Spatial Heteroskedasticity and Spill-over Effect on Volatility: A GARCH-like Model

Yang Yang

September 14, 2021

Abstract

We develop a GARCH-like model for capturing spatial heteroskedasticity and spill-over effect on volatility level. By proving the $\alpha-$mixing property of the DGP, we derive the maximum-likelihood estimator and prove some asymptotic properties. Monte Carlo simulations show good finite sample properties of the MLE. When applying this model to US housing market, we identify the spatial spillover effect on volatility of county level housing prices in Northeastern US which can not be explained by SAR model.

1 Introduction

This paper developed a GARCH-like model to study spatial heteroskedasticity and spill-over effect at volatility level. Spatial Heteroskedasticity comes from heterogeneity. On one hand, demographic and economic differences across regions can make people have different preferences, risk aversion levels and mobilities. Thus, people in different regions will have different responds and decisions when they are facing same policy, economic shock and environmental change, such like air pollution. On the other hand, as each region contains multiple individuals and legal entities, such like companies, there exist heterogeneity inside a region. Thus, considering treatment effect and macroeconomic performance, the heterogeneities will probably cause heteroskedasticity across different regions. More over, due to information spread, labor mobility or similar geographic characteristics, geographically closer regions or regions with closer economic correlation will have spill over effect to each other. The spill-over effect will also show up on volatility level, not only the spatial process itself which captured by SAR models. If we could capture both the spatial heteroskedasticity and the spill-over effect at volatility, it would be helpful to understand treatment effect of a policy better, or understand co-movement among asset prices better, and then make better policies or trading strategies.There are growing interests considering spatial correlation in many fields, not only in traditional regional economics. For example, in asset pricing area, Steven Kou. etc (2017) introduces spatial correlation into traditional CAPM, and develops a spatial arbitrage pricing theory which can explain cross-sectional stock returns in eurozone. Robert J. Richmond (2019) uses trade network to uncover the exposure to global risk and tries to explain the currency risk premium. A model which can captures spill-over effect and spatial heteroskedasticity would be helpful in empirical research on how risk spread across different markets.

In time series, ARCH and GARCH models are well established and empirically very useful models to capture the autoregressive conditional heteroskedasticity structure and predict volatility of macroeconomic variables, especially for financial asset returns and derivative prices. However, in spatial econometric literature, although the SAR is well established, it can only capture the spill-over effect on the return level, but not the volatility level. In previous literatures, there are some paper discussed the spatial heteroskedasticity issue. Bera and Simlai (2005) formats a spatial ARCH model with conditional setting and give an application on Boston housing price. Sato and Matsuda (2017) suggests a similar model and applies it in Tokyo land market. Caporin and Paruolo (2006), as well as Borovkova and Lopuhaa (2012), formulates GARCH like spatial models. However, the asymptotic theory of in those papers are not well-established. Kelejian and Prucha (2006) develops an HAC estimator for regressions with spatial heteroskedasticity error terms. But this paper does not focused on the structure of spatial heteroskedasticity. Thus, in this paper, we try to develop an unconditional spatial GARCH-like model to capture both the spatial heteroskedasticity and spill-over effect at the volatility level. Moreover, we try to give a formal discussion on MLE estimator and its asymptotic properties. Besides, by analyzing the annual change of housing price indexes in 240 northeastern US counties, we find out that SAR model is not enough to capture the spatial correlation among housing prices across regions, and our model can be a good supplemental method to capture the effects which SAR can not handle.

In the following part of this paper, Section 2 is the model setting and the proof of $\alpha-$mixing property which will be useful to develop further results on asymptotic theory. Section 3 is about maximum-likelihood estimation, including the built up of likelihood function, concentrated likelihood method and also the proof of identification, consistency and asymptotic Normality for a special case. Section 4 is Monte Carlo simulation for finite sample performance of the maxim-likelihood estimators. Section 5 is an Lagrangian multiplier test developed to differentiate the general case and a special case. Finally, Section 6 is the application in housing market.

2 Model Setting and $\alpha-$mixing Property

2.1 Data Generating Process

Suppose n individual spatial unites in an economy are located in a region $D_n \subset \mathbb{R}^d$, where the cardinality of D_n is $|D_n| = n$. For convenience, we name these n units as $1, 2, \cdots, n$. The distance between individuals i and j is denotes by d_{ij}. For regularity, we need the following assumption:

Assumption 1:

$d_{ij} \geqslant 1$ for any $i \neq j$.

The data generating process is defined by the following two equations:

$$\begin{aligned} u_{i,n} &= \sqrt{h_{i,n}}\varepsilon_{i,n} \\ logh_{i,n} &= \phi\sum_{j=1}^{n} w_{ij,n} logu_{j,n}^2 + \theta\sum_{j=1}^{n} w_{ij,n} logh_{j,n} + \alpha \end{aligned}$$

where $\varepsilon_{i,n}$ are i.i.d random process with mean 0 and variance 1, and $w_{ij,n}$ are spatial correlation between individual i and j. This model can be viewed as a spatial analog of GARCH or EGARCH in some sense, but without conditional setting. However, there is no conditional variance setting here. The $h_{i,n}$ is an indicator of the unconditional variance term. For particular $\varepsilon_{i,n}$ process, it has clear correlation between unconditional variance which will be discussed later.

After an easy transformation, we can get the following equation:

$$logu_{i,n}^2 = (\phi+\theta)\sum_{j=1}^{n} w_{ij,n} logu_{j,n}^2 + log\varepsilon_{i,n}^2 - \theta\sum_{j=1}^{n} w_{ij,n} log\varepsilon_{j,n}^2 + \alpha \tag{1}$$

Denote $logu_n^2 \equiv (logu_{1,n}^2, \cdots, logu_{n,n}^2)^{'}$, $log\varepsilon_n^2 \equiv (log\varepsilon_{1,n}^2, \cdots, log\varepsilon_{n,n}^2)^{'}$, $W_n \equiv (w_{ij})_{n\times n}$, $1_n \equiv (1, \cdots, 1)^{'}$, we can write the following form for $logu_n^2$ process:

$$[I_n - (\phi+\theta)W_n]logu_n^2 = (I_n - \theta W_n)log\varepsilon_n^2 + \alpha 1_n \tag{2}$$

To make sure this equation to be spatially stable for $logu_n^2$, we need $I_n-(\phi+\theta)W_n$ invertible. In the following section, I will put more restrictive assumptions on this matrix for further properties. Just focused on the form itself, it is easy to see the $\left\{logu_{i,n}^2\right\}_{i\in D_n}$ is a SARMA process but with a non-zero drift term. We will use this result to extend properties of $\left\{logu_{i,n}^2\right\}_{i\in D_n}$ to $\{u_{i,n}\}_{i\in D_n}$. In the remaining part of this paper, when we allow $\theta \neq 0$, we call it a GARCH-like model. when $\theta \equiv 0$ as a special case, we call it an ARCH-like model where $\left\{logu_{i,n}^2\right\}_{i\in D_n}$ is a SAR process. In general, this two type of model are both motivated by exponential stochastic volatility model in time series literature.

When $\sigma_\varepsilon^2 \neq 1$, we have

$$\begin{aligned}
\log u_{i,n}^2 &= (\phi+\theta)\sum_{j=1}^n w_{ij,n}\log u_{j,n}^2 + \log\varepsilon_{i,n}^2 - \theta\sum_{j=1}^n w_{ij,n}\log\varepsilon_{j,n}^2 + \alpha \\
&= (\phi+\theta)\sum_{j=1}^n w_{ij,n}\log u_{j,n}^2 + \log\left(\frac{\varepsilon_{i,n}}{\sigma}\right)^2 - \theta\sum_{j=1}^n w_{ij,n}\log\left(\frac{\varepsilon_{j,n}}{\sigma}\right)^2 \\
&+ \ln\sigma^2 - \theta\sum_{j=1}^n w_{ij,n}\ln\sigma^2 + \alpha \\
&= (\phi+\theta)\sum_{j=1}^n w_{ij,n}\log u_{j,n}^2 + \log\left(\frac{\varepsilon_{i,n}}{\sigma}\right)^2 - \theta\sum_{j=1}^n w_{ij,n}\log\left(\frac{\varepsilon_{j,n}}{\sigma}\right)^2 \\
&+ (1-\theta)\ln\sigma^2 + \alpha
\end{aligned}$$

This shows that the effect of σ_ε^2 is obsorbed by α. Thus, normalize $\sigma_\varepsilon^2 = 1$ can still generate the same DGP.

2.2 $\alpha-$mixing Property

In Jenish and Prucha (2009), they introduced mixing concepts from time series literatures to arbitrary random field and built up some very useful asymptotic results. One of the most important mixing concept is $\alpha-$mixing coefficient defined as following:

Definition 1 ($\alpha-$mixing coefficient):

For two sub-σ-fields $\mathcal{A}$ and $\mathcal{B}$, the α-mixing coefficient, also called strong mixing coefficient, between $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$\alpha(\mathcal{A},\mathcal{B}) \equiv \sup(|\Pr(A\cap B) - \Pr(A)\Pr(B)| : A\in\mathcal{A}, B\in\mathcal{B})$$

For any two random variables(vectors) Y and Z, $\alpha(Y,Z) \equiv \alpha(\sigma(Y),\sigma(Z))$, where $\sigma(Y)$ and $\sigma(Z)$ are the σ-fields generated by Y and Z.

Based on this concept, we can define the concept of $\alpha-$mixing random field for spatial setting:

Definition 2 ($\alpha-$mixing random field):

Let $X = \{X_{i,n} : i \in D_n, n \in \mathbb{N}\}$ be a triangular array of random vectors on a probability space $(\Omega, \mathcal{F}, Pr)$. For $U, V \subseteq D_n$, denote $\sigma_n(U) \equiv \sigma(X_{i,n} : i \in U)$ the σ-field generated by the random *vectors $X_{i,n}$ in U. For simplicity, $\alpha_n(U,V) \equiv \alpha(\sigma_n(U),\sigma_n(V))$. The notion of α-mixing coefficient for X is*

$$\alpha_{k,l,n}(r) \equiv \sup_{U,V\subseteq D_n} \{\alpha_n(U,V) : |U| \leqslant k, |V| \leqslant l, d(U,V) \geqslant r\}$$

where $d(U,V) \equiv \inf\{d_{ij} : i \in U, j \in V\}$ is the distance between U and V. Denote $\alpha_{k,l}(r) \equiv \sup_n \alpha_{k,l,n}(r)$. X is said to be α-mixing iff for any $k,l \in \mathbb{N}, \lim_{r\to\infty} \alpha_{k,l}(r) = 0$.

From the definitions above, we can see that the dependence between any two different regions are descending to zero as their distance increasing. The remaining part is to find out whether our process $\{u_{i,n}\}_{i\in D_n}$ is $\alpha-$mixing or not.

With noticing that $u_{i,n} = |u_{i,n}|\, sgn(u_{i,n}) = exp(\frac{1}{2}logu_{i,n}^2)sgn(\varepsilon_{i,n})$, and the function $f(a,b) = exp(\frac{1}{2}a)sgn\,(b)$ is measurable. Since $\alpha-$mixing property is preserved measurable functions which is also proved in Jenish and Prucha (2009), if we could get that $\{(logu_{i,n}^2, \varepsilon_{i,n})^{'}\}_{i\in D_n}$ is α-mixing, then $\{u_{i,n}\}_{i\in D_n}$ is α-mixing.

As $\{\varepsilon_{i,n}\}_{i\in D_n}$ is an i.i.d random process with mean 0 and variance 1, we have $\{(log\varepsilon_{i,n}^2, \varepsilon_{i,n})\}_{i\in D_n}$ is an i.i.d random vector. Denote $G_n \equiv [I_n - (\phi+\theta)W_n]^{-1}(I_n - \theta W_n) \equiv (g_{ij})_{n\times n}$, $C_n \equiv \alpha[I_n - (\phi+\theta)W_n]^{-1}1_n \equiv (c_{1,n},\cdots,c_{n,n})^{'}$, we can get

$$\begin{pmatrix} logu_{i,n}^2 \\ \varepsilon_{i,n} \end{pmatrix} = \begin{pmatrix} c_{i,n} \\ 0 \end{pmatrix} + \sum_{j=1,j\neq i}^{n} \begin{pmatrix} g_{ij} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} log\varepsilon_{j,n}^2 \\ \varepsilon_{j,n} \end{pmatrix} + \begin{pmatrix} g_{ii} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} log\varepsilon_{i,n}^2 \\ \varepsilon_{i,n} \end{pmatrix}$$

Denote $z_{i,n} \equiv (logu_{i,n}^2, \varepsilon_{i,n})^{'}$, $e_{i,n} \equiv (log\varepsilon_{i,n}^2, \varepsilon_{j,n})^{'}$, $Z_n \equiv (z_{1,n}^{'},\cdots,z_{n,n}^{'})^{'}$, $e_n \equiv (e_{1,n}^{'},\cdots,e_{n,n}^{'})^{'}$, $\widetilde{G}_{ij,n} \equiv \begin{pmatrix} g_{ij} & 0 \\ 0 & 0 \end{pmatrix}$ when $i \neq j$, $\widetilde{G}_{ii,n} \equiv \begin{pmatrix} g_{ii} & 0 \\ 0 & 1 \end{pmatrix}$, $\widetilde{G}_n \equiv \begin{pmatrix} \widetilde{G}_{11,n} & \cdots & \widetilde{G}_{in,n} \\ \vdots & \ddots & \vdots \\ \widetilde{G}_{n1,n} & \cdots & \widetilde{G}_{nn,n} \end{pmatrix}$, then we have

$$Z_n = \widetilde{G}_n e_n + (c_{1,n}, 0, c_{2,n}, 0, \cdots, c_{n,n}, 0)^{'}$$

If $\{\begin{pmatrix} \widetilde{z}_{i,n} \\ (c_{i,n},0)^{'} \end{pmatrix}\}_{i\in D_n} \equiv \{\begin{pmatrix} z_{i,n} - (c_{i,n},0)^{'} \\ (c_{i,n},0)^{'} \end{pmatrix}\}_{i\in D_n}$ is α-mixing, then $\{z_{i,n}\}_{i\in D_n}$ is α-mixing since $f(x,y) = x+y$ is measurable. First, we can focus on the stochastic part:

$$\widetilde{Z}_n \equiv (\tilde{z}_{1,n}^{'},\cdots,\tilde{z}_{n,n}^{'})^{'} = \widetilde{G}_n e_n$$

In Xu and Lee (2019), they have a detailed discussion on the α-mixing property of this type of spatial processes (Section 3.1, the spatial autoregressive model). With proper assumptions on $\widetilde{G}_n$ and e_n, we can make sure $\widetilde{Z}_n$ is α-mixing. First, we need the following assumption for e_n:

Assumption 2:

(1) The distribution of $\varepsilon_{i,n}$ is absolute continuous.

(2) For any $i \in D_n$ and n, $\|e_n\|_{L^2} \equiv \sup_{j\in D_n} \max\left\{\left\|log\varepsilon_{j,n}^2\right\|_{L^2}, \|\varepsilon_{j,n}\|_{L^2}\right\} < \infty$.

Since the density of e_n are concentrated at $\left(\ln x^2, x\right) \in \mathbb{R}^2$ and $f_e\left(\ln x^2, x\right) = f_\varepsilon\left(x\right)$, the absolute continuity of distribution of e_n is ensured. So, e_n satisfies the Assumption 2.3 and 2.4 in Xu and Lee (2019).

Now, we need to focus on the matrix $\widetilde{G}_n$. Recall the form of $\widetilde{G}_n$:

$$\widetilde{G}_n = \begin{pmatrix} \widetilde{G}_{11,n} & \cdots & \widetilde{G}_{1n,n} \\ \vdots & \ddots & \vdots \\ \widetilde{G}_{n1,n} & \cdots & \widetilde{G}_{nn,n} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} g_{11,n} & 0 \\ 0 & 1 \end{pmatrix} & \cdots & \begin{pmatrix} g_{1n,n} & 0 \\ 0 & 0 \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \begin{pmatrix} g_{n1,n} & 0 \\ 0 & 0 \end{pmatrix} & \cdots & \begin{pmatrix} g_{nn,n} & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix}$$

Notice that by switching rows and columns with the same procedure, $\widetilde{G}_n$ can be transferred to $\begin{pmatrix} G_n & 0 \\ 0 & I_n \end{pmatrix}$. Thus, we can write $\widetilde{G}_n$ as:

$$\widetilde{G}_n = T_{k_n} \cdots T_2 T_1 \begin{pmatrix} G_n & 0 \\ 0 & I_n \end{pmatrix} T_1 T_2 \cdots T_{k_n} = K_n^{'} \begin{pmatrix} G_n & 0 \\ 0 & I_n \end{pmatrix} K_n$$

where $K_n \equiv T_1 \cdots T_{k_n}$, and for any $j \in \{1, \cdots, k_n\}$, $T_j = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$ which equals to I_n but with $(T_j)_{mn} = (T_j)_{nm} = 1$ and $(T_j)_{nn} = (T_j)_{mm} = 0$ for a pair of $m \neq n$. Easy to check, T_j has some important properties: $|T_j| = -1$, $T_j^{'} = T_j^{-1} = T_j$. Thus, K_n matrix has the following properties: $K_n^{'} = K_n^{-1} = T_k \cdots T_1$, $|K_n| = (-1)^k$. Thus, we can easily get the following property of $\widetilde{G}_n$ with assuming G_n is invertible:

$$\begin{aligned} \widetilde{G}_n^{-1} &= \left[K_n^{'} \begin{pmatrix} G_n & 0 \\ 0 & I_n \end{pmatrix} K_n \right]^{-1} \\ &= K_n^{\prime} \begin{pmatrix} G_n^{-1} & 0 \\ 0 & I_n \end{pmatrix} K_n \\ &= \begin{pmatrix} \begin{pmatrix} (G_n^{-1})_{11,n} & 0 \\ 0 & 1 \end{pmatrix} & \cdots & \begin{pmatrix} (G_n^{-1})_{1n,n} & 0 \\ 0 & 0 \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \begin{pmatrix} (G_n^{-1})_{n1,n} & 0 \\ 0 & 0 \end{pmatrix} & \cdots & \begin{pmatrix} (G_n^{-1})_{nn,n} & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix} \end{aligned}$$

Here are some additional notations. For any finite dimensional vector x, $|x| \equiv (x^{'}x)^{1/2}$ is the Euclidian norm, $\|x\|_1 \equiv \sum_i |x_i|$, and $\|x\|_\infty \equiv \max_i |x_i|$. For any finite dimensional $n \times n$ matrix $A \equiv (a_{ij})$, $\|A\|_\infty \equiv \sup_{i=1,\cdots,n} \sum_{j=1}^n |a_{ij}|$ is the row sum norm, and $\|A\|_1 \equiv \sup_{i=1,\cdots,n} \sum_{i=1}^n |a_{ij}|$ is the column sum norm. For any $Q \subset D_n$ and any $s > 0$, denote $Q^s = \{i \in D_n : d(i, Q) < s\}$, which is the set of individuals located within distance s from the set Q, and $\mathcal{F}_{Q,n} = \sigma(\{e_{j,n} : j \in Q\})$ is the σ-field generated by disturbance $e_{j,n}$ for individuals in Q.

Now, we need to put some additional assumptions on G_n and W_n:

Assumption 3:

For any two regions i, j, we have the following properties for their correlation $w_{ij,n}$:

(1) for every i, we have $w_{ii,n} = 0$;

(2) when $d_{ij} > \bar{d}_0$, $w_{ij,n} = 0$, where $\bar{d}_0$ is a constant greater than 1.

This is the setting which does not allow direct correlations between individuals far away form each other.

Assumption 4:

$G_n \equiv [I_n - (\phi+\theta)W_n]^{-1}(I_n - \theta W_n)$ has the following properties for given real number θ and ϕ:

(1) G_n is invertible, which requires $I_n - (\phi+\theta)W_n$ and $I_n - \theta W_n$ both invertible;

(2) $\xi = \|(\phi+\theta)W_n\|_\infty < 1$;

(3) $\zeta = \|\theta W_n\|_\infty < 1$.

With Assumption 3 and Assumption 4, we can decompose G_n as the following when $\phi+\theta \neq 0$:

$$G_n = \sum_{l=0}^{\infty}[(\phi+\theta)W_n]^l(I_n - \theta W_n) = \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty}[(\phi+\theta)W_n]^l - \frac{\theta}{\phi+\theta}\sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty}[(\phi+\theta)W_n]^{l+1}$$

where $\lfloor x \rfloor$ is the largest integer that is less than or equal to x.

Then, we can get

$$\begin{aligned}
|g_{ij,n}| &= \left| \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty}[(\phi+\theta)W_n]^l_{ij} - \frac{\theta}{\phi+\theta}\sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty}[(\phi+\theta)W_n]^{l+1}_{ij} \right| \\
&\leqslant \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty}\left[\|(\phi+\theta)W_n\|^l_\infty + \frac{|\theta|}{|\theta+\phi|}\|(\phi+\theta)W_n\|^{l+1}_\infty\right] \\
&\leqslant \left(1 + \frac{|\theta|}{|\theta+\phi|}\xi\right)\frac{\xi^{\lfloor d_{ij}/\bar{d}_0 \rfloor}}{1-\xi}
\end{aligned}$$

From Lemma A.1 in Jenish and Prucha (2009), for any $l \geqslant 0$ and for any $i \in D_n$, $\left|\{j \in D_n : l\bar{d}_0 \leqslant d_{ij} \leqslant (l+1)\bar{d}_0\}\right| \leqslant C_1(l+1)^{d-1}$ for some constant $C_1 > 0$. The index d is the dimension of the space D_n. If we consider time series model as a special case, $d = 1$; for spatial case, $d = 2$. Thus, we have the following result:

$$\begin{aligned}
\sup_n \|G_n\|_1 &= \sup_{j \in D_n, n}\sum_{i=1}^{n}|g_{ij,n}| \\
&= \sup_{j,n}\sum_{l=0}^{\infty}\sum_{i \in D_n : l\bar{d}_0 < d_{ij} < (1+l)\bar{d}_0}|g_{ij,n}| \\
&\leqslant \sum_{l=0}^{\infty}C_1(l+1)^{d-1}(1 + \frac{|\theta|}{|\theta+\phi|}\xi)\frac{\xi^l}{1-\xi} < \infty
\end{aligned}$$

As $G_n^{-1} = (I_n - \theta W_n)^{-1}[I_n - (\phi+\theta)W_n]$, we can get $\sup_n \left\|G_n^{-1}\right\|_1 < \infty$ following the similar procedure above by expending $(I_n - \theta W_n)^{-1}$ as we have the assumption $\zeta = \|\theta W_n\| < 1$. We can also get $\lim_{s\to\infty}\sup_n \sup_{j \in Q}[\sum_{r \in D_n \setminus Q^s}\left|(G_n^{-1})_{r,j}\right|] = 0$ by Assumption 3 for the spatial matrix setting. When $\theta+\phi = 0$, $G_n = I_n - \theta W_n$ which also satisfies the above results. Then, we have the following statement for $\widetilde{G}_n$ based on the properties of G_n:

Lemma 1:

The matrix $\widetilde{G}_n$ satisfies the Assumption 2.2 in Xu and Lee (2019):

(1) $\widetilde{G}_n$ is invertible;

(2) $\sup_n \left\|\widetilde{G}_n\right\|_1 < \infty$, and $\sup_n \left\|\widetilde{G}_n^{-1}\right\|_1 < \infty$;

(3) $\lim_{s\to\infty}\sup_n \sup_{j\in Q;q=1,2}[\sum_{r\in D_n\setminus Q^s;t=1,2}\left|(\widetilde{G}_n^{-1})_{tq,rj}\right|] = 0$ where $(\widetilde{G}_n^{-1})_{tq,rj}$ is the $(t,q)^{th}$ element of $(r,j)^{th}$ block of 2×2 sub-matrix of $\widetilde{G}_n^{-1}$.

Proof:

(1) As $\widetilde{G}_n = T_{k_n}\cdots T_2T_1\begin{pmatrix} G_n & 0 \\ 0 & I_n\end{pmatrix}T_1T_2\cdots T_{k_n} = K_n^{'}\begin{pmatrix} G_n & 0 \\ 0 & I_n\end{pmatrix}K_n$, since G_n and K_n are invertible, $\widetilde{G}_n$ is invertible since $\left|\widetilde{G}_n\right| = |G_n| \neq 0$;

(2) From the form of $\widetilde{G}_n$ and $\widetilde{G}_n^{-1}$, we can get that

$$\sup_n \left\|\widetilde{G}_n\right\|_1 = \sup_n(\|G_n\|_1, 1) = \max(1, \sup_n \|G_n\|_1) < \infty$$

$$\sup_n \left\|\widetilde{G}_n^{-1}\right\|_1 = \sup_n(\left\|G_n^{-1}\right\|_1, 1) = \max(1, \sup_n \left\|G_n^{-1}\right\|_1) < \infty$$

(3) From the form of $\widetilde{G}_n$, we have

$$\begin{aligned}&\lim_{s\to\infty}\sup_n \sup_{j\in Q;q=1,2}[\sum_{r\in D_n\setminus Q^s;t=1,2}\left|(\widetilde{G}_n^{-1})_{tq,rj}\right|] \\ &= \lim_{s\to\infty}\sup_n \sup_{j\in Q}[\sum_{r\in D_n\setminus Q^s}\left|(G_n^{-1})_{r,j}\right| = 0 \quad \square\end{aligned}$$

Denote $A_s \equiv \sup_{i\in D_n,n}\sum_{k=1}^2\sum_{q=1}^2\sum_{j\in D_n:d_{ij}\geqslant s}\widetilde{g}_{kq,ij,n}^2$, where $\widetilde{g}_{pq,ij,n}$ comes from a new notation $\widetilde{G}_{ij,n} \equiv (\widetilde{g}_{pq,ij,n})_{p,q=1}^2$. Thus, we can get the following result:

$$\begin{aligned}A_s &= \sup_{i\in D_n,n}\sum_{j\in D_n:d_{ij}\geqslant s}\widetilde{g}_{ij,n}^2 \\ &\leqslant \sup_{i\in D_n,n}\sum_{l=\lfloor s/\bar{d}_0\rfloor}\sum_{j\in D_n:l\leqslant d_{ij}/\bar{d}_0<l+1}\widetilde{g}_{ij,n}^2 \\ &\leqslant \sum_{l=\lfloor s/\bar{d}_0\rfloor} C_1(l+1)^d(1+\frac{|\theta|}{|\theta+\phi|}\xi)^2\frac{\xi^{2l}}{(1-\xi)^2} \\ &\leqslant \frac{C_1(1+\frac{|\theta|}{|\theta+\phi|}\xi)^2}{(1-\xi)^2}\int_{\lfloor s/\bar{d}_0\rfloor}^{\infty} x^{d+1}\xi^{2(x-2)}dx \\ &= O(s^{d-1}\xi^{2s/\bar{d}_0})\end{aligned}$$

where the last inequality is coming from the following limit: $\lim_{a\to\infty}\frac{\int_a^\infty x^{d-1}\xi^x dx}{a^{d-1}\xi^a} = -\frac{1}{ln\xi}$ by L'Hospital's rule.

Thus, with all the assumptions and results above, we can apply Theorem 1(2) in Xu and Lee (2019), the α-mixing coefficients of the random field $\{\tilde{z}_{i,n}\}_{i\in D_n} \equiv \{z_{i,n} - (c_{i,n}, 0)^{'}\}_{i\in D_n}$ satisfy:

$$\alpha_{k,l}(r) \leqslant C_2 min(k,l) A_{r/2}^{1/2} \leqslant C_3 min(k,l)(r^{d-1}\xi^{r/\bar{d}_0})^{1/3}$$

for any positive integers k, l and any $r \geqslant 2s_0$. C_2 and C_3 are some large enough positive constants. As $\lim_{r\to\infty} r^a \xi^{br} = 0$ for any positive a and b since $\xi < 1$, we have $\lim_{r\to\infty} \alpha_{k,l}(r) = 0$. Thus, $\{\tilde{z}_{i,n}\}_{i\in D_n}$ is α-mixing when we do not direct correlations between individuals far away.

Theorem 1: *Under Assumption 1, 2, 3 and 4,* $\{u_{i,n}\}_{i\in D_n}$ *is* $\alpha-$*mixing.*

Proof:

With the previous results, the last thing need to prove is that $\{\psi_{i,n}\}_{i\in D_n} \equiv \{\begin{pmatrix} \widetilde{z}_{i,n} \\ (c_{i,n}, 0)^{'} \end{pmatrix}\}_{i\in D_n}$ is α-mixing, where for $\forall n$, $c_{i,n}$'s are fixed constants given θ, ϕ and W_n. Thus, for any event $A, B \in \sigma(\{(x_{i,n}, y_{i,n})_{i\in D_n}, x, y_{i,n} \in \mathbb{R}\})$ and n,

$$\Pr(\widetilde{z}_{i,n} \in A, (c_{i,n},0)^{'} \in B) = \begin{cases} 0 & if \{(c_{i,n},0)^{'} = ((\alpha[I_n - (\phi+\theta)W_n]^{-1}1_n)_i, 0)^{'}, \forall i\} \notin B \\ \Pr(\widetilde{z}_{i,n} \in A) & else \end{cases}$$

$$= \Pr(\widetilde{z}_{i,n} \in A)\Pr((c_{i,n},0)^{'} \in B)$$

since $\Pr((c_{i,n},0)^{'} \in B) = \begin{cases} 0 & if \{(c_{i,n},0)^{'} = ((\alpha[I_n - (\phi+\theta)W_n]^{-1}1_n)_i, 0)^{'}, \forall i\} \notin B \\ 1 & else \end{cases}$. Thus, $\{\tilde{z}_{i,n}\}_{i\in D_n}$ and $\{(c_{i,n},0)\}_{i\in D_n}^{'}$ are independent. Thus for any event $A_1 \otimes B_1, A_2 \otimes B_2 \in \sigma(\{\psi_{i,n}\}_{i\in D_n}) \equiv \sigma(\{\tilde{z}_{i,n}\}_{i\in D_n}) \otimes \sigma(\{(c_{i,n},0)\}_{i\in D_n}^{'})$ and n, if event $\{(c_{i,n},0)^{'} = ((\alpha[I_n - (\phi+\theta)W_n]^{-1}1_n)_i, 0)^{'}, \forall i\}$ contains in C and D, then we have

$$\begin{aligned} &\Pr((A_1 \otimes B_1) \cap (A_2 \otimes B_2)) - \Pr(A_1 \otimes B_1)\Pr(A_2 \otimes B_2) \\ = &\begin{cases} 0 & if \{(c_{i,n},0)^{'} = ((\alpha[I_n - (\phi+\theta)W_n]^{-1}1_n)_i, 0)^{'}, \forall i\} \notin B_1 \cap B_2 \\ \Pr(A_1 \cap A_1) - \Pr(A_1)\Pr(A_2) & else \end{cases} \end{aligned}$$

Thus, for any sub σ-field $\mathcal{A}, \mathcal{B} \subset \sigma(\{\psi_{i,n}\}_{i\in D_n})$, we have

$$\begin{aligned} \alpha(\mathcal{A},\mathcal{B}) &\equiv \sup(|\Pr((A_1 \otimes B_1) \cap (A_2 \otimes B_2)) - \Pr(A_1 \otimes B_1)\Pr(A_2 \otimes B_2)| : A_1 \otimes B_1 \in \mathcal{A}, A_2 \otimes B_2 \in \mathcal{B}) \\ &\leqslant \sup(|\Pr(A_1 \cap A_2) - \Pr(A_1 \cap A_2)|, A_1, A_2 \in \sigma(\{\tilde{z}_{i,n}\}_{i\in D_n})) \end{aligned}$$

Since $\{\tilde{z}_{i,n}\}_{i\in D_n}$ is α-mixing, $\{\begin{pmatrix} \widetilde{z}_{i,n} \\ (c_{i,n}, 0)^{'} \end{pmatrix}\}_{i\in D_n}$ is α-mixing with the α-mixing coefficient $\alpha_{k,l}^{'}(r) \leqslant \alpha_{k,l}(r)$, for any positive integer k, l and any $r \geqslant 1$. As the $\alpha-$mixing property is preserved by measurable transformations, we have $u_{i,n} = \exp\left(logu_{i,n}^2\right) sgn\left(\varepsilon_{i,n}\right)$ is mixing. $\square$

Thus, we have $\{z_{i,n}\}_{i\in D_n} \equiv \{(logu_{i,n}^2, \varepsilon_{i,n})^{'}\}_{i\in D_n}$ and $\{u_{i,n}\}_{i\in D_n}$ are α-mixing under measurable function claims before. Further more, from Jenish and Prucha (2009), we can get that the α-mixing coefficients for $\{z_{i,n}\}_{i\in D_n}$ and $\{u_{i,n}\}_{i\in D_n}$ are also less than $\alpha_{k,l}(r)$, thus bounded from above by $C_3 min(k,l)(r^{d-1}\xi^{r/\bar{d}_0})^{1/3}$ or $C_4 \min(k,l) r^{(d-2\alpha)/3}$ under Assumption 3 or Assumption 3'. The $\alpha-$mixing property and the upper-bound of mixing coefficients will be essential to build up asymptotic property for potential estimators.

