Threshold spatial autoregressive model

Li, Kunpeng

27 June 2022

Threshold spatial autoregressive model*

Kunpeng Li

International School of Economics and Management
Capital University of Economics and Business

June 26, 2022

Abstract

This paper considers the estimation and inferential issues of threshold spatial autoregressive model, which is a hybrid of threshold model and spatial autoregressive model. We consider using the quasi maximum likelihood (QML) method to estimate the model. We prove the tightness and the Hájek-Rényi type inequality for a quadratic form, and establish a full inferential theory of the QML estimator under the setup that threshold effect shrinks to zero along with an increasing sample size. We consider the hypothesis testing on the presence of threshold effect. Three super-type statistics are proposed to perform this testing. Their asymptotic behaviors are studied under the Pitman local alternatives. A bootstrap procedure is proposed to obtain the asymptotically correct critical value. We also consider the hypothesis testing on the threshold value equal to some prespecified one. We run Monte carlo simulations to investigate the finite sample performance of the QML estimators and find that the QML estimators have good performance.

Key Words: Spatial autoregressive models, Spillover effects, Threshold effect, Maximum likelihood estimation, Inferential theory.

JEL: C31; C33.

*This paper was presented at Tsinghua University, Shanghai University of Finance and Economics and Capital University of Economics and Business. The author would like to thank Xiaohong Chen, Lung-fei Lee, Liangjun Su, Aman Ullah and the participants of the seminars in these universities for their helpful comments. The author's email address: kunpenglithu@126.com.

1 Introduction

This paper considers the estimation and inference issues of threshold spatial autoregressive (TSAR) models, which is a hybrid of threshold model and spatial autoregressive (SAR) model. Both threshold model and SAR model have over forty-year histories, and gain great popularities in empirical studies. Threshold models are widely used in scenarios in the presence of asymmetric effects or multiple equilibria. As a typical example, it is well documented in the literature that the economic structures may switch between recession and expansion. Putting the threshold variable to be an indicator of the economic state, and one regime denotes recession, the other regime expansion, threshold models effectively capture the asymmetric relationships of economic variables in recession and in expansion in a unified way (see, e.g., Marcucci and Quagliariello (2009)). Another typical application of threshold model is to characterize the multiple equilibria of growth rates across the countries. It is argued in the economic growth theories that there exist multiple balanced growth pathes. One can use threshold model to study this multiple-path feature by making a one-to-one correspondence of model regime and growth path, see Masanjala and Papageorgiou (2004) for a related study.

Spatial models are widely used to characterize spatial interactions, or quantify spillover effects. In some studies such as tax competition or expenditure competition among local federal governments, spatial interactions are believed to play an important role to explain the observed government behavior. Spatial models are also used to study peer effects due to the seminal work of Manski (1993), which derives a spatial econometric specification from a linear-in-mean utility function. Calvó-Armengol, Patacchini and Zenou (2009) further prove that the spatial specification can be derived from a quadratic utility function. Since Manski (1993), spatial models are one of primary tool to investigate social network, see Bramoullé, Djebbari and Fortin (2009), Lin (2010), among others.

Both threshold models and spatial econometrics have experienced substantial theoretical advancement over last four decades. Threshold models can date back to Tong (1978). Early studies of this active branch focus on threshold autoregressive (TAR) models with a fixed threshold effect, see Tsay (1989, 1998), Chan (1993), Chan and Tsay (1998). An undesirable consequence of this fixed-value assumption is that the limiting distribution of the estimator for threshold value depends on many nuisance parameters, which causes trouble to statistical inference. In the influential work of Hansen (2000), he considers a cross-sectional regression model under the shrinking assumption, that is threshold effect shrinks to zero along the sample size with a suitable convergence rate. After Hansen (2000), various threshold models have been investigated. Caner and Hansen (2004) and Seo and Shin (2016) consider endogeneity issue in threshold models, and propose the instrumental variable (IV) method or the generalized method of moments (GMM). Su and Xu (2017) consider a quantile regression with threshold effect. Miao, Li and Su (2020) consider threshold panel data model in presence of interactive effects. Spatial econometrics is another active branch in the econometric literature. Early developments on spatial models have been summarized in several monographes, such as Cliff and Ord (1973), Anselin (1988), etc. Kelejian and Prucha (1998, 1999) propose the IV and GMM methods to address the endogeneity issue arising from the spatial term. Lee (2004) considers the quasi maximum likelihood (QML) estimation and establishes the associated asymptotic theory. Other QML studies on spatial models include Yu et al. (2008), Lee and Yu (2010), Li (2018), to name a few.

In this paper, we consider QML method to estimate a TSAR model. Following Hansen (2000), we consider the shrinking-value framework. We build up a full statistical theory on the QML estimators under this framework. In the TSAR model, the theoretical analysis will encounter several new theoretical challenges. First, we have to show the tightness and the Hájek-Rényi type inequality of a quadratic form. It is well known in spatial econometrics that the asymptotic representation of the resultant QML estimator is the sum of a linear part and a quadratic part. It is also well known that the asymptotic theory of discontinuous models such as structural change model or threshold model relies critically on the concept of tightness and the Hájek-Rényi type inequality. Therefore, establishing the tightness and the Hájek-Rényi type inequality of a quadratic form is extremely demanded in the current theoretical analysis. In this paper, we present a result on the expression of the fourth moments of a quadratic form, which lays the base to address this challenge. Second, the theoretical analysis need to address a complicated correlation structure among the spatial terms. Specifically, the spatial term $\ddot{Y}_i = \sum_{j=1}^{N} w_{ij} Y_j$ would have a very complicated correlation with $\ddot{Y}_j$ for any pair (i, j) in the TSAR model, where w_{ij} is the (i, j)-th element of the spatial weights matrix W and Y_i is the i-th observation of the dependent variable. As seen below, this correlation depends on the entire explanatory variables and threshold variables, as well as the spatial weight matrix W, in a nonlinear way. The complicated correlation structure increases the difficulty of theoretical analysis since the usual tools, such as the law of large number and the central limit theorem, can be applied only when some correlation assumption is satisfied. Third, we need to address the identification issue related with the TSAR model carefully. Consider the pure TSAR model which includes only the spatial term as regressor. Unlike the pure SAR model, which uses the variance information to identify the coefficient of spatial term, we need to use the same information to identify three parameters in the TSAR model, i.e., the coefficient of spatial term, the coefficient of spatial term with threshold effect and the threshold value. Note that the threshold value enters into model in a different way with the other two parameters. This increases the difficulty of identification issue in the model. In addition, the model is nonlinear with the coefficient of spatial term, which makes the usual arguments of consistency in linear regression models, such as Hansen (2000), not applicable in the current setup. Some new arguments are needed and we develop the new arguments for the analytical purpose.

Our theoretical analysis indicates that the QML estimators for the slope coefficients and the variance of errors possess the oracle property, that is, the final limiting distributions are the same with those of the infeasible QML estimators obtained as if the threshold value is observed a priori. As regard to the limiting distribution of the QML estimator of threshold value, we show that it takes a different form from that found in classical models. More specifically, the final limiting distribution depends on the skewness and kurtosis of errors, due to the misspecification on the distribution of errors. If the errors are normally distributed, the limiting expression can be simplified and reduces to the classical one. We also consider the hypothesis testing issues related to the TSAR model. Three super-type statistics are proposed to perform the hypothesis testing on the presence of threshold effect. Based on the tightness of the quadratic form established in the paper, we study the asymptotic properties of the three statistics. Our analysis indicates that the limiting distribution relies on the nuisance parameters. We therefore propose a bootstrap procedure to obtain the critical value. We also propose a likelihood ratio (LR) statistic to test that the threshold value is equal to some prespecified one. The LR statistic has an advantage that the adjusting constant reduces to one when the errors are normally distributed.

Although there are numerous studies on SAR models and threshold models separately, only a few recent studies consider this combined topic. Deng (2018) considers a TSAR model. She proposes the two-stage least square method to estimate the model. But no the asymptotic results are established. Li (2018) considers the QML method to estimate a hybrid model of structural change and panel SAR model. To my knowledge, the above two studies are the only existing studies close to this paper.

This paper is organized as follows. Section 2 gives a formal description on the proposed TSAR model. The likelihood function is formulated and the related QML estimators are defined. Section 2 also lists the assumptions needed for the subsequent analysis. Section 3 presents theoretical results on the QML estimators, including the consistencies, convergence rates and limiting distributions. The implications of the theoretical results are also discussed. Section 4 considers the hypothesis testing issues related with the TSAR model. Section 5 concludes the paper. All the theoretical materials are delivered in the appendix. Throughout the paper, we use $a \vee b$ and $a \wedge b$ to denote the respective maximum and minimum values of a and b. For any $N \times N$ matrix M, $\|M\| = \sqrt{M'M}$ is the Frobenius norm. $\|M\|_1 = \max_{1 \leq j \leq N} \sum_{i=1}^{N} |m_{ij}|$ the column sum norm, $\|M\|_\infty = \max_{1 \leq i \leq N} \sum_{j=1}^{N} |m_{ij}|$ the row sum norm, where m_{ij} is the (i, j)th element of M. In addition, we use $|M|$ to denote the matrix which is obtained by taking the absolute value of the corresponding entry of M. For any two $N \times N$ matrices M_1 and M_2, we say $M_1 \leq M_2$ if $m_{1,ij} \leq m_{2,ij}$, where $m_{1,ij}$ and $m_{2,ij}$ are the respective (i, j)th entries of M_1 and M_2.

2 Model specification

The model considered in this paper is the following TSAR model

$$Y_i = \begin{cases} (\rho^* + \varrho^*) \sum_{j=1}^{N} w_{ij} Y_j + X_i'(\beta^* + \Delta^*) + e_i & \text{if } q_i \leq \gamma^*, \\ \rho^* \sum_{j=1}^{N} w_{ij} Y_j + X_i'\beta^* + e_i & \text{if } q_i > \gamma^*, \end{cases} \tag{2.1}$$

where Y_i is the dependent variable and X_i the k-dimensional explanatory variables for unit i. $W = [w_{ij}]_{N \times N}$ is an $N \times N$ prespecified exogenous spatial weights matrix with all its diagonal elements being 0. q_i is the threshold variable which can be an element of X_i. Hereafter, we use $d_j(\gamma_a)$ to denote $\mathbb{1}(q_j \leq \gamma_a)$, and $d_j(\gamma_a, \gamma_b)$ to denote $d_j(\gamma_a) - d_j(\gamma_b)$, where $\mathbb{1}(\cdot)$ is an indicator function, which takes value 1 if the expression in the brackets is satisfied, and 0 otherwise. Throughout the paper, we use the symbols with asterisk to denote the underlying true values. The above model can be reparameterized as

$$Y_i = \begin{cases} \rho_1^* \sum_{j=1}^{N} w_{ij} Y_j + X_i'\beta_1^* + e_i & \text{if } q_i \leq \gamma^*, \\ \rho_2^* \sum_{j=1}^{N} w_{ij} Y_j + X_i'\beta_2^* + e_i & \text{if } q_i > \gamma^*. \end{cases} \tag{2.2}$$

where $\rho_1^* = \rho^* + \varrho^*$,$\beta_1^* = \beta^* + \Delta^*$ and $\rho_2^* = \rho^*$ and $\beta_2^* = \beta^*$. If the spatial term $\sum_{j=1}^{N} w_{ij} Y_j$ does not exist, the model is a standard threshold linear regression model, which is thoroughly studied in many previous studies, for example, Chan (1993), Hansen (2000). If the model does not include the threshold effect, it is a traditional spatial autoregressive model, which is investigated by Kelejian and Prucha (1999), Lee (2004), etc.

