. However, it would be generate a contradiction due to Assumption 10.

B Proofs of Theorem 2.

The asymptotic distribution of the QMLE $\hat{\theta}_{b,n}$ is derived from the Taylor expansion $\frac{\partial \ln L_{b,n}(\hat{\theta}_{b,n})}{\partial \theta_{b,n}} = 0$.

$$\sqrt{2n}(\hat{\theta}_{b,n} - \theta_{b,n0}) = -\left\{\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \theta_{b,n}\partial \theta'_{b,n}}\right\}^{-1} \frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}} + o_p(1). \tag{A7}$$

where $\tilde{\theta}_{b,n}$ lies between $\hat{\theta}_{b,n}$ and $\theta_{b,n0}$. The asymptotic distribution of the QMLE $\hat{\theta}_{b,n}$, Theorem 2, holds only if

$$\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}} \xrightarrow{d} (0, \Sigma_{\theta_{b,n}} + \Omega_{\theta_{b,n}}) \tag{A8}$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \theta_{b,n}\partial \theta'_{b,n}} - E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta'_{b,n}}\right) \xrightarrow{p} 0 \tag{A9}$$

(1) Derivation of (A8). By Lemma C.1 and Lemma A.10 in Lee (2004), at $\theta_{b,n0}$, the first-order derivatives of the log-likelihood function in (12) only involve both linear and quadratic functions of $V_{b,n}$. Then, using central limit theorem, the variance matrix of the score vector in $\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}}$ is

$$E\left(\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}} \cdot \frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}}\right) = -E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta'_{b,n}}\right) + \Omega_{\theta_{b,n}},$$

where

$$-E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta'_{b,n}}\right) = \begin{pmatrix} \frac{1}{2n\sigma_0^2}X'_{b,n}\mathcal{W}_{2n}X_{b,n} & * & * & * \\ \frac{1}{2n\sigma_0^2}(X_{b,n}\beta_{b,0})'G'_{b,n}\odot \mathbf{e}_1\mathcal{W}_{2n}X_{b,n} & \mathbf{T}_{22,1} & * & * \\ \frac{1}{2n\sigma_0^2}(X_{b,n}\beta_{b,0})'G'_{b,n}\mathbf{e}_2\odot \mathcal{W}_{2n}X_{b,n} & 0 & \mathbf{T}_{22,2} & * \\ 0 & \frac{1}{2n\sigma_0^2 tr(G'_{b,n}\odot \mathbf{e}_1)} & \frac{1}{2n\sigma_0^2 tr(G'_{b,n}\odot \mathbf{e}_2)} & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

where $\mathbf{T}_{22,1} = \frac{1}{2n\sigma^2_{b,0}}(X_{b,n}\beta_{b,0})'G'_{b,n}\odot \mathbf{e}_1\mathcal{W}_{2n}\mathbf{e}_1\odot G_{b,n}(X_{b,n}\beta_{b,0}) + \frac{1}{2n}[tr(\mathbf{e}_1\odot G'_{b,n}\mathbf{e}_1\odot G_{b,n}) + tr(\mathbf{e}_1\odot G_{b,n}\mathbf{e}_1\odot G_{b,n})]$ and $\mathbf{T}_{22,2} = \frac{1}{2n\sigma^2_{b,0}}(X_{b,n}\beta_{b,0})'G'_{b,n}\odot \mathbf{e}_2\mathcal{W}_{2n}\mathbf{e}_2\odot G_{b,n}(X_{b,n}\beta_{b,0}) + \frac{1}{2n}[tr(\mathbf{e}_2\odot G'_{b,n}\mathbf{e}_2\odot G_{b,n}) + tr(\mathbf{e}_2\odot G_{b,n}\mathbf{e}_2\odot G_{b,n})]$, is the average Hessian matrix which is non-singular due to Assumption 10, and