3 Maximum-Likelihood Estimation

3.1 Preliminary Results and Regularity Conditions

To write down the likelihood function of $\{u_{i,n}\}_{i\in D_n}$, we need to the following result:

Theorem 2:

The projection between $\varepsilon_n = (\varepsilon_{1,n}, \cdots, \varepsilon_{n,n})'$ and $u_n = (u_{1,n}, \cdots, u_{n,n})'$ given by the spatial GARCH-like model is a bijection.

Proof:

First, we can write $h_{i,n}$ as functions of $\{\varepsilon_{i,n}\}_{i\in D_n}$ or $\{u_{i,n}\}_{i\in D_n}$, and then we can represent ε_n and u_n by each other.

1. Direction $u \to \varepsilon$:

The second equation has a matrix form:

$$logh_n = \phi(I_n - \theta W_n)^{-1} W_n logu_n^2 + \alpha(I_n - \theta W_n)^{-1} 1_n$$

From this form, we can directly write $h_{i,n}$ as function of $u_{i,n}$'s:

$$h_{i,n} = \exp\{(\alpha \sum_{j=1}^{n} ((I_n - \theta W_n)^{-1})_{ij}\} \prod_{j=1}^{n} u_{j,n}^{2(\phi(I_n - \theta W_n) W_n)_{ij}}$$

Thus, since $\varepsilon_{i,n} = h_{i,n}^{-\frac{1}{2}} u_{i,n}$, we can get

$$\begin{aligned} \varepsilon_{i,n} &= \exp\left\{-\frac{1}{2}\alpha \sum_{j=1}^{n} ((I_n - \theta W_n)^{-1})_{ij}\right\} \prod_{j=1}^{n} |u_{j,n}|^{-(\phi(I_n - \theta W_n)^{-1} W_n)_{ij}} u_{i,n} \\ &\equiv \exp\{A_i(\alpha, \theta)\} \prod_{j=1}^{n} |u_{j,n}|^{B_{ij}(\theta, \phi)} u_{i,n} \end{aligned} \tag{3}$$

where A_i is the sum of ith row elements of $A_n \equiv -\frac{1}{2}\alpha(I_n - \theta W_n)^{-1}$, and B_{ij} is the (i, j) th element of $B_n \equiv -\phi(I - \theta W_n)^{-1} W_n$.

2. Direction $\varepsilon \to u$:

Since $logu_{i,n}^2 = logh_{i,n} + log\varepsilon_{i,n}^2$, after replacing $logu_{j,n}^2$ in the second equation, we can get

$$logh_n = \phi[I_n - (\phi + \theta)W_n]^{-1} W_n log\varepsilon_n^2 + \alpha[I_n - (\phi + \theta)W_n]^{-1} 1_n$$

From this form, we can directly write $h_{i,n}$ as function of $\varepsilon_{i,n}$'s:

$$h_{i,n} = \exp\{\alpha \sum_{j=1}^{n} ([I_n - (\phi + \theta)W_n]^{-1})_{ij}\} \prod_{j=1}^{n} \varepsilon_{j,n}^{2(\phi[I_n - (\phi+\theta)W_n]^{-1} W_n)_{ij}}$$

Thus, replacing $h_{i,n}$ in the first equation, we can get

$$\begin{aligned} u_{i,n} &= \exp\{\frac{1}{2} \sum_{j=1}^{n} ([I_n - (\phi + \theta)W_n]^{-1})_{ij}\} \prod_{j=1}^{n} |\varepsilon_{j,n}|^{(\phi[I_n - (\phi+\theta)W_n]^{-1} W_n)_{ij}} \varepsilon_{i,n} \\ &\equiv \exp\{C_i(\alpha, \phi, \theta)\} \prod_{j=1}^{n} |\varepsilon_{j,n}|^{D_{ij}} \varepsilon_{i,n} \end{aligned} \tag{4}$$

where C_i is the sum of ith row elements of $C_n \equiv \frac{1}{2}\alpha[I_n - (\phi+\theta)W_n]^{-1}$, and D_{ij} is the (i,j) th element of $D_n \equiv \phi[I_n - (\phi+\theta)W_n]^{-1}W_n$.

By showing these two directions, we can see the mapping between u_n and ε_n is a bijection. $\square$

From the results above, we can evaluate the mean of $u_{i,n}$:

$$\begin{aligned} E(u_{i,n}) &= E(\exp\{C_i(\alpha,\phi,\theta)\} \prod_{j=1}^{n} |\varepsilon_{j,n}|^{D_{ij}} \varepsilon_{i,n}) \\ &= \exp\{C_i(\alpha,\phi,\theta)\} \prod_{j=1, j\neq i}^{n} E(|\varepsilon_{j,n}|^{D_{ij}}) E\left(|\varepsilon_{i,n}|^{D_{ii}} \varepsilon_{i,n}\right) \end{aligned}$$

Since we assume $\varepsilon_{i,n} \sim (0,1)$ i.i.d, if the density function of $\varepsilon_{i,n}$ is symmetric by $y-$axis,

$$E(|\varepsilon_{i,n}|^{D_{ii}} \varepsilon_{i,n}) = \int_{-\infty}^{\infty} |\varepsilon_{i,n}|^{D_{ii}} \varepsilon_{i,n} dF(\varepsilon) = 0$$

when $\int_0^\infty \varepsilon_{j,n}^{D_{ii}+1} dF(\varepsilon) < \infty$ since the function inside the integral is an odd function. Then, if $E(|\varepsilon_{j,n}|^{D_{ij}}) = \int_0^\infty \varepsilon_{j,n}^{D_{ij}} < \infty$ for $\forall j$, then we can get $E(u_{i,n}) = 0$.

Look back the matrix D_n, based on previous assumptions, we can decompose D_n as

$$D_n = \phi[I_n - (\phi+\theta)W_n]^{-1}W_n = \phi \sum_{k=1}^{\infty} (\phi+\theta)^k W_n^{k+1}$$

For each element D_{ij}, we have $D_{ij} = \phi \sum_{k=1}^{\infty} (\phi+\theta)^k (W_n^{k+1})_{ij}$. For a typical spatial matrix, since each element is non-negative, each element of W_n^{k+1} is non-negative. When ϕ and $\phi+\theta$ are non-negative, we have $D_{ij} \geqslant 0$. For $\varepsilon_{i,n}$ with symmetric density function or $\varepsilon_{i,n} \sim N(0,1)$ i.i.d, since $\int_0^\infty \varepsilon^p dF(\varepsilon) < \infty$ for any positive p, we have $E(\varepsilon_{i,n}) = 0$ for $\forall i$. In this case, with sufficiently higher order moments exist, we can also insure finite variance:

$$Var(u_{i,n}) = E(u_{i,n}^2) = \exp\{2C_i(\alpha,\phi,\theta)\} \prod_{j=1, j\neq i}^{n} E(\varepsilon_{j,n}^{2D_{ij}}) E\left(\varepsilon_{i,n}^{2+2D_{ii}}\right) < \infty$$

However, for other values of ϕ and θ, things become much more complicated since $\int_0^\infty \varepsilon^p dF(\varepsilon) < \infty$ does not always hold when $p < 0$. Take standard normal distribution as an example,

$$\begin{aligned} \int_0^\infty x^p \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} dx &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \sqrt{2}(\sqrt{2}y)^p e^{-y^2} dy \\ &= \frac{1}{\sqrt{2\pi}} 2^{\frac{1+p}{2}} \int_0^\infty y^{p-1} e^{-y^2} dy \\ &= (2\pi)^{-\frac{1}{2}} 2^{-p} \Gamma(\frac{p+1}{2}) \end{aligned}$$

Since the Gamma function is undefined at non-positive integers, we need to avoid the set $p \in \{-1,-2,-3,\cdots\}$ to make sure $\int_0^\infty \varepsilon^p dF(\varepsilon) < \infty$. Similar to the analysis before for G_n, we have

$$
\begin{aligned}
|D_{ij}| &= \frac{\phi}{\phi+\theta} \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty} [(\phi+\theta)W_n]^{l+1} \\
&\leqslant \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty} \frac{|\phi|}{|\theta+\phi|} \left\|(\phi+\theta)W_n\right\|_{\infty}^{l+1} \\
&\leqslant \sup_{(\theta,\phi)\in\Theta} \frac{|\phi|}{|\theta+\phi|} \frac{\xi^{\lfloor d_{ij}/\bar{d}_0 \rfloor}}{1-\xi}
\end{aligned}
$$

when $\phi+\theta \neq 0$ and $|D_{ij}| = |\phi w_{ij,n}|$ when $\phi+\theta = 0$.

The only way to make sure every D_{ij} does not meet negative integer is to make sure $|D_{ij}| < 1$. For most commonly used spatial weighting matrix, row-normalized W_n, $\xi = |\theta+\phi| < 1$. Easy to see $|D_{ij}| < 1$ is satisfied when $\phi+\theta = 0$. When $\phi+\theta \neq 0$, we have

$$
|D_{ij}| \leqslant \frac{|\phi|}{|\theta+\phi|} \frac{|\theta+\phi|}{1-|\theta+\phi|} < 1 \Longrightarrow |\theta+\phi| + |\phi| < 1
$$

But this condition can not make sure $Var(u_{i,n}) < \infty$. Even if the expectation equals to 0, we need to make sure $\int_0^{\infty} \varepsilon_{i,n}^{2D_{ij}} dF(\varepsilon) < \infty$, which can be make sure when $|\theta+\phi| + |\phi| < \frac{1}{2}$.

When we have $\varepsilon_{i,n} \sim Uniform(\sqrt{3}, \sqrt{3})$ i.i.d with row normalized spatial matrix, this condition can also make $E(u_{i,n}) = 0$ and $Var(u_{i,n}) < \infty$ since $\int_0^{\sqrt{3}} x^p dx < \infty$ iff $p > -1$. But this condition can only make sure second order moment exist, for higher order moments, the parameter space will be more restrictive to make sure the existence. This will potentially make asymptotic theory fail. Also, negative $\theta+\phi$ is somewhat counter-intuitive. When $\theta+\phi < 0$, it means the externality one area to another at volatility level is negative in general, but no matter we have positive or negative externality at mean level, it is natural to have a positive impact on volatility to adjunct area. Thus, in the following part of this paper, we will only consider the case when ϕ and $\theta+\phi$ are non-negative.

Combining the arguments above with previous assumptions for α-mixing property, it is convenient to limit our spatial weighting matrix in the row-normalized and matrix class, so that we can have a clear parameter space for θ and ϕ. Thus, we put the following assumption:

Assumption 5:

(1) W_n is row-normalized matrix, i.e. $\sum_{j=1}^{n} w_{ij,n} = 1$ for $\forall i = 1, \cdots, n$;

(2) The induced parameter space for θ and ϕ is $\Theta = \{(\alpha,\theta,\phi) : \alpha \in \mathbb{R}, -1 < \theta < 1, 0 \leqslant \phi < 1, 0 \leqslant \phi+\theta < 1\}$;

(3) $\varepsilon_{i,n} \overset{iid}{\sim} (0,1)$ with $E\left(\varepsilon_{i,n}^{k}\right) < \infty$ for $\forall k > 0$.

This condition can make sure the Assumption 2 holds, since $E\log\varepsilon_{i,n}^{2k} \leqslant \log E\varepsilon_{i,n}^{2k} < \infty$ for $\forall k > 0$ by Jenssen's Inequality.

A special case is when $\varepsilon_{i,n} \overset{iid}{\sim} N(0,1)$, we can get the following result:

$$
\begin{aligned}
\frac{E(h_{i,n})}{Var(u_{i,n})} &= \frac{\exp\{2C_i(\alpha,\phi,\theta)\}\prod_{j=1}^{n} E(\varepsilon_{i,n}^{2D_{ij}})}{\exp\{2C_i(\alpha,\phi,\theta)\}\prod_{j=1,j\neq i}^{n} E(\varepsilon_{i,n}^{2D_{ij}})E\left(\varepsilon_{i,n}^{2+2D_{ii}}\right)} \\
&= \frac{E(\varepsilon_{i,n}^{2D_{ii}})}{E(\varepsilon_{i,n}^{2D_{ii}+2})} \\
&= \frac{(2\pi)^{-\frac{1}{2}}2^{-\frac{1-2D_{ii}}{2}}\Gamma(\frac{1+2D_{ii}}{2})}{(2\pi)^{-\frac{1}{2}}2^{-\frac{-2D_{ii}-1}{2}\Gamma(\frac{3+2D_{ii}}{2})}} \\
&= \frac{\Gamma(\frac{1}{2}+D_{ii})}{2\Gamma(\frac{3}{2}+D_{ii})} \\
&= \frac{1}{2D_{ii}+1}
\end{aligned}
$$

Thus, we have

$$
Var(u_{i,n}) = (2D_{ii}+1)E(h_{i,n}) = [2(\phi[I_n-(\phi+\theta)W_n]^{-1}W_n)_{ii}+1]E(h_{i,n}) \tag{5}
$$

From this equation, we can see, the variance of $u_{i,n}$ is a linear function of the expectation of $h_{i,n}$. Once we get consistent estimator of α, ϕ and θ, since $h_{i,n}$ is a continuous function of the parameters, we get consistent estimator of $h_{i,n}$. Then we can get consistent estimator of $Var(u_{i,n})$. For general $\varepsilon_{i,n}$, since $\frac{E(h_{i,n})}{Var(u_{i,n})} = \frac{E(\varepsilon_{i,n}^{2D_{ii}})}{E(\varepsilon_{i,n}^{2D_{ii}+2})}$ always holds, we still have the linear relationship between $E(h_{i,n})$ and $Var(u_{i,n})$.

3.2 Likelihood Function and Estimation Procedure

From the equation (3), we can easily see that $\varepsilon_{i,n}$ is an odd function of $u_{i,n}$ and an even function of $\varepsilon_{i,n}$ for $\forall j \neq i$. Notice that $\varepsilon_{i,n}(u_n)$ is differentiable when $u_{j,n} \neq 0$ for $\forall j \in D_n$, $\frac{\partial \varepsilon_{i,n}}{\partial u_{j,n}}$ exists and has very good property: when $j = i$, it is an even function of $u_{i,n}$; when $j \neq i$, it is an even function of $u_{j,n}$ followed by properties of derivatives. Thus, we can consider the case when $\varepsilon_{j,n} > 0$, and then extend the result to the whole domain. For the function $f(x) = |x|^k$, the derivative is $f^{'}(x) = kx^{k-1}$ when $x > 0$. Due to $f(x)$ is even, $f^{'}(x)$ is odd, then we can get that $f^{'}(x) = sgn(x)kx^{k-1} = k\frac{|x|^k}{x}$ whenever $x \neq 0$. Similarly, for the function $g(x) = x\,|x|^k$, we have $g^{'}(x) = |x|^k$. Thus, we can get the following first order derivatives:

$$
\frac{\partial \varepsilon_{i,n}}{\partial u_{i,n}} = [1+B_{ii}(\theta,\phi)]\exp\{A_i(\alpha,\theta)\}\prod_{j=1}^{n}|u_{j,n}|^{B_{ij}(\theta,\phi)} = [1+B_{ii}(\theta,\phi)]\frac{\varepsilon_{i,n}}{u_{i,n}}
$$

$$
\frac{\partial \varepsilon_{i,n}}{\partial u_{k,n}} = B_{i,k}(\theta,\phi)\prod_{j=1}^{n}|u_{j,n}|^{B_{ij}(\theta,\phi)}\frac{1}{u_{k,n}} = \frac{B_{i,k}(\theta,\phi)\varepsilon_{i,n}}{u_{k,n}}, \forall k \neq n
$$

Thus, the Jacobean matrix from u to ε is

$$J_{u\to\varepsilon} = \begin{bmatrix} (1+B_{11})\frac{\varepsilon_{1,n}}{u_{1,n}} & \frac{B_{12}\varepsilon_{2,n}}{u_{1,n}} & \cdots & \frac{B_{1n}\varepsilon_{n,n}}{u_{i,n}} \\ \frac{B_{21}\varepsilon_{1,n}}{u_{2,n}} & (1+B_{22})\frac{\varepsilon_{2,n}}{u_{2,n}} & \cdots & \frac{B_{2n}\varepsilon_{2,n}}{u_{1,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{B_{n1}\varepsilon_{1,n}}{u_{n,n}} & \frac{B_{n2}\varepsilon_{2,n}}{u_{n,n}} & \cdots & (1+B_{nn})\frac{\varepsilon_{1,n}}{u_{n,n}} \end{bmatrix}$$

The determinant of $J_{u\to\varepsilon}$ is

$$\begin{aligned} \det(J_{u\to\varepsilon}) &= \prod_{i=1}^{n} \frac{\varepsilon_{i,n}}{u_{i,n}} \begin{vmatrix} 1+B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & 1+B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n1} & \cdots & 1+B_{nn} \end{vmatrix} \\ &= \prod_{i=1}^{n} h_{i,n}^{-\frac{1}{2}} \left| I_n - \phi(I_n-\theta W_n)^{-1} W_n \right| \end{aligned}$$

Notice that

$$\begin{aligned} \prod_{i=1}^{n} h_{i,n}^{-\frac{1}{2}} &= \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n} logh_{i,n} \right\} \\ &= \exp\left\{ -\frac{1}{2}[1_n^{'}\phi(I_n-\theta W_n)^{-1}W_n logu_n^2 + \alpha 1_n^{'}(I_n-\theta W_n)^{-1}1_n] \right\} \end{aligned}$$

Thus, the determinant of Jacobean matrix can be written as

$$\det(J_{u\to\varepsilon}) = \exp\left\{ -\frac{1}{2}[1_n^{'}\phi(I_n-\theta W_n)^{-1}W_n logu_n^2 + \alpha 1_n^{'}(I_n-\theta W_n)^{-1}1_n] \right\} \left| I_n - \phi(I_n-\theta W_n)^{-1} W_n \right|$$

Thus, by applying the density function of $N(0, I_n)$, the log-likelihood function of u_n is

$$\begin{aligned} \ln L_n(u_n;\alpha,\phi,\theta) &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\varepsilon_{i,n}^2(u_n) + \ln\left|\det(J_{u\to\varepsilon})\right| \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\exp\left\{ -\alpha\sum_{j=1}^{n}((I_n-\theta W_n)^{-1})_{ij} \right\} \prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}\, u_{i,n}^2 \\ &\quad - \frac{1}{2}[1_n^{'}\phi(I_n-\theta W_n)^{-1}W_n logu_n^2 + \alpha 1_n^{'}(I_n-\theta W_n)^{-1}1_n] + \ln\left| I_n - \phi(I_n-\theta W_n)^{-1} W_n \right| \end{aligned}$$

This log-likelihood functions seems very complicated, but by assuming W_n is a row-normalized matrix, i.e. $W_n 1_n = 1_n$. Thus for $\forall k \in \mathbb{N}$, we have $W_n^k 1_n = 1_n$. By this property, we can get

$$\sum_{j=1}^{n}((I_n-\theta W_n)^{-1})_{ij} = \sum_{j=1}^{n}(\sum_{k=0}^{\infty}(\theta W_n)^k)_{ij} = \sum_{k=0}^{\infty}\theta^k\sum_{j=1}^{n}(W_n^k)_{ij} = \sum_{k=0}^{\infty}\theta^k = \frac{1}{1-\theta}$$

and

$$1_n^{'}(I_n-\theta W_n)^{-1}1_n = 1_n^{'}\sum_{k=0}^{\infty}\theta^k W_n^k 1_n = n\sum_{k=0}^{\infty}\theta^k = \frac{n}{1-\theta}$$

Also, we have $\ln\left|I_n-\phi(I_n-\theta W_n)^{-1}W_n\right| = \ln|I_n-(\theta+\phi)W_n| - \ln|I_n-\theta W_n|$, then we can simplify the log-likelihood function and get

$$\begin{aligned}\ln L_n(u_n;\alpha,\phi,\theta) = &-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\exp\left\{-\frac{\alpha}{1-\theta}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}\ u_{i,n}^2 \\ &-\frac{1}{2}[1_n^{'}\phi(I_n-\theta W_n)^{-1}W_n log u_n^2 + \frac{n\alpha}{1-\theta}] + \ln|I_n-(\theta+\phi)W_n| - \ln|I_n-\theta W_n|\end{aligned} \tag{6}$$

To get the MLE, we need to maximize this log-likelihood function. But there is a computational issue: αcan take any value in $\mathbb{R}$, which will increase the computation burden. To make it easier, we can concentrate out α. Here is the first order condition for α:

$$\frac{\partial \ln L_n(u_n;\alpha,\phi,\theta)}{\partial\alpha} = \frac{1}{2(1-\theta)}\exp\left\{-\frac{\alpha}{1-\theta}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}\ u_{i,n}^2 - \frac{n}{2(1-\theta)} = 0 \tag{7}$$

From this FOC, we can write down the maximum likelihood estimator of α by the other estimators:

$$\hat{\alpha} = (1-\theta)\ln[\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}\ u_{i,n}^2] \tag{8}$$

Thus, by concentrating out α, we can get the following concentrated log-likelihood function with only ϕ and θ remained:

$$\begin{aligned}g_n(u_n;\phi,\theta) = &-\frac{1}{2}\{1_n^{'}\phi(I_n-\theta W_n)^{-1}W_n log u_n^2 + n\ln[\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}\ u_{i,n}^2]\} \\ &+\ln|I_n-(\theta+\phi)W_n| - \ln|I_n-\theta W_n| - \frac{n}{2}[\ln(2\pi)+1]\end{aligned} \tag{9}$$

This equation is much easier to be maximized. From Assumption 5, ϕ and θ are bounded. Thus, using local optimization method to find the maximum for (9) will much efficient than doing global search for (6).

3.3 Identification

Theorem 3:

Under Assumption 3, 4 and 5, when $\phi_0 \neq 0$, the spatial GARCH-like model can be identified.

Proof:

Let $\psi_0 = (\alpha_0,\phi_0,\theta_0)^{'}$ be the true parameter, and $\psi = (\alpha,\phi,\theta)^{'}$ be an arbitrary value of parameter in Θ defined in Assumption 5. Since $\ln x \leqslant x-1$ for any $x \geqslant 0$, we can also have $\ln\sqrt{x} \leqslant \sqrt{x}-1$. Thus, we have

$$
\begin{aligned}
E\ln\left[L_n(\psi)/L_n(\psi_0)\right] &\leqslant E\left(\sqrt{L_n(\psi)/L_n(\psi_0)}-1\right)\\
&= \int\left(\sqrt{L_n(\psi)/L_n(\psi_0)}-1\right)L_n(\psi_0)du_n\\
&= \int\sqrt{L_n(\psi)L_n(\psi_0)}du_n - 1\\
&= -\frac{1}{2}\int\left[\sqrt{L_n(\psi)}-\sqrt{L_n(\psi_0)}\right]^2 du_n \leqslant 0
\end{aligned}
$$

This implies in particular the information inequality that $E\ln L_n(\psi) \leqslant E\ln L_n(\psi_0)$ for all ψ. Thus, ψ_0 is a maximizer. Also, this inequality also implies that if $E\ln L_n(\psi) = E\ln L_n(\psi_0)$, $\ln L_n(\psi) = \ln L_n(\psi_0)$ almost surely. Assume there exist ψ_1 that $\ln L_n(\psi_1) = \ln L_n(\psi_0)$ almost surely, then we must have If

$$
\begin{aligned}
&-\frac{1}{2}\exp\left\{-\frac{\alpha_1}{1-\theta_1}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi_1(I_n-\theta_1W_n)^{-1}W_n)_{ij}}\,u_{i,n}^2 - \frac{1}{2}1_n^{'}\phi_1(I_n-\theta_1W_n)^{-1}W_n logu_n^2\\
&-\frac{n\alpha_1}{2(1-\theta_1)} + \ln\left|I_n - \phi_1(I_n-\theta_1W_n)^{-1}W_n\right|\\
\equiv &-\frac{1}{2}\exp\left\{-\frac{\alpha_0}{1-\theta_0}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi_0(I_n-\theta_0W_n)^{-1}W_n)_{ij}}\,u_{i,n}^2 - \frac{1}{2}1_n^{'}\phi_0(I_n-\theta_0W_n)^{-1}W_n logu_n^2\\
&-\frac{n\alpha_0}{2(1-\theta_0)} + \ln\left|I_n - \phi_0(I_n-\theta_0W_n)^{-1}W_n\right|
\end{aligned}
$$

for almost every possible value of u_n. Since u_n can take arbitrary values and W_n is invertible, we must have the following results:

$$
\begin{aligned}
&-\frac{1}{2}\exp\left\{-\frac{\alpha_1}{1-\theta_1}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi_1(I_n-\theta_1W_n)^{-1}W_n)_{ij}}\,u_{i,n}^2 - \frac{1}{2}1_n^{'}\phi_1(I_n-\theta_1W_n)^{-1}W_n logu_n^2\\
\equiv &-\frac{1}{2}\exp\left\{-\frac{\alpha_0}{1-\theta_0}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi_0(I_n-\theta_0W_n)^{-1}W_n)_{ij}}\,u_{i,n}^2 - \frac{1}{2}1_n^{'}\phi_0(I_n-\theta_0W_n)^{-1}W_n logu_n^2
\end{aligned}
\tag{10}
$$

and

$$
-\frac{n\alpha_1}{2(1-\theta_1)} + \ln\left|I_n - \phi_1(I_n-\theta_1W_n)^{-1}W_n\right| \equiv -\frac{n\alpha_0}{2(1-\theta_0)} + \ln\left|I_n - \phi_0(I_n-\theta_0W_n)^{-1}W_n\right| \tag{11}
$$

From (10), take $|u_{i,n}| = 1$ for $\forall i$, then $logu_n^2 = 0$ and $|u_{i,n}|^a = 0$ for $a \in \mathbb{R}$, then we have

$$
\frac{\alpha_1}{1-\theta_1} \equiv \frac{\alpha_0}{1-\theta_0} \tag{12}
$$

Then, take $|u_{i,n}| = 1$ for $\forall i \neq 1$, then (1) becomes

$$
\begin{aligned}
& -\frac{1}{2}\exp\left\{-\frac{\alpha_1}{1-\theta_1}\right\}\left\{u_{1,n}^2+\sum_{i=2}^{n}|u_{1,n}|^{-2\left(\phi_1(I_n-\theta_1 W_n)^{-1}W_n\right)_{i1}}\right\}-\frac{1}{2}\sum_{i=1}^{n}\left(\phi_1\left(I_n-\theta_1 W_n\right)^{-1}W_n\right)_{i1}\log u_{1,n}^2 \\
\equiv & -\frac{1}{2}\exp\left\{-\frac{\alpha_0}{1-\theta_0}\right\}\left\{u_{1,n}^2+\sum_{i=2}^{n}|u_{1,n}|^{-2\left(\phi_0(I_n-\theta_0 W_n)^{-1}W_n\right)_{i1}}\right\}-\frac{1}{2}\sum_{i=1}^{n}\left(\phi_0\left(I_n-\theta_0 W_n\right)^{-1}W_n\right)_{i1}\log u_{1,n}^2
\end{aligned}
$$

Since $u_{i,n}$ can take arbitrary non-zero value, we need to have $\left(\phi_0(I_n-\theta_0 W_n)^{-1}W_n\right)_{i1} = \left(\phi_1(I_n-\theta_1 W_n)^{-1}W_n\right)_{i1}$ for $\forall i$. Similarly, for $\forall j$, we can take $|u_{i,n}|=1$ for $\forall i \neq j$. By the same argument, we need to have $\left(\phi_0(I_n-\theta_0 W_n)^{-1}W_n\right)_{ij} = \left(\phi_1(I_n-\theta_1 W_n)^{-1}W_n\right)_{ij}$. Thus, we must have the following result:

$$
\phi_0(I_n-\theta_0 W_n)^{-1}W_n = \phi_1(I_n-\theta_1 W_n)^{-1}W_n
$$

which can be transformed to

$$
\left(\phi_0-\phi_1\right)W_n = \left(\phi_0\theta_1-\phi_1\theta_0\right)W_n^2 \tag{13}
$$

Since W_n is row-normalized, W_n^2 is also a row-normalized matrix. Thus, on one hand, $W_n \neq kW_n^2$ for any $k \neq 1$. On the other hand, suppose $W_n = W_n^2$, (υ,ξ) is a pair of eigenvalue and eigenvector of W_n, then we should have $\upsilon\xi = W_n\xi = W_n^2\xi = W_n(\nu\xi) = \upsilon W_n\xi = \upsilon^2\xi$, as $\xi \neq 0$, υ can only be 0 or 1. Then, $trace\,(W_n) = rank(W_n)$ since they are all equal to sum of eigenvalues of W_n. However, as $w_{ii}=0$ for $\forall i$, $trace(W_n)=0$ which obviously not equal to the rank. Thus, we can not have $W_n = W_n^2$. To make sure (13) holds, the only way is to have

$$
\phi_0 = \phi_1 \; and \; \phi_0\theta_1 = \phi_1\theta_0 \tag{14}
$$

By assuming $\phi_0 \neq 0$, we have both $\phi_0 = \phi_1$ and $\theta_0 = \theta_1$. Then from (12), we can get $\alpha_0 = \alpha_1$. Thus, we must have $\psi_1 = \psi_0$ which implies that ψ_0 is the unique maximizer, i.e. it can be identified. $\square$

When $\phi_0 = 0$, the model can not be identified since we can only get $\phi_1 = \phi_0 = 0$ from (14) and $\frac{\alpha_1}{1-\theta_1} = \frac{\alpha_0}{1-\theta_0}$ from (12), but α_0 and θ_0 can not be separately identified. However, in this case, as long as $\frac{\alpha_1}{1-\theta_1} = \frac{\alpha_0}{1-\theta_0}$, we actually have the same DGP. When $\phi_0 = 0$, we have

$$
\begin{aligned}
log u_n^2 &= \left(I_n-\theta_0 W_n\right)^{-1}\left(I_n-\theta_0 W_n\right) log\varepsilon_n^2 + \alpha_0(I_n-\theta_0 W_n)^{-1}1_n \\
&= log\varepsilon_n^2 + \alpha_0\sum_{k=0}^{\infty}\theta_0^k W_n^k 1_n \\
&= log\varepsilon_n^2 + \alpha_0\sum_{k=0}^{\infty}\theta_0^k 1_n \\
&= log\varepsilon_n^2 + \frac{\alpha_0}{1-\theta_0}1_n
\end{aligned}
$$

which implies any combination of α and θ will give the same model. But in this case, we no longer have heteroskedasticity and any spatial correlation since $u_{i,n}^2 = \exp\left\{\frac{\alpha_0}{1-\theta_0}\right\}\varepsilon_{i,n}^2$. In the following discussion, to avoid this case and meet compact condition, the following assumption is needed:

Assumption 6: *The parameter space* Θ *of* $\psi = (\alpha, \theta, \phi)^{'}$ *is a compact subset of* $\Theta^{'} = \{(\alpha, \theta, \phi)^{'} : \alpha \in \mathbb{R}, -1 < \theta < 1, 0 < \phi < 1, 0 \leqslant \phi + \theta < 1\}$.

For convenience, denote $Q_n(\psi) \equiv E[\ln L_n(\psi)]$. After build up identification condition, In the limit as n tends to infinity, we assume the identification in terms of limiting information inequality remains valid:

Assumption 7: $\liminf_{n\to\infty} \frac{1}{n}[Q_n(\psi_0) - Q_n(\psi)] > 0$ *for any* $\psi \neq \psi_0$.

Corollary 1:

For the ARCH-like model, i.e. $\theta_0 \equiv 0$ *case, the model can be identified when* $\phi_0 = 0$.