Let $W_{i\bullet} = (w_{i1}, w_{i2}, \ldots, w_{iN})$ be the ith row of W, $Y = (Y_1, Y_2, \ldots, Y_N)$, $X = (X_1', X_2', \ldots, X_N')$ and $e = (e_1, e_2, \ldots, e_N)'$. Further define

$$W_\gamma = \Big[W_{1\bullet}' d_1(\gamma), W_{2\bullet}' d_2(\gamma), \ldots, W_{N\bullet}' d_N(\gamma)\Big]'; \quad X_\gamma = \Big[X_1 d_1(\gamma), X_2 d_2(\gamma), \ldots, X_N d_N(\gamma)\Big]'.$$

We therefore have the following expression of matrix form

$$Y = \rho W Y + \varrho W_\gamma Y + X\beta + X_\gamma \Delta + e. \tag{2.3}$$

Let $D(\rho, \varrho, \gamma) = I_N - \rho W - \varrho W_\gamma$ and $\phi = (\beta', \Delta')'$. With the above expression, the log-likelihood function is

$$\mathcal{L}(\theta) = -\frac{1}{2}\ln\sigma^2 + \frac{1}{N}\ln|D(\rho, \varrho, \gamma)| - \frac{1}{2\sigma^2}L(\theta)'L(\theta), \tag{2.4}$$

with $\theta = (\rho, \varrho, \beta', \Delta', \sigma^2)'$ the parameters to be estimated and

$$L(\theta) = Y - \rho W Y - \varrho W_\gamma Y - X\beta - X_\gamma \Delta.$$

The QML estimator is then defined as $\widehat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}}\ \mathcal{L}(\theta)$ with $\Theta = \mathrm{Par}_\rho \times \mathrm{Par}_\rho \times \mathrm{Par}_\beta \times \mathrm{Par}_\beta \times \Gamma \times \mathrm{Par}_{\sigma^2}$, where $\mathrm{Par}_\rho, \mathrm{Par}_\beta, \Gamma$ and Par_{σ^2} are defined in Assumption C below.

Here we give a brief discussion on performing the QML estimation on the TSAR model. The QML estimation is implemented by the following two steps. First, for each given $\gamma \in \{q_1, q_2, \ldots, q_N\}$, the model is a spatial autoregressive model with two spatial weights matrices. We therefore invoke the QML estimation on spatial models to obtain $\widehat{\vartheta(\gamma)}$ with $\vartheta = (\rho, \varrho, \beta', \Delta', \sigma^2)$, see Li (2017) for the discussions on the QML estimation for multiple spatial weights matrices. Then, for such a γ, we use $\widehat{\vartheta(\gamma)}$ to calculate the corresponding likelihood value. Second, the estimator for γ is the q_i which maximizes the likelihood value. Once $\widehat{\gamma}$ is known, the other parameters are calculated according to $\widehat{\vartheta} = \widehat{\vartheta(\widehat{\gamma})}$.

The above estimation procedure is intuitive, but has an obvious drawback that the computation may be time-consuming when N is large since it involves the calculation of the determinant of an $N \times N$ matrix many times. To effectively reduce the computation cost, we suggest to use the two-stage method in Deng (2018) or more simpler IV method to calculate the N likelihood values in the first step. Then we pick out those q_i whose likelihood values are among the largest 5%. Only for these q_i, we apply the above QML estimation procedure.

To analyze the asymptotic properties of the QMLE, we make the following assumptions.

Assumption A: The errors $e_i (i = 1, 2, \ldots, N)$ are identically and independently distributed with mean zero and variance $\sigma^{*2} > 0$. In addition, we assume that $\mathbb{E}(|e_i|^8) < \infty$ for some $c > 0$.

Assumption B: The spatial weights matrix W is an exogenous spatial weights matrix whose diagonal elements are all zeros. In addition, $\|W\|_1 \vee \|W\|_\infty < \infty$ for all N.

Assumption C: The underlying true value $(\rho^*, \beta^{*\prime}, \gamma^*, \sigma^{*2})'$ is an interior point of the parameter space $\mathrm{Par}_\rho \times \mathrm{Par}_\beta \times \Gamma \times \mathrm{Par}_{\sigma^2}$, where Par_ρ and Par_β are both compact sets of $\mathbb{R}^1$ and $\mathbb{R}^p$, and $\Gamma = [\underline{\gamma}, \overline{\gamma}] \subset \mathbb{R}^1$, and Par_{σ^2} is a compact set which is bounded away from zero.

Assumption D: Let $\tilde{D}(\rho) = I_N - \rho W$. We assume that $\tilde{D}(\rho)$ is invertible over Par_ρ. In addition, $\|\tilde{D}^{-1}(\rho)\|_1$ and $\|\tilde{D}^{-1}(\rho)\|_\infty$ are bounded by some constant C for all N over Par_ρ.

Assumption E: $\psi^* = (\varrho^*, \Delta^{*\prime})' = N^{-\nu}(C_\varrho^*, C_\Delta^{*\prime})'$, where $C_\psi^* \equiv (C_\varrho^*, C_\Delta^{*\prime})' \neq 0_{k+1}$ is $(k+1)$-dimensional vector of constants and $0 < \nu < \frac{1}{2}$.

Assumption F: $\mathbb{E}(\|X_p\|^4|q_i = t) \le C_X$ and $\mathbb{E}(\|X_p\|^4|q_i = t, q_j = s) \le C_{XX}$ for all $1 \le i, j, p \le N$ and $t, s \in \Gamma$. In addition, $0 < c_f \le f_i(t) \le C_f$ and $0 < c_{ff} \le f_{ij}(t, s) \le C_{ff}$ for all $t, s \in \Gamma$, where $f_i(t)$ is the density function of q_i and $f_{ij}(t, s)$ the joint density function of (q_i, q_j).

Assumption G: (X_i, q_i) is an α-mixing process with $\sum_l^\infty \alpha_l^{\frac{c}{2+c}} < \infty$ with some $c > 0$.

Let $S = WD^{-1}(\rho, \varrho, \gamma)$, $S_\gamma = W_\gamma D^{-1}(\rho, \varrho, \gamma)$, and define S^*, S_γ^*, $S^\bullet$ and $S_\gamma^\bullet$ similarly as S, S_γ except that we use $D^* = D(\rho^*, \varrho^*, \delta^*)$ to replace $D(\rho, \varrho, \gamma)$ in S^* and S_γ^*, and use $D^\bullet = I_N - \rho^* W$ to replace $D(\rho, \varrho, \gamma)$ in $S^\bullet$ and $S_\gamma^\bullet$. Hereafter, we drop off the input arguments ρ, ϱ and γ from $D(\rho, \varrho, \gamma)$ for notational simplicity.

Assumption H: $f_{ij}(t, s)$ and $\mathbb{E}(\mathbf{X}_i\mathbf{X}_i'|q_i = t)$ are continuous at $(t, s) = (\gamma^*, \gamma^*)$ and $t = \gamma^*$ respectively, where $\mathbf{X}_i = [S_{i\bullet}^\bullet X\beta^*, X_i']'$ with $S_{i\bullet}^\bullet$ the i-th row of $S^\bullet$.

Further define two symmetric matrices $\mathcal{I}_a(\gamma)$ and $\mathcal{I}_b(\gamma)$:

$$\mathcal{I}_a(\rho, \varrho, \gamma) = \frac{1}{\mathrm{tr}\big[D^{\bullet -1\prime} D' D D^{\bullet -1}\big]} \begin{bmatrix} \mathrm{tr}(S^{\bullet\prime} S^\bullet) & \mathrm{tr}(S_\gamma^{\bullet\prime} S^\bullet) \\ * & \mathrm{tr}(S_\gamma^{\bullet\prime} S_\gamma^\bullet) \end{bmatrix} + \frac{1}{N} \begin{bmatrix} \mathrm{tr}(S^2) & \mathrm{tr}(S_\gamma S) \\ * & \mathrm{tr}(S_\gamma S_\gamma) \end{bmatrix}$$
$$- \frac{2}{\big[\mathrm{tr}(D^{\bullet -1\prime} D' D D^{\bullet -1})\big]^2} \begin{bmatrix} [\mathrm{tr}(S^{\bullet\prime} D D^{\bullet -1})]^2 & \mathrm{tr}(S_\gamma^{\bullet\prime} D D^{\bullet -1})\mathrm{tr}(S^{\bullet\prime} D D^{\bullet -1}) \\ * & [\mathrm{tr}(S_\gamma^{\bullet\prime} D D^{\bullet -1})]^2 \end{bmatrix},$$

and

$$\mathcal{I}_b(\gamma) = \frac{1}{N} \begin{bmatrix} \mathrm{tr}(S^{\bullet 2} + S^{\bullet\prime} S^\bullet) & \mathrm{tr}(S^\bullet S_\gamma^\bullet + S^{\bullet\prime} S_\gamma^\bullet) & 2\mathrm{tr}(S^\bullet)/\sigma^{*2} & \mathrm{tr}(S^\bullet S_{\gamma,\gamma^*}^\bullet + S^{\bullet\prime} S_{\gamma,\gamma^*}^\bullet) \\ * & \mathrm{tr}(S_\gamma^\bullet S_\gamma^\bullet + S_\gamma^{\bullet\prime} S_\gamma^\bullet) & 2\mathrm{tr}(S_\gamma^\bullet)/\sigma^{*2} & \mathrm{tr}(S_\gamma^\bullet S_{\gamma,\gamma^*}^\bullet + S_\gamma^{\bullet\prime} S_{\gamma,\gamma^*}^\bullet) \\ * & * & N/(2\sigma^{*4}) & 2\mathrm{tr}(S_{\gamma,\gamma^*}^\bullet)/\sigma^{*2} \\ * & * & * & \mathrm{tr}(S_{\gamma,\gamma^*}^\bullet S_{\gamma,\gamma^*}^\bullet + S_{\gamma,\gamma^*}^{\bullet\prime} S_{\gamma,\gamma^*}^\bullet) \end{bmatrix}.$$

Given the above notation, we make following assumption on the parameters identifications.

Assumption I: Let $\tau_{\min}(\cdot)$ denote the minimum eigenvalue of its input. One of the following assumption holds

Assumption I.1: (a) Matrix $\mathcal{I}_a(\rho, \varrho, \delta)$ is positive definite over the parameter space $\mathrm{Par}_\rho \times \mathrm{Par}_\varrho \times \Gamma$. (b) For all γ, $\tau_{\min}(\mathcal{I}_b(\gamma)) \ge C \min(1, |\gamma - \gamma^*|)$, and (c) $\tau_{\min}(\frac{1}{N}[X, X_\gamma]'[X, X_\gamma]) \ge C$.

Assumption I.2: For all γ, $\tau_{\min}(\frac{1}{N}U(\gamma)'U(\gamma)) \ge C$ and $\tau_{\min}(\frac{1}{N}R(\gamma)'R(\gamma)) \ge C \min(1, |\gamma - \gamma^*|)$ with $U(\gamma) = [S^\bullet X\beta^*, S_\gamma^\bullet X\beta^*, X, X_\gamma]$ and $R(\gamma) = [S_{\gamma,\gamma^*}^\bullet X\beta^*, X_{\gamma,\gamma^*}, U(\gamma)]$, where $S_{\gamma,\gamma^*}^\bullet = S_\gamma^\bullet - S_{\gamma^*}^\bullet$.