$$\Omega_{\theta_{b,n}} = \begin{pmatrix} 0 & * & * & * \\ \frac{\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^1_{b,iin}X_{b,in} & \Omega_{\theta_{b,22n}} & * & * \\ \frac{\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^2_{b,iin}X_{b,in} & 0 & \Omega_{\theta_{b,33n}} & * \\ \frac{\mu_3}{4n\sigma_0^6}l_{2n}\Sigma^{\frac{1}{2}}\mathbf{L}_{2n}X_{b,n} & \frac{\mu_3}{4n\sigma_0^6}l_{2n}\tilde{G}_{b,in}X_{b,n}\beta_{b,0} + \frac{\mu_3}{4n\sigma_0^6}tr(G_{b,n}) & \frac{\mu_4 - 3\sigma_0^4}{8n\sigma_0^8} & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

where

$$\Omega_{\theta_{b,22n}} = \frac{2\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^1_{b,iin}\tilde{G}^1_{b,in}X_{b,n}\beta_b + \frac{\mu_4 - 3\sigma_0^4}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^1_{b,iin}$$

$$\Omega_{\theta_{b,33n}} = \frac{2\mu_3}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^2_{b,iin}\tilde{G}^2_{b,in}X_{b,n}\beta_b + \frac{\mu_4 - 3\sigma_0^4}{2n\sigma_0^4}\sum_{i=1}^n \tilde{G}^2_{b,iin}$$

with $\tilde{G}^1_{b,n} = \Sigma^{\frac{1}{2}}\mathbf{L}_{2n}G_{b,n}\odot \mathbf{e}_1$ and $\tilde{G}^2_{b,n} = \Sigma^{\frac{1}{2}}\mathbf{L}_{2n}G_{b,n}\odot \mathbf{e}_2$. Specifically, $\Omega_{\theta_{b,n}} = 0$ as $V_{b,n}$ is normally distributed.

Hence, combining the central limit theorem for quadratic forms of double array (Kelejian and Prucha 2001) and Kolmogorov's central limit theorem,

$$\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}} \xrightarrow{d} (0, \Sigma_{\theta_{b,n}} + \Omega_{\theta_{b,n}}) \tag{A10}$$

where

$$\Sigma_{\theta_{b,n}} = -\lim_{n\to\infty} E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta_{b,n}'}\right) \quad \text{and}$$

$$\Omega_{\theta_{b,n}} = -\lim_{n\to\infty} \Omega_{\theta_{b,n}}$$

(2) Derivation of (A9).

For any $\tilde{\theta}_{b,n}$ which converges in probability to $\theta_{b,n0}$,

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \theta_{b,n}\partial \theta_{b,n}'} - E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta_{b,n}'}\right) \xrightarrow{p} 0 \tag{A11}$$

holds if and only if

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \theta_{b,n}\partial \theta_{b,n}'} - \frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta_{b,n}'} \xrightarrow{p} 0 \tag{A12}$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta_{b,n}'} - E\left(\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \theta_{b,n}\partial \theta_{b,n}'}\right) \xrightarrow{p} 0 \tag{A13}$$

hold.

As $\mathbf{K}_{2n}$ and $\Sigma_{b,n}$ are two non-stochastic bounded matrices, $\frac{X_{b,n}'\mathcal{W}_{2n}X_{b,n}}{2n} = O_P(1)$, $\frac{X_{b,n}'\mathcal{W}_{2n}\mathbf{e}_2 \odot W_{b,n}Y_{b,n}}{2n} = O_P(1)$ and $\tilde{\sigma}^2 \xrightarrow{p} \sigma_0^2$, By Lemma C.1,