Proof:

When $\theta_0 \equiv 0$, recall the proof for Theorem 3, we can get $\alpha_0 = \alpha_1$ directly, and also

$$(\phi_0 - \phi_1) W_n = (\phi_0 - \phi_1) W_n^2$$

By the discussion on W_n^2, this equation will only hold when $\phi_0 = \phi_1$ for any $\phi_0 \in [0, 1)$. Thus, the model can be identified when $\phi_0 = 0$. $\square$

3.4 $\alpha-$mixing Property of Two Induced Variables

In this section, we are going to do some preparing work for the proof of consistency. In Section 2.2, we prove $\{u_{i,n}\}_{i\in D_n}$ is an $\alpha-$mixing spatial process with some particular upper bond of the mixing coefficient under Assumption 1 to Assumption 4. Due to the high non-linear structure of our log-likelihood function, it is not enough to build up asymptotic theory based on the property of $\{u_{i,n}\}_{i\in D_n}$ itself. Define the following two variables:

$$v_{i,n}(\psi; u_n) = \left((I_n - \theta W_n)^{-1} W_n log u_n^2\right)_{i,n}$$

$$\kappa_{i,n}(\psi; u_n) = \exp\left\{-\frac{\alpha}{1-\theta}\right\} u_{i,n}^2 \prod_{j=1}^{n} |u_{j,n}|^{-2\left(\phi(I_n-\theta W_n)^{-1}W_n\right)_{ij}}$$

which are contained in the log-likelihood function. The remaining part of this section is to prove the jointly mixing of $\{v_{i,n}(\psi; u_n)\}_{i\in D_n}$ and $\{\kappa_{i,n}(\psi; u_n)\}$, and also derive the upper bond of their $\alpha-$mixing coefficients under Assumption 1 to Assumption 4.

Let $e_n = W_n log u_n^2$, then we have

$$\begin{pmatrix} e_{i,n} \\ \ln u_{i,n}^2 \end{pmatrix} = \sum_{j=1}^{n} \begin{pmatrix} w_{ij,n} \\ 0 \end{pmatrix} \ln u_{j,n}^2 + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ln u_{i,n}^2$$

Recall that for any $Q \subset D_n$, $s > 0$, denote $Q^s = \{i \in D_n : d(i, Q) < s\}$, i.e. the neighbor of Q with their distance less than s. Under Assumption 3, since $w_{ij,n} = 0$ when $d_{ij} > \bar{d}_0$,

$$\begin{pmatrix} e_{i,n} \\ \ln u_{i,n}^2 \end{pmatrix} = \sum_{j\in\{i\}^{\bar{d}_0}} \begin{pmatrix} w_{ij,n} \\ 0 \end{pmatrix} \ln u_{j,n}^2 + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ln u_{i,n}^2$$

$\forall\ U, V \subseteq D_n$, $|U| \leqslant k$, $|V| \leqslant l$, when $d(U, V) \geqslant r > 3\bar{d}_0$, by Jenish and Prucha (2009), as $\left|\{j \in D_n : d_{ij} \leqslant \bar{d}_0\}\right| \leqslant C_d (3\bar{d}_0)^d$, where C_d is a positive finite constant. then we have

$$\left|U^{\bar{d}_0}\right| \leqslant |U| C_d (3\bar{d}_0)^d \leqslant k C_d (3\bar{d}_0)^d$$

$$\left|V^{\bar{d}_0}\right| \leqslant |V| C_d \left(3\bar{d}_0\right)^d \leqslant l C_d \left(3\bar{d}_0\right)^d$$

For $i \in U$, we have $\sigma\left(\left\{\ln u_{j,n}^2 : j \in \{i\}^{\bar{d}_0}\right\}\right) \subseteq \sigma\left(\left\{\ln u_{j,n}^2 : j \in U^{\bar{d}_0}\right\}\right)$. Thus, $\forall A \in \sigma\left(\left\{(e_{i,n}, \ln u_{i,n})^{'}\right\}_{i\in U}\right)$, it implies $A \in \sigma\left(\left\{\ln u_{j,n}^2 : j \in U^{\bar{d}_0}\right\}\right)$. Similarly, For $i \in V$, we have $\sigma\left(\left\{\ln u_{j,n}^2 : j \in \{i\}^{\bar{d}_0}\right\}\right) \subseteq \sigma\left(\left\{\ln u_{j,n}^2 : j \in V^{\bar{d}_0}\right\}\right)$. Thus, $\forall B \in \sigma\left(\left\{(e_{i,n}, \ln u_{i,n})^{'}\right\}_{i\in V}\right)$, it implies $B \in \sigma\left(\left\{\ln u_{j,n}^2 : j \in V^{\bar{d}_0}\right\}\right)$. Also, $\forall i \in U^{\bar{d}_0}, j \in V^{\bar{d}_0}$, $d_{ij} > r - 2\bar{d}_0$. Thus, when $r > 3\bar{d}_0$, $d\left(U^{\bar{d}_0}, V^{\bar{d}_0}\right) > \frac{r}{3}$, we have

$$\begin{aligned}
\alpha_{k,l}^{(e,lnu)}(r) &= \sup_{U,V\subseteq D_n} \{\alpha_n(U,V) : |U| \leqslant k, |V| \leqslant l, d(U,V) \geqslant r\} \\
&\leqslant \sup_{U^{\bar{d}_0}, V^{\bar{d}_0}\subseteq D_n} \left\{\alpha_n(U^{\bar{d}_0}, V^{\bar{d}_0}) : |U^{\bar{d}_0}| \leqslant kC_d\left(3\bar{d}_0\right)^d, |V^{\bar{d}_0}| \leqslant lC_d\left(3\bar{d}_0\right)^d, d(U^{\bar{d}_0}, V^{\bar{d}_0}) \geqslant \frac{r}{3}\right\} \\
&= \alpha^{(\ln u)}_{kC_d\left(3\bar{d}_0\right)^d, lC_d\left(3\bar{d}_0\right)^d}\left(\frac{r}{3}\right) \\
&\leqslant C_3 C_d \left(3\bar{d}_0\right)^d min(k,l)\left(\frac{1}{3^{d-1}} r^{d-1} \xi^{r/\bar{d}_0}\right)^{1/3} \\
&\equiv C_3^{'} \min(k,l)\left(\frac{1}{3^{d-1}} r^{d-1} \xi^{r/\bar{d}_0}\right)^{1/3}
\end{aligned}$$

As $r \to \infty$, $\alpha_{k,l}^{(e,lnu)}(r) \to 0$ as $\xi < 1$. Thus, $\left\{\left(e_{i,n}, \ln u_{i,n}^2\right)^{'}\right\}_{i\in D_n}$ is $\alpha-$mixing.

Let $v_n = (v_{1,n}, \cdots, v_{n,n})^{'}$, $\log\kappa = \left(\ln\kappa_{1,n}^2, \cdots, \ln\kappa_{n,n}^2\right)^{'}$, we have

$$v_n = (I_n - \theta W_n)^{-1} W_n \log u_n^2 = (I_n - \theta W_n)^{-1} e_n$$

$$\log\kappa_n = (I_n - \theta W_n)^{-1}\left[I_n - (\theta+\phi)W_n\right]\log u_n^2 - \alpha\left(I_n - \theta W_n\right)^{-1} 1_n$$

then we can write down

$$\begin{aligned}
\begin{pmatrix} v_{i,n}(\psi) \\ \ln\kappa_{i,n}(\psi) \end{pmatrix} &= \sum_{j=1}^n \begin{pmatrix} (I_n - \theta W_n)^{-1}_{ij,n} & 0 \\ 0 & \left((I_n - \theta W_n)^{-1}\left[I_n - (\theta+\phi)W_n\right]\right)_{ij,n} \end{pmatrix} \begin{pmatrix} e_{j,n} \\ \ln u_{j,n}^2 \end{pmatrix} \\
&\quad - \begin{pmatrix} 0 \\ \alpha\left(I_n - \theta W_n\right)^{-1}_{ij,n} \end{pmatrix} \\
&= \sum_{j=1}^n \begin{pmatrix} A_{1,ij}(\psi) & 0 \\ 0 & A_{2,ij}(\psi) \end{pmatrix} \begin{pmatrix} e_{j,n} \\ \ln u_{j,n}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ \alpha_i(\psi) \end{pmatrix}
\end{aligned}$$

Let $\iota_{i,n}(\psi) = \left(v_{i,n}(\psi), \ln\kappa_{i,n}(\psi)\right)^{'}$, $\widetilde{A}_{ij} = \begin{pmatrix} A_{1,ij}(\psi) & 0 \\ 0 & A_{2,ij}(\psi) \end{pmatrix}$, $\mathfrak{I}_n = \left(\iota_{1,n}^{'}, \cdots, \iota_{n,n}^{'}\right)^{'}$, $\widetilde{A} = \begin{pmatrix} \widetilde{A}_{11} & \cdots & \widetilde{A}_{1n} \\ \vdots & \ddots & \vdots \\ \widetilde{A}_{n1} & \cdots & \widetilde{A}_{nn} \end{pmatrix}$, $\widetilde{\varepsilon}_{j,n} = \left(e_{j,n}, \ln u_{j,n}^2\right)^{'}$, $\widetilde{\varepsilon} = \left(\widetilde{\varepsilon}_{1,n}^{'}, \cdots, \widetilde{\varepsilon}_{n,n}^{'}\right)^{'}$, $\widetilde{\alpha}_{i,n} = \left(0, \alpha_i(\psi)\right)^{'}$, $\widetilde{\alpha}_n = \left(\widetilde{\alpha}_{1,n}^{'}, \cdots, \widetilde{\alpha}_{n,n}^{'}\right)^{'}$, then

$$\mathfrak{I}_n = \widetilde{A}\widetilde{\varepsilon}_n + \widetilde{\alpha}_n$$

Similar to the proof in Section 2.2, $\widetilde{A}$ can be written as

$$\widetilde{A} = T_{l_n} \cdots T_2 T_1 \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} T_1 T_2 \cdots T_{l_n}$$

where T_{l_n} is the same type of matrix defined as before.

With Assumption 3 and Assumption 5, we have

$$\sup_n \|A_1\|_\infty = \sup_n \left\| (I_n - \theta W_n)^{-1} \right\|_\infty \leqslant \frac{1}{1-|\theta|}$$

$$\sup_n \|A_2\|_\infty \leqslant \sup_n \left\| (I_n - \theta W_n)^{-1} \right\|_\infty \sup_n \|I_n - (\theta+\phi) W_n\|_\infty \leqslant \frac{1+\theta+\phi}{1-|\theta|}$$

For the spatial GARCH-like case and the spatial ARCH-like case, due to different property of $(I_n - \theta W_n)^{-1} [I_n - (\theta+\phi)W_n]$ when $\theta = 0$ or not, we need to discuss the $\alpha-$mixing property separately for

Case 1: $\theta \neq 0$ (GARCH-like)

$$|A_{1,ij}| = \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty} \theta^l (W_n)_{ij}^l \leqslant \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty} \theta^l \leqslant \frac{\theta^{\lfloor d_{ij}/\bar{d}_0 \rfloor}}{1-\theta}$$

$$\begin{aligned}
|A_{2,ij}| &= \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty} \theta^l (W_n)_{ij}^l - (\phi+\theta) \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor - 1}^{\infty} \theta^l (W_n)_{ij}^{l+1} \\
&\leqslant \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor}^{\infty} \theta^l + (\phi+\theta) \sum_{l=\lfloor d_{ij}/\bar{d}_0 \rfloor - 1}^{\infty} \theta^{l+1} \\
&\leqslant \frac{1}{1-\theta} \left(1 + \frac{\phi+\theta}{\theta}\right) \theta^{\lfloor d_{ij}/\bar{d}_0 \rfloor}
\end{aligned}$$

Thus, by Lemma A.1 in Jenish and Prucha (2009),

$$\begin{aligned}
\sup_n \|A_1\|_1 = \sup_{j \in D_n, n} \sum_{j=1}^{n} |A_{1,ij}| \\
&\leqslant \sup_{j,n} \sum_{l=0}^{\infty} \sum_{i \in D_n : l\bar{d}_0 < d_{ij} < (1+l)\bar{d}_0} |A_{1,ij}| \\
&\leqslant \sum_{l=0}^{\infty} C_1 (l+1)^{d-1} \frac{\theta^{\lfloor d_{ij}/\bar{d}_0 \rfloor}}{1-\theta} < \infty
\end{aligned}$$

$$
\begin{aligned}
\sup_n \|A_2\|_1 &= \sup_{j \in D_n, n} \sum_{j=1}^{n} |A_{2,ij}| \\
&\leqslant \sup_{j,n} \sum_{l=0}^{\infty} \sum_{i \in D_n : l\bar{d}_0 < d_{ij} < (1+l)\bar{d}_0} |A_{2,ij}| \\
&\leqslant \sum_{l=0}^{\infty} C_1 (l+1)^{d-1} \frac{1}{1-\theta} \left(1 + \frac{\phi+\theta}{\theta}\right) \theta^{\lfloor d_{ij}/\bar{d}_0 \rfloor} < \infty
\end{aligned}
$$

Similarly, we can prove $\sup_n \left\|A_1^{-1}\right\|_1 < \infty$, $\sup_n \left\|A_2^{-1}\right\|_1 < \infty$. Also, we have $\lim_{s\to\infty} \sup_n \sup_{j\in Q} \sum_{i\in D_n\setminus Q^s} \left|\left(A_1^{-1}\right)_{ij}\right| =$ 0 and $\lim_{s\to\infty} \sup_n \sup_{j\in Q} \sum_{i\in D_n\setminus Q^s} \left|\left(A_2^{-1}\right)_{ij}\right| = 0$ based on Assumption 3. Thus, $\sup_n \left\|\widetilde{A}\right\|_1 < \infty$, $\sup_n \left\|\widetilde{A}^{-1}\right\|_1 < \infty$, and $\lim_{s\to\infty} \sup_n \sup_{j\in Q; q=1,2} \left[\sum_{r\in D_n\setminus Q^s; t=1,2} \left|\left(\widetilde{A}^{-1}\right)_{tq,rj}\right|\right] = 0$.

Let $C^{''} = \max\left\{\frac{1}{1-\theta}, \frac{1}{1-\theta}\left(1+\frac{\phi+\theta}{\theta}\right)\right\}$, then we have

$$
\sup_{1\leqslant p\leqslant 2, 1\leqslant q\leqslant 2, n} \left|\widetilde{A}_{pq,ij,n}\right| \leqslant \sup_n \left\{|A_{1,ij}|, |A_{2,ij}|\right\} \leqslant C^{''} \theta^{\lfloor d_{ij}/\bar{d}_0 \rfloor}
$$

where $C^{''}\theta^{\lfloor d_{ij}/\bar{d}_0 \rfloor}$ is a non-increasing function of d_{ij}.

Also, the $\alpha-$mixing coefficient of $(e_{i,n}, \ln u_{i,n})^{'}$ satisfies that

$$
\begin{aligned}
\sum_{r=1}^{\infty} r^{d-1} \alpha_{1,1}^{(e,\ln u)}(r)^{\delta/(2+\delta)} &\leqslant \sum_{r=1}^{\lfloor 3\bar{d}_0 \rfloor} r^{d-1} \alpha_{1,1}^{(e,\ln u)}(r)^{\delta/(2+\delta)} + \sum_{r=\lfloor 3\bar{d}_0 \rfloor}^{\infty} r^{d-1} \left(\frac{1}{3^{d-1}} r^{d-1} \xi^{r/\bar{d}_0}\right)^{\delta/3(2+\delta)} \\
&= \sum_{r=1}^{\lfloor 3\bar{d}_0 \rfloor} r^{d-1} \alpha_{1,1}^{(e,\ln u)}(r)^{\delta/(2+\delta)} + \sum_{r=1}^{\infty} \left(\frac{1}{3}\right)^{d-1} r^{[1+\delta/3(2+\delta)](d-1)} \xi^{r\delta/3(2+\delta)\bar{d}_0} < \infty
\end{aligned}
$$

for any positive δ since $\xi < 1$.

Since we proved $\sup_{n,i\in D_n} E\left|\ln u_{i,n}^2\right|^p < \infty$ for any positive integer p, we have

$$
\begin{aligned}
E|e_{i,n}|^p &= E\left|\sum_{j=1}^{n} w_{ij,n} \ln u_{j,n}^2\right|^p \\
&\leqslant (\|W_n\|_\infty)^p \sup_{n,i\in D_n} E\left|\ln u_{i,n}^2\right|^p \\
&\leqslant \sup_{n,i\in D_n} E\left|\ln u_{i,n}^2\right|^p < \infty
\end{aligned}
$$

thus $\|\widetilde{\varepsilon}_{i,n}\|_{L^p} = \sup_{i\in D_n, n} \max\left\{E|e_{i,n}|^p, E\left|\ln u_{i,n}^2\right|^p\right\} < \infty$. Since the density of $\varepsilon_{i,n}$ is absolute continuous on $\mathbb{R}$, and the mapping from u_n and ε_n is bijection, we can easily get absolute continuity of the density function of $\widetilde{e}_{i,n}$. By Lemma A.1 in Jenish and Prucha (2009), $|\{j \in D_n : d_{ij} \leqslant l\}| \leqslant C_d l^d$ for all $l \geqslant 1$ for some constant $C_d > 0$.Thus, based on Theorem 2 in Xu and Lee (2019), the $\alpha-$mixing coefficient of $\iota_{i,n}$ satisfies:

$$
\begin{aligned}
\alpha_{k,l}^{(\iota)}(r) &\leqslant C_{P_2}\min(k,l)\left(C^{''}\theta^{\lfloor r/3\bar{d}_0\rfloor}\left[\frac{1}{1-|\theta|}+\frac{1+\theta+\phi}{1-|\theta|}\right]\right)^{1/3} \\
&+\alpha_{C_dk(r/3)^d,C_dl(r/3)^d}^{(e,\ln u)}\left(\frac{r}{3}\right) \\
&\leqslant C^{'''}\theta^{\lfloor r/3\bar{d}_0\rfloor/3}+C_3^{'}\min\left\{C_dk(r/3)^d,C_dl(r/3)^d\right\}\left(\frac{1}{3^{2(d-1)}}r^{d-1}\xi^{r/\bar{d}_0}\right)^{1/3} \\
&= C^{'''}\theta^{\lfloor r/3\bar{d}_0\rfloor/3}+C_3^{'}\min\{k,l\}C_d3^{-\frac{7}{3}}r^{\frac{13}{3}}\xi^{r/3\bar{d}_0} \qquad (15)
\end{aligned}
$$

for $r>9\bar{d}_0$ where C_{P_2} and $C^{'''}$ are finite constants depend on θ and ϕ.

For any k and l, we have $\lim_{r\to\infty}\alpha_{k,l}^{(\iota)}(r)=0$. Thus, $v_{i,n}(\psi)$ and $\ln\kappa_{i,n}(\psi)$ are jointly $\alpha-$mixing.

Case 2: $\theta=0$ (ARCH-like)

When $\theta=0$, we have

$$v_n(\psi)=\log u_n^2=e_n$$

$$\log\kappa_n=(I_n-\phi W_n)\log u_n^2-\alpha 1_n$$

Then, the matrices A_1 and A_2 becomes simpler: $A_1=I_n$ and $A_2=I_n-\phi W_n$. Properties of A_1 are easy to see. For A_2, $|A_{2,ij}|=0$ when $d_{ij}>\bar{d}_0$. By Lemma A.1 in Jenish and Prucha (2009), we have

$$
\begin{aligned}
\sup_n\|A_2\|_1 &\leqslant 1+\phi\sup_n\sum_{i=1}^n w_{ij,n} \\
&\leqslant 1+\phi\sum_{i\in D_n:d(i,j)\leqslant\bar{d}_0} w_{ij,n} \\
&\leqslant 1+\phi C_d\bar{d}_0^d<\infty
\end{aligned}
$$

Then, we have

$$\sup_{1\leqslant p\leqslant 2,1\leqslant q\leqslant 2,n}\left|\widetilde{A}_{pq,ij,n}\right|=\begin{cases}1 & d_{ij}\leqslant\bar{d}_0\\ 0 & d_{ij}>\bar{d}_0\end{cases}\equiv g(d_{ij})$$

$$\sup_{i\in D_n,n}\sum_{j\in D_n\backslash Q^s}g(d_{ij})=0\ when\ s>\bar{d}_0$$

With all other results similar to GARCH-like case, for $r>9\bar{d}_0$, we have

$$
\begin{aligned}
\alpha_{k,l}^{(\iota)}(r) &\leqslant \alpha_{C_dk(r/3)^d,C_dl(r/3)^d}^{(e,\ln u)}\left(\frac{r}{3}\right) \\
&\leqslant \min\left\{C_dk(r/3)^d,C_dl(r/3)^d\right\}\left(\frac{1}{3^{2(d-1)}}r^{d-1}\xi^{r/\bar{d}_0}\right)^{1/3} \\
&= \min\{k,l\}C_d3^{-\frac{7}{3}}r^{\frac{13}{3}d}\xi^{r/3\bar{d}_0} \qquad (16)
\end{aligned}
$$

For any k and l, we have $\lim_{r\to\infty}\alpha_{k,l}^{(\iota)}(r)=0$. Thus, $v_{i,n}(\psi)$ and $\ln\kappa_{i,n}(\psi)$ are jointly $\alpha-$mixing.

3.5 Consistency of the MLE

The next step is to discuss the consistency of estimators. Let

$$
\begin{aligned}
q_n(\psi) &= \frac{1}{n} \ln L_n(\psi) \\
&= -\frac{1}{2}\ln(2\pi) - \frac{1}{2n}\exp\left\{-\frac{\alpha}{1-\theta}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2\left(\phi(I_n-\theta W_n)^{-1}W_n\right)_{ij}} u_{i,n}^2 - \frac{\alpha}{2(1-\theta)} \\
&- \frac{\phi}{2n}1_n^{'}\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2 + \frac{1}{n}\ln|I_n-(\theta+\phi)W_n| - \frac{1}{n}\ln|I_n-\theta W_n|
\end{aligned}
$$

To proof consistency, we can prove the uniform convergence in probability of $q_n(\psi)$: $\sup_{\psi\in\Theta}|q_n(\psi) - Eq_n(\psi)| = o_p(1)$. Denote the following functions:

$$
q_{1,n}(\psi) = -\frac{\alpha}{2(1-\theta)} + \frac{1}{n}\ln|I_n-(\theta+\phi)W_n| - \frac{1}{n}\ln|I_n-\theta W_n| \tag{17}
$$

$$
q_{2,n}(\psi) = \frac{\phi}{n}1_n^{'}\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2 \tag{18}
$$

$$
q_{3,n}(\psi) = \frac{1}{n}\exp\left\{-\frac{\alpha}{1-\theta}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2\left(\phi(I_n-\theta W_n)^{-1}W_n\right)_{ij}} u_{i,n}^2 \tag{19}
$$

It is easy to see that the uniform convergence in probability of $q_n(\psi)$ follows from the uniform convergence in probability of the three functions above. The easiest part is $q_{1,n}$ since it is a non-stochastic function of ψ. Thus, for all n, we should have $q_{3,n}(\psi) = Eq_{3,n}(\psi)$. Thus, $\sup_{\psi\in\Theta}|q_{3,n}(\psi) - Eq_{3,n}(\psi)| = 0$ which implies we always have the uniform convergence of $q_{3,n}(\psi)$. So, the key point is to show uniform convergence in probability of $q_{2,n}(\psi)$ and $q_{3,n}(\psi)$. Following Newey (1991), we are going to prove point-wise convergence and conditions in Corollary 3.1 to show uniform convergence.

Uniform convergence of $q_{2,n}(\psi)$:

For $q_{2,n}(\psi)$, for any ψ, we have

$$
\begin{aligned}
&|q_{2,n}(\psi) - Eq_{2,n}(\psi)| \\
=&\left|\frac{1}{n}\phi 1_n^{'}\left(I_n-\theta W_n\right)^{-1}W_n\left(logu_n^2 - Elogu_n^2\right)\right| \\
=&\left|\frac{1}{n}\phi\sum_{j=1}^{n}\left(\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2 - E\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2\right)_{i,n}\right| \\
\leqslant&\left|\frac{1}{n}\sum_{i=1}^{n}\left(\left(\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2\right)_{i,n} - E\left(\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2\right)_{i,n}\right)\right| \\
\equiv&\left|\frac{1}{n}\sum_{i=1}^{n}\left(v_{i,n}(\psi;u_n) - Ev_{i,n}(\psi;u_n)\right)\right|
\end{aligned}
$$

since $\phi\in(0,1)$.

By Assumption 3 and 5, we have

$$
\begin{aligned}
\left|\left((I_n-\theta W_n)^{-1}W_n\right)_{ij}\right| &= \left|\sum_{l=\lfloor d_{ij}/\bar{d}_0\rfloor}^{\infty}[\theta^l W_n^{l+1}]_{ij}\right| \\
&\leqslant \sum_{l=\lfloor d_{ij}/\bar{d}_0\rfloor}^{\infty}\left\|\theta^l W_n^{l+1}\right\|_{\infty} \\
&= \sum_{l=\lfloor d_{ij}/\bar{d}_0\rfloor}^{\infty}|\theta|\,\theta^l \\
&= \frac{|\theta|^{\lfloor d_{ij}/\bar{d}_0\rfloor}}{1-|\theta|}
\end{aligned}
$$

Recall Lemma A.1 in Jenish and Prucha (2009), we can get

$$
\begin{aligned}
\sup_n\left\|(I_n-\theta W_n)^{-1}W_n\right\|_1 &= \sup_{j\in D_n,n}\sum_{i=1}^{n}\left|\left((I_n-\theta W_n)^{-1}W_n\right)_{ij}\right| \\
&= \sup_{j\in D_n,n}\sum_{l=0}^{\infty}\sum_{i\in D_n:l\bar{d}_0<d_{ij}<(1+l)\bar{d}_0}\left|\left((I_n-\theta W_n)^{-1}W_n\right)_{ij}\right| \\
&\leqslant \sum_{l=0}^{\infty}C_1(l+1)^{d-1}\frac{\theta^l}{1-\theta} = C_5(\theta)<\infty
\end{aligned}
$$

Also, from Assumption 5, we have

$$
\begin{aligned}
\left\|(I_n-\theta W_n)^{-1}W_n\right\|_{\infty} &= \left\|\sum_{l=0}^{\infty}\theta^l W_n^{l+1}\right\|_{\infty} \\
&= \sum_{l=0}^{\infty}\left|\theta^l\right|\left\|W_n^{l+1}\right\|_{\infty} \\
&= \frac{1}{1-|\theta|}
\end{aligned}
$$

Since we have the following equation:

$$
\begin{aligned}
logu_n^2 &= [I_n-(\phi_0+\theta_0)W_n]^{-1}(I_n-\theta_0 W_n)log\varepsilon_n^2+\alpha_0[I_n-(\phi_0+\theta_0)W_n]^{-1}1_n \\
&= G_n(\phi_0,\theta_0)log\varepsilon_n^2+\frac{\alpha_0}{1-\phi_0-\theta_0}
\end{aligned}
$$

By Lemma A.3 in Xu and Lee (2015b) , for any $p \in N^{+}$, as $\|G_n(\phi_0,\theta_0)\|_\infty = \frac{1+|\theta_0|}{1-\phi_0-\theta_0}$, we have

$$\begin{aligned}
E\left|\log u_{i,n}^2\right|^p &= E\left|\sum_{j=1}^{n}\left(G_n(\phi_0,\theta_0)\right)_{ij}\log\varepsilon_{j,n}^2 + \frac{\alpha_0}{1-\phi_0-\theta_0}\right|^p \\
&\leqslant E\left\{\left(\sum_{j=1}^{n}\left|\left(G_n(\phi_0,\theta_0)\right)_{ij}\log\varepsilon_{j,n}^2\right|\right) + \left|\frac{\alpha_0}{1-\phi_0-\theta_0}\right|\right\}^p \\
&= E\sum_{r=0}^{p}\binom{p}{r}\left(\sum_{j=1}^{n}\left|\left(G_n(\phi_0,\theta_0)\right)_{ij}\log\varepsilon_{j,n}^2\right|\right)^r\left|\frac{\alpha_0}{1-\phi_0-\theta_0}\right|^{p-r} \\
&= \sum_{r=0}^{p}\binom{p}{r}\left|\frac{\alpha_0}{1-\phi_0-\theta_0}\right|^{p-r} E\left(\sum_{j=1}^{n}\left|\left(G_n(\phi_0,\theta_0)\right)_{ij}\log\varepsilon_{j,n}^2\right|\right)^r \\
&\leqslant \sum_{r=0}^{p}\binom{p}{r}\left|\frac{\alpha_0}{1-\phi_0-\theta_0}\right|^{p-r}\left(\frac{1+|\theta_0|}{1-\phi_0-\theta_0}\right)^r E\left(\left|\log\varepsilon_{j,n}^2\right|^r\right) \\
&< \infty
\end{aligned}$$

Then we have

$$\begin{aligned}
\left|v_{i,n}\left(\psi,u_n\right)\right| &= \left|\sum_{j=1}^{n}\left(\left(I_n-\theta W_n\right)^{-1}W_n\right)_{ij}\log u_{j,n}^2\right| \\
&\leqslant \left\|\left(I_n-\theta W_n\right)^{-1}W_n\right\|_\infty \sup_{j\in D_n}\left|\log u_{j,n}^2\right| \\
&= \frac{1}{1-|\theta|}\sup_{j\in D_n}\left|\log u_{j,n}^2\right|
\end{aligned}$$

Thus, we also have $\sup_{\Theta}\sup_n\sup_{i\in D_n}E\left|v_{i,n}\right|^p < \infty$ for $\forall p \in N^{+}$, which indicates uniform integrability:

$$\lim_{k\to\infty}\sup_n\sup_{i\in D_n}E\left[\left|v_{i,n}(\psi;u_n)\right|1_{(|v_{i,n}(\psi;u_n)|>k)}\right] = 0$$

Recall that the results (15) and (16) , as mixing property preserves under measurable transformation, we have

$$
\begin{aligned}
\sum_{m=1}^{\infty} m^{d-1}\alpha_{1,1}^{v}(m) &\leqslant \sum_{m=1}^{\lfloor 9\bar{d}_0 \rfloor} m\bar{\alpha}_{1,1}^{v}(m) + \sum_{m=\lfloor 9\bar{d}_0 \rfloor+1}^{\infty} m\bar{\alpha}_{1,1}^{v}(m) \\
&\leqslant \sum_{m=1}^{\lfloor 9\bar{d}_0 \rfloor} m + C^{'''} \sum_{m=\lfloor 9\bar{d}_0 \rfloor+1}^{\infty} m\theta^{\lfloor r/3\bar{d}_0 \rfloor/3} \\
&+ \sum_{m=\lfloor 9\bar{d}_0 \rfloor+1}^{\infty} mC_3^{'}C_d 3^{-\frac{7}{3}} m^{\frac{13}{3}} \xi^{r/3\bar{d}_0} \\
&\leqslant m\left\lfloor 9\bar{d}_0 \right\rfloor + C^{'''} \sum_{m=\lfloor 9\bar{d}_0 \rfloor+1}^{\infty} m\theta^{r/9\bar{d}_0} \\
&+ \sum_{m=\lfloor 9\bar{d}_0 \rfloor+1}^{\infty} C_3^{'}C_d 3^{-\frac{7}{3}} m^{\frac{16}{3}} \xi^{r/3\bar{d}_0} < \infty
\end{aligned} \tag{20}
$$

when $\theta \neq 0$ and

$$
\begin{aligned}
\sum_{m=1}^{\infty} m^{d-1}\alpha_{1,1}^{v}(m) &\leqslant \sum_{m=1}^{\lfloor 9\bar{d}_0 \rfloor} m\bar{\alpha}_{1,1}^{v}(m) + \sum_{m=\lfloor 9\bar{d}_0 \rfloor+1}^{\infty} m\bar{\alpha}_{1,1}^{v}(m) \\
&\leqslant m\left\lfloor 9\bar{d}_0 \right\rfloor + \sum_{m=\lfloor 9\bar{d}_0 \rfloor+1}^{\infty} C_d 3^{-\frac{7}{3}} m^{\frac{16}{3}} \xi^{r/3\bar{d}_0} < \infty
\end{aligned} \tag{21}
$$

when $\theta = 0$.

By Theorem 3 in Jenish and Prucha (2009), we have

$$
\frac{1}{n}\sum_{i=1}^{n}(v_{i,n} - Ev_{i,n}) \xrightarrow{L_1} 0
$$

Thus, $q_{2,n}(\psi) - Eq_{2,n}(\psi) \xrightarrow{L_1} 0$ directly follows from the LLN above. Thus, we get the point-wise convergence of $q_{2,n}(\psi)$.