We give some discussions on the above assumptions in sequence. Assumption A is about regression errors. It assumes that the errors are independent and identically distributed over the cross section. This assumption, which is often made in the studies of the QML estimation on spatial models such as Lee (2004), Yu et al. (2008), Lee and Yu (2010), is important to our theoretical analysis since the asymptotic representation of the QML estimator is a linear-quadratic form of errors. In addition, our analysis involves the term $\mathbb{E}[e'Ae - \mathrm{tr}(A)]^4$, whose final expression is greatly simplified under the independent assumption. Requiring $\mathbb{E}(e_i^8) < C$ in the current assumption, instead of $\mathbb{E}(|e_i|^{4+c}) < C$ for some $c > 0$ as in the previous studies is due to the proof of tightness for the quadratic form. Assumption B is standard in spatial econometrics. Zero diagonal elements in spatial weights matrix preclude one obvious identification issue in spatial models. In addition, Assumption B precludes the possibility of endogenous spatial weights matrix (Qu and Lee (2015)). Extension to allow the endogeneity of spatial weights matrix is interesting but beyond the scope of this paper. An important implication of Assumption B is that $\sup_{\gamma\in\Gamma}\|W_\gamma\|_1 \vee \|W_\gamma\|_\infty < \infty$ due to the fact $|d_i(\gamma)| \le 1$ for all γ and i. Assumption C is about the parameters space. This assumption is standard in the econometric literature. We note that the parameters space for β can be extended to the whole $\mathbb{R}^k$ since the objective function is a negative quadratic form of β, so the maximum value cannot be achieved at very large β. Assumption D has two purposes. Invertibility of $D^\bullet(\rho)$ guarantees that the spatial model is well-defined. The finiteness of the two norms of $D^{\bullet-1}(\rho)$ is for the tractability of theoretical analysis. This kind of assumption is originated from Kelejian and Prucha (1998), and are widely adopted in the subsequent spatial studies, see, e.g., Lee (2004) and Yu et al. (2008).

Assumption E specifies the shrinking rate of threshold effect, which is standard in the threshold studies, see Hansen (2000), Caner and Hansen (2004), etc. Intuitively speaking, if the shrinking rate is too fast, say $\nu \ge \frac{1}{2}$, we would have identification issue for γ since threshold effect is too weak. On the other hand, if the shrink does not occur ($\nu = 0$), the final limiting distribution would depend on many nuisance parameters, which causes trouble to statistical inference on γ. Assumption F assumes the finiteness of the fourth conditional moments. Loosely speaking, this assumption is imposed to deal with nonlinear dependence of Y on the entire X_i and q_i. Note that $W_{i\bullet}Y$ involves the whole X, naturally we need $\mathbb{E}(\|X_p\|^4|q_i = t) \le C$ for all i and p instead of $\mathbb{E}(\|X_i\|^4|q_i = t) \le C$ for all i as in the classical analysis. Assumption G assumes mixing process on (X_i, q_i), which is often made in the econometric studies and is very general to allow a rich class of stochastic processes. Assumption H is a regularity condition. It is plausible since it often holds except for some peculiar cases.

Assumption I is the identification condition (IC) for the underlying parameters. Following the tradition of spatial econometrics (Lee (2004), Lee and Yu (2010), Yu et al. (2008)), we consider two types of identification strategy, which come from the fact that the presence of spatial term has both mean and variance consequences on the dependent variable. Assumption I.1 (a) is the IC for ρ and ϱ, using the information contained in variance. Assumption I.1 (b) is the IC for γ and Assumption I.1 (c) the IC for β and Δ. Once the identification of there parameters are achieved, the variance of errors, σ^2, is automatically identified. The assumption on $U(\gamma)$ is the IC for the ρ, ϱ, β and Δ, using the mean information and the assumption on $R(\gamma)$ is the IC for γ. Comparatively speaking, the variance identification strategy has the advantage of more generality since it allows the parameters to be identified when the mean of dependent variables is zero, which occurs in a pure spatial model with no exogenous explanatory variables, but has the disadvantage that the condition is not intuitive. For this reason, we give more detailed discussions on Assumption I.1 here. To see the intuition behind the matrix $\mathcal{I}_a(\gamma)$, consider the case that the underlying true model is $Y = \rho^* WY + e$, but unfortunately we use the TSAR model $Y = \rho W + \varrho W_\gamma + e$ to estimate the parameters. We expect that, for any γ, the consistency of $\widehat{\rho}$ and $\widehat{\varrho}$ will not be affected in this case. Let $\widetilde{\mathcal{L}}(\rho, \varrho, \sigma^2, \gamma)$ be the likelihood function function in this simple case. It can be shown that, with a constant adjustment,

$$\mathbb{E}_Q\widetilde{\mathcal{L}}(\rho, \varrho, \sigma^2, \gamma) = -\frac{1}{2}\ln\sigma^2 + \frac{1}{N}\ln|I_N - \rho W - \varrho W_\gamma| + \frac{1}{2}\ln\sigma^{*2} - \frac{1}{N}\ln|D^\bullet| + \frac{1}{2} - \frac{\sigma^{*2}}{2N\sigma^2}\mathrm{tr}\Big[D^{\bullet-1}(I_N - \rho W - \varrho W_\gamma)'(I_N - \rho W - \varrho W_\gamma)D^{\bullet-1}\Big].$$

where $\mathbb{E}_Q$ denotes the expectation conditional on $(q_1, q_2, \ldots, q_N)$. After concentrating out σ^2, the resultant function $\mathbb{E}_Q\widetilde{\mathcal{L}}(\rho, \varrho, \sigma^2, \gamma)$, denoted by $F(\rho, \varrho, \gamma)$, is

$$F(\rho,\varrho,\gamma) = \frac{1}{2N}\ln\left|D^{\bullet-1}(I_N - \rho W - \varrho W_\gamma)'(I_N - \rho W - \varrho W_\gamma)D^{\bullet-1}\right| \\ - \frac{1}{2}\ln\left|\frac{1}{N}\mathrm{tr}\Big[D^{\bullet-1}(I_N - \rho W - \varrho W_\gamma)'(I_N - \rho W - \varrho W_\gamma)D^{\bullet-1}\Big]\right|.$$

By the Jensen's inequality $\frac{1}{N}\sum_{i=1}^N \ln\lambda_i \leq \ln(\frac{1}{N}\sum_{i=1}^N \lambda_i)$ for any positive values λ_i, we have $F(\rho,\varrho,\gamma) \leq 0$ for all ρ, ϱ and γ. However, for each γ, the Taylor's expansion on (ρ, ϱ) around ρ^* and ϱ^* (note that $\varrho^* = 0$) gives

$$F(\rho,\varrho,\gamma) = -\frac{1}{2}\Big[\rho - \rho^*, \varrho - \varrho^*\Big]\mathcal{I}_a(\ddot{\rho}_\gamma, \ddot{\varrho}_\gamma, \gamma)\begin{bmatrix}\rho - \rho^* \\ \varrho - \varrho^*\end{bmatrix},$$

due to the fact that $F(\rho^*, \varrho^*, \gamma) = 0$ and $\frac{\partial F(\rho^*,\varrho^*,\gamma)}{\partial\rho} = 0$, $\frac{\partial F(\rho^*,\varrho^*,\gamma)}{\partial\varrho} = 0$, where $\ddot{\rho}_\gamma$ is some value between ρ^* and ρ, which depends on γ, and $\ddot{\varrho}_\gamma$ is defined similarly. Given that $F(\rho,\varrho,\gamma) \leq 0$ for all ρ, ϱ, γ, we expect that $\mathcal{I}_a(\ddot{\rho}_\gamma, \ddot{\varrho}_\rho, \gamma)$ is semi-positive definite. Assumption I.1 (a) is imposed along this direction, requiring that $\mathcal{I}_a(\rho,\varrho,\gamma)$ is positive definite for all ρ, ϱ and γ to achieve the identification of ρ^* and $\varrho^* = 0$.

3 Theoretical results

Before presenting the theoretical results of the QMLE, we first give two propositions. These two propositions serve as the base for the subsequent theoretical analysis. In addition, they have their own independent interests.

Proposition 3.1 *Let $\{u_i\}_{i=1}^N$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(|u_i|^4) < \infty$, and $z_i(\gamma)$ be a vector random function of γ, which is independent with u_j for all i and j. For each $\gamma_a < \gamma_b$, there exists a non-negative function $z_i^\star(\gamma_a, \gamma_b)$ which is a decreasing function of γ_a and an increasing function of γ_b such that (i) $|z_i(\gamma_a) - z_i(\gamma_b)| \leq z_i^\star(\gamma_a,\gamma_b)$; (ii) $\mathbb{E}(\|z_i^\star(\gamma_a,\gamma_b)\|^r) \lesssim |\gamma_a - \gamma_b|$ for $r = 1, 2, 4$. Then we have*

$$\mathbb{E}^*\Big[\sup_{|s-t|\leq\delta}\frac{1}{\sqrt{N}}\Big\|\sum_{i=1}^N u_i[z_i(s) - z_i(t)]\Big\|\Big]^4 \lesssim \delta.$$

where $\mathbb{E}^*$ denote the expectation with respect to outer probability measure, and $\lesssim$ denotes "is bounded by up to a universal constant".

Remark 3.1 The proof of Proposition 3.1 essentially relies on the chain arguments, which are used in the empirical processes theory, to deal with the supremum over a compact set. For a detailed reference on the chain arguments, see the proof of Theorem 2.2.4 in Van Der Vaart and Wellner (1996). The current proof is different from that of Van Der Vaart and Wellner (1996) in that they make the separable assumption on the empirical process to bridge the difference of the supremum over the original set and its dense subset. In contrast, due to the speciality of indicator function, we make use of the monotonous property to deal with the difference. The result in Proposition 3.1 plays an important role in theoretical analysis of the current paper. Let

$$J_N(\gamma) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} u_i z_i(\gamma).$$

The preceding proposition implies the following stochastic equicontinuous condition: for any $\epsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{N \to \infty} \mathbb{P}^* \left(\sup_{|s-t| \leq \delta} \|J_N(s) - J_N(t)\| > \epsilon \right) \to 0.$$

where $\mathbb{P}^*$ denotes the outer probability measure. So the tightness of $J_N(\gamma)$ is obtained. This tightness result, together with finite dimensional weak convergence, would lead to the weak convergence of stochastic process $J_N(\gamma)$ to a Gaussian process under the uniform metric (or the Skorohod metric if measurability is concerned). This result, further equipped with the argmax continuous mapping theorem (see Theorem 3.2.2 of Van Der Vaart and Wellner (1996)), would deliver the final limiting distribution of $\widehat{\gamma}$.

In addition, if we choose $t = \underline{\gamma}$, $s = \gamma$ and $\delta = \overline{\gamma} - \underline{\gamma}$, we would have

$$\sup_{\gamma \in \Gamma} \frac{1}{\sqrt{N}} \Big\| \sum_{i=1}^{N} u_i z_i(\gamma) \Big\| = O_{p^*}(1).$$

The above result is frequently used in the proof of consistency. □

Proposition 3.2 *Let $u = (u_1, u_2, \ldots, u_N)'$ be N-dimensional random vector with u_i being independent and identically distributed variables such that $\mathbb{E}(u_i) = 0$ and $\mathbb{E}(u_i^2) = 1$. Let $\kappa_p^* = \mathbb{E}(u_i^p)$ for $p = 3, 4, \ldots, 8$. For any $N \times N$ symmetric matrix A, we have*

$$\mathbb{E}\Big[|u'Au - \mathbb{E}(u'Au)|^4\Big] = 12[\mathrm{tr}(A^2)]^2 + 48\mathrm{tr}(A^4) + \aleph(A),$$

where

$$\begin{aligned}
\aleph(A) = {} & \varsigma_1 \Big[12 l_N'(A \circ A) l_N \mathrm{tr}(A \circ A) + 96\mathrm{tr}[(I_N \circ A)A^3] + 48 l_N'(I_N \circ A^2)^2 l_N\Big] \\
& + \varsigma_2 \cdot 6\mathrm{tr}(A \circ A \circ A^2) + \varsigma_3 \mathrm{tr}(A \circ A \circ A \circ A) \\
& + \varsigma_4 \Big[48 l_N'(I_N \circ A)A^2(I_N \circ A) l_N + 96 l_N(A \circ A)A(I_N \circ A) l_N + 96\mathrm{tr}[A(A \circ A)A]\Big] \\
& + \varsigma_5 \Big[3[\mathrm{tr}(A \circ A)]^2 + 24 l_N'(I_N \circ A)(A \circ A)(I_N \circ A) l_N + 8 l_N'(A \circ A \circ A \circ A) l_N\Big] \\
& + \varsigma_6 \Big[24 l_N'(I_N \circ A)A(I_N \circ A \circ A) l_N + 32 l_N'(I_N \circ A)(A \circ A \circ A) l_N\Big]
\end{aligned}$$

where $\varsigma_1 = \kappa_4^* - 3$, $\varsigma_2 = \kappa_6^* - 15\kappa_4^* - 10\kappa_3^{*2} + 30$, $\varsigma_3 = \kappa_8^* - 28\kappa_6^* - 56\kappa_5^*\kappa_3^* - 35\kappa_4^{*2} + 420\kappa_4^* + 560\kappa_3^{*2} - 630$, $\varsigma_4 = \kappa_3^{*2}$, $\varsigma_5 = (\kappa_4^* - 3)^2$ and $\varsigma_6 = \kappa_3^*(\kappa_5^* - 10\kappa_3^*)$.