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \beta_b \partial \beta_b'} - \frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \beta_b \partial \beta_b'}$$
$$= (\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}^2})\frac{X_{b,n}'\mathcal{W}_{2n}X_{b,n}}{2n} = o_p(1).$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \beta_b \partial \lambda_b'} - \frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \beta_b \partial \lambda_b'}$$
$$= (\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}^2})\left(\frac{X_{b,n}'\mathcal{W}_{2n}\mathbf{e}_1 \odot W_{b,n}Y_{b,n}}{2n} \quad \frac{X_{b,n}'\mathcal{W}_{2n}\mathbf{e}_2 \odot W_{b,n}Y_{b,n}}{2n}\right)$$
$$= o_p(1).$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \beta_b \partial \sigma^2} - \frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \beta_b \partial \sigma^2}$$
$$= (\frac{1}{\sigma_0^4} - \frac{1}{\tilde{\sigma}^4})\frac{X_{b,n}'\mathcal{W}_{2n}V_{b,n}}{2n} + \frac{1}{2\tilde{\sigma}^4}\frac{X_{b,n}'\mathcal{W}_{2n}X_{b,n}}{2n}$$
$$\times (\tilde{\beta}_b - \beta_{b,0}) + \frac{1}{\tilde{\sigma}^4}\frac{X_{b,n}'\mathcal{W}_{2n}(\tilde{\lambda}_b - \lambda_{b,0}) \otimes X_{b,n}}{2n}$$
$$= o_p(1).$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \lambda_b \partial \lambda_b'} - \frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \lambda_b \partial \lambda_b'} = \begin{pmatrix} -2\frac{tr(\mathbf{e}_1 G_{b,n}^3(\bar{\lambda}_b))}{2n} & 0 \\ 0 & -2\frac{tr(\mathbf{e}_2 G_{b,n}^3(\bar{\lambda}_b))}{2n} \end{pmatrix}\lambda_b'$$
$$+ (\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}^2})\begin{pmatrix} \frac{Y_{b,n}'\mathcal{W}_{b,n}' \odot \mathbf{e}_1\mathcal{W}_{2n}\mathbf{e}_1 \odot \mathcal{W}_{b,n}Y_{b,n}}{2n} & 0 \\ 0 & \frac{Y_{b,n}'\mathcal{W}_{b,n}' \odot \mathbf{e}_2\mathcal{W}_{2n}\mathbf{e}_2 \odot \mathcal{W}_{b,n}Y_{b,n}}{2n} \end{pmatrix} = o_p(1)$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial \sigma^2 \partial \lambda_b} - \frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial \sigma^2 \partial \lambda_b} = \left(\frac{Y_{b,n}'W_{b,n}' \otimes \mathbf{e}_1\mathcal{W}_{2n}X_{b,n}}{2n\sigma^4}(\tilde{\beta}_b - \tilde{\beta}_{b0}) \quad \frac{Y_{b,n}'W_{b,n}' \otimes \mathbf{e}_2\mathcal{W}_{2n}X_{b,n}}{2n\sigma^4}(\tilde{\beta}_b - \tilde{\beta}_{b0})\right)$$
$$+ \left(\frac{Y_{b,n}'W_{b,n}' \otimes \mathbf{e}_1\mathcal{W}_{2n}(\tilde{\lambda}_b - \tilde{\lambda}_{b,0}) \otimes W_{b,n}Y_{b,n}}{2n\sigma^4} \quad \frac{Y_{b,n}'W_{b,n}' \otimes \mathbf{e}_2\mathcal{W}_{2n}(\tilde{\lambda}_b - \tilde{\lambda}_{b,0}) \otimes W_{b,n}Y_{b,n}}{2n\sigma^4}\right)$$
$$+ (\frac{1}{\sigma_0^4} - \frac{1}{\tilde{\sigma}^4})\left(\frac{Y_{b,n}'W_{b,n}' \otimes \mathbf{e}_1\mathcal{W}_{2n}V_{b,n}}{2n\sigma^4} \quad \frac{Y_{b,n}'W_{b,n}' \otimes \mathbf{e}_2\mathcal{W}_{2n}V_{b,n}}{2n\sigma^4}\right) = o_p(1)$$

and

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\tilde{\theta}_{b,n})}{\partial\sigma^2\partial\lambda_b} - \frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\sigma^2\partial\sigma^2}$$
$$= \frac{1}{2}(\frac{1}{\tilde{\sigma}^4} - \frac{1}{\sigma_0^4}) + (\frac{1}{\sigma_0^6} - \frac{1}{\tilde{\sigma}^6})\frac{V'_{b,n}\mathcal{W}_{2,n}V_{b,n}}{2n} + o_P(1)$$
$$= o_p(1).$$