For $\forall\psi, \psi^{'} \in \Theta$, as ϕ,$\phi^{'} \in (0,1)$

$$
\begin{aligned}
&\left|q_{2,n}(\psi)-q_{2,n}(\psi^{'})\right| \\
=&\left|\frac{1}{n}\phi 1_n^{'}\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2-\frac{1}{n}\phi^{'}1_n^{'}\left(I_n-\theta^{'}W_n\right)^{-1}W_n logu_n^2\right| \\
\leqslant&\left|\frac{1}{n}\left(\phi-\phi^{'}\right)1_n^{'}\left(I_n-\theta W_n\right)^{-1}W_n logu_n^2\right| \\
+&\left|\frac{\phi^{'}}{n}\left[1_n^{'}\left(I_n-\theta W_n\right)^{-1}W_n-1_n^{'}\left(I_n-\theta^{'}W_n\right)^{-1}W_n\right]logu_n^2\right| \\
\leqslant&\left|\phi-\phi^{'}\right|\frac{1}{n}\sum_{i=1}^{n}\left|v_{i,n}\left(\psi;u_n\right)\right|+\frac{1}{n}\sum_{i=1}^{n}\left|v_{i,n}\left(\psi;u_n\right)-v_{i,n}\left(\psi^{'};u_n\right)\right|
\end{aligned}
$$

By uniform integrability of $v_{i,n}$, we have $\sup_n\sup_{i\in N}E\left|v_{i,n}\left(\psi;u_n\right)\right|\leqslant C_7\left(\psi,\psi_0\right)$ which is a positive constant depend on ψ and ψ_0, thus $\frac{1}{n}\sum_{i=1}^{n}E\left|v_{i,n}\left(\psi;u_n\right)\right|=O(1)$. Focused on the second term in the inequality above, we have

$$
\begin{aligned}
&\left|v_{i,n}\left(\psi;u_n\right)-v_{i,n}\left(\psi^{'};u_n\right)\right| \\
=&\left|\sum_{j=1}^{n}\left(\left(I_n-\theta W_n\right)^{-1}W_n\right)_{ij}\log u_{i,n}^2-\sum_{j=1}^{n}\left(\left(I_n-\theta^{'}W_n\right)^{-1}W_n\right)_{ij}\log u_{j,n}^2\right| \\
\leqslant&\sum_{j=1}^{n}\left|\left(\left(I_n-\theta W_n\right)^{-1}W_n\right)_{ij}-\left(\left(I_n-\theta^{'}W_n\right)^{-1}W_n\right)_{ij}\right|\left|\log u_{j,n}^2\right| \\
=&\sum_{j=1}^{n}\left|\sum_{l=0}^{n}\left[\theta^l-\left(\theta^{'}\right)^l\right]\left(W_n^l\right)_{ij}\right|\left|\log u_{j,n}^2\right| \\
\leqslant&\sum_{j=1}^{n}\left|\sum_{l=0}^{n}\left[\theta^l-\left(\theta^{'}\right)^l\right]\right|\left|\log u_{j,n}^2\right| \\
=&\left|\frac{1}{1-\theta}-\frac{1}{1-\theta^{'}}\right|\sum_{j=1}^{n}\left|\log u_{j,n}^2\right|
\end{aligned}
$$

From uniform integrability of $\log u_{j,n}^2$, by the same argument for $v_{i,n}$, we have $\frac{1}{n}\sum_{i=1}^{n}\left|Ev_{i,n}\left(\psi;u_n\right)\right|=O(1)$. Thus, we can further get

$$
\begin{aligned}
&\left|q_{2,n}(\psi)-q_{2,n}(\psi^{'})\right| \\
\leqslant &\left|\phi-\phi^{'}\right| \frac{1}{n}\sum_{i=1}^{n}\left|v_{i,n}\left(\psi;u_n\right)\right| + \frac{1}{n}\sum_{i=1}^{n}\left|v_{i,n}\left(\psi;u_n\right)-v_{i,n}\left(\psi^{'};u_n\right)\right| \\
\leqslant &\left|\phi-\phi^{'}\right| \frac{1}{n}\sum_{i=1}^{n}\left|v_{i,n}\left(\psi;u_n\right)\right| + \left|\frac{1}{1-\theta}-\frac{1}{1-\theta'}\right| \frac{1}{n}\sum_{j=1}^{n}\left|\log u_{j,n}^2\right| \\
\leqslant &\left(\left|\phi-\phi^{'}\right| + \left|\frac{\theta-\theta'}{(1-\theta)(1-\theta')}\right|\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left|v_{i,n}\left(\psi;u_n\right)\right| + \frac{1}{n}\sum_{j=1}^{n}\left|\log u_{j,n}^2\right|\right) \\
\leqslant &\left(\left|\phi-\phi^{'}\right| + \widetilde{M}\left|\theta-\theta^{'}\right|\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left|v_{i,n}\left(\psi;u_n\right)\right| + \frac{1}{n}\sum_{j=1}^{n}\left|\log u_{j,n}^2\right|\right) \\
\equiv &h_1\left(\left|\psi-\psi^{'}\right|\right)B_1\left(u_n\right)
\end{aligned}
$$

where$\widetilde{M}=\sup_{\psi,'\psi\in\Theta}\frac{1}{(1-\theta)\left(1-\theta'\right)}$. Then we have $EB\left(u_n\right)=\frac{1}{n}\sum_{i=1}^{n}E\left|v_{i,n}\left(\psi;u_n\right)\right|+\frac{1}{n}\sum_{j=1}^{n}E\left|\log u_{j,n}^2\right|=O(1)+O\left(1\right)=O\left(1\right)$ uniformly in Θ, and $\lim_{\psi-\psi'\to 0}h_1\left(\left|\psi-\psi^{'}\right|\right)=\lim_{\psi-\psi'\to 0}\left(\left|\phi-\phi^{'}\right|+\widetilde{M}\left|\theta-\theta^{'}\right|\right)=0$. The By Corollary 3.1, combined with the point-wise convergence and compact parameter space, we have $Eq_{2,n}(\psi)$ is equicontinuous and uniform convergence of $q_{2,n}\left(\psi\right)$: $\sup_{\theta\in\Theta}\left|q_{2,n}\left(\psi\right)-Eq_{2,n}\left(\psi\right)\right|=o_p\left(1\right)$.

Uniform Convergence of $q_{3,n}\left(\psi\right)$:

Define the following function for $\forall i$:

$$
\kappa_{i,n}(\psi;u_n)=\exp\left\{-\frac{\alpha}{1-\theta}\right\}u_{i,n}^2\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2\left(\phi\left(I_n-\theta W_n\right)^{-1}W_n\right)_{ij}}
$$

With the true parameter ψ_0, $\kappa_{i,n}(\psi_0)=\varepsilon_{i,n}^2$ by recalling the equation (3). Given arbitrary ψ, we can get the following equation by transform the equation (2):

$$
\begin{aligned}
log\kappa_n &= \left(I_n-\theta W_n\right)^{-1}\left[I_n-\left(\theta+\phi\right)W_n\right]logu_n^2-\alpha\left(I_n-\theta W_n\right)^{-1}1_n \\
&= \left(I_n-\theta W_n\right)^{-1}\left[I_n-\left(\theta+\phi\right)W_n\right]logu_n^2-\frac{\alpha}{1-\theta}1_n \\
&= \left(I_n-\theta W_n\right)^{-1}\left[I_n-\left(\theta+\phi\right)W_n\right][I_n-(\phi_0+\theta_0)W_n]^{-1}(I_n-\theta_0W_n)log\varepsilon_n^2 \\
&+ \alpha_0\left(I_n-\theta W_n\right)^{-1}\left[I_n-\left(\theta+\phi\right)W_n\right][I_n-(\phi_0+\theta_0)W_n]^{-1}1_n-\frac{\alpha}{1-\theta}1_n \\
&\equiv \widetilde{K}_n\left(\psi_0,\psi\right)log\varepsilon_n^2+\widetilde{h}_n\left(\psi_0,\psi\right)
\end{aligned}
$$

where $log\kappa_n=\left(\log\kappa_{1,n}(\psi),\cdots,\log\kappa_{n,n}\left(\psi\right)\right)^{'}$. By transforming the equation above, we have

$$
\kappa_{i,n}=\exp\left(\sum_{j=1}^{n}\left(\widetilde{K}_n\left(\psi_0,\psi\right)\right)_{ij}\log\varepsilon_{j,n}^2+\widetilde{h}_{i,n}\left(\psi_0,\psi\right)\right)
$$

Since $\widetilde{K}_n = (I_n - \theta W_n)^{-1} [I_n - (\theta+\phi) W_n] [I_n - (\phi_0+\theta_0)W_n]^{-1}(I_n - \theta_0 W_n)$, we have

$$
\begin{aligned}
\left\|\widetilde{K}_n\right\|_\infty &\leqslant \left\|(I_n - \theta W_n)^{-1}\right\|_\infty \left\|(I_n - (\theta+\phi) W_n)^{-1}\right\|_\infty \left\|(I_n - (\phi_0+\theta_0)W_n)^{-1}\right\|_\infty \left\|(I_n - \theta_0 W_n)^{-1}\right\|_\infty \\
&= \frac{1}{1-|\theta|}\left(1-(\theta+\phi)\right)\frac{1}{1-(\theta_0+\phi_0)}\left(1-|\theta_0|\right) \\
&= \frac{1-|\theta_0|}{1-|\theta|}\frac{1-\theta-\phi}{1-\theta_0-\phi_0}
\end{aligned}
$$

And also we have

$$
\begin{aligned}
&\alpha_0 (I_n - \theta W_n)^{-1} [I_n - (\theta+\phi) W_n] [I_n - (\phi_0+\theta_0)W_n]^{-1} 1_n \\
=&\alpha_0 (I_n - \theta W_n)^{-1} [I_n - (\theta+\phi) W_n] \sum_{l=0}^{\infty} (\phi_0+\theta_0)^l W_n^l 1_n \\
=&\frac{\alpha_0}{1-\phi_0-\theta_0}\left(\sum_{l=1}^{\infty}\theta^l W_n^l - (\theta+\phi)\sum_{l=0}^{\infty}\theta^l W_n^{l+1}\right)1_n \\
=&\frac{\alpha_0}{1-\phi_0-\theta_0}\frac{1-\theta-\phi}{1-\theta}1_n
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\sup_n \sup_{i\in D_n} \kappa_{i,n} &= \exp\left\{\frac{\alpha_0}{1-\phi_0-\theta_0}\frac{1-\theta-\phi}{1-\theta} - \frac{\alpha}{1-\theta}\right\}\sup_n \sup_{i\in D_n} \exp\left(\sum_{j=1}^{n}\left(\widetilde{K}_n(\psi_0,\psi)\right)_{ij}\log\varepsilon_{j,n}^2\right) \\
&\leqslant \exp\left\{\frac{\alpha_0}{1-\phi_0-\theta_0}\frac{1-\theta-\phi}{1-\theta} - \frac{\alpha}{1-\theta}\right\}\sup_n \sup_{i\in D_n} \exp\left(\left\|\widetilde{K}_n(\psi_0,\psi)\right\|_\infty \left|\log\varepsilon_{i,n}^2\right|\right) \\
&\leqslant \exp\left\{\frac{\alpha_0}{1-\phi_0-\theta_0}\frac{1-\theta-\phi}{1-\theta} - \frac{\alpha}{1-\theta}\right\}\sup_n \sup_{i\in D_n} \exp\left(\frac{1-|\theta_0|}{1-|\theta|}\frac{1-\theta-\phi}{1-\theta_0-\phi_0}\left|\log\varepsilon_{i,n}^2\right|\right)
\end{aligned}
$$

For $\forall k \in \mathbb{N}+$ and random variable X, we have

$$
\begin{aligned}
&E\left(\exp\left\{k\left|\log X^2\right|\right\}\right) \\
=&\int_{\mathbb{R}} \exp\left\{k\left|\log x^2\right|\right\} dF(x) \\
=&\int_{\{|x|\geqslant 1\}} \exp\left\{k\log x^2\right\} dF(x) + \int_{\{|x|<1\}} \exp\left\{-k\log x^2\right\} dF(x) \\
=&\int_{\{|x|\geqslant 1\}} x^{2k} dF(x) + \int_{\{|x|<1\}} \frac{1}{x^{2k}} dF(x)
\end{aligned}
$$

Then, we can build up an upper bound of $E\kappa_{i,n}^k$:

$$
\begin{aligned}
&\sup_{\psi\in\Theta}\sup_{n}\sup_{i\in D_n} E\kappa_{i,n}^{k}\\
\leqslant &\sup_{\psi\in\Theta}\exp\left\{\frac{\alpha_0}{1-\phi_0-\theta_0}\frac{1-\theta-\phi}{1-\theta}-\frac{\alpha}{1-\theta}\right\}\sup_{n}\sup_{i\in D_n} E\exp\left(\frac{1-|\theta_0|}{1-|\theta|}\frac{1-\theta-\phi}{1-\theta_0-\phi_0}\left|\log\varepsilon_{i,n}^{2}\right|\right)\\
\leqslant &\sup_{\psi\in\Theta}\exp\left\{\frac{\alpha_0}{1-\phi_0-\theta_0}\frac{1-\theta-\phi}{1-\theta}-\frac{\alpha}{1-\theta}\right\}\sup_{n}\sup_{i\in D_n}\{\int_{\{|x|\geqslant 1\}} x^{2k\frac{1-|\theta_0|}{1-|\theta|}\frac{1-\theta-\phi}{1-\theta_0-\phi_0}}dF_{\varepsilon}(x)\\
+&\int_{\{|x|<1\}} x^{-2k\frac{1-|\theta_0|}{1-|\theta|}\frac{1-\theta-\phi}{1-\theta_0-\phi_0}}dF_{\varepsilon}(x)\}
\end{aligned}
$$

To make sure this upper bond exist, we need an additional assumption for $\varepsilon_{i,n}$:

Assumption 8:

The distribution function of $\varepsilon_{i,n}$, $F_{\varepsilon}(x)$, satisfies the following condition

$$
\int_{\{|x|<1\}} x^{-k}dF_{\varepsilon}(x)<\infty
$$

for any positive integer k.

This assumption seems contradict to the Normality assumption, since $N(0,1)$ does not meet this requirement. In fact, any distribution with non-zero density at $x=0$ would violate Assumption 8. But it does not mean the asymptotic of the estimator based on this assumption is useless. In fact, for any finite sample without zero observation, this model can still be applied. The detailed discussion will be showed in Section 4.1, where you can see this estimator still has a good performance when the epsilon is Normal.

Combined with other assumptions, we have $\sup_{\psi\in\Theta}\sup_n\sup_{i\in D_n}E\epsilon_{i,n}^{k}<\infty$. Then, we have uniform integrability of $\kappa_{i,n}(\psi;u_n)$.

$$
\lim_{k\to\infty}\sup_{n}\sup_{i\in D_n} E\left[|\kappa_{i,n}(\psi;u_n)|\,1_{(|\epsilon_{i,n}(\psi;u_n)|>k)}\right]=0
$$

Since the upper bond of the $\alpha-$mixing coefficient of $\{\kappa_{i,n}(\psi)\}_{i\in D_n}$ is the same as $\{v_{i,n}(\psi)\}_{i\in D_n}$, the inequality (20) and (21) also hold for $\{\kappa_{i,n}(\psi)\}_{i\in D_n}$. Thus, by Theorem 3 in Jenish and Prucha (2009),

$$
q_{3,n}(\psi)-Eq_{3,n}(\psi)=\frac{1}{n}\sum_{j=1}^{n}\left(\kappa_{i,n}(\psi;u_n)-E\kappa_{i,n}(\psi;u_n)\right)\xrightarrow{L_1}0
$$

Thus, we have the point-wise convergence for $q_{3,n}(\psi)$.

For $\forall\psi,\psi^{'}\in\Theta$, let the maximum value of $\exp\left\{-\frac{\alpha}{1-\theta}\right\}$ be M, then

$$
\begin{aligned}
&\left|q_{3,n}(\psi)-q_{3,n}(\psi^{'})\right|\\
=&\frac{1}{n}\left|\sum_{i=1}^{n}\kappa_{i,n}(\psi;u_n)-\sum_{i=1}^{n}\kappa_{i,n}(\psi^{'};u_n)\right|\\
\leqslant&\frac{1}{n}\sum_{i=1}^{n}\left|\kappa_{i,n}(\psi;u_n)-\kappa_{i,n}(\psi^{'};u_n)\right|
\end{aligned}
$$

Since we have

$$
\begin{aligned}
&\left|\kappa_{i,n}(\psi;u_n)-\kappa_{i,n}(\psi^{'};u_n)\right| \\
=&\left|u_{i,n}^2\left(\prod_{j=1}^{n}|u_{j,n}|^{-2\left(\phi(I_n-\theta W_n)^{-1}W_n\right)_{ij}}-\prod_{j=1}^{n}|u_{j,n}|^{-2\left(\phi^{'}\left(I_n-\theta^{'}W_n\right)^{-1}W_n\right)_{ij}}\right)\right| \\
=&\left|u_{i,n}^2\prod_{j=1}^{n}|u_{j,n}|^{-2\left(\phi(I_n-\theta W_n)^{-1}W_n\right)_{ij}}\left(1-\prod_{j=1}^{n}|u_{j,n}|^{2\left(\phi(I_n-\theta W_n)^{-1}W_n\right)_{ij}-2\left(\phi^{'}\left(I_n-\theta^{'}W_n\right)^{-1}W_n\right)_{ij}}\right)\right| \\
=&\left|\kappa_{i,n}(\psi;u_n)\left(1-\prod_{j=1}^{n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\left(\theta^{'}\right)^l\right)\left(W_n^l\right)_{ij}}\right)\right| \\
\leqslant&\left|\left(1-\prod_{j=1}^{n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\left(\theta^{'}\right)^l\right)\left(W_n^l\right)_{ij}}\right)\right||\kappa_{i,n}(\psi;u_n)|
\end{aligned}
$$

WLOG assume $\phi^{'}<\phi$ and $\theta^{'}<\theta$, $|u_{j,n}|\geqslant 1$ for $\forall n\leqslant m_n$, $|u_{j,n}|<1$ for $\forall n>m_n$ with $m_n\in[0,n]$, we have

$$
\prod_{j=m_n+1}^{n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\theta^l\right)\left(W_n^l\right)_{ij}}\leqslant\prod_{j=1}^{n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\theta^l\right)\left(W_n^l\right)_{ij}}\leqslant\prod_{j=1}^{m_n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\theta^l\right)\left(W_n^l\right)_{ij}}
$$

The LHS and RHS will goes to 1 as $\psi^{'}\to\psi$, then

$$
\lim_{\psi^{'}\to\psi}\left(1-\prod_{j=1}^{n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\theta^l\right)\left(W_n^l\right)_{ij}}\right)=0
$$

For different situations, we can get the same conclusion by constructing upper and lower bonds in similar ways. Also, similar to the arguments for $\log u_{i,n}$ and $z_{i,n}$, from uniform integrability of $\kappa_{i,n}$, we can get $\frac{1}{n}\sum_{i=1}^{n}E\left|\kappa_{i,n}(\psi;u_n)\right|$. Then, we have

$$
\begin{aligned}
&\left|q_{3,n}(\psi)-q_{3,n}(\psi^{'})\right| \\
\leqslant&\frac{1}{n}\sum_{i=1}^{n}|\kappa_{i,n}(\psi;u_n)|\left|1-\prod_{j=1}^{n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\left(\theta^{'}\right)^l\right)\left(W_n^l\right)_{ij}}\right| \\
\leqslant&\left(\frac{1}{n}\sum_{i=1}^{n}|\kappa_{i,n}(\psi;u_n)|\right)\left(\sup_{i\in D_n}\left|1-\prod_{j=1}^{n}|u_{j,n}|^{2\sum_{l=0}^{\infty}\left(\phi\theta^l-\phi^{'}\left(\theta^{'}\right)^l\right)\left(W_n^l\right)_{ij}}\right|\right) \\
\equiv&B_2\left(u_n\right)h_2\left(\left|\psi-\psi^{'}\right|\right)
\end{aligned}
$$

where $EB_2\left(u_n\right)=\frac{1}{n}\sum_{i=1}^{n}E\left|\epsilon_{i,n}(\psi;u_n)\right|=O\left(1\right)$ uniformly in Θ and $\lim_{\psi-\psi^{'}\to 0}h_2\left(\left|\psi-\psi^{'}\right|\right)=0$. By Corollary 3.1, combined with the point-wise convergence and compact parameter space, we have

$Eq_{3,n}(\psi)$ is equicontinuous and uniform convergence of $q_{3,n}(\psi)$: $\sup_{\theta\in\Theta}|q_{3,n}(\psi)-Eq_{3,n}(\psi)| = o_p(1)$. Thus, we have the consistency:

Theorem 4: *Under Assumption 1-8, the MLE estimators are consistent.*

Proof:

From the arguments before, we have uniform convergence of $q_{2,n}(\psi)$ and $q_{3,n}(\psi)$ defined by (18) and (19), thus

$$\begin{aligned}\sup_{\psi\in\Theta}|q_n(\psi)-Eq_n(\psi)| &= \sup_{\psi\in\Theta}\left|-\frac{1}{2}(q_{2,n}(\psi)-Eq_{2,n}(\psi))-\frac{1}{2}(q_{3,n}(\psi)-Eq_{3,n}(\psi))\right| \\ &\leqslant \frac{1}{2}\left\{\sup_{\psi\in\Theta}|q_{2,n}(\psi)-Eq_{2,n}(\psi)|+\sup_{\psi\in\Theta}|q_{3,n}(\psi)-Eq_{3,n}(\psi)|\right\} \\ &= \frac{1}{2}o_p(1)+\frac{1}{2}o_p(1)=o_p(1)\end{aligned}$$

Combined with the identification, compact parameter space Θ and measurability, we have consistency for MLE estimators. □

Thus, we proved that consistency when does not allow direct correlations between individuals far away form each other, i.e. Assumption 3 holds.

3.6 Asymptotic Distribution of MLE Estimators for ARCH-like Model

In this section, we derive the asymptotic distribution of MLE estimator for the spatial ARCH-like model, which is the special case when $\theta_0 = 0$. For the general spatial GARCH-like case, there are some difficulties to apply LLN and CLT since both the first order derivatives and the second order derivatives would contain much more complex nonlinear terms which would be hard to discuss their $\alpha-$mixing property and uniform $L^{2+\delta}$ integrability for some positive δ. So, here we just try to handle an easier case when the FOC and SOC's can be represented as measurable functions of $\kappa_{i,n}$ and $v_{i,n}$, which we had proved their jointly $\alpha-$mixing properties and derived the upper bond of their mixing coefficients. There might be some other ways instead of using LLN and CLT for mixing processes to get the asymptotic, which might solve the difficulties for general GARCH-like case.

Recall the log-likelihood function for ARCH-like model:

$$\begin{aligned}\ln L_{ns}(u_n;\alpha,\phi) = &-\frac{n}{2}\ln(2\pi)-\frac{1}{2}e^{-\alpha}\sum_{i=1}^{n}u_{i,n}^2\left[\prod_{j=1}^{n}|u_{j,n}|^{-2w_{ij}}\right]^{\phi} \\ &-\frac{1}{2}[\phi 1_n' W_n log u_n^2+n\alpha]+\ln|I_n-\phi W_n|\end{aligned}$$

We can get the first order derivatives of parameters:

$$\frac{\partial \ln L_{ns}}{\partial\alpha} = \frac{1}{2}e^{-\alpha}\sum_{i=1}^{n}u_{i,n}^2\left[\prod_{j=1}^{n}|u_{j,n}|^{-2w_{ij}}\right]^{\phi}-\frac{n}{2}$$

$$\frac{\partial \ln L_{ns}}{\partial \phi} = -\frac{1}{2}e^{-\alpha}\sum_{i=1}^{n} u_{i,n}^2 \left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]^{\phi} \ln\left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]$$
$$-\frac{1}{2}1_n' W_n log u_n^2 - tr\left((I_n - \phi W_n)^{-1} W_n\right)$$

From previous proof for consistency, we defined two new variables: $v_{i,n}$ and $\epsilon_{i,n}$. In the simplified ARCH-like model, as $\theta \equiv 0$, we can write them as the following:

$$v_{i,n}(u_n) = \left(W_n log u_n^2\right)_{i,n} = \sum_{j=1}^{n} w_{ij} \ln u_{j,n}^2 = \ln\left[\prod_{j=1}^{n} |u_{j,n}|^{2w_{ij}}\right]$$

$$\kappa_{i,n}(\psi; u_n) = e^{-\alpha} u_{i,n}^2 \left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]^{\phi}$$

Then, we can rewrite the first as the following:

$$\frac{\partial \ln L_{ns}}{\partial \alpha} = \frac{1}{2}\sum_{i=1}^{n} \kappa_{i,n} - \frac{n}{2}$$

$$\frac{\partial \ln L_{ns}}{\partial \phi} = \frac{1}{2}\sum_{i=1}^{n} \kappa_{i,n} v_{i,n} - \frac{1}{2}\sum_{i=1}^{n} v_{i,n} - tr\left((I_n - \phi W_n)^{-1} W_n\right)$$

By the results before, we have

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_{ns}(\psi_0)}{\partial \psi} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \frac{1}{2}\kappa_{i,n}(\psi_0) - \frac{1}{2} \\ -\frac{1}{2}\kappa_{i,n}(\psi_0) v_{i,n} - \frac{1}{2}v_{i,n} - \frac{1}{n}tr\left((I_n - \phi W_n)^{-1} W_n\right)\end{pmatrix}$$

To get asymptotic distribution, we need the following two assumptions:
Assumption 8: *ψ_0 is in the interior of the parameter space Θ.*
Assumption 9:
$\Sigma_0 = \lim_{n\to\infty} \Sigma_n$ *exists and is nonsingular, where*

$$\Sigma_n = \frac{1}{n}Var\left(\sum_{i=1}^{n}\left(\frac{1}{2}\kappa_{i,n}(\psi_0) - \frac{1}{2}, -\frac{1}{2}\kappa_{i,n}(\psi_0) v_{i,n} - \frac{1}{2}v_{i,n} - \frac{1}{n}tr\left((I_n - \phi W_n)^{-1} W_n\right)\right)'\right)$$

Then we have the following proposition:
Proposition 1:
Under Assumption 1-9,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \frac{1}{2}\kappa_{i,n}(\psi_0) - \frac{1}{2} \\ -\frac{1}{2}\kappa_{i,n}(\psi_0) v_{i,n} - \frac{1}{2}v_{i,n} - \frac{1}{n}tr\left((I_n - \phi_0 W_n)^{-1} W_n\right)\end{pmatrix} \xrightarrow{d} N(0, \Sigma_0)$$

Proof:

The asymptotic mean zero and variance matrix $\sum_0$ are followed by the consistency and Assumption 9. The key point here is to prove the joint normality. By definition of multivariate Normal distribution, random $(X,Y)^{'}$ is multivariate Normal iff $\forall a,b\in\mathbb{R}$, $aX+bY$ is a Normal random variable. Thus, from the following, we are focusing on prove the asymptotic Normality of the following linear combination:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{a\left[\frac{1}{2}\kappa_{i,n}\left(\psi_0\right)-\frac{1}{2}\right]+b\left[-\frac{1}{2}\kappa_{i,n}\left(\psi_0\right)v_{i,n}-\frac{1}{2}v_{i,n}-\frac{1}{n}tr\left(\left(I_n-\phi_0W_n\right)^{-1}W_n\right)\right]\right\},\forall a,b\in\mathbb{R}$$

As we proved before, $\{v_{i,n}\}_{i\in D_n}$ and $\{\kappa_{i,n}\}_{i\in D_n}$ are jointly $\alpha-$mixing with their mixing coefficient satisfy inequality (16). Thus any measurable function of $\{v_{i,n},\kappa_{i,n}\}$ are have $\alpha-$mixing coefficients with the same upper bond , so does $\left\{a\left[\frac{1}{2}\kappa_{i,n}\left(\psi_0\right)-\frac{1}{2}\right]+b\left[\kappa_{i,n}\left(\psi_0\right)v_{i,n}-v_{i,n}-\frac{1}{n}tr\left(\left(I_n-\phi_0W_n\right)^{-1}W_n\right)\right]\right\}_{i\in D_n}$ for $\forall a,b\in\mathbb{R}$.

Notice that

$$\frac{1}{n}tr\left(\left(I_n-\phi_0W_n\right)^{-1}W_n\right)=\frac{1}{n}tr\left(\sum_{l=0}^{\infty}\phi_0^lW_n^{l+1}\right)=\frac{1}{n}\sum_{l=0}^{\infty}\phi_0^ltr\left(W_n^{l+1}\right)\leqslant\sum_{l=0}^{\infty}\phi_0^l=\frac{1}{1-\phi_0}$$

Then, we have

$$\begin{aligned}
&\left|a\left[\frac{1}{2}\kappa_{i,n}\left(\psi_0\right)-\frac{1}{2}\right]+b\left[-\frac{1}{2}\kappa_{i,n}\left(\psi_0\right)v_{i,n}-\frac{1}{2}v_{i,n}-\frac{1}{n}tr\left(\left(I_n-\phi_0W_n\right)^{-1}W_n\right)\right]\right|\\
=&\left|\frac{1}{2}a\kappa_{i,n}\left(\psi_0\right)-\frac{1}{2}bv_{i,n}-\frac{1}{2}b\kappa_{i,n}\left(\psi_0\right)v_{i,n}-\frac{1}{2}a-\frac{b}{n}tr\left(\left(I_n-\phi_0W_n\right)^{-1}W_n\right)\right|\\
\leqslant&\left|\frac{1}{2}a\kappa_{i,n}\left(\psi_0\right)\right|+\left|\frac{1}{2}bv_{i,n}\right|+\left|\frac{1}{2}b\kappa_{i,n}\left(\psi_0\right)v_{i,n}\right|+\frac{1}{2}\left|a\right|+\frac{|b|}{1-\phi_0}\\
\equiv&\widetilde{A}\left|\kappa_{i,n}\left(\psi_0\right)\right|+\widetilde{B}\left|v_{i,n}\right|+\widetilde{C}\left|\kappa_{i,n}\left(\psi_0\right)\right|\left|v_{i,n}\right|+\widetilde{D}
\end{aligned}$$

where $\widetilde{A}$, $\widetilde{B}$,$\widetilde{C}$ and $\widetilde{D}$ are finite positive constants depend on ψ_0, a and b.