Remark 3.2 Proposition 3.2 has its own independent interest. It has been used in a wide range of econometric issues, including the moments of the Durbin-Watson statistic (Bao and Ullah, 2007), high dimensional covariance testing (Chen, Zhang and Zhong, 2010), etc. Previous studies derive this type of result either under the quasi-normality assumption on errors (see, e.g., Pukelsheim

1980), or with the positive definiteness condition on A (see, e.g., Chen, Zhang and Zhong, 2010). Apparently, the previous results are not suitable to the current analysis, given Assumption A on errors and the quadratic form in spatial models. Bao and Ullah (2010) derives $\mathbb{E}(e'Ae)^4$ under a general non-normal condition. Even with this result, deriving Proposition 3.2 is a non-trivial task since we still need to know the explicit expression of $\mathbb{E}(e'Ae)^3$.

Proposition 3.2 plays an important role to establish some basic results of this paper such as the tightness of the quadratic form, the Hájek-Rényi type inequality for the quadratic form, and so forth. To see this point clear, let us first review part of theoretical results in Hansen (2000). To analyze the convergence rate of $\widehat{\gamma}$, Hansen (2000) shows the following theoretical result that there exists a constant $C^{\dagger}$ such that for all positive values η and ϵ,

$$\mathbb{P}\left(\sup_{\frac{C^{\dagger}}{a_N}\leq|\gamma-\gamma^*|\leq B}\frac{\|J_N(\gamma)-J_N(\gamma^*)\|}{\sqrt{a_N}|\gamma-\gamma^*|}>\eta\right)\leq\epsilon$$

with $J_N(\gamma)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i d_i(\gamma)e_i$ and $a_N=N^{1-2\nu}$, see Lemma A.8 of Hansen (2000). The above result is loosely equivalent to

$$\mathbb{E}\Big[\sup_{\frac{C^{\dagger}}{a_N}\leq|\gamma-\gamma^*|\leq B}|J_N(\gamma)-J_N(\gamma^*)|\Big]\lesssim|\gamma-\gamma^*|.$$

Although Proposition 3.1 gives the tightness result in terms of expectation, the bound there is not tight. Since $\mathbb{E}\|N^{-1/2}\sum_{i=1}^{N}u_i[z_i(s)-z_i(t)]\|^4\lesssim|t-s|^2$ for any given s and t, we may expect that the bound for the supremum over $|t-s|\leq\delta$ should be δ^2. But the bound in Proposition 3.2 is δ instead. This loose bound arises from the maximal inequality used in the proof. However, as shown in Hansen (2000), obtaining a tight bound is important to derive the convergence rate of $\widehat{\gamma}$. So the tightness arguments of expectation-version are not helpful to derive the wanted result. On the other hand, according to the discussions on page 57 in Billingsley (1968), the tightness arguments of probability-version entails no real loss, due to the asymptotic independence of the increment of the stochastic process. To use Theorem 12.2 of Billingsley (1968), the fourth moment of partial sum is needed. This is the reason why Proposition 3.2 is basic for our theoretical analysis since it gives an explicit expression of the fourth moment of the quadratic form. $\square$

Theorem 3.1 *Let $\widehat{\theta}$ be the maximizer of the likelihood function (2.4). Under Assumptions A-I, as $N\to\infty$, we have $N^{\nu}(\widehat{\vartheta}-\vartheta^*)=o_p(1)$ and $\widehat{\gamma}-\gamma^*=o_p(1)$, where $\vartheta=(\rho,\varrho,\beta',\Delta',\sigma^2)$.*

Remark 3.3 Theorem 3.1 presents the global properties of the QML estimators. It claims that the QML estimators would be in any neighborhood of the underlying true values. In addition, it also delivers preliminary convergence rates of the QML estimators for the regression coefficients and the variance of errors. Theorem 3.1 is proved under the following trick. The first step is to show that whatever value γ takes, the consistency of $\widehat{\rho}$ and $\widehat{\varrho}$ can still be guaranteed. This is possible under the assumption of shrinking threshold effect. Consider model (2.1), which we rewrite as

$$Y_i=\rho^*W_{i\bullet}Y+\varrho^*W_{i\bullet}d_i(\gamma^*)+\beta^{*\prime}X+\Delta^{*\prime}X_i d_i(\gamma)+e_i.$$

If γ jumps a big step from γ^* to $\gamma^{\dagger}$, we see that such a change has a trivial effect on the observed Y_i since ϱ^* and Δ^* diminish to zero with a growing N under Assumption E. This is in contrast with another fact that big jumps of other parameters would cause a large change of Y_i. With these two facts, even γ may not be correctly determined, we still has the chance to correctly estimate other parameters in term of consistency. Given the consistency of other parameters, we are able to ignore the troublesome terms of smaller order. Therefore, we next show the results of Theorem 3.1 by multiplying N^ν on both the sides of the proceeding equation. Note that the change of γ now has a nontrivial effect on Y_i since $N^\nu(\varrho^*, \Delta^*) = (C^*_\varrho, C^*_\Delta)$ are constants. In addition, the consistency of other parameters now is strengthened to $N^\nu(\widehat{\vartheta} - \vartheta^*) = o_p(1)$.

Theorem 3.1 is sufficient to the subsequent analysis. To see this, for any place where $\widehat{\vartheta}$ appears, we can decompose it as $\widehat{\vartheta} - \vartheta^*$ and ϑ^*. Although some components of ϑ^* shrink to zero, according to Theorem 3.1, term $\widehat{\vartheta} - \vartheta^*$ is of smaller order relative to ϑ^*. As a consequence, we can use ϑ^* to replace $\widehat{\vartheta}$ in the asymptotic analysis. □

Theorem 3.2 *Under Assumptions A-I, we have* $\widehat{\gamma} - \gamma^* = O_p(N^{2\nu-1})$.

Remark 3.4 Establishing the convergence rate of $\widehat{\gamma}$ requires considerable amount of work. To appreciate the difficulties, we note that the current model contains $W_{i\bullet}Y$ as a regressor, which is equal to $W_{i\bullet}(I_N - \rho^* W - \varrho^* W_{\gamma^*})^{-1}(X\beta^* + X_{\gamma^*}\Delta^* + e)$. From this expression, we see that $W_{i\bullet}Y$ depends nonlinearly on the entire $(X_i, q_i)_{i=1}^N$. Even (X_i, q_i) is a mixing process, we cannot conclude that $W_{i\bullet}Y$ is a mixing process too. As a result, the correlations pattern for $W_{i\bullet}Y$ becomes quite complicated. However, it is obvious, according to the analysis of Hansen (2000), that for a given γ, the asymptotic result

$$\mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N X_i X_i' d_i(\gamma, \gamma^*) - \mathbb{E}[X_i X_i' d_i(\gamma, \gamma^*)]\Big]^2 \lesssim \frac{1}{N}|\gamma - \gamma^*| \tag{3.1}$$

is important for Theorem 3.2. The mixing process assumption on (X_i, q_i) in Hansen (2000) guarantees the above result. However, the same arguments cannot be applied to the current model since, as we argued before, $W_{i\bullet}Y$ is determined within the model and may be not a mixing process. So some new arguments are needed. In this paper, to address this issue, we decompose $(W_{i\bullet}Y)^2$ into two parts $[S^\bullet_{i\bullet}(X\beta^* + e)]^2$ with $S^\bullet_{i\bullet} = W_{i\bullet}(I_N - \rho^* W)^{-1}$ and the remaining term, denoted by $\mathcal{R}_i$. For the first term, note that $S^\bullet_{i\bullet}(X\beta^* + e)$ is a linear combination of X_i and e_i, we can show

$$\mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N \big(S^\bullet_{i\bullet}(X\beta^* + e)\big)^2 d_i(\gamma, \gamma^*) - \mathbb{E}\big[\big(S^\bullet_{i\bullet}(X\beta^* + e)\big)^2 d_i(\gamma, \gamma^*)\big]\Big]^2 \lesssim \frac{1}{N}|\gamma - \gamma^*|.$$

For the remaining term, because of $\varrho^* = O(N^{-\nu})$ and $\Delta^* = O(N^{-\nu})$, we can show that $\mathbb{E}[\frac{1}{N}\sum_{i=1}^N \mathcal{R}_i - \mathbb{E}(\mathcal{R}_i)]^2 \lesssim N^{-2\nu}|\gamma - \gamma^*|^2$. The above two results give

$$\mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N (W_{i\bullet}Y)^2 d_i(\gamma, \gamma^*) - \mathbb{E}[(W_{i\bullet}Y)^2 d_i(\gamma, \gamma^*)]\Big]^2 \lesssim \frac{1}{N}|\gamma - \gamma^*| + \frac{1}{N^{2\nu}}|\gamma - \gamma^*|^2.$$

Although the preceding display is weaker than (3.1), it is enough for our theoretical analysis on the convergence rate of $\widehat{\gamma}$. □

Now we present the limiting distribution of $\widehat{\vartheta}$. We first introduce the following notations for ease of exposition. Define $\Sigma_N(\gamma_a,\gamma_b)$ and $\Omega_N(\gamma_a,\gamma_b)$ as

$$\Sigma_N(\gamma_a,\gamma_b) = \frac{1}{N\sigma^{*2}}\mathcal{X}_{\gamma_a}'\mathcal{X}_{\gamma_b} + \frac{1}{N}\begin{bmatrix} \mathrm{tr}(S^{\bullet 2}) & \mathrm{tr}(S^{\bullet}S^{\bullet}_{\gamma_b}) & 0 & 0 & \mathrm{tr}(S^{\bullet})/\sigma^{*2} \\ \mathrm{tr}(S^{\bullet}_{\gamma_a}S^{\bullet}) & \mathrm{tr}(S^{\bullet}_{\gamma_a}S^{\bullet}_{\gamma_b}) & 0 & 0 & \mathrm{tr}(S^{\bullet}_{\gamma_a})/\sigma^{*2} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \mathrm{tr}(S^{\bullet})/\sigma^{*2} & \mathrm{tr}(S^{\bullet}_{\gamma_b})/\sigma^{*2} & 0 & 0 & N/(2\sigma^{*4}) \end{bmatrix}$$