Thus, the convergence of (A12) holds. Besides, the equation (A13) also holds. Because it is straightforward by showing that linear functions and quadratic functions of $V_{b,n}$, deviated from their means, e.g., $\frac{X'_{b,n}G_{b,n}\otimes e_1\mathbf{K}_{2n}V_{b,n}}{2n}$, $\frac{X'_{b,n}G_{b,n}\otimes e_2\mathbf{K}_{2n}V_{b,n}}{2n}$ and $\frac{V'_{b,n}\mathbf{K}_{2n}V_{b,n}}{2n} - \sigma_0^2\,(tr(\mathbf{e}_1\odot G_{b,n})tr(\mathbf{e}_2\odot G_{b,n})$, are all $o_p(1)$.

C Main Lemmas and Their Proofs

Lemma C.1 *The first order derivative of the joint log-likelihood function (13) at $\theta_{b,n0}\odot$ is*

$$\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial\theta_{b,n}} = \begin{pmatrix} H_1 \\ H_2 \\ H_3 \end{pmatrix},$$

where

$$H_1 = \frac{1}{\sigma_0^2\sqrt{2n}}X'_{b,n}\mathcal{W}_{2n}V_{b,n}$$

$$H_2 = \frac{1}{\sigma_0^2\sqrt{2n}}\left((\mathbf{e}_1\odot G_{b,n}X_{b,n}\beta_{b,0})'\mathcal{W}_{2n}V_{b,n}\right.$$
$$\left.\times\ (\mathbf{e}_2\odot G_{b,n}X_{b,n}\beta_{b,0})'\mathcal{W}_{2n}V_{b,n}\right) - \frac{1}{\sigma_0^2\sqrt{2n}}$$
$$\times\left(\left(V'_{b,n}G_{b,n}\odot\mathbf{e}_1\mathcal{W}_{2n}V_{b,n}\ \ V'_{b,n}G_{b,n}\odot\mathbf{e}_1\mathcal{W}_{2n}V_{b,n}\right)'\right.$$
$$\left.-\ \sigma_0^2\left(tr(\mathbf{e}_1\odot G_{b,n})\ tr(\mathbf{e}_2\odot G_{b,n})\right)'\right.$$

$$H_3 = \frac{1}{2\sigma_0^4\sqrt{2n}}(V'_{b,n}\mathcal{W}_{2n}V_{b,n} - 2n\sigma_0^2).$$

Proof Since

$$\frac{\partial \ln V_{b,n}(\delta_b)}{\beta_b} = -X_{b,n},$$
$$\frac{\partial \ln V_{b,n}(\delta_b)}{\lambda_b} = -(\mathbf{e}_1\odot W_{b,n}Y_{b,n},\mathbf{e}_2\odot W_{b,n}Y_{b,n}),$$
$$\frac{\partial \ln\|S_{b,n}(\lambda_b)\|}{\lambda_b} = -(tr(\mathbf{e}_1\odot W_{b,n}S^{-1}_{b,n}(\lambda_b)),$$
$$tr(\mathbf{e}_2\odot W_{b,n}S^{-1}_{b,n}(\lambda_b)))$$

where $\mathbf{e}_1$ is an $2n\times 2n$ block diagonal matrix whose the first block diagonal element is an $n\times n$ dimensional identity matrix and the second block diagonal element is an $n\times n$ dimensional zero matrix, and $\mathbf{e}_2$ is an $2n\times 2n$ block diagonal matrix whose the first block diagonal element is an $n\times n$ dimensional zero matrix, and second block diagonal element is an $n\times n$ dimensional identity matrix.