Then, we have

$$
\begin{aligned}
&\left|a\left[\frac{1}{2}\kappa_{i,n}(\psi_0)-\frac{1}{2}\right]+b\left[\frac{1}{2}\kappa_{i,n}(\psi_0)v_{i,n}-\frac{1}{2}v_{i,n}-\frac{1}{n}tr\left((I_n-\phi_0W_n)^{-1}W_n\right)\right]\right|^3\\
\leqslant&\left\{\widetilde{A}\left|\kappa_{i,n}(\psi_0)\right|+\widetilde{B}\left|v_{i,n}\right|+\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+\widetilde{D}\right\}^3\\
=&\left\{\widetilde{A}\left|\kappa_{i,n}(\psi_0)\right|+\widetilde{B}\left|v_{i,n}\right|\right\}^3+\left\{\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+\widetilde{D}\right\}^3\\
&+3\left\{\widetilde{A}\left|\kappa_{i,n}(\psi_0)\right|+\widetilde{B}\left|v_{i,n}\right|\right\}^2\left\{\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+\widetilde{D}\right\}\\
&+3\left\{\widetilde{A}\left|\kappa_{i,n}(\psi_0)\right|+\widetilde{B}\left|v_{i,n}\right|\right\}\left\{\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+\widetilde{D}\right\}^2\\
=&\widetilde{A}^3\left|\kappa_{i,n}(\psi_0)\right|^3+\widetilde{B}^3\left|v_{i,n}\right|^3+3\widetilde{A}^2\widetilde{B}\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|+3\widetilde{A}\widetilde{B}^2\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|^2\\
&+\widetilde{C}^3\left|\kappa_{i,n}(\psi_0)\right|^3\left|v_{i,n}\right|^3+\widetilde{D}^3+3\widetilde{C}^2\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|^2+3\widetilde{C}\widetilde{D}^2\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|\\
&+3\left\{\widetilde{A}^2\left|\kappa_{i,n}(\psi_0)\right|^2+2\widetilde{A}\widetilde{B}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+\widetilde{B}^2\left|v_{i,n}\right|^2\right\}\left\{\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+\widetilde{D}\right\}\\
&+3\left\{\widetilde{A}\left|\kappa_{i,n}(\psi_0)\right|+\widetilde{B}\left|v_{i,n}\right|\right\}\left\{\widetilde{C}^2\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|^2+2\widetilde{C}\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+\widetilde{D}^2\right\}\\
=&\widetilde{A}^3\left|\kappa_{i,n}(\psi_0)\right|^3+\widetilde{B}^3\left|v_{i,n}\right|^3+3\widetilde{A}^2\widetilde{B}\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|+3\widetilde{A}\widetilde{B}^2\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|^2\\
&+\widetilde{C}^3\left|\kappa_{i,n}(\psi_0)\right|^3\left|v_{i,n}\right|^3+\widetilde{D}^3+3\widetilde{C}^2\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|^2+3\widetilde{C}\widetilde{D}^2\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|\\
&+3\widetilde{A}^2\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|^3\left|v_{i,n}\right|+6\widetilde{A}\widetilde{B}\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|^2+3\widetilde{B}^2\widetilde{C}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|^2\\
&+3\widetilde{A}^2\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|^2+6\widetilde{A}\widetilde{B}\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|+3\widetilde{B}^2\widetilde{D}\left|v_{i,n}\right|^2\\
&+3\widetilde{A}\widetilde{C}^2\left|\kappa_{i,n}(\psi_0)\right|^3\left|v_{i,n}\right|^2+6\widetilde{A}\widetilde{C}\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|+3\widetilde{A}\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|\\
&+3\widetilde{B}\widetilde{C}^2\left|\kappa_{i,n}(\psi_0)\right|^2\left|v_{i,n}\right|^3+6\widetilde{B}\widetilde{C}\widetilde{D}\left|\kappa_{i,n}(\psi_0)\right|\left|v_{i,n}\right|^2+3\widetilde{B}\widetilde{D}^2\left|v_{i,n}\right|
\end{aligned}
\tag{22}
$$

Recall that $|v_{i,n}|\leqslant\sup_{j\in D_n}\left|\log u_{j,n}^2\right|\leqslant\sup_{j\in D_n}C_6(\psi_0)\left|\log\varepsilon_{i,n}^2\right|+\left|\frac{\alpha_0}{1-\phi_0}\right|$ and $\kappa_{i,n}(\psi_0)=\varepsilon_{i,n}^2$, hence we have $\sup_{\Theta}\sup_n\sup_{i\in D_n}E\left|v_{i,n}\right|^k<\infty$ and $\sup_n\sup_{i\in D_n}E\kappa_{i,n}^k(\psi_0)<\infty$ for $\forall k>0$. Notice that for any positive k and l

$$
\begin{aligned}
&E\left|\kappa_{i,n}(\psi_0)\right|^k\left|v_{i,n}\right|^l\\
=&\left|Cov\left(\left|\kappa_{i,n}(\psi_0)\right|^k,\left|v_{i,n}\right|^l\right)+E\left|\kappa_{i,n}(\psi_0)\right|^kE\left|v_{i,n}\right|^l\right|\\
\leqslant&\left|Cov\left(\left|\kappa_{i,n}(\psi_0)\right|^k,\left|v_{i,n}\right|^l\right)\right|+E\left|\kappa_{i,n}(\psi_0)\right|^kE\left|v_{i,n}\right|^l\\
\leqslant&\sqrt{Var\left(\left|\kappa_{i,n}(\psi_0)\right|^k\right)Var\left(\left|v_{i,n}\right|^l\right)}+E\left|\kappa_{i,n}(\psi_0)\right|^kE\left|v_{i,n}\right|^l\\
\leqslant&\sqrt{E\left|\kappa_{i,n}(\psi_0)\right|^{2k}E\left|v_{i,n}\right|^{2l}}+E\left|\kappa_{i,n}(\psi_0)\right|^kE\left|v_{i,n}\right|^l
\end{aligned}
$$

Thus, combined all the results before, denote

$$
\omega_{i,n}(a,b)=a\left[\frac{1}{2}\kappa_{i,n}(\psi_0)-\frac{1}{2}\right]+b\left[\kappa_{i,n}(\psi_0)v_{i,n}-\frac{1}{2}v_{i,n}-\frac{1}{n}tr\left((I_n-\phi_0W_n)^{-1}W_n\right)\right]
$$

we have $\sup_n\sup_{i\in D_n}E\left|\omega_{i,n}\right|^3<\infty$ since each of the terms in (1) is finite. Thus, we have $L^{2+\delta}$

uniform integrability for $\omega_{i,n}(a,b)$, i.e.

$$\lim_{k\to\infty}\sup_n \sup_{i\in D_n} E\left[\left|\omega_{i,n}(\psi;u_n)\right|^{2+\delta} 1_{(|\omega_{i,n}(\psi;u_n)|>k)}\right]=0$$

for $\forall 0<\delta<1$ and $\forall a,b>0$.

The next thing is to check the inequalities of $\alpha-$mixing coefficient. Define $\bar{\alpha}_{k,l}(r)=\sup_n \alpha_{k,l,n}(r)$ following Definition 1 in Jenish and Prucha (2009). Then, we have the following results hold:

$$\begin{aligned}
&\sum_{m=1}^{\infty}\bar{\alpha}_{1,1}(m)\,m^{[d(2+\delta)/\delta]-1}\\
=&\sum_{m=1}^{\lfloor 9\bar{d}_0\rfloor}\bar{\alpha}_{1,1}(m)\,m^{[2(2+\delta)/\delta]-1}+\sum_{m=\lfloor 9\bar{d}_0\rfloor+1}^{\infty}\bar{\alpha}_{1,1}(m)\,m^{[2(2+\delta)/\delta]-1}\\
\leqslant&\sum_{m=1}^{\lfloor 9\bar{d}_0\rfloor}m^{[2(2+\delta)/\delta]-1}+\sum_{m=\lfloor 9\bar{d}_0\rfloor+1}^{\infty}C_d 3^{-\frac{7}{3}}m^{\frac{13}{6}}\xi^{r/3\bar{d}_0}m^{[2(2+\delta)/\delta]-1}<\infty
\end{aligned}$$

$$\begin{aligned}
&\sum_{m=1}^{\infty}m^{d-1}\bar{\alpha}_{k,l}(m)\\
\leqslant&\sum_{m=1}^{\lfloor 9\bar{d}_0\rfloor}m\bar{\alpha}_{k,l}(m)+\sum_{m=\lfloor 9\bar{d}_0\rfloor+1}^{\infty}m\bar{\alpha}_{k,l}(m)\\
\leqslant&\sum_{m=1}^{\lfloor 9\bar{d}_0\rfloor}m+\sum_{m=\lfloor 9\bar{d}_0\rfloor+1}^{\infty}\min\{k,l\}\,C_d 3^{-\frac{7}{3}}m^{\frac{19}{6}}\xi^{m/3\bar{d}_0}<\infty
\end{aligned}$$

$$\bar{\alpha}_{1,\infty}(m)=C_d 3^{-\frac{7}{3}}m^{\frac{19}{6}}\xi^{m/3\bar{d}_0}=O\left(m^{-2-\varepsilon}\right)$$

since $\lim_{m\to 0}\frac{m^{(d-1)/3}\xi^{m/\bar{d}_0}}{m^{-d-\varepsilon}}=\lim_{m\to 0}m^{4d/3-1/3+\varepsilon}\xi^{m/\bar{d}_0}=0$ by repeating L'Hospital Rule, where $k+l\leqslant 4$, $0<\delta<1$ and ε is some positive number. Thus, we can apply Corollary 1 and Theorem 1 in Jenish and Prucha (2009) to get the following CLT:

$$\sigma_n^{-1}\sum_{i=1}^{n}\left(\omega_{i,n}(a,b)-E\omega_{i,n}(a,b)\right)\xrightarrow{d}N(0,1)$$

where σ_n is the sample standard deviation of $\sum_{i=1}^{n}\left(\omega_{i,n}(a,b)-E\omega_{i,n}(a,b)\right)$. Thus, we proved the asymptotic Normality of the linear combination, then $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\frac{1}{2}\kappa_{i,n}(\psi_0)-\frac{1}{2}\\ \kappa_{i,n}(\psi_0)\,v_{i,n}(\psi_0)+\frac{1}{2}v_{i,n}-\frac{1}{n}tr\left((I_n-\phi_0W_n)^{-1}W_n\right)\end{pmatrix}$ is jointly Normal asymptotically. With Assumption 8 and 9, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\frac{1}{2}\kappa_{i,n}(\psi_0)-\frac{1}{2}\\ \kappa_{i,n}(\psi_0)\,v_{i,n}(\psi_0)-\frac{1}{2}v_{i,n}-\frac{1}{n}tr\left((I_n-\phi_0W_n)^{-1}W_n\right)\end{pmatrix}\xrightarrow{d}N(0,\Sigma_0)\quad\square$$

After the first order derivative, we have the following second order derivative:

$$\frac{\partial^2 \ln L_{ns}}{\partial \alpha^2} = -\frac{1}{2}e^{-\alpha}\sum_{i=1}^{n} u_{i,n}^2 \left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]^{\phi}$$

$$\frac{\partial^2 \ln L_{ns}}{\partial \phi \partial \alpha} = \frac{\partial^2 \ln L_{ns}}{\partial \alpha \partial \phi} = \frac{1}{2}e^{-\alpha}\sum_{i=1}^{n} u_{i,n}^2 \left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]^{\phi} \ln\left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]$$

$$\begin{aligned}\frac{\partial^2 \ln L_{ns}}{\partial \phi^2} &= -\frac{1}{2}e^{-\alpha}\sum_{i=1}^{n} u_{i,n}^2 \left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]^{\phi} \left\{\ln\left[\prod_{j=1}^{n} |u_{j,n}|^{-2w_{ij}}\right]\right\}^2 \\ &\quad - tr\left[(I_n - \phi W_n)^{-1} W_n\right]^2\end{aligned}$$

since

$$\begin{aligned}\frac{\partial (I_n - \phi W_n)^{-1} W_n}{\partial \theta} &= -\left(I_n - \phi W_n\right)^{-1} \frac{\partial (I_n - \phi W_n)}{\partial \theta}(I_n - \phi W_n)^{-1} W_n \\ &= \left(I_n - \phi W_n\right)^{-1} W_n (I_n - \phi W_n)^{-1} W_n \\ &= \left[\left(I_n - \phi W_n\right)^{-1} W_n\right]^2\end{aligned}$$

Similarly, we can use the variables $v_{i,n}$ and $\epsilon_{i,n}$ to represent them:

$$\frac{\partial^2 \ln L_{ns}}{\partial \alpha^2} = -\frac{1}{2}\sum_{i=1}^{n} \kappa_{i,n}$$

$$\frac{\partial^2 \ln L_{ns}}{\partial \phi \partial \alpha} = \frac{\partial^2 \ln L_{n}}{\partial \alpha \partial \phi} = -\frac{1}{2}\sum_{i=1}^{n} \kappa_{i,n} v_{i,n}$$

$$\frac{\partial^2 \ln L_{ns}}{\partial \phi^2} = -\frac{1}{2}\sum_{i=1}^{n} \kappa_{i,n} v_{i,n}^2 - tr\left[\left(I_n - \phi W_n\right)^{-1} W_n\right]^2$$

To show asymptotic Normality, we need to prove $\frac{1}{n}\left|\frac{\partial^2 \ln L_{ns}(\hat{\psi}_n)}{\partial \psi \partial \psi'} - E\frac{\partial^2 \ln L_{ns}(\psi_0)}{\partial \psi \partial \psi'}\right| \xrightarrow{p} 0$ where $\frac{\partial^2 \ln L_{ns}(\psi)}{\partial \psi \partial \psi'} = \begin{bmatrix} \frac{\partial^2 \ln L_{ns}}{\partial \alpha^2} & \frac{\partial^2 \ln L_{ns}}{\partial \phi \partial \alpha} \\ \frac{\partial^2 \ln L_{ns}}{\partial \phi \partial \alpha} & \frac{\partial^2 \ln L_{ns}}{\partial \phi^2} \end{bmatrix}$. To show it, we will show $\frac{1}{n}\left|\frac{\partial^2 \ln L_{ns}(\psi_0)}{\partial \psi \partial \psi'} - E\frac{\partial^2 \ln L_{ns}(\psi_0)}{\partial \psi \partial \psi'}\right| \xrightarrow{p} 0$ and $\frac{1}{n}\left|\frac{\partial^2 \ln L_{ns}(\hat{\psi}_n)}{\partial \psi \partial \psi'} - \frac{\partial^2 \ln L_{ns}(\psi_0)}{\partial \psi \partial \psi'}\right| \xrightarrow{p} 0$.

To show $\frac{1}{n}\left|\frac{\partial^2 \ln L_{ns}(\psi_0)}{\partial \psi \partial \psi'} - E\frac{\partial^2 \ln L_{ns}(\psi_0)}{\partial \psi \partial \psi'}\right| \xrightarrow{p} 0$, it is sufficient to show the following three results:

$$\frac{1}{n}\left|\sum_{i=1}^{n} \kappa_{i,n}(\psi_0) - \sum_{i=1}^{n} E\kappa_{i,n}(\psi_0)\right| \xrightarrow{p} 0$$

$$\frac{1}{n}\left|\sum_{i=1}^{n}\kappa_{i,n}\left(\psi_0\right)v_{i,n}-\sum_{i=1}^{n}E\kappa_{i,n}\left(\psi_0\right)v_{i,n}\right|\xrightarrow{p}0$$
$$\frac{1}{n}\left|\sum_{i=1}^{n}\kappa_{i,n}\left(\psi_0\right)v_{i,n}^2-\sum_{i=1}^{n}E\kappa_{i,n}\left(\psi_0\right)v_{i,n}^2\right|\xrightarrow{p}0$$

From our proof for Proposition 1, we show that any measurable function of $v_{i,n}$ and $\kappa_{i,n}$ are $\alpha-$mixing with mixing coefficient no larger than $C_3 min(k,l)(r^{d-1}\xi^{r/\bar{d}_0})^{1/3}$ under Assumption 3. Also, we show that $\sup_{\Theta}\sup_n\sup_{i\in D_n}E\left|\kappa_{i,n}\right|^k\left|v_{i,n}\right|^l<\infty$ for any positive integer k and l. Thus, from weak LLN in Jenish and Prucha (2009), we have the convergence in probability of the above three sequences. Thus, $\frac{1}{n}\left|\frac{\partial^2\ln L_{ns}(\psi_0)}{\partial\psi\partial\psi'}-E\frac{\partial^2\ln L_{ns}(\psi_0)}{\partial\psi\partial\psi'}\right|\xrightarrow{p}0$ follows.

Next, we need to show $\frac{1}{n}\left|\frac{\partial^2\ln L_{ns}(\hat{\psi}_n)}{\partial\psi\partial\psi'}-\frac{\partial^2\ln L_{ns}(\psi_0)}{\partial\psi\partial\psi'}\right|\xrightarrow{p}0$. First, as $\hat{\psi}_n\xrightarrow{p}\psi_0$, for the term $tr\left[\left(I_n-\phi W_n\right)^{-1}W_n\right]^2$, the only thing we need to check $\frac{d}{d\phi}\frac{1}{n}tr\left[\left(I_n-\phi W_n\right)^{-1}W_n\right]^2=\frac{2}{n}tr\left[\left(I_n-\phi W_n\right)^{-1}W_n\right]^3$ is bounded. A sufficient condition is that $\left(\left[\left(I_n-\phi W_n\right)^{-1}W_n\right]^3\right)_{ii}$ is uniformly bounded. This is obvious since $\left(I_n-\phi W_n\right)^{-1}W_n$ is a uniformly bounded matrix for all $\psi\in\Theta$ as we assumed. The remaining thing is to show convergence for the following terms: $\frac{1}{n}\sum_{i=1}^{n}\kappa_{i,n}$, $\frac{1}{n}\sum_{i=1}^{n}\kappa_{i.n}v_{i,n}$ and $\frac{1}{n}\sum_{i=1}^{n}\kappa_{i.n}v_{i,n}^2$. By taking first order derivatives, we have

$$\begin{aligned}
&\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}\\
=&\frac{\partial}{\partial\psi}\sum_{i=1}^{n}e^{-\alpha}u_{i,n}^2\left[\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2w_{ij}}\right]^{\phi}\\
=&\left(-\sum_{i=1}^{n}e^{-\alpha}u_{i,n}^2\left[\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2w_{ij}}\right]^{\phi},\sum_{i=1}^{n}e^{-\alpha}u_{i,n}^2\left[\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2w_{ij}}\right]^{\phi}\ln\left[\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2w_{ij}}\right]\right)'\\
=&\left(-\sum_{i=1}^{n}\kappa_{i,n},-\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}\right)'
\end{aligned}$$

Since $v_{i,n}$ does not depend on ψ, we can easily get $\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}=\left(-\sum_{i=1}^{n}\kappa_{i,n}v_{i,n},-\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}^2\right)'$ and $\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}^2=\left(-\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}^2,-\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}^3\right)'$. As we proved $\sup_{\Theta}\sup_n\sup_{i\in D_n}E\left|\kappa_{i,n}\right|^k\left|v_{i,n}\right|^l<\infty$, $\frac{1}{n}\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}$, $\frac{1}{n}\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}$ and $\frac{1}{n}\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}v_{i,n}^2$ are all $O_p\left(1\right)$ for $\forall\psi\in\Theta$. Then by Taylor expansion, as $\hat{\psi}_n\xrightarrow{p}\psi_0$:

$$\frac{1}{n}\left(\sum_{i=1}^{n}\kappa_{i,n}\left(\hat{\psi}_n\right)-\sum_{i=1}^{n}\kappa_{i,n}\left(\psi_0\right)\right)=\left[\frac{1}{n}\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}\left(\psi'\right)\right]'\left(\hat{\psi}_n-\psi_0\right)\xrightarrow{p}0$$

$$\frac{1}{n}\left(\sum_{i=1}^{n}\kappa_{i,n}\left(\hat{\psi}_n\right)v_{i,n}-\sum_{i=1}^{n}\kappa_{i,n}\left(\psi_0\right)v_{i,n}\right)=\left[\frac{1}{n}\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}\left(\psi''\right)v_{i,n}\right]'\left(\hat{\psi}_n-\psi_0\right)\xrightarrow{p}0$$

$$\frac{1}{n}\left(\sum_{i=1}^{n}\kappa_{i,n}\left(\hat{\psi}_n\right)v_{i,n}^2-\sum_{i=1}^{n}\kappa_{i,n}\left(\psi_0\right)v_{i,n}^2\right)=\left[\frac{1}{n}\frac{\partial}{\partial\psi}\sum_{i=1}^{n}\kappa_{i,n}\left(\psi^{''}\right)v_{i,n}^2\right]^{'}\left(\hat{\psi}_n-\psi_0\right)\xrightarrow{p}0$$

where $\psi^{'}$, $\psi^{''}$ and $\psi^{'''}$ are between ψ_0 and $\hat{\psi}_n$. Thus, we have $\frac{1}{n}\left|\frac{\partial^2\ln L_{ns}\left(\hat{\psi}_n\right)}{\partial\psi\partial\psi'}-\frac{\partial^2\ln L_{ns}(\psi_0)}{\partial\psi\partial\psi'}\right|\xrightarrow{p}0$.
Then, we have $\frac{1}{n}\left|\frac{\partial^2\ln L_{ns}\left(\hat{\psi}_n\right)}{\partial\psi\partial\psi'}-E\frac{\partial^2\ln L_{ns}(\psi_0)}{\partial\psi\partial\psi'}\right|\xrightarrow{p}0$.

Theorem 6: *Under Assumption 1-8,* $\sqrt{n}\left(\hat{\psi}_n-\psi_0\right)\xrightarrow{d}N\left(0,\Sigma_0^{-1}\right)$.

Proof:

By Taylor expansion, we have

$$\frac{1}{\sqrt{n}}\frac{\partial\ln L_{ns}\left(\hat{\psi}_n\right)}{\partial\psi}=0=\frac{1}{\sqrt{n}}\frac{\partial\ln L_{ns}\left(\psi_0\right)}{\partial\psi}+\frac{1}{n}\frac{\partial^2\ln L_{ns}(\bar{\psi})}{\partial\psi\partial\psi'}\sqrt{n}\left(\hat{\psi}_n-\psi_0\right)$$

where $\bar{\psi}$ is between ψ_0 and $\hat{\psi}_n$.

As $\frac{1}{\sqrt{n}}\frac{\partial\ln L_{ns}(\psi_0)}{\partial\psi}\xrightarrow{d}N\left(0,\Sigma_0\right)$ and $\frac{1}{n}\left|\frac{\partial^2\ln L_{ns}\left(\hat{\psi}_n\right)}{\partial\psi\partial\psi'}-E\frac{\partial^2\ln L_{ns}(\psi_0)}{\partial\psi\partial\psi'}\right|\xrightarrow{p}0$, we have

$$\begin{aligned}\sqrt{n}\left(\hat{\psi}_n-\psi_0\right)&=\left(-\frac{1}{n}\frac{\partial^2\ln L_{ns}(\bar{\psi})}{\partial\psi\partial\psi'}\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial\ln L_{ns}\left(\psi_0\right)}{\partial\psi}\\&\xrightarrow{d}N\left(0,\Sigma_0^{-1}\right)\end{aligned}$$

which follows the equality $E_\psi\left(\frac{\partial^2\ln L_{ns}}{\partial\psi\partial\psi'}\right)+E_\psi\left(\frac{\partial\ln L_{ns}}{\partial\psi}\frac{\partial\ln L_{ns}}{\partial\psi'}\right)=0$. □

4 Monte Carlo Simulation for MLE

4.1 Finite Sample performance with $\mathbf{N}\left(\mathbf{0},\mathbf{1}\right)$ Disturbances

In this section, we perform simulations to study the finite sample performance of the MLE and test its robustness when $\varepsilon_{i,n}\overset{iid}{\sim}N\left(0,1\right)$. One thing should be noticed that $N(0,1)$ does not satisfy the Assumption 8 which indicates existence of negative integer moments. The reciprocal distribution of $N(0,1)$ is bimodal which does have finite positive integer moments. However, for finite sample, as long as we do not have have any observation $u_{i,n}=0$, from equation (3) and (4), there should also have no $\varepsilon_{i,n}=0$. Then, we can pick up a small positive number $\underline{\varepsilon}$which is smaller than the minimum of $\left|\varepsilon_{i,n}\right|$, we can always regard our sample is sampled from a distribution without any density around zero. For example, for a sample from $N\left(0,1\right)$, we can view it as a sample from a distribution with the following density function:

$$f_X\left(x\right)=\begin{cases}\frac{1}{1-[\Phi(\underline{\varepsilon})-\Phi(-\underline{\varepsilon})]}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\} & |x|\geqslant\underline{\varepsilon}\\0 & |x|<\underline{\varepsilon}\end{cases}$$

The ratio of likelihood to each sample point remains the same as the $N\left(0,1\right)$, and this distribution remains symmetric, zero mean, and existence of any positive moments. Only two changes we need to pay attention to: first, since we do not any density in an open neighborhood of zero, this distribution satisfies Assumption 8; second, the variance of X is not 1 but still close to one. From Section 3.2, as we proved, when $\sigma^2 \neq 1$, the only change in the DGP is α, the parameter captures spatial heteroskedasticity and spill-over effect, ϕ and θ remains the same and can be correctly identified. Also, easy to see, $X \xrightarrow{d} N(0,1)$ as $\underline{\varepsilon} \to 0$. Thus, we can view the simulations with $N(0,1)$ as simulations with $\varepsilon_{i,n}$ with the density $f_X(x)$, which basically has same properties as Normal with slightly change round point zero. Another thing we need to notice is that for different samples we generated, the $\underline{\varepsilon}$'s are also different. Holding this point of view, we may not observe shrinking biases as sample size getting larger since the limiting distribution are not exactly the same for each sample. Nevertheless, if MLE works well, it should give small bias on ϕ and θ. For α, it may not be persistently close to α_0, since here the proper limiting distribution may not have $\sigma^2 = 1$. But it does not affect our identification of the DGP. For real empirical applications, since the empirical integer moments of $\hat{\varepsilon}_{i,n}$ would always exist when there is no zero observations, we can always assume the sampling distribution satisfy our Assumption 7 and 8.

In this section, we try two groups of true parameters with different values: $(\phi, \theta, \alpha) = (0.3, 0.4, 1)$ and $(\phi, \theta, \alpha) = (0.7, -0.3, -1)$. The estimation procedure is first maximizing the concentrated log-likelihood function to get $\hat{\phi}$ and $\hat{\theta}$, and then using the first order condition to back out $\hat{\alpha}$. The initial value is $(\phi_0, \theta_0) = (0.5, 0)$ for each round of simulation. For each simulation round, we generate data starting from $\varepsilon_{i,n} \overset{i.i.d}{\sim} N(0,1)$. We can obtain the mean and standard deviation of the estimators based on 1000 replications for each of the experiments.

The simulations use two different type of regions. In the first round, the regions are different US counties. The spatial weighting matrix is formed using US county adjacency file which is available at https://data.nber.org/data/county-adjacency.html. By randomly picking up n areas which is adjacent with at least one another area picked, for each county i, we first have

$$w_{ij} = \begin{cases} 0 & j = i \text{ or } j \text{ not adjacent with } i \\ 1 & j \text{ adjacent with } i \end{cases}$$

and then do row normalization. In the second round, it is a more unrealistic circular lake case where each region only has two neighbors. For region i, we have $w_{i,i+1} = w_{i,i-1} = \frac{1}{2}$ and others 0. Also, we have $w_{1,n} = w_{n,1} = \frac{1}{2}$. For each type of regions, we consider three different numbers of regions $n = 250$, $n = 750$ and $n = 1250$. One thing should be noticed: although in Section 3, we need uniformly bounded away from zero condition to make sure consistency, for any finite sample, since $\Pr(\varepsilon_{i,n}) = 0$, any sample without containing 0 in the observations automatically satisfies this condition. Thus, in this section, we do not put such condition for any simulation exercise.

From Table 1 and Table 2, we can see the bias are shrinking when $n$gets larger, and the standard deviations are getting smaller for all combination of weighting matrix and true parameters. The median, upper quantile and lower quantile of each estimator show the similar trend.

Table 1: GARCH-like with County Adjacency Matrix

n	true	$\phi = 0.3$	$\theta = 0.4$	$\alpha = 1$	$\phi = 0.7$	$\theta = -0.3$	$\alpha = -1$
250	mean	0.3114	0.3008	1.1241	0.6933	-0.3447	-1.1483
	std	0.1019	0.3029	0.6002	0.0886	0.2268	0.4729
	med	0.3079	0.3416	1.0969	0.6954	-0.3449	-1.0937
	$q_{0.25}$	0.2384	0.1080	0.7607	0.6356	-0.5155	-1.4433
	$q_{0.75}$	0.3821	0.5317	1.5688	0.7486	-0.1823	-0.8021
750	mean	0.3076	0.3540	1.0979	0.6952	-0.3138	-1.0578
	std	0.0630	0.1834	0.3588	0.0556	0.1471	0.2998
	med	0.3065	0.3778	1.0399	0.6944	-0.3105	-1.0392
	$q_{0.25}$	0.2636	0.2495	0.8553	0.6611	-0.4112	-1.2325
	$q_{0.75}$	0.3467	0.4798	1.2827	0.7340	-0.2149	-0.8532
1250	mean	0.3015	0.3830	1.0367	0.7007	-0.3184	-1.0467
	std	0.0497	0.1361	0.2582	0.0437	0.1147	0.2259
	med	0.3001	0.3930	1.0078	0.7020	-0.3160	-1.0266
	$q_{0.25}$	0.2656	0.2943	0.8556	0.6713	-0.3954	-1.1880
	$q_{0.75}$	0.3354	0.4783	1.1944	0.7321	-0.2371	-0.8888

Table 2: GARCH-like with Circular Lake Matrix

n	true	$\phi = 0.3$	$\theta = 0.4$	$\alpha = 1$	$\phi = 0.7$	$\theta = -0.3$	$\alpha = -1$
250	mean	0.3050	0.3610	1.0957	0.6935	-0.3103	-1.0385
	std	0.0650	0.1698	0.3491	0.0452	0.1001	0.2489
	med	0.3038	0.3678	1.0582	0.6934	-0.3154	-1.0385
	$q_{0.25}$	0.2632	0.2524	0.8574	0.6630	-0.3789	-1.2107
	$q_{0.75}$	0.3493	0.4762	1.2997	0.7251	-0.2423	-0.8732
750	mean	0.3019	0.3876	1.0288	0.6986	-0.3037	-1.0181
	std	0.0417	0.1024	0.2034	0.0267	0.0529	0.1393
	med	0.3020	0.3868	1.0235	0.6988	-0.3014	-1.0125
	$q_{0.25}$	0.2732	0.3191	0.8861	0.6800	-0.3408	-1.1042
	$q_{0.75}$	0.3306	0.4574	1.1548	0.7157	-0.2680	-0.9202
1250	mean	0.3004	0.3935	1.0151	0.6986	-0.3023	-1.0117
	std	0.0317	0.0778	0.1521	0.0207	0.0421	0.1112
	med	0.3006	0.3933	1.0158	0.6986	-0.3034	-1.0052
	$q_{0.25}$	0.2787	0.3389	0.9108	0.6850	-0.3298	-1.0804
	$q_{0.75}$	0.3227	0.4465	1.1119	0.7119	-0.2737	-0.9341

In small sample, however, the performance of the GARCH-like model does not work well. Comparing to the following simplified version ARCH-like model without parameter θ:

$$
\begin{aligned}
u_{i,n} &= \sqrt{h_{i,n}}\varepsilon_{i,n} \\
logh_{i,n} &= \phi \sum_{j=1}^{n} w_{ij,n} logu_{j,n}^2 + \alpha
\end{aligned}
$$

In Table 1 to Table 4, we can see the estimators in ARCH-like model have much smaller biases than the GARCH-like model. When $n = 100$, the simulated biases of ARCH-like model are already less than 10% for the true parameters, and show good significance (Table 3 and Table 4). However, even when $n = 200$, the GARCH-type model does not have a good performance, especially for ϕ and α, and the standard deviations are also very large comparing to the true values.

Also, considering computational efficiency, estimating ARCH-type model is much faster. Running on the same computer (13.3 inch MacBook Pro 2018, i5-8259u, 16G RAM, 512SSD, MacOS Catalina 10.15.2) with Matlab 2019b, both using *parpool* for all the four CPU cores, and using *fminbnd* (golden section search and parabolic interpolation, single variable) or *fmincon* (internal KKT method, multiple variables) with default options to maximize the concentrated log-likelihood function. Since the GARCH-type model has one more parameter, it is expected to take longer time but not so much. However, by using the internal timer of Matlab, we can see clearly that the time for each round of simulation is much longer for GARCH-like model, at least 3 times longer. Moreover, as the sample size increases, the computational time increases much faster for the GARCH type model.