and

$$\Omega_N(\gamma_a,\gamma_b) = \frac{\kappa_4 - 3\sigma^{*4}}{N\sigma^{*4}}\begin{bmatrix} \mathrm{tr}(S^{\bullet}\circ S^{\bullet}) & \mathrm{tr}(S^{\bullet}\circ S^{\bullet}_{\gamma_b}) & 0 & 0 & \mathrm{tr}(S^{\bullet})/(2\sigma^{*2}) \\ \mathrm{tr}(S^{\bullet}_{\gamma_a}\circ S^{\bullet}) & \mathrm{tr}(S^{\bullet}_{\gamma_a}\circ S^{\bullet}_{\gamma_b}) & 0 & 0 & \mathrm{tr}(S^{\bullet}_{\gamma_a})/(2\sigma^{*2}) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \mathrm{tr}(S^{\bullet})/(2\sigma^{*2}) & \mathrm{tr}(S^{\bullet}_{\gamma_b})/(2\sigma^{*2}) & 0 & 0 & N/(4\sigma^{*4}) \end{bmatrix}$$
$$+\frac{\kappa_3}{N\sigma^{*4}}\sum_{i=1}^{N}\begin{bmatrix} 2s^{\bullet}_{ii}\mu_i & 2s^{\bullet}_{ii}d_i(\gamma_b)\mu_i & s^{\bullet}_{ii}X_i' & s^{\bullet}_{ii}d_i(\gamma_b)X_i' & \frac{1}{2\sigma^{*2}}\mu_i \\ 2s^{\bullet}_{ii}d_i(\gamma_a)\mu_i & 2s^{\bullet}_{ii}d_i(\gamma_c)\mu_i & s^{\bullet}_{ii}d_i(\gamma_a)X_i' & s^{\bullet}_{ii}d_i(\gamma_c)X_i' & \frac{1}{2\sigma^{*2}}d_i(\gamma_a)\mu_i \\ s^{\bullet}_{ii}X_i & s^{\bullet}_{ii}d_i(\gamma_b)X_i & 0 & 0 & \frac{1}{2\sigma^{*2}}X_i \\ s^{\bullet}_{ii}d_i(\gamma_a)X_i & s^{\bullet}_{ii}d_i(\gamma_c)X_i & 0 & 0 & \frac{1}{2\sigma^{*2}}d_i(\gamma_a)X_i \\ \frac{1}{2\sigma^{*2}}\mu_i & \frac{1}{2\sigma^{*2}}d_i(\gamma_b)\mu_i & \frac{1}{2\sigma^{*2}}X_i' & \frac{1}{2\sigma^{*2}}d_i(\gamma_b)X_i' & 0 \end{bmatrix}$$

where $\mu_i = S^{\bullet}_{i\bullet}X\beta^*$, $\gamma_c = \gamma_a \wedge \gamma_b$ and $\mathcal{X}_\gamma = [WY, W_\gamma Y, X, X_\gamma, 0]$, and κ_3 and κ_4 are the respective third and fourth moment of e_{it}, i.e., $\kappa_3 = \mathbb{E}(e_{it}^3)$ and $\kappa_4 = \mathbb{E}(e_{it}^4)$. Let $\Sigma(\gamma_a,\gamma_b) = \operatorname*{plim}_{N\to\infty}\Sigma_N(\gamma_a,\gamma_b)$ and $\Omega(\gamma_a,\gamma_b) = \operatorname*{plim}_{N\to\infty}\Omega_N(\gamma_a,\gamma_b)$ for each pair (γ_a,γ_b). We have the following proposition.

Proposition 3.3 *Under Assumption A-I, we have*

$$\sup_{(\gamma_a,\gamma_b)\in\Gamma\times\Gamma}\left\|\Sigma_N(\gamma_a,\gamma_b) - \Sigma(\gamma_a,\gamma_b)\right\| + \left\|\Omega_N(\gamma_a,\gamma_b) - \Omega(\gamma_a,\gamma_b)\right\| = o_p(1).$$

Theorem 3.3 *Under Assumptions A-I, as $N\to\infty$, we have*

$$\sqrt{N}(\widehat{\vartheta} - \vartheta^*) \xrightarrow{d} N\left(0, \Sigma_\vartheta^{*-1}(\Sigma_\vartheta^* + \Omega_\vartheta^*)\Sigma_\vartheta^{*-1}\right),$$

where $\Sigma_\vartheta^* = \Sigma(\gamma^*,\gamma^*)$ and $\Omega_\vartheta^* = \Omega(\gamma^*,\gamma^*)$.

Remark 3.5 Theorem 3.3 presents the limiting distribution for the QML estimators for the regression coefficients and variance of errors. Like in the previous studies such as Hansen (2000), Caner and Hansen (2004), Su and Xu (2017), the QML estimators possess the oracle properties, that is, the final limiting distributions are the same as those of infeasible QML estimators which are obtained as if the threshold value γ^* is observed a priori. We emphasize that the above result is derived under the shrinking threshold effect framework. If threshold effect is a fixed value, we expect that the threshold value γ^* can be estimated more precisely since the signal for γ^* becomes stronger. Given that a case of less precise estimation on γ^* has already achieve the oracle property, the case of more precise estimation would certainly possess the oracle property too. As a result, we conclude that the results in Theorem 3.3 also hold under the fixed-value threshold effect setup.□

Remark 3.6 We can use the plug-in method to estimate the limiting variance. However, to make the estimators achieve better finite sample performance, and also to make the estimators valid even under the nonshrinking threshold framework, we suggest using the following estimator. Define $\mathcal{X}_{\hat{\gamma}} = [WY, W_{\hat{\gamma}}Y, X, X_{\hat{\gamma}}, 0]$ is an $N \times (2k+3)$ dimensional matrix. Σ_{ϑ}^* and Ω_{ϑ}^* can be consistently estimated by $\widehat{\Sigma_{\vartheta}^*}$ and $\widehat{\Omega_{\vartheta}^*}$, which are given as

$$\widehat{\Sigma_{\vartheta}^*} = \frac{1}{N\widehat{\sigma}^2}\mathcal{X}_{\hat{\gamma}}'\mathcal{X}_{\hat{\gamma}} + \frac{1}{N}\begin{bmatrix} \mathrm{tr}(\widehat{S^*}\widehat{S^*}) & \mathrm{tr}(\widehat{S^*}\widehat{S^*_{\hat{\gamma}}}) & 0 & 0 & \mathrm{tr}(\widehat{S^*})/\widehat{\sigma}^2 \\ \mathrm{tr}(\widehat{S^*}\widehat{S^*_{\hat{\gamma}}}) & \mathrm{tr}(\widehat{S^*_{\hat{\gamma}}}\widehat{S^*_{\hat{\gamma}}}) & 0 & 0 & \mathrm{tr}(\widehat{S^*_{\hat{\gamma}}})/\widehat{\sigma}^2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \mathrm{tr}(\widehat{S^*})/\widehat{\sigma}^2 & \mathrm{tr}(\widehat{S^*_{\hat{\gamma}}})/\widehat{\sigma}^2 & 0 & 0 & N/(2\widehat{\sigma}^4) \end{bmatrix}$$

and

$$\widehat{\Omega_{\vartheta}^*} = \frac{\widehat{\kappa}_4 - 3\widehat{\sigma}^4}{N\widehat{\sigma}^4}\begin{bmatrix} \mathrm{tr}(\widehat{S^*}\circ\widehat{S^*}) & \mathrm{tr}(\widehat{S^*}\circ\widehat{S^*_{\hat{\gamma}}}) & 0 & 0 & \mathrm{tr}(\widehat{S^*})/(2\widehat{\sigma}^2) \\ \mathrm{tr}(\widehat{S^*}\circ\widehat{S^*_{\hat{\gamma}}}) & \mathrm{tr}(\widehat{S^*_{\hat{\gamma}}}\circ\widehat{S^*_{\hat{\gamma}}}) & 0 & 0 & \mathrm{tr}(\widehat{S^*_{\hat{\gamma}}})/(2\widehat{\sigma}^2) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \mathrm{tr}(\widehat{S^*})/(2\widehat{\sigma}^2) & \mathrm{tr}(\widehat{S^*_{\hat{\gamma}}})/(2\widehat{\sigma}^2) & 0 & 0 & N/(4\widehat{\sigma}^4) \end{bmatrix}$$
$$+ \frac{\widehat{\kappa}_3}{N\widehat{\sigma}^4}\sum_{i=1}^{N}\begin{bmatrix} 2\widehat{s_{ii}^*}\widehat{\mu}_i & 2\widehat{s_{ii}^*}\widehat{d}_i\widehat{\mu}_i & \widehat{s_{ii}^*}X_i' & \widehat{s_{ii}^*}\widehat{d}_iX_i' & \frac{1}{2\widehat{\sigma}^2}\widehat{\mu}_i \\ 2\widehat{s_{ii}^*}\widehat{d}_i\widehat{\mu}_i & 2\widehat{s_{ii}^*}\widehat{d}_i\widehat{\mu}_i & \widehat{s_{ii}^*}\widehat{d}_iX_i' & \widehat{s_{ii}^*}\widehat{d}_iX_i' & \frac{1}{2\widehat{\sigma}^2}\widehat{d}_i\widehat{\mu}_i \\ \widehat{s_{ii}^*}X_i & \widehat{s_{ii}^*}\widehat{d}_iX_i & 0 & 0 & \frac{1}{2\widehat{\sigma}^2}X_i \\ \widehat{s_{ii}^*}\widehat{d}_iX_i & \widehat{s_{ii}^*}\widehat{d}_iX_i & 0 & 0 & \frac{1}{2\widehat{\sigma}^2}\widehat{d}_iX_i \\ \frac{1}{2\widehat{\sigma}^2}\widehat{\mu}_i & \frac{1}{2\widehat{\sigma}^2}\widehat{d}_i\widehat{\mu}_i & \frac{1}{2\widehat{\sigma}^2}X_i' & \frac{1}{2\widehat{\sigma}^2}\widehat{d}_iX_i' & 0 \end{bmatrix}$$

where $\widehat{S^*}$ and$\widehat{S^*_{\hat{\gamma}}}$ are the plug-in estimators for S^* and $S^*_{\gamma^*}$, i.e., $\widehat{S^*} = W(I_N - \widehat{\rho}W - \widehat{\varrho}W_{\hat{\gamma}})^{-1}$ and $\widehat{S^*_{\hat{\gamma}}} = W_{\hat{\gamma}}(I_N - \widehat{\rho}W - \widehat{\varrho}W_{\hat{\gamma}})^{-1}$. $\hat{\kappa}_3$ and $\hat{\kappa}_4$, the estimators for κ_3 and κ_4, are given as $\widehat{\kappa}_3 = \frac{1}{N}\sum_{i=1}^{N}\widehat{e}_i^3$ and $\widehat{\kappa}_4 = \frac{1}{N}\sum_{i=1}^{N}\widehat{e}_i^4$ with $\widehat{e}_i = Y_i - \widehat{\rho}W_{i\bullet}Y - \widehat{\varrho}W_{i\bullet}Y\mathbb{1}(q_i \le \widehat{\gamma}) - X_i'\widehat{\beta} - X_i'\widehat{\Delta}\mathbb{1}(q_i \le \widehat{\gamma})$. $\hat{\mu}_i = \widehat{S^*_{i\bullet}}(X\hat{\beta} + X_{\hat{\gamma}}\hat{\Delta})$ with $\widehat{S^*_{i\bullet}}$ the i-th column of $\widehat{S^*}$, and $\widehat{s_{ii}^*}$ is the i-th diagonal element of $\widehat{S^*}$, and $\widehat{d}_i = \mathbb{1}(q_i \le \widehat{\gamma})$. □

We further introduce the following notations for the ease of exposition on the following theorem. Define

$$\Psi_1^* = \frac{1}{N\sigma^{*2}}\sum_{i=1}^{N}\mathbb{E}(\mathbf{X}_i\mathbf{X}_i'|q_i = \gamma^*)f_i(\gamma^*) + \frac{1}{N}\sum_{i=1}^{N}(S_{i\bullet}^{\bullet}S_{i\bullet}^{\bullet\prime} + s_{ii}^{\bullet 2})f_i(\gamma^*)\mathbf{v}_{k+1}\mathbf{v}_{k+1}',$$
$$\Psi_2^* = \lim_{N\to\infty}\frac{\kappa_3}{N\sigma^{*4}}\sum_{i=1}^{N}\begin{bmatrix} 2s_{ii}^{\bullet}\mathbb{E}(S_{i\bullet}^{\bullet}X\beta^*|q_i = \gamma^*)f_i(\gamma^*) & s_{ii}^{\bullet}\mathbb{E}(X_i'|q_i = \gamma^*)f_i(\gamma^*) \\ s_{ii}^{\bullet}\mathbb{E}(X_i|q_i = \gamma^*)f_i(\gamma^*) & 0 \end{bmatrix},$$
$$\Psi_3^* = \lim_{N\to\infty}\frac{\kappa_4 - 3\sigma^{*4}}{N\sigma^{*4}}\sum_{i=1}^{N}s_{ii}^{\bullet 2}f_i(\gamma^*)\mathbf{v}_{k+1}\mathbf{v}_{k+1}',$$

where $\mathbf{v}_{k+1}$ is the first column of the $(k+1)$-dimensional identity matrix. $\mathbf{X}_i$ is defined in Assumption H in Section 2. $f_i(\cdot)$ is the density function of q_i, and κ_3 and κ_4 are the respective third and fourth moments of e_{it}. We have the following theorem on the estimator of the threshold location.