Hence, the first order derivatives of the joint log-likelihood function (13) at $\beta_{b,0}$, $\lambda_{b,0}$, and σ^2 are

$$\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial\beta_b} = \frac{1}{\sigma_0^2\sqrt{2n}}X'_{b,n}\mathcal{W}_{2n}V_{b,n},$$

$$\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial\lambda_b}$$
$$= \frac{1}{\sigma_0^2\sqrt{2n}}\left(\mathbf{e}_1\odot G_{b,n}X_{b,n}\beta_{b,0})'\mathcal{W}_{2n}V_{b,n}\right.$$
$$\left.\times(\mathbf{e}_2\odot G_{b,n}X_{b,n}\beta_{b,0})'\mathcal{W}_{2n}V_{b,n}\right) - \frac{1}{\sigma_0^2\sqrt{2n}}$$
$$\left(\left(V'_{b,n}G_{b,n}\odot\mathbf{e}_1\mathcal{W}_{2n}V_{b,n}\ \ V'_{b,n}G_{b,n}\odot\mathbf{e}_1\mathcal{W}_{2n}V_{b,n}\right)\right.$$
$$\left.-\sigma_0^2\left(tr(\mathbf{e}_1\odot G_{b,n})\ tr(\mathbf{e}_2\odot G_{b,n})\right)\right),$$

$$\frac{1}{\sqrt{2n}}\frac{\partial \ln L_{b,n}(\theta_{b,n0})}{\partial\sigma^2} = \frac{1}{2\sigma_0^4\sqrt{2n}}(V'_{b,n}\mathcal{W}_{2n}V_{b,n} - 2n\sigma_0^2).$$

□

Lemma C.2 *The second order derivative of the joint log-likelihood function (13) at $\theta_{b,n0}$ is*

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\theta_{b,n}\partial\theta_{b,n}'}=\begin{pmatrix} -\frac{1}{2n\sigma_0^2}X_{b,n}'\mathcal{W}_{2n}X_{b,n} & * & * & * \\ -\frac{1}{2n\sigma_0^2}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_1\mathcal{W}_{2n}X_{b,n} & -T_{22,1} & * & * \\ -\frac{1}{2n\sigma_0^2}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_2\mathcal{W}_{2n}X_{b,n} & 0 & -T_{22,2} & * \\ -\frac{1}{2n\sigma_0^4}V_{b,n}'\mathcal{W}_{2n}X_{b,n} & -\frac{1}{2n\sigma_0^4}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_1\mathcal{W}_{2n}V_{b,n} & -\frac{1}{2n\sigma_0^4}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_2\mathcal{W}_{2n}V_{b,n} & T_{22,3} \end{pmatrix}.$$

where

$$T_{22,1}=\frac{1}{2n}tr(\mathbf{e}_1\odot G_{b,n}^2)+\frac{1}{2n\sigma_0^2}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_1\mathcal{W}_{2n}\mathbf{e}_1\odot W_{b,n}Y_{b,n}$$

$$T_{22,2}=\frac{1}{2n}tr(\mathbf{e}_2\odot G_{b,n}^2)+\frac{1}{2n\sigma_0^2}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_2\mathcal{W}_{2n}\mathbf{e}_2\odot W_{b,n}Y_{b,n}$$

$$T_{22,3}=\frac{1}{2\sigma_0^4}-\frac{1}{2n\sigma_0^6}(V_{b,n}'\mathcal{W}_{2n}V_{b,n})$$

Proof

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\beta_b\partial\beta_b'}=-\frac{1}{2n\sigma_0^2}X_{b,n}'\mathcal{W}_{2n}X_{b,n},$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\beta_b\partial\lambda_b'}=\left(-\frac{1}{2n\sigma_0^2}X_{b,n}'\mathcal{W}_{2n}\mathbf{e}_1\odot W_{b,n}Y_{b,n}\quad -\frac{1}{2n\sigma_0^2}X_{b,n}'\mathcal{W}_{2n}\mathbf{e}_2\odot W_{b,n}Y_{b,n}\right),$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\beta_b\partial\sigma^2}=-\frac{1}{2n\sigma_0^4}X_{b,n}'\mathcal{W}_{2n}V_{b,n},$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\lambda_b\partial\lambda_b'}=-\begin{pmatrix}T_{22,1} & 0\\ 0 & T_{22,2}\end{pmatrix}$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial\sigma^2\partial\lambda_b}=\left(-\frac{1}{2n\sigma_0^4}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_1\mathcal{W}_{2n}V_{b,n}\quad -\frac{1}{2n\sigma_0^4}Y_{b,n}'W_{b,n}'\odot\mathbf{e}_2\mathcal{W}_{2n}V_{b,n}\right)$$

$$\frac{1}{2n}\frac{\partial^2 \ln L_{b,n}(\theta_{b,n0})}{\partial^2\sigma^2}=\frac{1}{2\sigma_0^4}-\frac{1}{2n\sigma_0^6}(V_{b,n}'\mathcal{W}_{2n}V_{b,n})$$