Table 3: Running Time (GARCH-like County)

n	true	$\phi = 0.3$	$\theta = 0.4$	$\alpha = 1$	$\phi = 0.7$	$\theta = -0.3$	$\alpha = -1$
50	mean	0.3335	0.0854	1.6771	0.6745	-0.4095	-1.4188
	std	0.1800	0.4451	0.9484	0.1877	0.3620	0.7931
	med	0.3260	00860	1.5993	0.6836	-0.4644	-1.3379
	$q_{0.25}$	0.1898	-0.2762	0.8593	0.5498	-0.6733	-2.0039
	$q_{0.75}$	0.4570	0.4560	2.4658	0.8095	-0.1902	-0.7972
	time		62.328s			57.980s	
100	mean	0.3240	0.2039	1.4284	0.6895	-0.3624	-1.2232
	std	0.1420	0.4039	0.8290	0.1318	0.2969	0.6247
	med	0.3153	0.2321	1.2909	0.6942	-0.3771	-1.1732
	$q_{0.25}$	0.2203	-0.1200	0.7497	0.6111	-0.5844	-1.6720
	$q_{0.75}$	0.4245	0.5417	2.0365	0.7788	-0.1731	-0.7464
	time		132.171s			124.714s	
150	mean	0.3203	0.2413	1.3422	0.6986	-0.3719	-1.1939
	std	0.1184	0.3540	0.7217	0.1113	0.2591	0.5475
	med	0.3154	0.2879	1.1961	0.6959	-0.3742	-1.1157
	$q_{0.25}$	0.2354	-0.0174	0.7871	0.6300	-0.5571	-1.5700
	$q_{0.75}$	0.4010	0.5137	1.8216	0.7705	-0.1959	-0.7776
	time		234.431s			218.823s	
200	mean	0.3201	0.2608	1.3001	0.6941	-0.3451	-1.1505
	std	0.1107	0.3284	0.6675	0.0996	0.2420	0.5150
	med	0.3148	0.3036	1.1686	0.6935	-0.3456	-1.0698
	$q_{0.25}$	0.2432	0.0347	0.7835	0.6322	-0.5252	-1.4896
	$q_{0.75}$	0.3953	0.5158	1.6839	0.7624	-0.1824	-0.7619
	time		377.395s			337.169s	

Table 4: Running Time (GARCH-like Circular Lake)

n	true	$\phi = 0.3$	$\theta = 0.4$	$\alpha = 1$	$\phi = 0.7$	$\theta = -0.3$	$\alpha = -1$
50	mean	0.3047	0.2373	1.4199	0.6689	-0.3560	-1.2646
	std	0.1371	0.3858	0.8865	0.1099	0.2069	0.5991
	med	0.3034	0.2486	1.2522	0.6683	-0.3893	-1.1953
	$q_{0.25}$	0.2096	-0.0803	0.7228	0.5944	-0.5073	-1.6556
	$q_{0.75}$	0.3967	0.5362	2.0601	0.7429	-0.2251	-0.8202
	time		63.308s			63.360s	
100	mean	0.3106	0.2961	1.2511	0.6859	-0.3346	-1.1346
	std	0.1012	0.2813	0.6265	0.0759	0.1528	0.4179
	med	0.3089	0.3102	1.1739	0.6835	-0.3491	-1.0865
	$q_{0.25}$	0.2416	0.1135	0.7807	0.6435	-0.4401	-1.3879
	$q_{0.75}$	0.3790	0.4963	1.6047	0.7372	-0.2381	-0.8435
	time		133.292s			136.759s	
150	mean	0.3042	0.3425	1.1424	0.6906	-0.3238	-1.0917
	std	0.0865	0.2345	0.4968	0.0627	0.1234	0.3518
	med	0.3089	0.3526	1.0687	0.6895	-0.3208	-1.0529
	$q_{0.25}$	0.2248	0.1913	0.7735	0.6484	-0.4154	-1.3096
	$q_{0.75}$	0.3628	0.5116	1.4312	0.7330	-0.2407	-0.8360
	time		238.069s			212.059s	
200	mean	0.3084	0.3519	1.1103	0.6944	-0.3232	-1.0763
	std	0.0787	0.2000	0.4102	0.0500	0.1079	0.2871
	med	0.3116	0.3585	1.0697	0.6949	-0.3231	-1.0546
	$q_{0.25}$	0.2548	0.2104	0.8035	0.6599	-0.4046	-1.2526
	$q_{0.75}$	0.3622	0.4982	1.3701	0.7290	-0.2535	-0.8742
	time		381.433s			333.561s	

Table 5: Running Time (ARCH-like County)

n	true	$\phi = 0.2$	$\alpha = 1$	$\phi = 0.5$	$\alpha = 0$	$\phi = 0.8$	$\alpha = -1$
50	mean	0.1655	0.9436	0.4440	-0.1475	0.7470	-1.5948
	std	0.1293	0.2224	0.1339	0.3936	0.0911	1.0189
	med	0.1613	0.9546	0.4592	-0.0851	0.7641	-1.3952
	$q_{0.25}$	0.0496	0.8016	0.3692	-0.3777	0.7045	-2.0865
	$q_{0.75}$	0.2581	1.0999	0.5399	0.1321	0.8126	-0.8699
	time	16.621s		16.260s		16.624s	
100	mean	0.1834	0.9767	0.4744	-0.0763	0.7794	-1.2264
	std	0.1013	0.1518	0.0885	0.2643	0.0498	0.5534
	med	0.1863	0.9891	0.4807	-0.0356	0.7868	-1.1541
	$q_{0.25}$	0.1108	0.8861	0.4206	-0.2203	0.7502	-1.5346
	$q_{0.75}$	0.2569	1.0785	0.5404	0.1040	0.8153	-0.8220
	time	19.897s		19.650s		20.038s	
150	mean	0.1879	0.9818	0.4812	-0.0541	0.7849	-1.1733
	std	0.0843	0.1173	0.0719	0.2084	0.0427	0.4761
	med	0.1901	0.9856	0.4868	-0.0319	0.7896	-1.1361
	$q_{0.25}$	0.1385	0.9055	0.4403	-0.1690	0.7617	-1.4277
	$q_{0.75}$	0.2397	1.0622	0.5321	0.0929	0.8149	-0.8358
	time	23.155s		23.082s		23.636s	
200	mean	0.1888	0.9851	0.4869	-0.0382	0.7885	-1.1313
	std	0.0792	0.1041	0.0628	0.1832	0.0372	0.4149
	med	0.1941	0.9934	0.4911	-0.0221	0.7921	-1.0907
	$q_{0.25}$	0.1379	0.9181	0.4465	-0.1526	0.7673	-1.3803
	$q_{0.75}$	0.2460	1.0595	0.5322	0.0867	0.8142	-0.8337
	time	27.686s		27.269s		28.184s	

Table 6: Running Time (ARCH-like Circular Lake)

n	true	$\phi = 0.2$	$\alpha = 1$	$\phi = 0.5$	$\alpha = 0$	$\phi = 0.8$	$\alpha = -1$
50	mean	0.1863	0.9540	0.4733	-0.0946	0.7745	-1.2784
	std	0.1019	0.2105	0.0876	0.2939	0.0573	0.6173
	med	0.1867	0.9585	0.4776	-0.0451	0.7830	-1.1686
	$q_{0.25}$	0.1090	0.8187	0.4204	-0.2431	0.7461	-1.5892
	$q_{0.75}$	0.2581	1.0957	0.5348	0.1068	0.8153	-0.8371
	time	16.287s		16.210s		16.554s	
100	mean	0.1913	0.9734	0.4880	-0.0379	0.7899	-1.1133
	std	0.0709	0.1477	0.0614	0.1858	0.0332	0.3496
	med	0.1918	0.9748	0.4925	-0.0264	0.7931	-1.0798
	$q_{0.25}$	0.1461	0.8791	0.4533	-0.1545	0.7703	-1.3262
	$q_{0.75}$	0.2354	1.0727	0.5284	0.0972	0.8143	-0.8701
	time	19.442s		19.561s		20.217s	
150	mean	0.1953	0.9837	0.4922	-0.0262	0.7933	-1.0718
	std	0.0613	0.1170	0.0490	0.1530	0.0267	0.2797
	med	0.1961	0.9895	0.4952	-0.0113	0.7962	-1.0473
	$q_{0.25}$	0.1549	0.9046	0.4619	-0.1181	0.7778	-1.2359
	$q_{0.75}$	0.2367	1.0646	0.5257	0.0807	0.8113	-0.8748
	time	23.911s		22.987s		24.471s	
200	mean	0.1956	0.9875	0.4943	-0.0230	0.7953	-1.0455
	std	0.0518	0.0952	0.0430	0.1345	0.0218	0.2360
	med	0.1961	0.9874	0.4938	-0.0099	0.7959	-1.0279
	$q_{0.25}$	0.1601	0.9201	0.4670	-0.1145	0.7804	-1.1981
	$q_{0.75}$	0.2337	1.0524	0.5241	0.0681	0.8107	-0.8792
	time	27.079s		26.840s		27.735s	

4.2 Performance of Variance Fitting with $\mathbf{N}(\mathbf{0},\mathbf{1})$ Disturbance

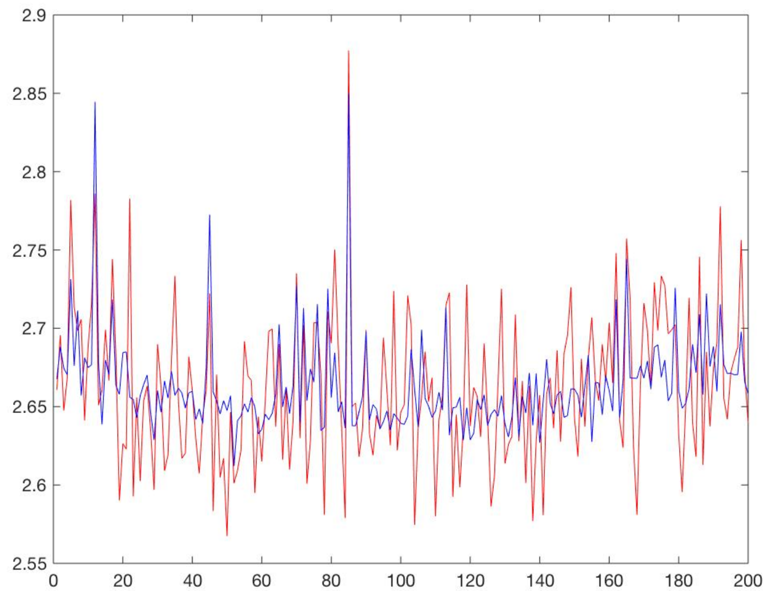
In Section 3.1, with $\varepsilon_{i,n} \overset{iid}{\sim} N(0,1)$, we got the relationship between $h_{i,n}$ and unconditional variance of $u_{i,n}$:

$$Var(u_{i,n}) = [2(\phi[I_n - (\phi+\theta)W_n]^{-1}W_n)_{ii} + 1]E(h_{i,n})$$

To see whether the MLE can capture the spatial structure of unconditional variance, we can calculate the empirical variance of each round of simulations, and we can calculate the mean of $[2(\hat{\phi}[I_n-(\hat{\phi}+\hat{\theta})W_n]^{-1}W_n)_{ii}+1]\hat{h}_{i,n}$ as an approximation of empirical expectation. For the simplified ARCH-like model, we use 200 regions and run 10000 rounds of simulations to get more precise results; for the GARCH-like model, due to larger sample requirement for consistency and much heavier computational burden, we use 1250 regions with 1000 rounds of simulations. The true parameters we use here are the same combinations in the previous parts. Again, to see whether the results are robust for different types of spatial correlation, we use both the county adjacency matrix and circular lake matrix, and repeat each simulation exercise for these two different situations. In each of the figures below, the red line is the empirical variance of simulated sample, and the blue line is our approximation. The x-axis is the regions, and the y-axis is the level of empirical variance and the approximation.

In general, our proxy can capture the trend of heteroskedasticity across each regions, however it underestimate the level of heteroskedasticity. In most of the figures, the simulated unconditional

Figure 1: ARCH-like County Adjacency Variance Fitting ($\phi_0 = 0.2, \alpha_0 = 1$)



variances are much more volatile than our proxy, even though our proxy recovers the average level of unconditional variance and relationship between the unconditional variance of each regions. Also, from the ARCH-like model results, we can see that when the spatial correlation are larger (as ϕ increases), our fitting performance are getting better. In Figure 3 and Figure 5, we can nearly perfectly fit unconditional variances of each region for county adjacency areas. The level of αseems not affect the performance of fitting a lot, since we couldn't see much differences of performance comparing Figure 7 and 9, as well as Figure 8 and 10.

An interesting observation is that the fitting performance for county adjacency areas are much better than the circular lake areas, both for ARCH-like model and GARCH-like model. In Figure 1 to 10, by setting the same true parameters, the differences of between simulated unconditional variance and our approximation for most regions are much smaller for county adjacency matrix. A potential reason may come from the symmetry of circular lake areas which make the heteroskedasticity level smaller than county adjacency areas. In the figures above, we can see the simulated unconditional variance look more "stationary" for circular lake areas than county adjacency areas. If true, it is coincident with the observation for ϕ. In general, our model will have a better fit for unconditional variance when the spatial correlation between areas are high asymmetric and the spill-over effect at volatility level is strong among different regions. In empirical research, since in most of the cases, we will consider geographic or economic distance between irregular regions, the good fitting performance is expected as the county adjacency areas showed.

Figure 2: ARCH-like Circular Lake Variance Fitting ($\phi_0 = 0.2, \alpha_0 = 1$)

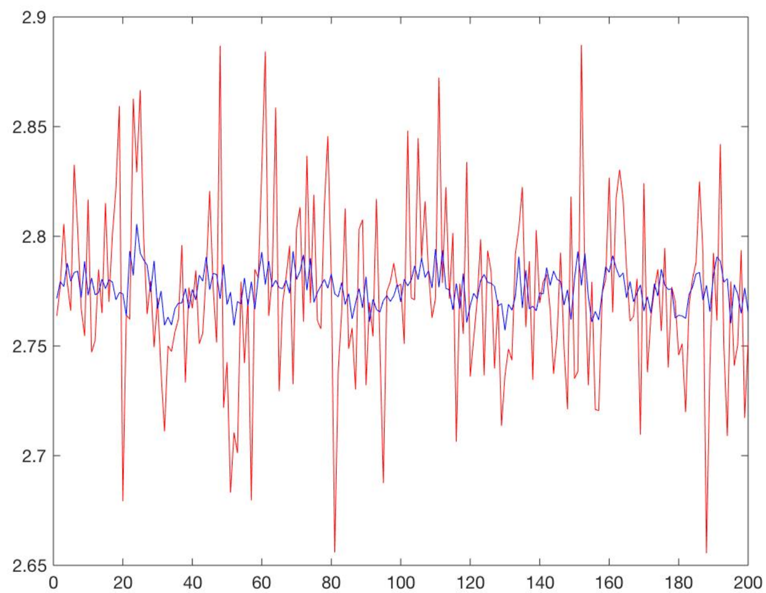


Figure 3: ARCH-like County Adjacency Variance Fitting ($\phi_0 = 0.5, \alpha_0 = 0$)

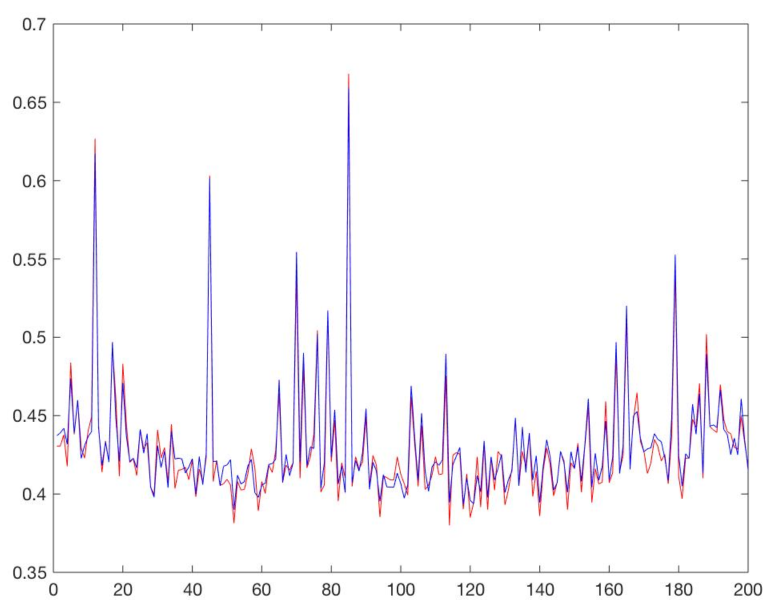
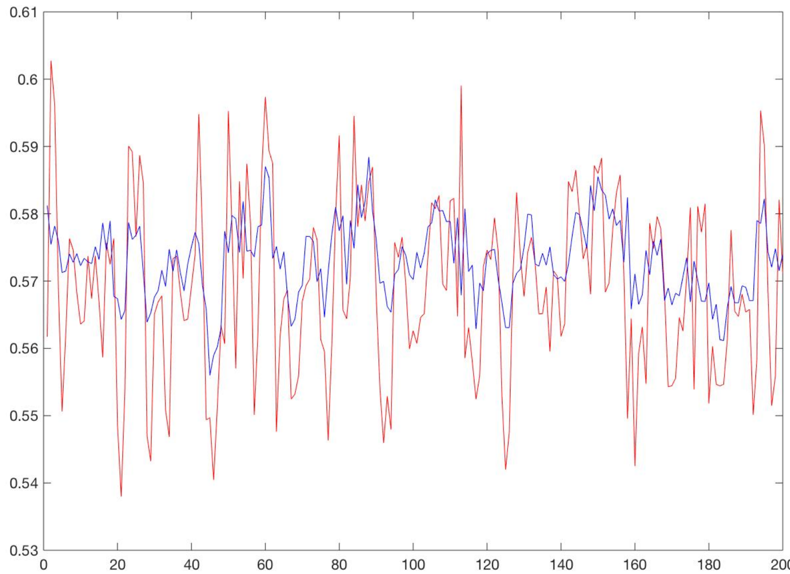


Figure 4: ARCH-like Circular Lake Variance Fitting ($\phi_0 = 0.5, \alpha_0 = 0$)



4.3 Distribution of Estimators with $\mathbf{N}(\mathbf{0}, \mathbf{1})$ Disturbance

Besides consistency, distribution of the estimators can help us to do hypothesis test and build up confidential intervals. For each round of simulation in the variance fitting section, we draw the histogram of the estimators with comparing to Normal distribution. In each of the figures below, the red line is the Normal density curve with the mean and variance equal to the sample mean and variance of estimators. For ARCH-like model, the left-one is for $\hat{\phi}$, and the right-one is for $\hat{\alpha}$. For GARCH-like model, from the left to right, they are for $\hat{\phi}$, $\hat{\theta}$ and $\hat{\alpha}$. Again, the sample size for ARCH-like and GARCH-like model are 200 and 1250 regions with 10000 and 1000 different samples.

Figure 5: ARCH-like County Adjacency Variance Fitting ($\phi_0 = 0.8, \alpha_0 = -1$)

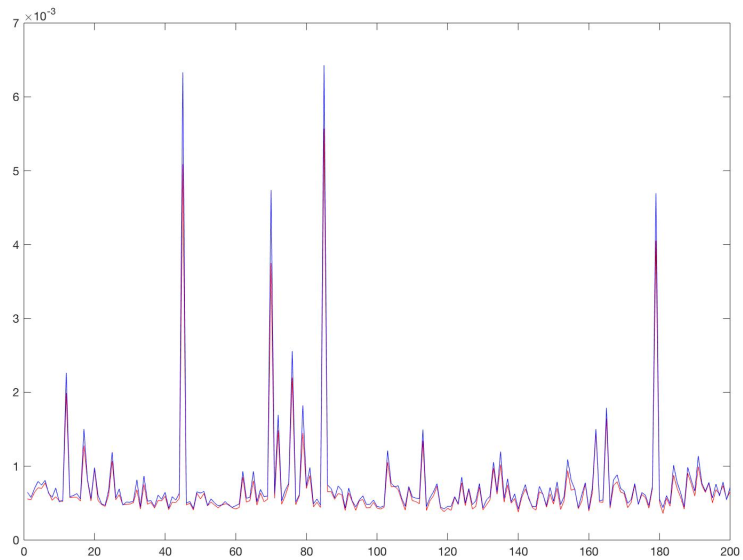


Figure 6: ARCH-like Circular Lake Variance Fitting ($\phi_0 = 0.8, \alpha_0 = -1$)

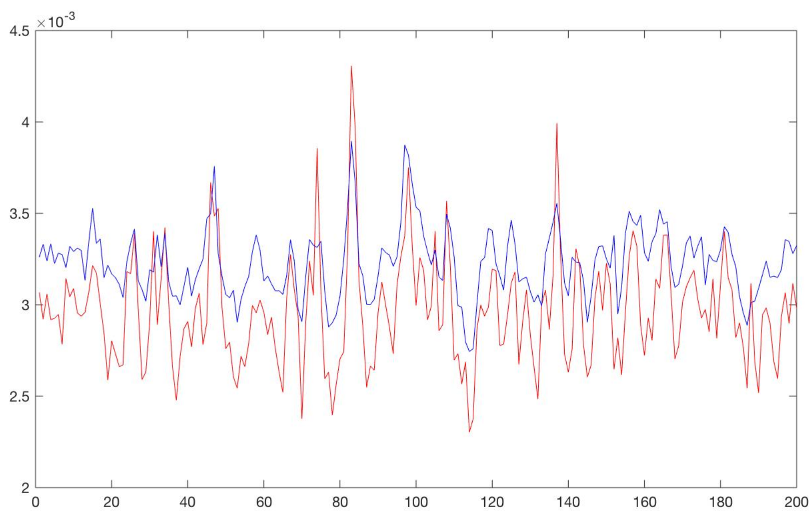


Figure 7: GARCH-like County Adjacency Variance Fitting ($\phi_0 = 0.3, \theta_0 = 0.4, \alpha_0 = 1$)

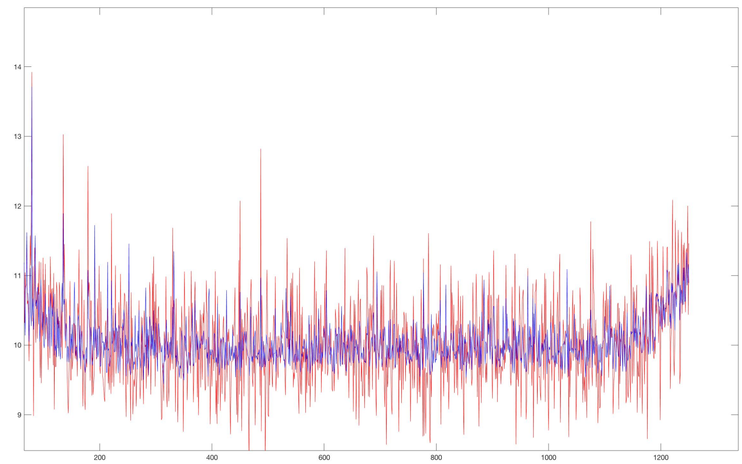


Figure 8: GARCH-like Circular Lake Variance Fitting ($\phi_0 = 0.3, \theta_0 = 0.4, \alpha_0 = 1$)

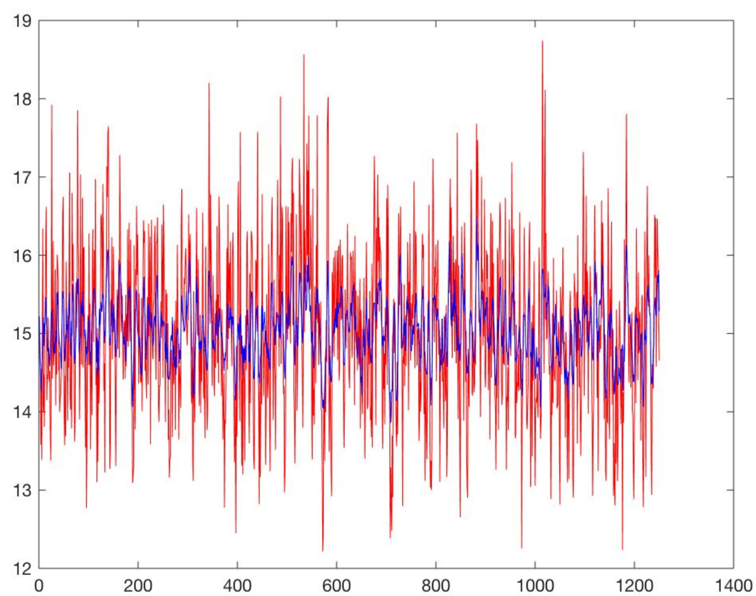


Figure 9: GARCH-like County Adjacency Variance Fitting ($\phi_0 = 0.7, \theta_0 = -0.3, \alpha_0 = -1$)

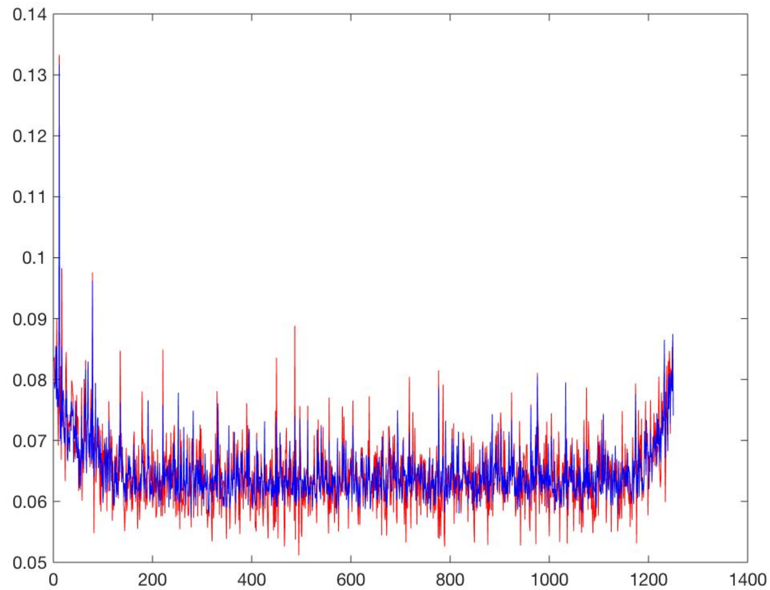


Figure 10: GARCH-like Circular Lake Variance Fitting ($\phi_0 = 0.7, \theta_0 = -0.3, \alpha_0 = -1$)

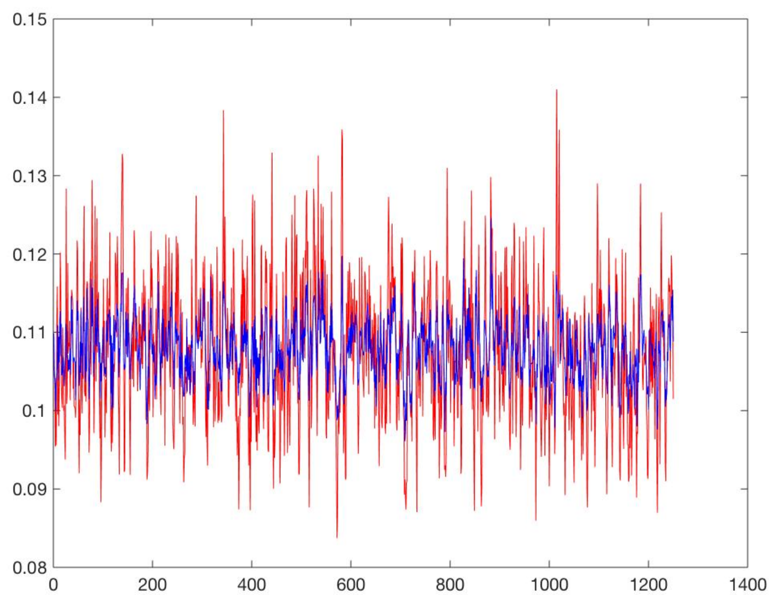


Figure 11: ARCH-like County Adjacency Distribution of Estimators ($\phi_0 = 0.2, \alpha_0 = 1$)

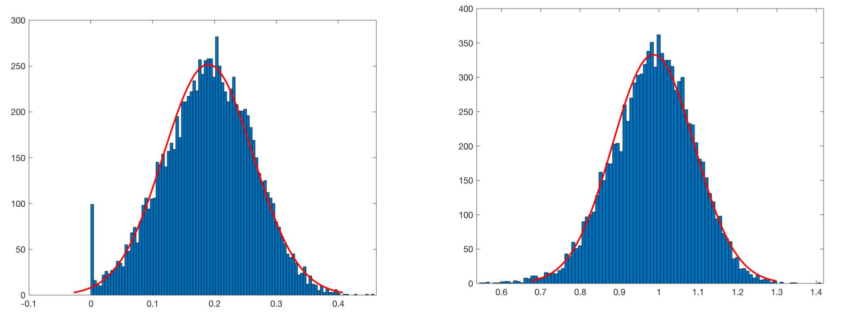
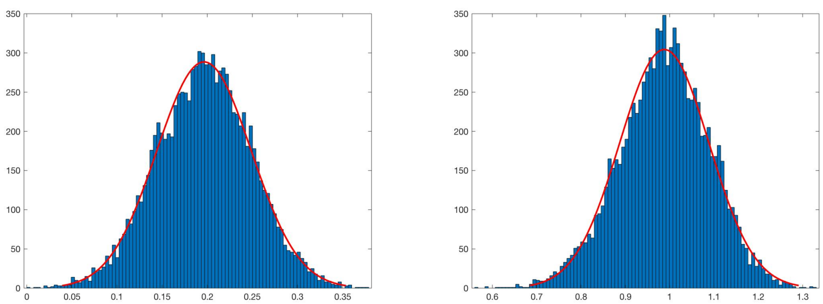


Figure 12: ARCH-like Circular Lake Distribution of Estimators ($\phi_0 = 0.2, \alpha_0 = 1$)



In some of the figures, the estimators showed an extraordinary large density on the tails, 0 and 1. This is due to the constrained optimization method used in maximizing the concentrated log-likelihood function. The most obvious case is in Figure 11, which has 100 times meet 0. This is due to the small true value $\phi_0 = 0$. However, comparing to 10000 times simulation, by using this empirical distribution of $\hat{\phi}$, we can still get the correct inference to reject $H_0 : \rho_0 = 0$ at 95% significance level since only 1% of the tail probability had been affected. A potential solution is to apply unconstrained optimization method, such as *fminsearch* and *fminunc* in Matlab. However, after we tried to implement them on ARCH-like model, they does give larger biased estimators and larger standard deviations. Due to this reason, we still present the results from constrained optimizers.

In general, for GARCH-like model, due to the large sample size (1250 regions), the distribution of the estimators are closer to Normal distribution (Figure 17 to 20). However, for ARCH-like model, we can clearly see the distribution of $\hat{\phi}$ and $\hat{\alpha}$ are not symmetric for both county adjacency areas and circular lake areas. From Figure 13 to 16, when ϕ is large, density functions of both estimators are negative-skewed with a long tail on the left. If we use Normal distribution to do statistical inference, we are likely to use wrong critical and get wrong conclusion.

Figure 13: ARCH-like County Adjacency Distribution of Estimators ($\phi_0 = 0.5, \alpha_0 = 0$)

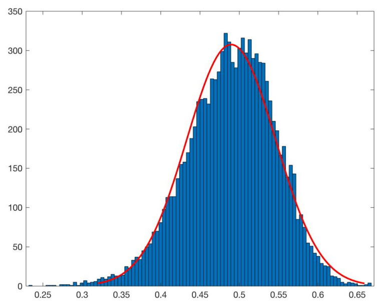


Figure 14: ARCH-like Circular Lake Distribution of Estimators ($\phi_0 = 0.5, \alpha_0 = 0$)

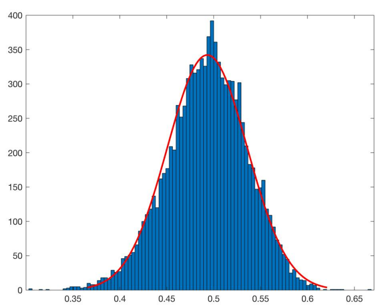


Figure 15: ARCH-like County Adjacency Distribution of Estimators ($\phi_0 = 0.8, \alpha_0 = -1$)

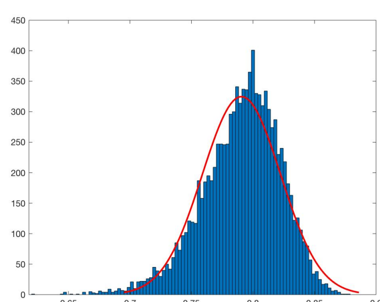


Figure 16: ARCH-like Circular Lake Distribution of Estimators ($\phi_0 = 0.8, \alpha_0 = -1$)

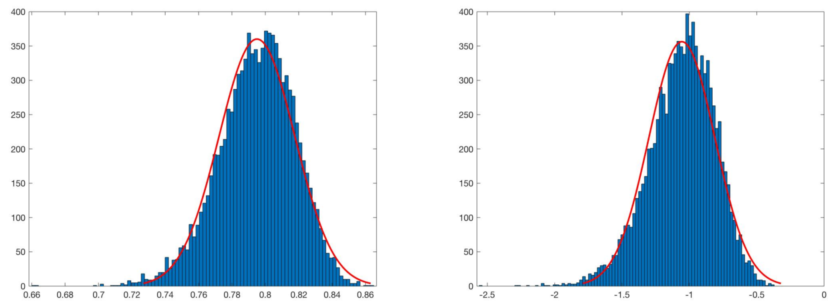
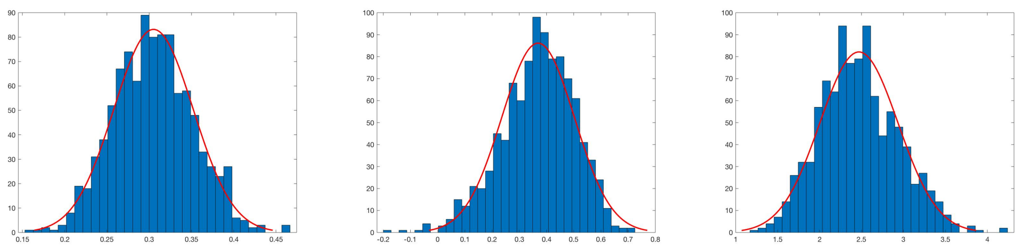


Figure 17: GARCH-like County Adjacency Distribution of Estimators ($\phi_0 = 0.3, \theta_0 = 0.4, \alpha_0 = 1$)



4.4 Finite Sample Performance for Non-Normal Disturbances

In this section, we try to see whether our MLE perform good for non-Normal residuals. Here, we consider three type of distributions: t-distribution, uniform distribution and Laplace distribution.