Theorem 3.4 *Under Assumptions A-I, as $N \to \infty$, we have*

$$N^{1-2\nu}(\widehat{\gamma}-\gamma^*) \xrightarrow{d} \frac{C_\psi^{*\prime}(\Psi_1^*+\Psi_2^*+\Psi_3^*)C_\psi^*}{(C_\psi^{*\prime}\Psi_1^* C_\psi^*)^2} \operatorname*{argmax}_s \Big[-\frac{1}{2}|s|+B(s)\Big],$$

where $B(s)$ is a two-sided Brownian motion on $(-\infty,\infty)$, which is defined as $B(s)=B_a(-s)$ for $s<0$ and $B(s)=B_b(s)$ for $v\geq 0$, where $B_a(\cdot)$ and $B_b(\cdot)$ are two independent Brownian motion processes on $[0,\infty)$ with $B_a(0)=B_b(0)=0$. If e_{it} is normally distributed, then

$$N^{1-2\nu}(\widehat{\gamma}-\gamma^*) \xrightarrow{d} \frac{1}{C_\psi^{*\prime}\Psi_1^* C_\psi^*} \operatorname*{argmax}_s \Big[-\frac{1}{2}|s|+B(s)\Big].$$

Remark 3.7 Theorem 3.4 indicates that the final limiting distribution of $\widehat{\gamma}$ is pivotal up to a scalar, which involves the skewness and kurtosis of the errors. Basically, one can use Theorem 3.4 to perform the hypothesis testing on γ. However, this test has some undesirable features, see Remark 4.1 below. We will turn back to this issue in the next section.

4 Hypothesis testing

In this section, we consider several hypothesis testing issues related to the TSAR model. This first, maybe the most important issue is to determine whether the traditional SAR model or the proposed TSAR model should be used in a particular empirical study. This is equivalent to testing the presence of threshold effect. More specifically, in the TSAR model (2.1), we consider the following null and alternative that

$$\mathcal{H}_0: \psi^*=(\varrho^*,\Delta^{*\prime})'=0 \text{ for any } \gamma^*\in\Gamma, \quad \text{vs.} \quad \mathcal{H}_1: \psi^*=(\varrho^*,\Delta^{*\prime})'=\frac{C_\psi^\diamond}{\sqrt{N}} \text{ for some } \gamma^*\in\Gamma.$$

As seen, the parameter γ is not identified under the null hypothesis. This is a typical example of "Davies problem" (Davies, 1977), which entails a non-standard hypothesis testing. There are many studies related with this testing, see Davies (1977), Hansen (1996), Andrews and Ploberger (1994), among others. Recently, Lee, Seo and Shin (2011) deal with the same hypothesis testing in a quite general framework, which can cover some nonregular objective functions such as maximum score method. The Lee, Seo and Shin's treatment is applicable in the TSAR model. However, a drawback of Lee, Seo and Shin's treatment is that the final asymptotic limiting distribution depends on many nuisance parameters. The resampling method, used to obtain the correct critical value, has to be designed carefully.

We propose the following supWald, supLM and supLR statistics, which are defined as

$$\begin{aligned}
\text{supWald} &= N\sup_{\gamma\in\Gamma}\widehat{\psi^u(\gamma)}'\Big[\mathbf{E}'\widehat{\Sigma^u_\vartheta(\gamma)}^{-1}[\widehat{\Sigma^u_\vartheta(\gamma)}+\widehat{\Omega^u_\vartheta(\gamma)}]\widehat{\Sigma^u_\vartheta(\gamma)}^{-1}\mathbf{E}\Big]^{-1}\widehat{\psi^u(\gamma)},\\
\text{supLM} &= N\sup_{\gamma\in\Gamma}\widehat{\mathcal{D}(\gamma)}'\Big(\mathbf{E}'\widehat{\Sigma^r_\vartheta(\gamma)}^{-1}\mathbf{E}\Big)\Big[\mathbf{E}'\widehat{\Sigma^r_\vartheta(\gamma)}^{-1}[\widehat{\Sigma^r_\vartheta(\gamma)}+\widehat{\Omega^r_\vartheta(\gamma)}]\widehat{\Sigma^r_\vartheta(\gamma)}^{-1}\mathbf{E}\Big]^{-1}\\
&\quad\times\Big(\mathbf{E}'\widehat{\Sigma^r_\vartheta(\gamma)}^{-1}\mathbf{E}\Big)\widehat{\mathcal{D}(\gamma)},\\
\text{supLR} &= 2N\sup_{\gamma\in\Gamma}\Big[\mathcal{L}(\widehat{\vartheta^u(\gamma)},\gamma)-\mathcal{L}(\widehat{\vartheta}^r,\gamma)\Big]-N\Big[\mathbf{E}'(\widehat{\vartheta}^r-\widehat{\vartheta^u(\gamma)})\Big]'\Big[\mathbf{E}'\widehat{\Sigma^u_\vartheta(\gamma)}^{-1}\mathbf{E}
\end{aligned}$$

$$+\mathbf{E}'\widehat{\Sigma_{\vartheta}^{u}(\gamma)}^{-1}\mathbf{E}\Big(\mathbf{E}'\widehat{\Sigma_{\vartheta}^{u}(\gamma)}^{-1}\widehat{\Omega_{\vartheta}^{u}(\gamma)}\widehat{\Sigma_{\vartheta}^{u}(\gamma)}^{-1}\mathbf{E}\Big)^{-1}\mathbf{E}'\widehat{\Sigma_{\vartheta}^{u}(\gamma)}^{-1}\mathbf{E}\Big]^{-1}\Big[\mathbf{E}'(\widehat{\vartheta}^{r}-\widehat{\vartheta^{u}(\gamma)})\Big].$$

with

$$\widehat{\mathcal{D}(\gamma)}=\begin{bmatrix}\frac{1}{\widehat{\sigma}^{r2}}Y'W_{\gamma}'(Y-\widehat{\rho}^{r}WY-X\widehat{\beta}^{r})-\frac{1}{N}\mathrm{tr}\Big[W_{\gamma}(I_N-\widehat{\rho}^{r}W)^{-1}\Big]\\ \frac{1}{\widehat{\sigma}^{r2}}X_{\gamma}'(Y-\widehat{\rho}^{r}WY-X\widehat{\beta}^{r})\end{bmatrix}.$$

Here the symbols with the superscript "u" and "r" denote the estimators related with *unrestricted* and *restricted* QML estimators, respectively. For a given γ, $\widehat{\vartheta^{u}(\gamma)}$ denote the unrestricted QML estimators supposing that γ is the underlying threshold value. $\widehat{\theta}^{r}=(\widehat{\rho}^{r},0,\widehat{\beta}^{r\prime},0_{1\times k},\widehat{\sigma}^{r2})'$ denote the restricted QML estimators. Note that under the null hypothesis, no threshold effect exists, so the restricted QML estimators are independent with γ. $\widehat{\Sigma_{\vartheta}^{u}(\gamma)}$ and $\widehat{\Sigma_{\vartheta}^{r}(\gamma)}$ denote the two estimators for $\Sigma_{\vartheta}(\gamma)$ using the unrestricted and restricted QML estimators, respectively. Here $\Sigma_{\vartheta}(\gamma)$ is defined in the same way as $\Sigma_{N,\vartheta}$ in Theorem 3.3 except that γ^{*} is replaced with a predetermined γ. $\widehat{\Omega_{\vartheta}^{u}(\gamma)}$ and $\widehat{\Omega_{\vartheta}^{r}(\gamma)}$ are defined in the same way as $\widehat{\Sigma_{\vartheta}^{u}(\gamma)}$ and $\widehat{\Sigma_{\vartheta}^{r}(\gamma)}$. $\mathbf{E}$ is the selection matrix, which select the coefficients related with threshold effect from the whole parameter ϑ. More specifically, $\mathbf{E}$ is defined as

$$\mathbf{E}=\begin{bmatrix}0 & 1 & \underset{1\times k}{0} & \underset{1\times k}{0} & 0\\ \underset{k\times 1}{0} & \underset{k\times 1}{0} & \underset{k\times k}{0} & I_k & 0\end{bmatrix}'.$$

Before we present theoretical results on three statistics.

Theorem 4.1 *Under Assumptions A-I, if the Pitman local alternative hypothesis holds, as $N\to\infty$, we have*

$$\mathrm{supWald}\xrightarrow{d}\sup_{\gamma\in\Gamma}\Big[\overline{\mathcal{G}(\gamma)}+\overline{\mathcal{Q}(\gamma)}C_{\psi}^{\diamond}\Big]'\mathcal{K}(\gamma,\gamma)^{-1}\Big[\overline{\mathcal{G}(\gamma)}+\overline{\mathcal{Q}(\gamma)}C_{\psi}^{\diamond}\Big],$$
$$\mathrm{supLM}\xrightarrow{d}\sup_{\gamma\in\Gamma}\Big[\overline{\mathcal{G}(\gamma)}+\overline{\mathcal{Q}(\gamma)}C_{\psi}^{\diamond}\Big]'\mathcal{K}(\gamma,\gamma)^{-1}\Big[\overline{\mathcal{G}(\gamma)}+\overline{\mathcal{Q}(\gamma)}C_{\psi}^{\diamond}\Big],$$
$$\mathrm{supLR}\xrightarrow{d}\sup_{\gamma\in\Gamma}\Big[\overline{\mathcal{G}(\gamma)}+\overline{\mathcal{Q}(\gamma)}C_{\psi}^{\diamond}\Big]'\mathcal{K}(\gamma,\gamma)^{-1}\Big[\overline{\mathcal{G}(\gamma)}+\overline{\mathcal{Q}(\gamma)}C_{\psi}^{\diamond}\Big],$$

where $\overline{\mathcal{G}(\gamma)}=\mathbf{E}'\Sigma^{-1}(\gamma,\gamma)\mathcal{G}^{\diamond}(\gamma)$ with $\mathcal{G}^{\diamond}(\gamma)$ being a mean-zero Gaussian process with covariance kernel $\mathbb{E}[\mathcal{G}^{\diamond}(\gamma_a)\mathcal{G}^{\diamond}(\gamma_b)']=\Sigma(\gamma_a,\gamma_b)+\Omega(\gamma_a,\gamma_b)$; $\overline{\mathcal{Q}(\gamma)}=\mathbf{E}'\Sigma^{-1}(\gamma,\gamma)\Sigma(\gamma,\gamma^{*})\mathbf{E}$, and the covariance kernel $\mathcal{K}(\gamma,\gamma)$ is

$$\mathcal{K}(\gamma,\gamma)=\mathbf{E}'\Sigma^{-1}(\gamma,\gamma)\Big[\Sigma(\gamma,\gamma)+\Omega(\gamma,\gamma)\Big]\Sigma^{-1}(\gamma,\gamma)\mathbf{E}.$$

Here the notations $\Sigma(\gamma,\gamma)$, $\Omega(\gamma,\gamma)$ and $\Sigma(\gamma,\gamma^{*})$ are defined before Theorem 3.3.