□

Acknowledgements We are grateful to the editor, associate editor and anonymous review for their valuable comments and suggestions, which have greatly improved the manuscript. This research was supported by the National Statistical Science Research Project (2023LY025), Operations Research and Management Innovation Team of Hebei University (IT2023C02), and the Innovation Capacity Enhancement Program-Science and Technology Platform Project (22567623H).

Author Contributions Jinjin Zhang, Kun Li and Aibing Ji conceived the methodology, Jinjin Zhang wrote the manuscript and software in close collaboration with both Kun Li and Aibing Ji. All authors reviewed the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflicts of Interest The authors declare that they have no conflict of interest.

Competing interests The authors declare no competing interests.

References

Anselin, L.: Spatial Econometrics: Methods and Models. Springer, Netherlands, Dordrecht (1988)

Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathware and Soft Computing **2**(2), 71–84 (1995)

Billard, L., Diday, E.: Regression analysis for interval-valued data. In: Kiers, H.A.L., Rasson, J.-P., Groenen, P.J.F., Schader, M. (eds.) Data Analysis, Classification, and Related Methods, Berlin, Heidelberg, pp. 369–374. Springer, Berlin Heidelberg (2000)

Billard, L., Diday, E.: Symbolic regression analysis. In: Jajuga, K., Sokołowski, A., Bock, H.-H. (eds.) Classification, Clustering, and Data Analysis, Berlin, Heidelberg, pp. 281–288. Springer, Berlin Heidelberg (2002)

Dawn, S., Bandyopadhyay, S.: IV-GNN: interval valued data handling using graph neural network. Appl. Intell. **53**(5), 5697–5713 (2023)

D'Urso, P., De Giovanni, L., Federico, L., Vitale, V.: Fuzzy clustering of spatial interval-valued data. Spatial Statistics **57**, 100764 (2023)

Freitas, W., Souza, R., Amaral, G., Moraes, R.: Regression applied to symbolic interval-spatial data. Appl. Intell. **54**, 1–21 (2024)

Freitas, W.W., de Souza, R.M., Amaral, G.J., De Bastiani, F.: Exploratory spatial analysis for interval data: A new autocorrelation index with covid-19 and rent price applications. Expert Syst. Appl. **195**, 116561 (2022)