Figure 18: GARCH-like Circular Lake Distribution of Estimators ($\phi_0 = 0.3, \theta_0 = 0.4, \alpha_0 = 1$)

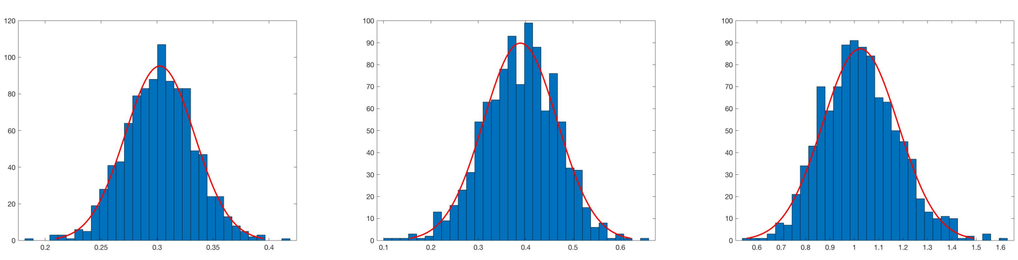


Figure 19: GARCH-like County Adjacency Distribution of Estimators
($\phi_0 = 0.7, \theta_0 = -0.3, \alpha_0 = -1$)

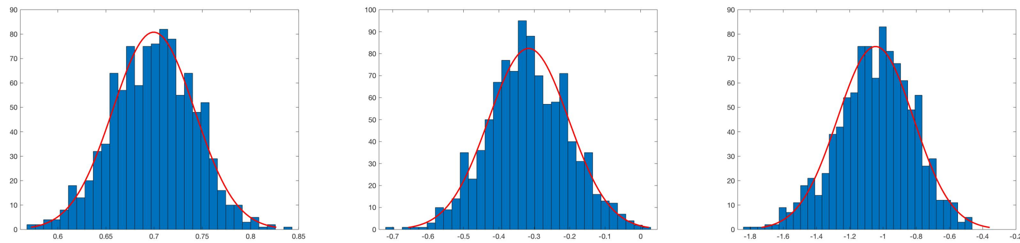
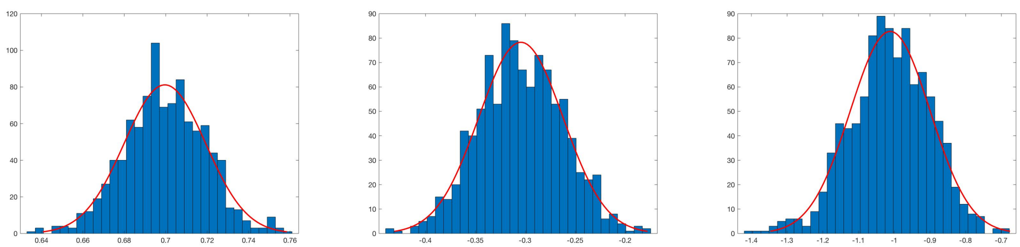


Figure 20: GARCH-like Circular Lake Distribution of Estimators ($\phi_0 = 0.7, \theta_0 = 0.3, \alpha_0 = -1$)



To make sure the true variance is 1, we normalized the $t\,(10)$, $Uniform[-1,1]$ and $Laplace(0,1)$ to $\sqrt{\frac{2}{3}}t(6)$, $Uniform[-\sqrt{3},\sqrt{3}]$ and $\frac{1}{\sqrt{2}}Laplace(0,1)$. With similar arguments as for $N\,(0,1)$ case, we do not need to pay much attention to the existence of negative moments. These three distribution represents three different type of residuals: 1. t-distribution has a fatter tail than Normal distribution; 2. uniform distribution has a compact support which is not common for a residual term; 3. Laplace distribution has more densities concentrated around zero. By exploring the small sample performances for these three different cases, we can have a better understand of potential applications of the MLE in empirical research. For each round of simulations, the spatial weighting matrices are the county adjacency matrices used in Section 4.1, as well as the initial values, parameter settings, and repetition times. Considering the performance of Normal case, we use different sample sizes for the ARCH-like model and GARCH-like model: for ARCH-like model, we use 100 and 200 regions; for GARCH-like model, we use 500 and 1250 regions. Here, we should notice that $\sqrt{\frac{4}{5}}t(10)$ does not meet Assumption 5 either, since moments with order higher than 10 does not exist. In Section 3.4, to get uniform integrability condition, we need to make sure the moment of order $2\frac{1-|\theta_0|}{1-|\theta|}\frac{1-\theta-\phi}{1-\theta_0-\phi_0}$ is finite, this condition would be satisfied for some parameter settings, but in general could not be guaranteed. So, the performance with t-distribution may highly depend on the parameter setting.

Table 7 and Table 8 show the performances for the ARCH-like model. Similar to $N\,(0,1)$ case, the small sample performance of $\hat{\phi}_{MLE}$ is good. For all the three distributions, the bias is small and the standard deviation of $\hat{\phi}$ is shrinking which showed increasing significance level. For $\hat{\alpha}$, there are larger biases, but as we argued before, this is a not a problem on identifying the DGP and capturing spatial correlation between different regions.

Table 7 : ARCH-like performance for non-Normal cases (100 regions)

distribution	true	$\phi = 0.2$	$\alpha = 1$	$\phi = 0.5$	$\alpha = 0$	$\phi = 0.8$	$\alpha = -1$
	mean	0.2040	0.9584	0.5023	-0.0291	0.8020	-1.0154
	std	0.1006	0.2058	0.0794	0.3021	-1.2922	-0.6773
$\sqrt{\frac{2}{3}}t(6)$	med	0.2059	0.9622	0.5094	-0.0236	0.8060	-0.9565
	$q_{0.25}$	0.1325	0.8209	0.4560	-0.2056	0.7784	-1.2922
	$q_{0.75}$	0.2756	1.0860	0.5557	0.1537	0.8304	-0.6773
	mean	0.1651	1.0034	0.4508	-0.0728	0.7607	-1.3282
	std	0.0601	0.0999	0.0514	0.1209	0.032	0.2726
$Uniform[-\sqrt{3}, \sqrt{3}]$	med	0.1634	1.0079	0.4566	-0.0602	0.7644	-1.2856
	$q_{0.25}$	0.1255	0.9379	0.4170	-0.1455	0.7422	-1.4915
	$q_{0.75}$	0.2060	1.0721	0.4889	0.0119	0.7832	-1.1339
	mean	0.2105	0.9635	0.5252	0.0152	0.8142	-0.8689
	std	0.1084	0.2410	0.0817	0.3363	0.0390	0.5502
$\frac{1}{\sqrt{2}}Laplace(0,1)$	med	0.2151	0.9646	0.5233	0.0389	0.8182	-0.7924
	$q_{0.25}$	0.1384	0.7987	0.4690	-0.1982	0.7914	-1.1999
	$q_{0.75}$	0.2827	1.1326	0.5796	0.2571	0.8428	-0.4799

Table 8 : ARCH-like performance for non-Normal cases (200 regions)

distribution	true	$\phi = 0.2$	$\alpha = 1$	$\phi = 0.5$	$\alpha = 0$	$\phi = 0.8$	$\alpha = -1$
	mean	0.1983	0.9685	0.5088	0.0035	0.8054	-0.9518
	std	0.0949	0.1616	0.0725	0.2578	0.0367	0.4631
$\sqrt{\frac{2}{3}}t(6)$	med	0.1997	0.9714	0.5123	0.0068	0.8072	-0.9359
	$q_{0.25}$	0.1312	0.8605	0.4610	-0.1662	0.7841	-1.2200
	$q_{0.75}$	0.2649	1.0685	0.5572	0.1728	0.8312	-0.6338
	mean	0.1706	1.0016	0.4532	-0.0775	0.7615	-1.3402
	std	0.0504	0.0665	0.0448	0.0942	0.0295	0.2626
$Uniform[-\sqrt{3}, \sqrt{3}]$	med	0.1722	1.0030	0.4569	-0.0739	0.7652	-1.3013
	$q_{0.25}$	0.1384	0.9552	0.4251	-0.1343	0.7440	-1.4986
	$q_{0.75}$	0.2031	1.0479	0.4824	-0.0073	0.7821	-1.1581
	mean	0.2171	0.9926	0.5240	0.0501	0.8181	-0.7912
	std	0.0937	0.1753	0.0736	0.2993	0.0369	0.5149
$\frac{1}{\sqrt{2}}Laplace(0,1)$	med	0.2231	0.9961	0.5272	0.0659	0.8210	-0.7400
	$q_{0.25}$	0.1552	0.8771	0.4745	-0.1437	0.7965	-1.0887
	$q_{0.75}$	0.2837	1.1100	0.5799	0.2662	0.8447	-0.4225

Table 9 and 10 show the performances for the GARCH-like model. For uniform distribution and Laplace distribution, the performance are disasters. For t-distribution, although $\hat{\phi}$ has small bias, the biases of $\hat{\theta}$ and $\hat{\alpha}$ is getting larger as the sample size increases. As the existence of higher order moments are not guaranteed, as sample size increases, the estimators would be driven away from the true parameters by the extreme samples. In general, the MLE estimator does not have a good performance comparing to the Normal case.

Table 9 : GARCH-like performance for non-Normal cases (500 regions)

n	true	$\phi = 0.3$	$\theta = 0.4$	$\alpha = 1$	$\phi = 0.7$	$\theta = -0.3$	$\alpha = -1$
	mean	0.2815	0.4549	0.8983	0.6561	-0.1619	-0.7886
	std	0.0674	0.1600	0.2876	0.0746	0.1718	0.3333
$\sqrt{\frac{2}{3}}t(6)$	med	0.2802	0.4549	0.8655	0.6595	-0.1587	-0.7484
	$q_{0.25}$	0.2345	0.3589	0.6896	0.6063	-0.2785	-0.9821
	$q_{0.75}$	0.3239	0.5661	1.0667	0.7072	-0.0444	-0.5527
	mean	0.4645	-0.4348	2.9930	0.7915	-0.7844	-1.8203
	std	0.0521	0.1183	0.2963	0.0481	0.0571	0.1268
$Uniform[-\sqrt{3}, \sqrt{3}]$	med	0.4641	-0.4576	2.9930	0.7909	-0.7866	-1.8204
	$q_{0.25}$	0.4291	-0.4946	2.8785	0.7581	-0.8218	-1.9045
	$q_{0.75}$	0.4978	-0.4176	3.0852	0.8239	-0.7518	-1.7358
	mean	0.2633	0.5616	0.7241	0.5691	0.1328	-0.2082
	std	0.0573	0.1050	0.1773	0.0767	0.1308	0.2501
$\frac{1}{\sqrt{2}}Laplace(0,1)$	med	0.2581	0.5665	0.7120	0.5722	0.1404	-0.1851
	$q_{0.25}$	0.2229	0.4932	0.6028	0.5169	0.0504	-0.3477
	$q_{0.75}$	0.3003	0.6351	0.8345	0.6157	0.2159	-0.0395

Table 10 : GARCH-like performance for non-Normal cases (1250 regions)

n	true	$\phi = 0.3$	$\theta = 0.4$	$\alpha = 1$	$\phi = 0.7$	$\theta = -0.3$	$\alpha = -1$
	mean	0.2833	0.4014	0.8656	0.6514	-0.1268	-0.7029
	std	0.0455	0.1069	0.1881	0.0510	0.1123	0.2049
$\sqrt{\frac{2}{3}}t(6)$	med	0.2813	0.4814	0.8471	0.6523	-0.1334	-0.6865
	$q_{0.25}$	0.2530	0.4014	0.7342	0.6185	-0.2019	-0.8355
	$q_{0.75}$	0.3132	0.5443	0.9831	0.6856	-0.0505	-0.5608
	mean	0.4745	-0.4714	2.9914	0.8072	-0.8070	-1.8225
	std	0.0340	0.0426	0.1022	0.0321	0.0322	0.0780
$Uniform[-\sqrt{3}, \sqrt{3}]$	med	0.4741	-0.4732	2.9998	0.8065	-0.8065	-1.8257
	$q_{0.25}$	0.4530	-0.4952	2.9483	0.7851	-0.8300	-1.8746
	$q_{0.75}$	0.4963	-0.4514	3.0496	0.8300	-0.7850	-1.7660
	mean	0.2644	0.5771	0.7035	0.5624	0.1641	-0.1174
	std	0.0378	0.0682	0.1136	0.0504	0.0841	0.1554
$\frac{1}{\sqrt{2}}Laplace(0,1)$	med	0.2635	0.5785	0.7019	0.5627	0.1656	-0.1069
	$q_{0.25}$	0.2391	0.5325	0.6231	0.5278	0.1059	-0.2112
	$q_{0.75}$	0.2895	0.6236	0.7780	0.5956	0.2255	-0.0171

5 Lagrangian Multiplier Test for $\mathbf{H_0} : \theta_0 = 0$

5.1 Asymptotic Distribution of The LM Statistic

From Section 4.1, we can see that the finite sample performance of MLE for the GARCH-like model is not good, since the bias is small only when we have more than 2000 observations. In contrast, the ARCH-like model converge much faster and has good performance when we only have 100 observations. Thus, for small sample applications, it is more proper to use the spatial ARCH-like model to capture the spatial heteroskedasticity and risk spill-over effect. However, we still interested in whether there exist GARCH-like spatial correlation of our data. Thus, in this section, we develop a Lagrangian multiplier test for $H_0 : \theta_0 = 0$, which can be used to test the existence of the spatial GARCH-like correlation in relative small sample application, with getting the MLE for spatial ARCH-like model, i.e constrained estimator when $\theta_0 = 0$.

Recall the log-likelihood function of spatial GARCH-like model:

$$\ln L_n(u_n;\alpha,\phi,\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\exp\left\{-\frac{\alpha}{1-\theta}\right\}\sum_{i=1}^{n}\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}\,u_{i,n}^2$$
$$-\frac{1}{2}[1_n^{'}\phi(I_n-\theta W_n)^{-1}W_n log u_n^2 + \frac{n\alpha}{1-\theta}] + \ln|I_n-(\theta+\phi)W_n| - \ln|I_n-\theta W_n|$$

Since $\frac{\partial(I_n-\theta W_n)^{-1}W_n}{\partial\theta} = (I_n-\theta W_n)^{-1}W_n(I_n-\theta W_n)^{-1}W_n$, we have

$$\begin{aligned}&\frac{\partial\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}}{\partial\theta}\\ =&-\phi\left\{\sum_{k=1}^{n}\left[(I_n-\theta W_n)^{-1}W_n\right]_{ik}^2\ln u_{k,n}^2\right\}\\ &\cdot\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}}\end{aligned}$$

$$\begin{aligned}&\frac{\partial 1_n^{'}\phi(I_n-\theta W_n)^{-1}W_n log u_n^2}{\partial\theta}\\ =&1_n^{'}\phi\left[(I_n-\theta W_n)^{-1}W_n\right]^2 log u_n^2\end{aligned}$$

The we have

$$\frac{\partial\ln L_n}{\partial\alpha} = \frac{1}{2(1-\theta)}\exp\left\{-\frac{\alpha}{1-\theta}\right\}\sum_{i=1}^{n}u_{i,n}^2\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}} - \frac{n}{2(1-\theta)}$$

$$\begin{aligned}\frac{\partial\ln L_n}{\partial\phi} =& \frac{1}{2}\exp\left\{-\frac{\alpha}{1-\theta}\right\}\sum_{i=1}^{n}u_{i,n}^2\left\{\sum_{k=1}^{n}\left[(I_n-\theta W_n)^{-1}W_n\right]_{ik}\ln u_{k,n}^2\right\}\\ &\cdot\prod_{j=1}^{n}|u_{j,n}|^{-2(\phi(I_n-\theta W_n)^{-1}W_n)_{ij}} - \frac{1}{2}1_n^{'}(I_n-\theta W_n)^{-1}W_n log u_n^2\\ &-tr\left(\left[I_n-(\phi+\theta)W_n\right]^{-1}W_n\right)\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \ln L_n}{\partial \theta} &= \frac{\alpha}{2\left(1-\theta\right)^2} \exp\left\{-\frac{\alpha}{1-\theta}\right\} \sum_{i=1}^{n} u_{i,n}^2 \prod_{j=1}^{n} \left|u_{j,n}\right|^{-2\left(\phi\left(I_n-\theta W_n\right)^{-1} W_n\right)_{ij}} - \frac{n\alpha}{2\left(1-\theta\right)^2} \\
&+ \frac{1}{2}\phi \exp\left\{-\frac{\alpha}{1-\theta}\right\} \sum_{i=1}^{n} u_{i,n}^2 \left\{\sum_{k=1}^{n} \left[(I_n-\theta W_n)^{-1} W_n\right]_{ik}^2 \ln u_{k,n}^2\right\} \\
&\cdot \prod_{j=1}^{n} \left|u_{j,n}\right|^{-2\left(\phi\left(I_n-\theta W_n\right)^{-1} W_n\right)_{ij}} - \frac{1}{2} 1_n^{'} \phi \left[(I_n-\theta W_n)^{-1} W_n\right]^2 log u_n^2 \\
&- tr\left(\left[I_n - \left(\phi+\theta\right) W_n\right]^{-1} W_n\right) + tr\left(\left(I_n-\theta W_n\right)^{-1} W_n\right)
\end{aligned}
$$

When $\theta = 0$, we have

$$
\frac{\partial \ln L_n\left(\alpha,\phi,0\right)}{\partial \alpha} = \frac{1}{2} e^{-\alpha} \sum_{i=1}^{n} u_{i,n}^2 \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\phi} - \frac{n}{2}
$$

$$
\begin{aligned}
\frac{\partial \ln L_n\left(\alpha,\phi,0\right)}{\partial \phi} &= \frac{1}{2} e^{-\alpha} \sum_{i=1}^{n} u_{i,n}^2 \left[\sum_{k=1}^{n} w_{ij} \ln u_{k,n}^2\right] \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\phi} \\
&- \frac{1}{2} 1_n^{'} W_n log u_n^2 - tr\left(\left(I_n - \phi W_n\right)^{-1} W_n\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ln L_n\left(\alpha,\phi,0\right)}{\partial \theta} &= \frac{1}{2}\alpha e^{-\alpha} \sum_{i=1}^{n} u_{i,n}^2 \prod_{j=1}^{n} \left|u_{j,n}\right|^{-2\phi w_{ij}} - \frac{1}{2} n\alpha \\
&+ \frac{1}{2}\phi e^{-\alpha} \sum_{i=1}^{n} u_{i,n}^2 \left[\sum_{k=1}^{n} \left(W_n^2\right)_{i,k} \ln u_{k,n}^2\right] \prod_{j=1}^{n} \left|u_{j,n}\right|^{-2\phi w_{ij}} \\
&- \frac{1}{2}\phi 1_n^{'} W_n^2 log u_n^2 - tr\left(\left(I_n - \phi W_n\right)^{-1} W_n\right)
\end{aligned}
$$

To simplify notations, recall $\kappa_{i,n}\left(\psi\right) = e^{-\alpha} u_{i,n}^2 \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\phi}$ and $v_{i,n} = \sum_{j=1}^{n} w_{ij} \ln u_{j,n}^2$, define the following two variable:

$$
z_{i,n} = \left(W_n^2 log u_n^2\right)_{i,n} = \sum_{j=1}^{n} \left(W_n^2\right)_{ij} \ln u_{j,n}^2
$$

and

$$
\gamma_{i,n} = \left(W_n^3 log u_n^2\right)_{i,n} = \sum_{j=1}^{n} \left(W_n^3\right)_{ij} \ln u_{j,n}^2
$$

Under $H_0 : \theta = 0$, let our estimator from ARCH-like (restricted) model as $\left(\bar{\alpha}, \bar{\phi}\right)$, denote $\bar{\psi} = \left(\bar{\alpha}, \bar{\phi}, 0\right)$, we always have

$$\frac{\partial \ln L_n\left(\bar{\psi}\right)}{\partial \alpha} = \frac{\partial \ln L_n\left(\bar{\psi}\right)}{\partial \phi} \equiv 0$$

which implies

$$\sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) = n$$

and

$$\sum_{i=1}^{n} \left(\kappa_{i,n}\left(\bar{\psi}\right) v_{i,n} - v_{i,n}\right) = 2tr\left(\left(I_n - \bar{\phi} W_n\right)^{-1} W_n\right)$$

Using these two equations, we can simplify $\frac{\partial \ln L_n(\alpha,\phi,0)}{\partial \theta}$ as:

$$\frac{\partial \ln L_n\left(\alpha, \phi, 0\right)}{\partial \theta} = \frac{1}{2}\bar{\phi}\sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) z_{i,n} - \frac{1}{2}\bar{\phi}\sum_{i=1}^{n} z_{i,n} - tr\left(\left(I_n - \phi W_n\right)^{-1} W_n\right)$$

Then, the LM test will be derived from the asymptotic distribution of $\frac{\partial \ln L_n\left(\bar{\psi}\right)}{\partial \theta}$. Also, we can get

$$\begin{aligned}
\frac{\partial^2 \ln L_n\left(\bar{\psi}\right)}{\partial \alpha^2} &= -\frac{1}{2}e^{-\bar{\alpha}}\sum_{i=1}^{n} u_{i,n}^2 \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}\bar{\phi}}\right] \\
&= -\frac{1}{2}\sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) = -\frac{n}{2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_n\left(\bar{\psi}\right)}{\partial \phi^2} &= -\frac{1}{2}e^{-\bar{\alpha}}\sum_{i=1}^{n} u_{i,n}^2 \left[\sum_{k=1}^{n} w_{ik}\ln u_{k,n}^2\right]^2 \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}} \\
&- tr\left[\left(I_n - \bar{\phi}W_n\right)^{-1} W_n\right]^2 \\
&= -\frac{1}{2}\sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) v_{i,n}^2 - tr\left[\left(I_n - \bar{\phi}W_n\right)^{-1} W_n\right]^2
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 L_n\left(\bar{\psi}\right)}{\partial \theta^2} &= \left(\bar{\alpha} - \frac{1}{2}\bar{\alpha}^2\right) e^{-\bar{\alpha}} \sum_{i=1}^{n} u_{i,n}^2 \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}} - n\bar{\alpha} \\
&- \bar{\alpha}\bar{\phi} e^{-\bar{\alpha}} \sum_{i=1}^{n} u_{i,n}^2 \left[\sum_{k=1}^{n} \left(W_n^2\right)_{i,k} \ln u_{k,n}^2\right] \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}} \\
&- \frac{1}{2}\bar{\phi}^2 e^{-\bar{\alpha}} \sum_{i=1}^{n} u_{i,n}^2 \left[\sum_{k=1}^{n} \left(W_n^2\right)_{i,k} \ln u_{k,n}^2\right]^2 \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}} \\
&+ \bar{\phi} e^{-\bar{\alpha}} \sum_{i=1}^{n} u_{i,n}^2 \left[\sum_{k=1}^{n} \left(W_n^3\right)_{i,k} \ln u_{k,n}^2\right] \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}} \\
&- \bar{\phi} 1_n^{'} W_n^3 log u_n^2 - tr\left[\left(I_n - \bar{\phi} W_n\right)^{-1} W_n\right]^2 \\
&= \left(\bar{\alpha} - \frac{1}{2}\bar{\alpha}^2\right) \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) - n\bar{\alpha} - \bar{\alpha}\bar{\phi} \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) z_{i,n} \\
&- \frac{1}{2}\bar{\phi}^2 \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) z_{i,n}^2 + \bar{\phi} \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) \gamma_{i,n} - \bar{\phi} \sum_{i=1}^{n} \gamma_{i,n} \\
&- tr\left[\left(I_n - \bar{\phi} W_n\right)^{-1} W_n\right]^2 + tr\left(W_n^2\right) \\
&= -\frac{1}{2}\bar{\alpha}^2 n - \bar{\alpha}\bar{\phi} \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) z_{i,n} - tr\left[\left(I_n - \bar{\phi} W_n\right)^{-1} W_n\right]^2 \\
&- \frac{1}{2}\bar{\phi}^2 \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) z_{i,n}^2 + \bar{\phi} \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) \gamma_{i,n} - \bar{\phi} \sum_{i=1}^{n} \gamma_{i,n} + tr\left(W_n^2\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 L_n\left(\bar{\psi}\right)}{\partial \alpha \partial \phi} &= -\frac{1}{2} e^{-\bar{\alpha}} \sum_{i=1}^{n} u_{i,n}^2 \left(\sum_{k=1}^{n} w_{ij} \ln u_{k,n}^2\right) \left[\prod_{j=1}^{n} \left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}} \\
&= -\frac{1}{2} \sum_{i=1}^{n} \kappa_{i,n}\left(\bar{\psi}\right) v_{i,n}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 L_n\left(\bar{\psi}\right)}{\partial\alpha\partial\theta} &= \frac{1}{2}\left(1-\bar{\alpha}\right)e^{-\bar{\alpha}}\sum_{i=1}^{n}u_{i,n}^2\left[\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}}-\frac{1}{2}n \\
&-\frac{1}{2}\bar{\phi}e^{-\bar{\alpha}}\sum_{i=1}^{n}u_{i,n}^2\left[\sum_{k=1}^{n}\left(W_n^2\right)_{i,k}\ln u_{k,n}^2\right]\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2\bar{\phi}w_{ij}} \\
&= \frac{1}{2}\left(1-\bar{\alpha}\right)\sum_{i=1}^{n}\kappa_{i,n}\left(\bar{\psi}\right)-\frac{1}{2}n-\frac{1}{2}\bar{\phi}\sum_{i=1}^{n}\kappa_{i,n}\left(\bar{\psi}\right)z_{i,n} \\
&= -\frac{1}{2}\bar{\alpha}n-\frac{1}{2}\bar{\phi}\sum_{i=1}^{n}\kappa_{i,n}\left(\bar{\psi}\right)z_{i,n}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 L_n\left(\bar{\psi}\right)}{\partial\phi\partial\theta} &= -\frac{1}{2}\bar{\alpha}e^{-\bar{\alpha}}\sum_{i=1}^{n}u_{i,n}^2\left(\sum_{k=1}^{n}w_{ij}\ln u_{k,n}^2\right)\left[\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2w_{ij}}\right]^{\bar{\phi}} \\
&+\frac{1}{2}e^{-\bar{\alpha}}\sum_{i=1}^{n}u_{i,n}^2\left[\sum_{k=1}^{n}\left(W_n^2\right)_{i,k}\ln u_{k,n}^2\right]\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2\bar{\phi}w_{ij}} \\
&-\frac{1}{2}\bar{\phi}e^{-\bar{\alpha}}\sum_{i=1}^{n}u_{i,n}^2\left[\sum_{k=1}^{n}\left(W_n^2\right)_{i,k}\ln u_{k,n}^2\right]\left(\sum_{k=1}^{n}w_{ij}\ln u_{k,n}^2\right)\prod_{j=1}^{n}\left|u_{j,n}\right|^{-2\bar{\phi}w_{ij}} \\
&-\frac{1}{2}1_n^{'}W_n^2 log u_n^2-tr\left[\left(I_n-\bar{\phi}W_n\right)^{-1}W_n\right]^2 \\
&= -\frac{1}{2}\bar{\alpha}\sum_{i=1}^{n}\kappa_{i,n}\left(\bar{\psi}\right)v_{i,n}+\frac{1}{2}\sum_{i=1}^{n}\kappa_{i,n}\left(\bar{\psi}\right)z_{i,n}-\frac{1}{2}\bar{\phi}\sum_{i=1}^{n}\kappa_{i,n}\left(\bar{\psi}\right)v_{i,n}z_{i,n} \\
&-\frac{1}{2}\sum_{i=1}^{n}z_{i,n}-tr\left[\left(I_n-\bar{\phi}W_n\right)^{-1}W_n\right]^2
\end{aligned}
$$

Since $\frac{\partial^2 L_n\left(\bar{\psi}\right)}{\partial\alpha\partial\theta}$ and $\frac{\partial^2 L_n\left(\bar{\psi}\right)}{\partial\phi\partial\theta}$ are asymptotically converge to zero when $\bar{\psi}\xrightarrow{p}\psi_0$, by $\bar{\psi}$ is $\sqrt{n}-$convergence, we can use the conventional LM statistic which can be simplified as:

$$
\begin{aligned}
LM_{\theta=0} &= \frac{\partial\ln L(\bar{\psi})}{\partial\psi'}\left[-E\left(\frac{\partial^2\ln L(\bar{\psi})}{\partial\psi\partial\psi'}\right)\right]^{-1}\frac{\partial\ln L(\bar{\psi})}{\partial\psi} \\
&= -\left[\frac{\partial\ln L(\bar{\psi})}{\partial\theta}\right]^2\left[E\left(\frac{\partial^2\ln L(\bar{\psi})}{\partial\psi\partial\psi'}\right)\right]_{33}^{-1}
\end{aligned}
$$

The remaining work to derive the LM statistic is to prove the joint asymptotic Normality of $\frac{\partial\ln L_n\left(\bar{\psi}\right)}{\partial\theta}$.

Since we can write

$$
\frac{\partial\ln L_n\left(\bar{\psi}\right)}{\partial\theta} = \frac{1}{2}\sum_{i=1}^{n}\left[\bar{\alpha}\kappa_{i,n}\left(\bar{\psi}\right)+\bar{\phi}\kappa_{i,n}\left(\bar{\psi}\right)z_{i,n}-\bar{\phi}z_{i,n}-\bar{\alpha}\right]-tr\left(\left(I_n-\bar{\phi}W_n\right)^{-1}W_n\right)
$$

Let $\widetilde{\omega}_{i,n}\left(\bar{\psi}\right)=\bar{\alpha}\kappa_{i,n}\left(\bar{\psi}\right)+\bar{\phi}\kappa_{i,n}\left(\bar{\psi}\right)z_{i,n}-\bar{\phi}z_{i,n}-\bar{\alpha}$, similar to the discussion for property of $\omega_{i,n}\left(a,b\right)$, we can get $L^{2+\delta}$ uniform integrability for $\widetilde{\omega}_{i,n}\left(\bar{\psi}\right)$, i.e.

$$\lim_{k\to\infty}\sup_{n}\sup_{i\in D_n}E\left[\left|\widetilde{\omega}_{i,n}(\bar{\psi})\right|^{2+\delta}1_{\left(\left|\widetilde{\omega}_{i,n}(\bar{\psi})\right)\right|>k)}\right]=0$$

for $\forall\ 0<\delta<1$.