According to Theorem 4.1, under the null hypothesis, we have

$$\mathrm{supWald}\xrightarrow{d}\sup_{\gamma\in\Gamma}\overline{\mathcal{G}(\gamma)}'\mathcal{K}(\gamma,\gamma)^{-1}\overline{\mathcal{G}(\gamma)}.$$

The same limiting result also holds for supLM and supLR. We note that for a given γ, the expression $\overline{\mathcal{G}(\gamma)}'\mathcal{K}(\gamma,\gamma)^{-1}\overline{\mathcal{G}(\gamma)}$ follows a chi-square distribution with $k+1$ degrees of freedom. But the limiting distribution of $\sup_{\gamma\in\Gamma}\overline{\mathcal{G}(\gamma)}'\mathcal{K}(\gamma,\gamma)^{-1}\overline{\mathcal{G}(\gamma)}$ involves nuisance parameters contained in

$\Sigma(\gamma,\gamma)$, $\Omega(\gamma,\gamma)$ and $\mathcal{K}(\gamma,\gamma)$. So we have to resort to the sampling method to obtain the critical values.

The resampling method here is a little troublesome because we need to guarantee that the generated pseudo errors have the same second, third and fourth moments with the original errors since the final limiting results involve skewness and kurtosis. We use the method proposed by Fleishman (1978) to deal with this issue, and propose the following procedures based on the estimators $\widehat{\vartheta} = \widehat{\vartheta}^u(\widehat{\gamma}) = (\widehat{\rho}, \widehat{\varrho}, \widehat{\beta}', \widehat{\Delta}', \widehat{\sigma}^2)'$ for the unrestricted model. Take B as a large integer. For each $b = 1, 2, \ldots, B$,

(a) Generate independent and identically distributed random variables $\{u_i^b\}_{i=1}^N$, satisfying $\mathbb{E}(u_i) = 0$, $\mathbb{E}(u_i^2) = \widehat{\sigma}^2$, $\mathbb{E}(u_i^3) = \frac{1}{N}\sum_{i=1}^N \widehat{e}_i^3$ and $\mathbb{E}(u_i^4) = \frac{1}{N}\sum_{i=1}^N \widehat{e}_i^4$. We take the method proposed by Fleishman (1978) to generate u_i^b. More specifically, let ε_i be a standard normal variable, we generate a genetic u_i by $u_i = \widehat{\sigma}[\mathbf{a}\varepsilon_i + \mathbf{b}(\varepsilon_i^2 - 1) + \mathbf{c}\varepsilon_i^3]$, where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are the solution of the following equations system

$$\begin{aligned} \frac{\frac{1}{N}\sum_{i=1}^N \widehat{e}_i^4 - 3\widehat{\sigma}^4}{\widehat{\sigma}^4} &= 24\Big[\mathbf{ac} + \mathbf{b}^2(1 + \mathbf{a}^2 + 28\mathbf{ac}) + \mathbf{c}^2(12 + 48\mathbf{ac} + 141\mathbf{b}^2 + 225\mathbf{c}^2)\Big], \\ \frac{\frac{1}{N}\sum_{i=1}^N \widehat{e}_i^3}{\widehat{\sigma}^3} &= 2\mathbf{b}(\mathbf{a}^2 + 24\mathbf{ac} + 105\mathbf{c}^2 + 2), \\ 2 &= 2\mathbf{a}^2 + 12\mathbf{ac} + \frac{1}{(\mathbf{a}^2 + 24\mathbf{ac} + 105\mathbf{c}^2 + 2)^2}\Big[\frac{\frac{1}{N}\sum_{i=1}^N \widehat{e}_i^3}{\widehat{\sigma}^3}\Big]^2 + 30\mathbf{c}^2. \end{aligned}$$

(b) Let $\widehat{S} = W(I_N - \widehat{\rho}W)^{-1}$ and $\widehat{S}_\gamma = W_\gamma(I_N - \widehat{\rho}W)^{-1}$. Set

$$F_N^b(\gamma) = \frac{1}{\sqrt{N}\widehat{\sigma}^2}\begin{bmatrix} \widehat{\beta}'X'\widehat{S}'u^b + u^{b\prime}\widehat{S}u^b - \widehat{\sigma}^2\frac{1}{N}\mathrm{tr}(\widehat{S}) \\ \widehat{\beta}'X'\widehat{S}_\gamma'u^b + u^{b\prime}\widehat{S}_\gamma u^b - \widehat{\sigma}^2\frac{1}{N}\mathrm{tr}(\widehat{S}_\gamma) \\ X'u^b \\ X_\gamma'u^b \\ (u^{b\prime}u^b - N\widehat{\sigma}^2)/(2\widehat{\sigma}^2) \end{bmatrix}$$

(c) Compute the quantity

$$\widehat{SF}_N^b = \max_{\gamma\in\{q_i\}\cap\Gamma} F_N^b(\gamma)'\widehat{\Sigma_\vartheta^u(\gamma)}^{-1}\mathbf{E}\Big[\mathbf{E}'\widehat{\Sigma_\vartheta^u(\gamma)}^{-1}[\widehat{\Sigma_\vartheta^u(\gamma)}+\widehat{\Omega_\vartheta^u(\gamma)}]\widehat{\Sigma_\vartheta^u(\gamma)}^{-1}\mathbf{E}\Big]^{-1}\mathbf{E}\widehat{\Sigma_\vartheta^u(\gamma)}^{-1}F_N^b(\gamma).$$

Let $\widehat{c}_{1-\alpha}^B$ be the empirical $(1-\alpha)$-quantile of the sample $\{\widehat{SF}_N^1, \ldots, \widehat{SF}_N^B\}$, where $\alpha \in (0,1)$ is the nominal size. One should reject the null hypothesis if supWald (or supLM, supLR) is larger than $\widehat{c}_{1-\alpha}^B$. We have the following theorem on the bootstrap validity.

Theorem 4.2 *Under Assumptions A-I, as $N, B \to \infty$, we have that*

(i) $\mathbb{P}(\text{supWald} > \widehat{c}_{1-\alpha}^B) \to \alpha$, $\mathbb{P}(\text{supLM} > \widehat{c}_{1-\alpha}^B) \to \alpha$, *and* $\mathbb{P}(\text{supLR} > \widehat{c}_{1-\alpha}^B) \to \alpha$ *under* $\mathcal{H}_0$,

(ii) $\mathbb{P}(\text{supWald} > \widehat{c}_{1-\alpha}^B) \to 1$, $\mathbb{P}(\text{supLM} > \widehat{c}_{1-\alpha}^B) \to 1$, *and* $\mathbb{P}(\text{supLR} > \widehat{c}_{1-\alpha}^B) \to 1$ *under* $\mathcal{H}_1$ *as* $\|C_\psi^\diamond\| \to \infty$.

Now we consider the hypothesis testing $\mathcal{H}_0 : \gamma = \gamma^*$ and the alternative $\mathcal{H}_1 : \gamma \neq \gamma^*$ for a given γ^*. Following Hansen (2000), we consider the following likelihood ratio statistics,

$$\mathrm{LR}(\gamma^*) = N\Big[\mathcal{L}(\widehat{\theta}_{\hat{\gamma}}, \widehat{\gamma}) - \mathcal{L}(\widehat{\theta}_{\gamma^*}, \gamma^*)\Big].$$

The asymptotic result of $\mathrm{LR}(\gamma^*)$ is given in the following theorem.

Theorem 4.3 *Under Assumptions A-I, as $N \to \infty$, under the null $\mathcal{H}_0$, we have*

$$\mathrm{LR}(\gamma^*) \xrightarrow{d} \mathcal{C}_\psi^* \hbar, \qquad \text{with} \qquad \mathcal{C}_\psi^* = \frac{C_\psi^{*\prime}(\Psi_1^* + \Psi_2^* + \Psi_3^*)C_\psi^*}{C_\psi^{*\prime}\Psi_1^* C_\psi^*}.$$

and $\hbar = \max_{s\in\mathbb{R}}[-\frac{1}{2}|s| + B(s)]$. The distribution of $\hbar$ is $\mathbb{P}(\hbar \le x) = (1 - e^{-x})^2$.

Remark 4.1 Theorem 3.4 gives the final limiting distribution for $\mathrm{LR}(\gamma^*)$, which is used to do the statistical inference on γ^*. In this viewpoint, the above theorem looks a little bit redundant since Theorem 3.4 can also perform this work. However, there are two reasons that we prefer to use the above theorem. First, if the e_i is normally distributed, the adjusting constant $\frac{C^{*\prime}(\Psi_1^*+\Psi_2^*+\Psi_3^*)C^*}{C^{*\prime}\Psi_1^* C^*}$ reduces to one in the above theorem, but the adjusting constant in Theorem 3.4 is $\frac{1}{C^{*\prime}\Psi_1^* C^*}$, which still need a consistent estimation. Since any estimation would introduce an error, the statistical inference based on the above theorem would be more accurate in this sense. Second, the distribution of $\max_{s\in\mathbb{R}}[-\frac{1}{2}|s|+B(s)]$ has an explicit expression, whose critical values can be easily calculated. This is in contrast with the distribution $\mathrm{argmax}_{s\in\mathbb{R}}[-\frac{1}{2}|s| + B(s)]$ in Theorem 3.4, whose critical values are relatively difficult to obtain.

Remark 4.2 The parameter $\mathcal{C}_\psi^*$ can be estimated by the following way. Let $\mathbf{Z} = [WY, X]$ and $\mathbf{z}_i$ be the ith element of $\mathbf{Z}$. Let $\widehat{\mathbf{e}}_{i1} = (\widehat{\psi}'\mathbf{z}_i)^2$ and

$$\widehat{\mathbf{e}}_{i2} = \widehat{\mathbf{e}}_{i1} + \frac{\widehat{\kappa}_3}{\widehat{\sigma}^4}\Big[2\widehat{\varrho}^2\widehat{s_{ii}^{\bullet}}\widehat{S_{i\bullet}^{\bullet}}X\widehat{\beta} + 2\widehat{\varrho}\widehat{s_{ii}^{\bullet}}X_i'\widehat{\Delta}\Big].$$

We consider the following regression

$$\widehat{\mathbf{e}}_{ij} = \widehat{c}_{j0} + \widehat{c}_{j1}q_i + \widehat{c}_{j2}q_i^2 + \widehat{\varepsilon}_{ji}.$$

Then we estimate $\mathcal{C}_\psi^*$ as

$$\widehat{\mathcal{C}_\psi^*} = \frac{\widehat{c}_{10} + \widehat{c}_{11}\hat{\gamma} + \widehat{c}_{12}\hat{\gamma}^2 + \widehat{\varrho}^2\Big[\frac{1}{N}\sum_{i=1}^N \widehat{s_{ii}^{\bullet 2}}\Big]}{\widehat{c}_{20} + \widehat{c}_{21}\hat{\gamma} + \widehat{c}_{22}\hat{\gamma}^2 + \widehat{\varrho}^2(1 + \frac{\hat{\kappa}_4 - 3\hat{\sigma}^4}{\hat{\sigma}^4})\Big[\frac{1}{N}\sum_{i=1}^N \widehat{s_{ii}^{\bullet 2}}\Big]}.$$

5 Simulations

We run Monte Carlo simulations to investigate the finite sample performance of the QML estimators and the related statistics. We first consider the performance of the QML estimators. The data are generated according to

$$Y_i = \alpha^* + \rho^* \sum_{j=1}^N w_{ij}Y_j + \varrho^* \sum_{j=1}^N w_{ij}Y_j d_i(\gamma^*) + X_i\beta^* + X_i\Delta^* d_i(\gamma^*) + \sigma^* e_i,$$

with $(\rho^*, \alpha^*, \beta^*, \sigma^{*2}) = (0.5, 1, 1, 0.25)$, $(\varrho^*, \Delta^*) = N^{-0.2}(0.4, 0.6)$ and $\gamma^* = 0.2$. The spatial weights matrices used in simulations are "q ahead and q behind" spatial weights matrix as in Kelejian and Prucha (1999), which is obtained as follows: all the units are arranged in a circle and each unit is affected only by the q units immediately before it and immediately after it with equal weight. Following Kelejian and Prucha (1999), we normalize the spatial weights matrix by letting the sum of each row equal to 1. X_i generated independently from $N(0, 1)$ for each i, and $q_i = X_i + 0.5\varepsilon_i$ with $\varepsilon_i \sim N(0, 1)$. We consider three types of errors e_i: (i) the standard normal distribution, $N(0, 1)$; (ii) the normalized chi-square distribution with 2 degrees of freedom, $\frac{1}{2}(\chi^2(2) - 2)$; and (iii) the normalized t-distribution with 9 degrees of freedom, $\sqrt{\frac{7}{9}}t(9)$.