Giordani, P.: Lasso-constrained regression analysis for interval-valued data. Adv. Data Anal. Classif. **9**, 5–19 (2015)
Guo, J., Qu, X.: Spatial interactive effects on housing prices in shanghai and beijing. *Regional Science and Urban Economics 76*, 147–160 (2019). Spatial Econometrics: New Methods and Applications
Han, A., Hong, Y., Wang, S.: Autoregressive conditional models for interval-valued time series data. In *The 3rd international conference on singular spectrum analysis and its applications*, Volume 27 (2012)
Han, A., Hong, Y., Wang, S., Yun, X.: A vector autoregressive moving average model for interval-valued time series data. In *Essays in honor of Aman Ullah*, pp. 417–460. Emerald Group Publishing Limited (2016)
Huang, T.: A spatial durbin model for interval-valued data with t-distribution. J. Stat. Comput. Simul. **94**(18), 4037–4071 (2024)
Kelejian, H.H., Prucha, I.R.: On the asymptotic distribution of the moran i test statistic with applications. Journal of Econometrics **104**(2), 219–257 (2001)
Körner, R., Näther, W.: On the variance of random fuzzy variables. In *Statistical modeling, analysis and management of fuzzy data*, pp. 25–42. Springer (2002)
Körner, R.: On the variance of fuzzy random variables. Fuzzy Sets Syst. **92**(1), 83–93 (1997)
Lee, L.: Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica **72**, 1899–1925 (2004)
Li, Q., Zhang, J., Ji, A.: Time-varying coefficient spatial panel interval-valued models and applications. Commun. Nonlinear Sci. Numer. Simul. **148**, 108880 (2025)
Li, Q., Zheng, R., Ji, A., Ma, H.: Fixed effects spatial panel interval-valued autoregressive models and applications. Spatial Statistics **65**, 100875 (2025)
Li, Y., He, X., Liu, X.: Fuzzy multiple linear least squares regression analysis. Fuzzy Sets Syst. **459**, 118–143 (2023)
Liang, X., Gao, J., Gong, X.: Semiparametric spatial autoregressive panel data model with fixed effects and time-varying coefficients. Journal of Business & Economic Statistics **40**(4), 1–52 (2021)
Lima Neto, E.A., de Carvalho, F.D.A.T.: Nonlinear regression applied to interval-valued data: Pattern Anal. Appl. **20**, 809–824 (2017)
Lima Neto, E., D. A., F. D. A. de Carvalho,: Centre and range method for fitting a linear regression model to symbolic interval data. Computational Statistics & Data Analysis **52**(3), 1500–1515 (2008)
Moore, R.: *Methods and Applications of Interval Analysis*. Studies in Applied and Numerical Mathematics. Society for Industrial and Applied Mathematics (1979)
Neto, E.D.A.L., Carvalho, F.D.A.T.D.: Centre and range method for fitting a linear regression model to symbolic interval data. Computational Statistics & Data Analysis **52**(3), 1500–1515 (2008)
Ni, L., Wang, X.C., Chen, X.M.: A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data. Transportation Research Part C: Emerging Technologies **86**, 510–526 (2018)
Näther, W.: Regression with fuzzy random data. Computational Statistics & Data Analysis **51**(1), 235–252 (2006). The Fuzzy Approach to Statistical Analysis
Sinova, B., Aelst, S.V.: A spatial-type interval-valued median for random intervals. Statistics **52**(3), 479–502 (2018)
Sinova, B., Ángeles Gil, M., López, M.T., Van Aelst, S.: A parameterized l2 metric between fuzzy numbers and its parameter interpretation. Fuzzy Sets Syst. **245**, 101–115 (2014). Theme: Fuzzy Intervals and Fuzzy Analysis
Souza, L.C., Souza, R.M., Amaral, G.J., Silva Filho, T.M.: A parametrized approach for linear regression of interval data. Knowl.-Based Syst. **131**, 149–159 (2017)
Sun, Y., Han, A., Hong, Y., Wang, S.: Threshold autoregressive models for interval-valued time series data. Journal of Econometrics **206**(2), 414–446 (2018). Special issue on Advances in Econometric Theory: Essays in honor of Takeshi Amemiya
Sun, Y., Zhang, X., Hong, Y., Wang, S.: Asymmetric pass-through of oil prices to gasoline prices with interval time series modelling. Energy Economics **78**, 165–173 (2019)
Trutschnig, W., González-Rodríguez, G., Colubi, A., Ángeles Gil, M.: A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Inf. Sci. **179**(23), 3964–3972 (2009)
Xu, M., Qin, Z.: A bayesian parametrized method for interval-valued regression models. Stat. Comput. **33**(3), 67 (2023)
Yii, K.-J., Tan, C.-T., Ho, W.-K., Kwan, X.-H., Nerissa, F.-T.S., Tan, Y.-Y., Wong, K.-H.: Land availability and housing price in china: Empirical evidence from nonlinear autoregressive distributed lag (nardl). Land Use Policy **113**, 105888 (2022)
Zhao, Y.-Y., Ge, L.-L., Liu, Y.: Estimation of panel data partially linear time-varying coefficient models with cross-sectional spatial autoregressive errors. Stat. Pap. **66**(1), 1–37 (2025)