The next thing to do is to prove the jointly $\alpha-$mixing of $\left\{\left(z_{i,n},\kappa_{i,n}\left(\bar{\psi}\right)\right)'\right\}_{i\in D_n}$. First, since we have

$$\begin{pmatrix} z_{i,n} \\ \ln u_{i,n}^2 \end{pmatrix}=\sum_{j=1}^{n}\begin{pmatrix} \left(W_n^2\right)_{ij} \\ 0 \end{pmatrix}\ln u_{j,n}^2+\begin{pmatrix} 0 \\ 1 \end{pmatrix}\ln u_{i,n}^2$$

Recall that for any $Q\subset D_n$, $s>0$, denote $Q^s=\{i\in D_n: d(i,Q)<s\}$. Under Assumption 3, since $w_{ij,n}=0$ when $d_{ij}>\bar{d}_0$. Then for any $i,j,k\in D_n$, when $d_{ij}>2\bar{d}_0$, since $d_{ik}+d_{kj}\geqslant d_{ij}$, either w_{ik} or w_{kj} is 0. So, for $d_{ij}>2\bar{d}_0$, we have $\left(W_n^2\right)_{ij}=\sum_{k=1}^{n}w_{ik}w_{kj}=0$. Thus, we have

$$\begin{pmatrix} z_{i,n} \\ \ln u_{i,n}^2 \end{pmatrix}=\sum_{j\in\{i\}^{2\bar{d}_0}}\begin{pmatrix} \left(W_n^2\right)_{ij} \\ 0 \end{pmatrix}\ln u_{j,n}^2+\begin{pmatrix} 0 \\ 1 \end{pmatrix}\ln u_{i,n}^2$$

Similar to the discussion for jointly mixing of $\left\{\left(e_{i,n},\ln u_{i,n}^2\right)'\right\}_{i\in D_n}$, when $r>6\bar{d}_0$, $d\left(U^{2\bar{d}_0},V^{2\bar{d}_0}\right)>\frac{r}{3}$, we have

$$\begin{aligned}
\alpha_{k,l}^{(z,lnu)}(r)&=\sup_{U,V\subseteq D_n}\{\alpha_n(U,V):|U|\leqslant k,|V|\leqslant l,d(U,V)\geqslant r\}\\
&\leqslant\sup_{U^{\bar{d}_0},V^{\bar{d}_0}\subseteq D_n}\left\{\alpha_n(U^{\bar{d}_0},V^{\bar{d}_0}):|U^{\bar{d}_0}|\leqslant kC_d\left(6\bar{d}_0\right)^d,|V^{\bar{d}_0}|\leqslant lC_d\left(6\bar{d}_0\right)^d,d(U^{\bar{d}_0},V^{\bar{d}_0})\geqslant\frac{r}{3}\right\}\\
&=\alpha^{(\ln u)}_{kC_d\left(6\bar{d}_0\right)^d,lC_d\left(6\bar{d}_0\right)^d}\left(\frac{r}{3}\right)\\
&\leqslant C_3C_d\left(6\bar{d}\right)^d min(k,l)\left(\frac{1}{3^{d-1}}r^{d-1}\xi^{r/\bar{d}_0}\right)^{1/3}\\
&\equiv C_3'\min(k,l)2^d\left(\frac{1}{3^{d-1}}r^{d-1}\xi^{r/\bar{d}_0}\right)^{1/3}
\end{aligned}$$

By the following representation:

$$\begin{aligned}
\begin{pmatrix} z_{i,n} \\ \ln\kappa_{i,n}(\bar{\psi}) \end{pmatrix}&=\sum_{j=1,j\neq i}^{n}\begin{pmatrix} 0 & 0 \\ 0 & \left(I_n-\bar{\phi}W_n\right)_{ij,n} \end{pmatrix}\begin{pmatrix} z_{j,n} \\ \ln u_{j,n}^2 \end{pmatrix}\\
&+\begin{pmatrix} 1 & \\ & \left(I_n-\bar{\phi}W_n\right)_{ii,n} \end{pmatrix}\begin{pmatrix} z_{i,n} \\ \ln u_{i,n}^2 \end{pmatrix}-\begin{pmatrix} 0 \\ \bar{\alpha} \end{pmatrix}
\end{aligned}$$

Similar to the discussion of jointly mixing of $\left\{\left(v_{i,n}\left(\psi\right),\ln\kappa_{i,n}\left(\psi\right)\right)'\right\}_{i\in D_n}$ for the ARCH-like case, the matrix $\begin{pmatrix} I_n & \\ & I_n-\bar{\phi}W_n \end{pmatrix}$ satisfies the condition for Theorem 2 in Xu and Lee (2019). Then, for $r>18\bar{d}_0$, we have

$$\alpha_{k,l}^{(z,\ln\kappa)}(r) \leqslant \alpha_{C_d k(r/3)^d, C_d l(r/3)^d}^{(z,\ln u)}\left(\frac{r}{3}\right)$$
$$\leqslant \min\left\{C_d k(r/3)^d, C_d l(r/3)^d\right\} 2^d \left(\frac{1}{3^{2(d-1)}} r^{d-1}\xi^{r/\bar{d}_0}\right)^{1/3}$$
$$= \min\{k,l\} C_d 2^d 3^{\frac{2}{3}-\frac{5}{3}d} r^{\frac{4}{3}d-\frac{1}{2}} \xi^{r/3\bar{d}_0}$$

For any k and l, we have $\lim_{r\to\infty}\alpha_{k,l}^{(\iota)}(r) = 0$. Thus, $z_{i,n}(\psi)$ and $\ln\kappa_{i,n}(\psi)$ are jointly $\alpha-$mixing. Then, for any fixed $\bar{\psi}$, $\left\{\widetilde{\omega}_{i,n}\left(\bar{\psi}\right)\right\}_{i\in D_n}$ is a $\alpha-$mixing random field with the same upper bond of $\left\{\left(z_{i,n}(\psi), \ln\kappa_{i,n}(\psi)\right)'\right\}_{i\in D_n}$.

To apply CLT in Jenish and Prucha (2009), the $\alpha-$mixing coefficient need to have some other properties. We can check that they are all satisfied for $d = 2$ in our case:

$$\sum_{m=1}^{\infty}\bar{\alpha}_{1,1}(m)\, m^{[d(2+\delta)/\delta]-1}$$
$$= \sum_{m=1}^{\lfloor 18\bar{d}_0\rfloor}\bar{\alpha}_{1,1}(m)\, m^{[2(2+\delta)/\delta]-1} + \sum_{m=\lfloor 18\bar{d}_0\rfloor+1}^{\infty}\bar{\alpha}_{1,1}(m)\, m^{[2(2+\delta)/\delta]-1}$$
$$\leqslant \sum_{m=1}^{\lfloor 18\bar{d}_0\rfloor} m^{[2(2+\delta)/\delta]-1} + \sum_{m=\lfloor 18\bar{d}_0\rfloor+1}^{\infty} 4C_d 3^{-\frac{7}{3}} m^{\frac{13}{6}}\xi^{r/3\bar{d}_0} m^{[2(2+\delta)/\delta]-1} < \infty$$

$$\sum_{m=1}^{\infty} m^{d-1}\bar{\alpha}_{k,l}(m)$$
$$\leqslant \sum_{m=1}^{\lfloor 18\bar{d}_0\rfloor} m\bar{\alpha}_{k,l}(m) + \sum_{m=\lfloor 18\bar{d}_0\rfloor+1}^{\infty} m\bar{\alpha}_{k,l}(m)$$
$$\leqslant \sum_{m=1}^{\lfloor 18\bar{d}_0\rfloor} m + \sum_{m=\lfloor 18\bar{d}_0\rfloor+1}^{\infty}\min\{k,l\}\, 4C_d 3^{-\frac{7}{3}} m^{\frac{19}{6}}\xi^{m/3\bar{d}_0} < \infty$$

$$\bar{\alpha}_{1,\infty}(m) = 4C_d 3^{-\frac{7}{3}} m^{\frac{19}{6}}\xi^{m/3\bar{d}_0} = O\left(m^{-2-\varepsilon}\right)$$

for $0 < \delta < 1$, $k + l \leqslant 4$ and some $\varepsilon > 0$, since $0 < \xi < 1$. With the boundedness of $tr\left(\left(I_n - \bar{\phi}W_n\right)^{-1} W_n\right)$, let $\sigma_{g_\theta} \equiv Var\left(\frac{\partial \ln L_n\left(\bar{\alpha},\bar{\phi},0\right)}{\partial\theta}\right)$, by Theorem 1 and Corollary 1 in Jenish and Prucha (2009), we have

$$\sigma_{g_\theta}^{-1}\frac{\partial \ln L_n\left(\bar{\psi}\right)}{\partial\theta} \xrightarrow{d} N(0,1)$$

Then by similar arguments in Section 3.5, we can get the jointly asymptotic Normality of $\frac{\partial \ln L_n\left(\bar{\alpha},\bar{\phi},0\right)}{\partial\theta}$. Then, when H_0 is true, we have

$$LM_{\theta=0} = -\left[\frac{\partial \ln L(\bar{\psi})}{\partial \theta}\right]^2 \left[E\left(\frac{\partial^2 \ln L(\bar{\psi})}{\partial \psi \partial \psi'}\right)\right]_{33}^{-1} \xrightarrow{d} \chi^2(1)$$

with getting the constraint estimator $\left(\bar{\alpha}, \bar{\phi}\right)$.

5.2 Monte Carlo Simulation for the LM test

To see how this LM test works in finite sample, we implement some Monte Carlo simulation exercise. All the parameter settings are the same as previous Monte Carlo sessions. For each parameter setting and sample size, we replicate 10,000 rounds simulation exercises. Table 11 and Table 13 reports the test size when $H_0 : \theta_0 = 0$ holds. Table 13 and Table 14 reports the test power when $H_1 : \theta \neq 0$ holds. Table 15 shows the simulated critical value for 10%, 5% and 1% significant level. The residual process are $\varepsilon_{i,n} \overset{iid}{\sim} N(0, 11)$. The critical values we use here are $\chi^2_{0.95} = 3.8415$ (Table 1 and Table 2) and $\chi^2_{0.9} = 2.7055$ (Table 3 and Table 4) with considering two different spatial correlations: circular lake and county adjacent. In each simulation exercise, we use the empirical information matrix $\frac{\partial^2 \ln L(\bar{\psi})}{\partial \psi \partial \psi'}$ instead of $E\left(\frac{\partial^2 \ln L(\bar{\psi})}{\partial \psi \partial \psi'}\right)$ or $E\left(\frac{\partial^2 \ln L(\psi_0)}{\partial \psi \partial \psi'}\right)$. Since it is calculated by using empirical information matrix $\frac{\partial^2 \ln L_n(\hat{\psi})}{\partial \psi \partial \psi'}$ which may not necessary negative semidefinite, it is possible to get negative value even though the limiting distribution is $\chi^2(1)$. Simulated critical values are showed in Table 15.

From Table 11 and Table 13, we can see the test size converges to the theoretical value when sample size increases in general. With small sample size, the LM test has larger chance to over-reject the true models. However, from Table 12 and Table 14, we can see that the power of the LM test highly depend on the true parameter values and the spatial correlations. Although the test power increases as the sample size increases, the test powers in circular lake situation are much larger than the county adjacent situation. Also, combine the results from Table 11 to Table 14, the LM test seems more useful when ϕ_0 is larger, considering both the size and power. When ϕ_0 is small, LM test will have slightly larger chance to make Type I error, but very large chance to make Type II error with some particular spatial correlation settings, even with 500 samples. This conclusion can be confirmed by the simulated critical values in Table 15 . The critical values for large ϕ_0 case converge to the critical values of $\chi^2(1)$ much faster.

The potential reason for the bad performance when ϕ_0 is small may comes from the model setting itself of the spatial GARCH-like model. Recall Section 3.3, when $\phi_0 = 0$, α_0 and θ_0 can not be separately identified and the spatial GARCH-like model does not generate any heteroskedasticity and spill-over effect on volatility level. When our sample size is small, although the true ϕ_0 might not be zero, as the ϕ_0 is getting smaller, the chance to get a $\hat{\phi}$ nearly zero will getting larger. In this case, the identification nearly fails, and the LM statistic becomes meaningless since the empirical FOC of θ and the empirical information matrix will become very noisy. Also, as in applications, we never know the value of true ϕ_0, if $\hat{\phi}$ is very small and insignificant, we can simply say there is no spatial heteroskedasticity and spill-over effect on volatility level. In this case, the effect of GARCH-like term are almost absorbed by α, and the result from the noisy LM test is not important in general.

Another thing not showed here is the performance when residuals are not Normal. Similar to the performance of MLE for GARCH-like model, this LM test does not work well. Even though the MLE for spatial ARCH-like model performs good for other type of residuals which is showed in Table 7 and Table 8, the performance in Table 9 and Table 10 when introducing GARCH-like term is a disaster. Thus, although our LM statistic is derived under the assumption $\theta_0 = 0$, the FOC of

θ and the information matrix will have the wrong form for other distribution. It will make our LM statistic meaningless at all.

Table 11: Test Size with $N(0,1)$ Residuals ($\chi^2_{0.95} = 3.8415$)

n	correlation	$\alpha_0 = 1, \phi_0 = 0.2$	$\alpha_0 = 0, \phi_0 = 0.5$	$\alpha_0 = -1, \phi_0 = 0.8$
50	Circular Lake	0.091	0.121	0.105
	County Adjacent	0.064	0.085	0.089
200	Circular Lake	0.084	0.086	0.070
	County Adjacent	0.086	0.070	0.053
500	Circular Lake	0.072	0.056	0.049
	County Adjacent	0.070	0.048	0.059

Table 12 : Test Power with $N(0,1)$ Residuals ($\chi^2_{0.95} = 3.8415$)

n	correlation	$\alpha_0 = 1, \phi_0 = 0.3, \theta_0 = 0.4$	$\alpha_0 = -1, \phi_0 = 0.7, \theta_0 = -0.3$
50	Circular Lake	0.115	0.374
	County Adjacent	0.098	0.187
200	Circular Lake	0.435	0.790
	County Adjacent	0.241	0.267
500	Circular Lake	0.847	0.982
	County Adjacent	0.397	0.405

Table 13: Test Size with $N(0,1)$ Residuals ($\chi^2_{0.9} = 2.7055$)

n	correlation	$\alpha_0 = 1, \phi_0 = 0.2$	$\alpha_0 = 0, \phi_0 = 0.5$	$\alpha_0 = -1, \phi_0 = 0.8$
50	Circular Lake	0.095	0.153	0.166
	County Adjacent	0.065	0.161	0.121
200	Circular Lake	0.154	0.124	0.108
	County Adjacent	0.129	0.112	0.107
500	Circular Lake	0.1220	0.116	0.102
	County Adjacent	0.136	0.103	0.109

Table 14 : Test Power with $N(0,1)$ Residuals ($\chi^2_{0.9} = 2.7055$)

n	correlation	$\alpha_0 = 1, \phi_0 = 0.3, \theta_0 = 0.4$	$\alpha_0 = -1, \phi_0 = 0.7, \theta_0 = -0.3$
50	Circular Lake	0.180	0.4630
	County Adjacent	0.119	0.252
200	Circular Lake	0.565	0.853
	County Adjacent	0.285	0.372
500	Circular Lake	0.904	0.994
	County Adjacent	0.5010	0.553

Table 15 : Simulated Critical Values with $N(0,1)$ Residuals

n	correlation	$\alpha_0 = 1, \phi_0 = 0.2$			$\alpha_0 = 0, \phi_0 = 0.5$			$\alpha_0 = -1, \phi_0 = 0.8$		
		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
50	Circular Lake	2.48	5.26	23.69	4.44	8.79	43.78	3.95	6.64	17.42
	County Adjacent	1.65	3.37	14.06	3.85	6.49	27.05	3.42	5.22	12.19
200	Circular Lake	3.93	7.15	31.61	3.15	4.74	10.18	2.81	4.15	8.04
	County Adjacent	3.28	6.05	24.49	3.06	4.41	8.78	2.72	3.87	6.70
500	Circular Lake	3.27	5.00	12.39	2.81	4.17	7.63	2.74	3.98	7.09
	County Adjacent	3.47	5.82	22.21	2.81	4.14	7.11	2.78	3.98	6.88

6 An Application In U.S. Housing Market

6.1 Data Description

As a financial asset, risk management in housing market is important. Different from other financial asset, since real estate asset is adhere to a particular region, the spatial correlation among different regions can not be ignored. In existing literature, SAR model are used to capture the spatial correlation in housing research, e.g. Basu and Thibodeau (1998), Yong Tu, etc (2007). However, those SAR models only capture the spatial correlation at return level, but not volatility. Also, existing literatures are more focused on the price itself, but return rate especially the excess return compare to average housing market return, is more important to investors to make decisions. Some existing literature focused on the excess returns, e.g. Barry(1980), Draper and Findlay (1982), with CAPM models similar to analyzing stock returns. However, they totally ignored the spatial correlation among different regions. Our spatial heteroskedasticity model would be a proper model to capture the spatial correlation among the excess returns on housing market, and might be helpful to improve risk management in housing investment decision.

From Federal Housing Finance Agency, we can access the annual house price indexes at county level. The annual percentage change of HPI can be viewed as the average investment return. Although the HPI does not perfectly match the return of housing market, since rents and taxes are not included, and it counts totally price, not price per $sq.ft$, since county is a relatively large area which contains different types of houses, tradings in one year should be a good mixture of different types. Consider the Northeastern United States, it contains Washington, D.C. and 11 states: Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island and Vermont. In total, there are 245 counties in this area. However, since the data of New York county, NY (Manhattan, FIPS: 36061), which is the core area of New York city, are not available after 2014, . Thus, the time window we pick up is before 2014. Due to lack of trading data, other 5 counties namely, Sullivan County, PA (FIPS: 42113), Cameron County, PA(FIPS: 42023), Forest County, PA (FIPS: 42053), Juniata County, PA (FIPS: 42067), and Hamilton County, NY (FIPS:36061), are not included in our data set. Population size of these 5 counties are very small, with less than 30,000 in Juniata county, PA and less than 10,000 in the other four. We can expect that the house trading be inactive and not have significant impact on neighbor counties. Thus, 240 counties are included in our sample region.

The time window we use is from 2006 to 2014. One reason to use this time windows is the limitation of data accessibility as we stated before. The other reason is that we want to see whether the spatial correlation among the return volatility is affected by the business cycle and other time variant economics/demographic conditions. By NBER's US Business Cycle Expansions and Contractions research profile, December 2007 and June 2009 are the last peak and most recent trough, thus 2006~2014 covers several contraction and expansion periods. Also, one direct source of financial crisis in 2008 is subprime mortgage, which is closely correlated with the performance and risk of housing market. Focusing on this period, we can have a closer look on how spatial correlations among regions are interacting with the business cycle.

Table 16 are summary statistics of Annual Growth Rate of HPI in the selected 240 counties. We can clearly see that housing markets perform very different across regions. Even during the economic recession, some regions still have over 6% of annual return rate, which is much better than the stock market at the same time. Conversely, even during economic expansion and recover periods, the HPI in some areas still drop more than 4% annually. In each year, the volatility among different regions are large. Compare to the average annual growth rate, standard deviation is pretty large, which might indicate the heteroskedasticity across regions.

Table 16: Summary Statistics of Annual Growth Rate of HPI in 240 counties

	2006	2007	2008	2009	2010	2011	2012	2013	2014
Mean	7.09	2.07	-1.38	-4.45	-3.08	-2.28	-1.72	0.21	1.36
Minimum	-4.15	-4.25	-9.5	-19.83	-11.91	-9.53	-8.34	-6.99	-6.08
Maximum	32.27	12.75	10.39	6.70	6.54	7.15	6.50	8.08	11.88
s.t.d	4.05	3.00	3.33	3.97	2.81	2.65	2.39	2.02	2.86

6.2 Empirical Strategy

Since we want to estimate the spatial spill over effect on volatility, and to see whether it is affected by economic and geographic dynamics, running spatial ARCH-like model year by year is a reasonable strategy. Let $\bar{R}_t$ be the average annual growth rate of HPI across the 240 counties in year t, the excess return of region i in year t is

$$ER_{i,t} = R_{i,t} - \bar{R}_t$$

On one hand, since $E\left(ER_{i,t}\right) = 0$, it meets the requirement to apply the spatial ARCH-like model. On the other hand, as people care more about their relative performance of investment rather than the return itself, the correlation on the volatilities of excess returns would be much more helpful in real estate trading and risk management. Thus, we estimate the following model:

$$\begin{aligned} ER_{i,t} &= \sqrt{h_{i,t}}\varepsilon_{i,t}, \varepsilon_{i,t} \overset{iid}{\sim} (0,1) \\ logh_{i,t} &= \phi_t \sum_{j=1}^{n} w_{ij} logu_{j,t}^2 + \alpha_t \end{aligned}$$

for $t = 2006, \cdots, 2014$. The spatial correlation we consider here is county adjacent, i.e.

$$w_{ij} = \begin{cases} \frac{1}{n_i} & if\ i\ and\ j\ are\ neighbors \\ 0 & else \end{cases}$$

where n_i is the total number of neighbors of county i. Notice that we do not consider a county itself as its own neighbor, so $w_{ii} = 0$ for $\forall i$.

The estimation procedure follows the concentrated MLE estimation procedure introduced in previous chapters. To get the standard deviation of the estimator more precisely, instead of using the finite sample approximations of the asymptotic variance we derived in Section 3.5, we follow the residual bootstrap strategy introduced in Freedman and Peters (1984a) and Brock, etc (1992) which is widely used in finance and housing literatures. Once we get $\hat{\alpha}$ and $\hat{\phi}$, we can back-out $\hat{\varepsilon}_{i,t}$, and resample from $\hat{\varepsilon}_{i,t}$ and form a new artificial excess return process using resampled estimated residual terms and original estimators. Then, redo the MLE estimation on this new artificial process, we can get estimators $\widetilde{\alpha}$ and $\widetilde{\phi}$. By repeating the resampling process, we can get the standard deviation of those estimators. $z-$statistic can be constructed by the estimator divided by the bootstrap standard deviation. Since we proved the asymptotic normality of MLE estimator for the spatial ARCH-like model, and $\varepsilon_{i,t}$ is assumed to be $i.i.d$, this procedure is valid, and the $z-$statistic is Normal distributed. With this standard deviation, we can see how significant our estimators are. In this paper, we resampled 10,000 times to make sure precision. We also compute the Cox-Smell Pseudo R^2 for the concentrated likelihood function comparing to $\rho = 0$ case, which would be helpful to see how much the spatial spill-over effect helps to explain the total variance of excess return. In addition, due to our sample size, it is not suitable to directly estimate the spatial GARCH-like model, but there might be GARCH-like type spatial heteroskedasticity in the housing market return process. Here, we use the LM test proposed before to test whether the GARCH-like effect exist or not. In the following tables, we will report the LM statistic and its p-value.

Here are the results which showed in Table 17:

Table 17: Estimation Results For Excess Return

	2006	2007	2008	2009	2010	2011	2012	2013	2014
$\hat{\alpha}_t$	2.11	1.84	2.15	1.84	1.62	1.53	1.54	1.35	1.86
	(.17)	(.12)	(.16)	(.16)	(.13)	(.12)	(.11)	(.13)	(.12)
z_α	12.46	15.96	13.74	11.72	12.95	12.78	13.68	10.27	15.52
$\hat{\phi}_t$	.38	.32	.17	.41	.37	.39	.32	.34	.25
	(.078)	(.070)	(.082)	(.072)	(.076)	(.075)	(.076)	(.087)	(.076)
z_ϕ	4.84	4.65	2.01	5.76	4.90	5.22	4.15	3.89	3.32
Pseudo R^2 (Cox-Snell)	.15	.12	.03	.14	.16	.13	.10	.12	.08
LM statistic	2.43	0.78	11.71	-0.70	4.31	3.18	0.00	7.25	0.04
p-value	0.12	0.38	0.00	1	0.04	0.07	0.97	0.01	0.84

From Table 17, $\hat{\phi}_t$ for every sample year are significant at 1% level except 2008, but still with 5% significant level for 2008. Also, we can see that in most of the years, $\hat{\phi}_t$ is greater than 0.3. As we derived before, $Var\left(u_{i,n}\right) = \left[2\phi\left(\left(I_n - \phi W_n\right)^{-1} W_n\right)_{ii} + 1\right] E\left(h_{i,n}\right)$, from the correlation between $logh_{i,t}$ and $logu_{j,t}^2$, the magnitude of $\hat{\phi}$ indicates that when the variances of excess return increase by 10% for neighbors of county i, the variance of excess return of county i will be expected to increase by at least 3%. This result shows a significant spill-over effect among neighbor regions on the risk in housing market, both statistically and economically. Also, we can see that $\hat{\phi}_t$ is pretty persistent across years. The only significant drop is in 2008, since the global financial crisis happened and the economic environment of the US dramatically changed, the spill-over effect were relatively not that important. But still, it could not be ignored. Since we only capture the spatial correlation among neighbor counties, and do not allow direct correlation among counties apart from each other, their might be other types of spill-over effect on volatility. Also, there might be some non-geographic correlation among each county, such like labor force mobility and commodity tradings. However, due to lack of data for inter-county trades and labor mobility among counties, it is hard to measure the effect of economic correlation. Thus, the Cox-Snell Pseudo R^2 does not give an evidence that our model can explain most part of the variance. But only with the spill-over effect among neighbors, the likelihood function improves a lot. In traditional real estate finance literature, especially which discuss the risk and volatility, we focused more on mortgage and collateral channel, e.g. Cooper (2013). However, this is not enough due to our result. Mortgage and collateral channel are focused on specific individuals or areas, and do not contain any interaction among each regions. The LM test results reject $H_0 : \theta_0 = 0$ in 2008, 2010 and 2013, which indicates that there may exist spatial GARCH-like term.

6.3 Comparison with SAR model

As we pointed out, SAR model can only capture the spatial correlation at return level. However, intuitively, spill-over effect at return level will definitely have large impact on volatility, since the sources of externality are similar when consider the same spatial correlation structure. To see whether it is necessary to consider the spill-over effect on volatility level separately, we implement the following regressions on the excess returns: first fit the excess return processes with the SAR model, get the residual terms, and then fit the residuals with the spatial ARCH-like model. If the

SAR model can fully captures the spill-over effects, in the second round residual regressions, $\hat{\phi}_t$ should be small and insignificant. Table 18 and Table 19 are the estimation results:

Table 18: Estimation Results for SAR Fitting

	2006	2007	2008	2009	2010	2011	2012	2013	2014
$\hat{\rho}_t$	.75	.64	.82	.78	.66	.69	.72	.39	.51
	(.072)	(.089)	(.059)	(.066)	(.086)	(.082)	(.077)	(.12)	(.11)
t_ρ	10.33	7.16	13.90	11.85	7.70	8.44	9.33	3.19	4.81
R^2	.51	.37	.63	.63	.43	.46	.48	.11	.21

Table 19: Estimation Results for SAR Residuals

	2006	2007	2008	2009	2010	2011	2012	2013	2014
$\hat{\alpha}_t$	1.87	1.63	1.42	1.71	1.48	1.26	1.13	1.32	1.83
	(.23)	(.13)	(.16)	(.17)	(.17)	(0.13)	(.14)	(.13)	(.13)
z_α	8.30	12.62	9.11	10.06	8.78	9.74	8.34	10.08	13.87
$\hat{\phi}_t$	.31	.25	.10	.21	.14	.37	.12	.34	.15
	(.14)	(.094)	(.11)	(.11))	(.090)	(.086)	(.086)	(.082)	(.090)
z_ϕ	2.16	2.89	1.04	1.92	1.31	4.05	1.43	4.16	1.66
Pseudo R^2 (Cox-Snell)	.068	.011	.0059	.056	.047	.10	.0087	.13	.022
LM statistic	6.15	0.15	-0.15	10.89	-0.70	0.60	0.42	10.27	0.40
p-value	.01	.70	1	.00	1	.44	.52	.00	.53

In Table 18, similar to previous literatures, SAR model captures the spatial correlation and spill-over effect of housing market returns pretty well. However, after eliminate the first order effect, the residual terms still have strong spatial correlation on volatility level which is showed in Table 19. In 2006, 2007, 2009, 2011 and 2013, $\hat{\phi}_t$'s are still large and significant at 5% level. For other years, although statistically not very significant, $\hat{\phi}_t$'s are closer or larger than 0.1 which still shows some spatial correlation on the volatility level. Comparing to Table 2, $\hat{\phi}_t$ reduce after eliminate the first order effect, but the levels vary a lot for different years. In 2009, $\hat{\phi}_t$ decreases 0.2 after eliminate the SAR effect, however, it remains the same in 2013 and only decreases 0.02 in 2011. This indicates that for US housing market, the spill-over effect on volatility level can be partially explained by the spill-over effect on the return level, but the explanation power varies across time. The LM test reject $H_0 : \theta_0 = 0$ in 2006, 2009 and 2013, which is different from the result when directly estimate the spatial ARCH-like model on the excess return processes. It also indicates that the spill-over effect on the return level and volatility level may have different sources through the same spatial network, which could not be captured by SAR model. Combine the R^2 of SAR estimation and Cox-Snell Pseudo R^2 of the spatial ARCH-like estimation, the SAR model and spatial ARCH-like model have good explanation power on spatial externality though adjacent county networks.

Moreover, by comparing $\hat{\rho}_t$ and $\hat{\phi}_t$ from Table 17 to 19, the spatial correlation at return level and volatility level do not have a clear correlation. For example, in 2008, $\hat{\rho}_t$ is the largest among the 9-year time window, however $\hat{\phi}_t$ is the smallest and most insignificant one. Contrary, in 2013, $\hat{\phi}_t$ is significant and large, both fitting the original excess return and fitting the SAR residual terms. This indicates that the spatial correlation on return and volatility level are driven by different time varying factors. To fully learn the interaction between time-series dynamics and spatial spill-over effects on housing returns, a panel model is needed. But the result would be a starting point. Not only for housing market, but also some other financial markets may have similar characteristics. For example, the return of municipal bonds are highly correlated with the economic performance in a particular city or state, thus spill-over effect among neighbors on risk would also be large.

International stock market may be a more significant example. Asian financial crisis in 1997 and European debt crisis in 2009 are both good examples on how risk transmit from one or a few economies to neighbor economies. By similar empirical strategy in this paper, we would expect a significant $\hat{\phi}$ for each time period when using their excess stock returns. This model would be very helpful to test the spill-over effect on volatility in a particular area.

References

[1] Anil K. Bera and Pradosh Simlai, 2005, "Testing spatial autoregressive model and a formulation of spatial ARCH(SARCH) model with applications", *Econometric Society World Congress, London*

[2] Borovkova, S. and R. Lopuhaa, 2012, "Spatial GARCH: A spatial approach to multivariate volatility modeling". *Working paper*

[3] Caporin, M. and P. Paruolo, 2006, "GARCH models with spatial structure". *SIS Statistica*, 447–450

[4] Daniel Cooper, "House Price Fluctuations: The Role of Housing Wealth as Borrowing Collateral", *The Review of Economics and Statistics*, October 2013, Vol. 95, No. 4

[5] David A. Freedman and Stephen C. Peters, "Bootstrapping a Regression Equation: Some Empirical Results", *Journal of the American Statistical Association*, Mar. 1984, Vol. 79, No. 385

[6] Dennis W. Draper and M. Chapman Findlay, "Capital Asset Pricing and Real Estate Valuation", *Real Estate Economics*, June 1982, Vol. 10

[7] Harry H. Kelejian and Ingmar R. Prucha, 2006, "HAC estimation in a spatial framework". *Journal of Econometrics*, 140 131-154

[8] Nazgul Jenish, Ingmar R.Prucha, 2009, "Central Limit Theorems and Uniform Law of Large Numbers for Arrays of Random Fields". *Journal of Econometrics*, 150 86-98

[9] Peter J.Berry, "Capital Asset Pricing and Farm Real Estate", *American Journal of Agricultural Economics*, Aug., 1980, Vol. 62, No. 3

[10] Peter M.Lee, 2012, "Bayesian Statistics: An Introduction", fourth edition, Wiley.

[11] Sabyasachi Basu and Thomas G. Thibodeau, "Analysis of Spatial Autocorrelation in House Prices", *Journal of Real Estate Finance and Economics*, 1998, Vol: 17:1

[12] Sato, T. and Y. Matsuda, 2017, "Spatial Autoregressive Conditional Heteroskedasticity Models", *Journal of the Japan Statistical Society*, 47, 221–236

[13] Steven Kou, Xianhua Peng, Haowen Zhong, 2017, "Asset Pricing with Spatial Interactions", *Management Science*, 64(5), 2083-2101

[14] Richard J. Richmond, 2019, "Trade Network Centrality and Currency Risk Premia", *The Journal of Finance*, vol, No. 0

[15] Whitney K. Newey,1991, "Uniform Convergence in Probability and Stochastic Equicontinuity", *Econometrica*, Vol. 59, 1161-1167

[16] Wiliam Brock, Josef Lakonishok and Blake LeBaron, “Simple Technical Trading Rules and the Stochastic Properties of Stock Returns”, *The Journal of Finance*, Dec., 1992, Vol. 47, No. 5

[17] Xingbai Xu, Lung-fei Lee, 2019, "Conditions for $\alpha-$mixing Linear Random Fields: with Applications to Spatial Econometric Models". *Working paper*

[18] Xingbai Xu, Lung-fei Lee, 2015b, "Maximum Likelihood Estimation of a Spatial Autoregressive Tobit Model". *Journal of Econometrics*, 186 1-18

[19] Yong Tu, Hua Sun and Shi-Ming Yu, “Spatial Autocorrelations and Urban Housing Market Segmentation”, *Journal of Real Estate Finance and Economics*, 2007, Vol: 34