We use the empirical bias (Bias), standard deviation (SD), and coverage probability (CP) of the 95% confidence interval for the corresponding true value to measure the performance of the QML estimators for the regression coefficients and the variance of disturbance. For the estimator of the threshold location, we only report the bias and CP. Note that because of the nonstandard hypothesis testing on γ^*, SD is not a good indicator of the performance in this situation.

Table 1 below presents the simulation results under the sample size $N = 100, 200, 300, 500$ and 700. The results are obtained with 1000 repetitions. Overall, we find that the performance of all the estimators converges to what our theory predicts in large samples. The estimator of σ^2 performs a little inferiorly in comparison with other parameters. Overall, all the QML estimators have good performance in finite samples.

6 Conclusion

This paper considers the estimation and inference issues on the TSAR model. We propose using the QML method to estimate the TSAR model. We build up a relatively complete asymptotic theory on the QML estimators, including the consistencies, convergence rates and limiting distributions. We also consider the hypothesis testing issues related with the proposed TSAR model. The present paper has many interesting extensions. First, we may consider the panel threshold spatial autoregressive model along the spirit of Li (2018). In the panel data setup, the incidental parameters issue is additionally introduced and should be carefully addressed. Second, we may consider relaxing the exogeneity assumption on the regressors and threshold variable along the spirit of Seo and Shin (2016). These interesting extensions will be pursued in the future work.

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4), 821-856.

Andrews, D. W., and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6), 1383-1414.

Anselin, L. (1988). Spatial econometrics: methods and models. *The Netherlands: Kluwer Academic Publishers.*

Table 1: The performance of the QML estimator under different error distribution

	$\boldsymbol{\rho}$			$\boldsymbol{\varrho}$			$\boldsymbol{\beta}$			$\boldsymbol{\Delta}$			$\boldsymbol{\sigma^2}$			$\boldsymbol{\gamma}$	
N	Bias	SD	CP	Bias	SD	CP	Bias	SD	CP	Bias	SD	CP	Bias	SD	CP	Bias	CP
$\boldsymbol{e_i \sim N(0,1)}$																	
100	-0.0254	0.0696	91.3%	0.0179	0.0599	91.4%	0.0104	0.1090	90.1%	0.0108	0.1465	89.2%	-0.0185	0.0325	85.6%	-0.0706	96.7%
200	-0.0132	0.0489	92.7%	0.0119	0.0428	92.7%	0.0057	0.0779	91.6%	0.0068	0.1037	89.7%	-0.0089	0.0241	90.8%	-0.0371	95.3%
300	-0.0114	0.0400	93.8%	0.0096	0.0351	93.6%	0.0071	0.0640	91.7%	0.0009	0.0847	91.9%	-0.0058	0.0199	91.8%	-0.0208	95.8%
500	-0.0070	0.0311	93.9%	0.0058	0.0273	96.2%	0.0047	0.0495	93.3%	0.0011	0.0655	92.5%	-0.0036	0.0157	92.9%	-0.0102	95.5%
700	-0.0057	0.0264	94.7%	0.0054	0.0232	94.6%	0.0065	0.0420	94.1%	-0.0043	0.0555	94.4%	-0.0021	0.0133	94.9%	-0.0072	95.7%
$\boldsymbol{e_i \sim (\chi^2(2)-2)/2}$																	
100	-0.0294	0.0699	90.8%	0.0180	0.0624	90.6%	0.0069	0.1081	89.5%	0.0156	0.1454	88.7%	-0.0164	0.0560	80.8%	-0.0634	92.3%
200	-0.0151	0.0494	93.2%	0.0131	0.0450	93.4%	0.0098	0.0777	92.8%	0.0034	0.1037	91.6%	-0.0084	0.0439	86.6%	-0.0365	93.8%
300	-0.0106	0.0401	92.8%	0.0101	0.0367	94.4%	0.0080	0.0634	94.6%	0.0000	0.0845	91.8%	-0.0052	0.0374	89.3%	-0.0199	94.5%
500	-0.0046	0.0312	94.9%	0.0053	0.0287	94.4%	0.0053	0.0492	93.5%	-0.0014	0.0653	92.2%	-0.0046	0.0295	90.9%	-0.0111	94.2%
700	-0.0074	0.0267	93.7%	0.0060	0.0245	94.7%	0.0049	0.0422	93.8%	-0.0013	0.0557	94.0%	-0.0016	0.0260	92.1%	0.0052	93.9%
$\boldsymbol{e_i \sim \sqrt{\frac{7}{9}}t(9)}$																	
100	-0.0315	0.0715	92.7%	0.0167	0.0628	94.3%	0.0181	0.1140	91.8%	-0.0122	0.1484	90.7%	-0.0126	0.0572	80.8%	0.0089	94.4%
200	-0.0153	0.0497	95.4%	0.0105	0.0449	96.1%	0.0103	0.0789	93.3%	-0.0087	0.1041	92.2%	-0.0063	0.0446	86.2%	-0.0015	95.4%
300	-0.0120	0.0405	93.5%	0.0083	0.0367	94.7%	0.0082	0.0641	92.8%	-0.0053	0.0847	92.2%	-0.0051	0.0374	89.7%	0.0002	94.1%
500	-0.0071	0.0313	93.7%	0.0072	0.0287	95.8%	0.0083	0.0497	94.4%	-0.0071	0.0655	93.1%	-0.0048	0.0297	91.0%	0.0068	93.9%
700	-0.0064	0.0266	94.2%	0.0060	0.0244	95.2%	0.0033	0.0420	94.8%	0.0003	0.0554	94.2%	-0.0030	0.0255	92.8%	0.0027	94.4%

Bao, Y., and Ullah, A. (2010). Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference*, 140(5), 1193-1205.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1), 41-55.

Billingsley, P. (1968). Convergence of probability measures. *John Wiley & Sons*.

Caner, M., and Hansen, B. E. (2004). Instrumental variable estimation of a threshold model. *Econometric Theory*, 20(5), 813-843.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4), 1239-1267.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The annals of statistics*, 21(1), 520-533.

Chan, K. S., and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2), 413-426.

Chen, S. X., Zhang, L. X., and Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490), 810-819.

Cliff, A. D., and Ord, J. K. (1973) Spatial autocorrelation, *London: Pion Ltd.*

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2), 247-254.

Deng, Y. (2018). Estimation for the spatial autoregressive threshold model. *Economics Letters*, 171, 172-175.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4), 521-532.

Hall, P., and Heyde, C. C. (1980). Martingale limit theory and its application. *Academic press*.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2), 413-430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3), 575-603.

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

Kelejian, H. H., and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2), 509-533.

Kim, J., and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, 18(1), 191-219.

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

Lee, L. F., and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2), 165-185.

Lee, S., Seo, M. H., and Shin, Y. (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association*, 106(493), 220-231.

Li, K. (2017). Fixed-effects dynamic spatial panel data models and impulse response analysis. *Journal of Econometrics*, 198(1), 102-121.

Li, K. (2018). Spatial panel data models with structual change. *MPRA working paper No.85388*.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4), 825-860.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3), 531-542.

Marcucci, J., and Quagliariello, M. (2009). Asymmetric effects of the business cycle on bank credit risk. *Journal of Banking & Finance*, 33(9), 1624-1635.

Masanjala, W. H., and Papageorgiou, C. (2004). The Solow model with CES technology: nonlinearities and parameter heterogeneity. *Journal of Applied Econometrics*, 19(2), 171-201.

Miao, K., Li, K., and Su, L. (2020). Threshold panel data model with interactive effects. *Journal of Econometrics*, 219 (1), 137-170.

Pukelshim, F. (1980). Multilinear estimation of skewness and kutosis in linear models. *Metrika*, 27(1), 103-113.

Seo, M. H., and Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2), 169-186.

Su, L., and Xu, P. (2017). Common threshold in quantile regressions with an application to pricing for reputation. *Econometric Reviews*, 1-37.

Tong, H. (1978). On a threshold model. In: Chen, C, (ed.) *Pattern Recognition and Signal Processing. NATO ASI Series E: Applied Sc.* (29). Sijthoff & Noordhoff, Netherlands, pp. 575-586. ISBN 9789028609785

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American statistical association*, 84(405), 231-240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, 93(443), 1188-1202.

Ullah, A. (2004). Finite sample econometrics. *Oxford University Press*.

Utev, S., and Peligrad, M. (2003). Maximal inequalities and an invariance principle for a class of weakly dependent random variables. *Journal of Theoretical Probability*, 16(1), 101-115.

Van Der Vaart A. W., and Wellner J. A. (1996). *Weak Convergence and Empirical Processes with Application to Statistics*. Springer.

Wiens, D. P. (1992). On moments of quadratic forms in non-spherically distributed variables. *Statistics: a journal of statistical and applied statistics*, 23(3), 265-270.

Yu, J., de Jong, R., and Lee, L. F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1), 118-134.

Table 1: The symbols list

Symbol	Definition	Symbol	Definition
ρ	Regression coefficient	ϱ	Regression coefficient
β	Regression coefficient	Δ	Regression coefficient
σ^2	Variance of error	γ	Threshold value
Γ	Parameters Space for γ	W	Spatial weights Matrix
X	Exogenous regressors	$D(\rho,\varrho,\gamma)$	$I_N-\rho W-\varrho W_\gamma$
D^*	$I_N-\rho^* W-\varrho^* W_{\gamma^*}$	$D^\bullet$	$I_N-\rho^* W$
S^*	WD^{*-1}	S^*_γ	$W_\gamma D^{*-1}$
S^*_{γ,γ^*}	$S^*_\gamma - S^*_{\gamma^*}$	$S^\bullet$	$WD^{\bullet-1}$
$S^\bullet_\gamma$	$W_\gamma D^{\bullet-1}$	$S^\bullet_{\gamma,\gamma^*}$	$S^\bullet_\gamma - S^\bullet_{\gamma^*}$
θ	$(\rho,\varrho,\beta',\Delta',\sigma^2,\gamma)'$	ϑ	$(\rho,\varrho,\beta',\Delta',\sigma^2)$
ω	$(\rho,\varrho,\beta',\Delta')'$	φ	$(\rho,\varrho,\sigma^2)'$
ψ	$(\varrho,\Delta')'$	δ	(ρ,β)
ϕ	$(\rho,\varrho)'$	μ_i	$\sum_{j=1}^N s^\bullet_{ij} X_j \beta^*$
$f_i(t)$	Density function of q_i	$f_{ij}(t,s)$	Density function of (q_i,q_j)
$U(\gamma)$	$[S^\bullet X\beta^*, S^\bullet_\gamma X\beta^*, X, X_\gamma]$	$R(\gamma)$	$[S^\bullet_{\gamma,\gamma^*} X\beta^*, X_{\gamma,\gamma^*}]$
$Z(\gamma,\gamma^*)$	$[W_{\gamma,\gamma^*} Y, X_{\gamma,\gamma^*}]$	κ_d	the d-th moment of e_i