DSSR

Discussion Paper No. 78

Spatial GARCH Models

Takaki Sato
Yasumasa Matsuda

March , 2018

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Spatial GARCH Models

Takaki Sato * Yasumasa Matsuda †

Abstract

This study proposes a spatial extension of time series generalized autoregressive conditional heteroscedasticity (GARCH) models. We call the spatial extended GARCH models as spatial GARCH (S-GARCH) models. S-GARCH models specify conditional variances given simultaneous observations, which constitutes a good contrast with time series GARCH models that specify conditional variances given past observations. The S-GARCH model are transformed into a spatial autoregressive moving-average (SARMA) model and the parameters of the S-GARCH model are estimated by a two step procedure. First step estimation is the quasi maximum likelihood (QML) estimation method and consistency and asymptotic normality of the proposed QML estimators are given. Second step is estimation of an intercept term by the estimator derived from another QML to avoid bias in first step and consistency of the estimator is shown. We demonstrate empirical properties of the model by simulation studies and real data analyses of land price data in Tokyo areas. We find the estimators have small bias regardless of distributions of error terms from simulation studies and real data analyses show that spatial volatility in land price has global spillover and volatility clustering, namely units with higher spatial volatility are clustered in some specific districts like time series financial data.

Keywords: GARCH model, Spatial ARMA model, Quasi maximum likelihood, areal data, spatial volatility.

1 Introduction

Volatility models for time series financial data have developed with their application in academia and the financial industry. The seminal work by Engle (1982) introduces the autoregressive conditional heteroscedasticity (ARCH) model and Bollerslev (1986) proposes a extension known as the generalized ARCH (GARCH) model. These models are widely used to model and forecast volatility of univariate time series data for calculation of the price of options or value at risk of a financial position in risk management. Subsequently, Multivariate extensions of univariate models are proposed by Bollerslev et al (1988), Bollerslev (1990) and Engle and Kroner (1995) for modeling dynamic relationships between volatility of multiple asset returns. A major challenge of multivariate volatility modeling is to overcome the curse of dimensionality; there are $n(n+1)/2$ variances and covariances for n-dimensional asset return series. One solution for the problem is consider simpler structures of covariance matrices to reduce parameters.

The ideas of spatial econometrics have been applied to volatility models in recent years. Two main objectives of the applications are to reduce parameters in covariance matrices and to extend time series volatility models to spatial models for spatial data. Caporin and Paruolo (2008) and Borovkova and Lopuhaa (2012) have applied the ideas of spatial econometrics to time series multivariate GARCH models from the former view point. On the other hand, Yan (2007) and Robinson (2009) have done a spatial extension of stochastic volatility models which are another kind of volatility models and Sato and Matsuda (2017) have extend time series ARCH models to spatial ARCH (S-ARCH) models from both view points.

This paper aims to extend S-ARCH models to spatial generalized ARCH (S-GARCH) models. The S-GARCH model have two interesting features. Firstly, volatility at a point or an area in map is specified by surrounding observations in the S-ARCH model, whereas that of the S-GARCH model is characterized by surrounding observations and volatility. Thus, the S-GARCH model captures global spatial spillover in

*Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan. takaki.sato.s4@dc.tohoku.ac.jp

†Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan. matsuda@econ.tohoku.ac.jp

volatility in spatial data. Secondly, the S-GARCH model can be transformed into a spatial autoregressive moving average (SARMA) model. This means the existence condition of the S-GARCH model are easily established.

Parameters in the S-GARCH model are estimated by the quasi-maximum likelihood (QML) estimation method and we show the QMLE estimators have consistency and asymptotic normality. Two estimation methods are basically used in spatial econometrics literature. First one is the moment method. Kelejian and Robinson (1993) and Kelejian and Prucha (1997, 1998) propose two stage least squares estimation methods and Lee (2007) propose the generalized method of moments (GMM) for spatial autoregressive (SAR) models and spatial autoregressive models which also have a spatial autoregressive process in disturbances (SARAR). Moreover, Dogan and Taspinar (2013) consider the GMM methodology for spatial autoregressive models with moving average disturbances (SARMA). Second one is the QML estimation method. Lee (2004), Yu et al (2008), Su and Yang (2015) propose the QML estimation method for SAR models and spatial dynamic panel (SDP) models and also Yang (2015) shows M-estimator based on the QML for SDP models which have spatial autoregressive process in both dependent variables and disturbances. However, asymptotic properties of the QML estimator for SARMA models has not been discussed. As mentioned above, S-GARCH models can be transformed into SARMA models. Therefore, we show asymptotic properties of the QML estimator for SARMA models.

This paper proceeds as follows. Section 2 introduces the S-GARCH model and discusses characteristics of the model. Estimation methods for the model and asymptotic properties of the estimators are derived in Section 3. Section 4 examines empirical properties of the model by applying to simulated and land price data in Tokyo area. Section 5 concludes the paper. All the proofs are collected in the Appendix.

2 Model specification

We consider the S-GARCH model of the form

$$\begin{aligned} y_i &= \sqrt{h_i}\varepsilon_i, \\ \log h_i &= \lambda \sum_{j=1}^{n} w_{i,j} \log h_j + \rho \sum_{j=1}^{n} w_{i,j} \log y_j^2 + \alpha + \boldsymbol{z}_i' \delta, \\ i &= 1, \ldots, n, \end{aligned}$$

where y_i is an areal data, $\sqrt{h_i}$ is volatility, ε_i is an independent and identically distributed (i.i.d) random variable with zero mean and variance 1, $\boldsymbol{z}_i$ is $(k \times 1)$ non-stochastic regressors, and $w_{i,j}$ is a spatial weight which is predetermined and quantifies a closeness from area i to area j. Parameters in this model are $(\lambda, \rho, \alpha, \delta')'$. Scalar parameters λ and ρ characterizes the simultaneous effect, α is an intercept term and δ is usual regression coefficients.

The S-GARCH model is different from the time series GARCH model is the following two points. First one is a description of volatility. Spatial volatility in the S-GARCH model is described by observations and volatility at all other units, on the other hands time series volatility is defined by past observations and volatility following the flow of time. Although the descriptions of time series and spatial volatility are different, we have found in this paper that they have the similar properties. For instance, volatility clustering exists, namely large changes tend to be followed by large changes and small changes tend to be followed by small changes. This is a stylized feature of financial time series data and land price data also has the similar property that a large change at one area leads to large changes at surrounding areas.

Second one is the log transformation of volatility. Log transformation is used to ensure the existence of areal data y_i. If we define non logarithmic volatility, it would be difficult to guarantee the existence condition unlike that of time series GARCH models that can be derived from Markov process theories (Fan and Yao (2003)). On the other hand, the log transformation of volatility makes it much easier to show the existence condition because the S-GARCH model can be transformed into the spatial autoregressive moving average (SARMA) model as shown below and the existence condition of the SARMA model is already known.

Let us introduce the following SARMA transformation of the S-GARCH model. Denoting $\log \boldsymbol{y}^2 = (\log y_1^2, \ldots, \log y_n^2)'$, $\log \boldsymbol{h} = (\log h_1, \ldots, \log h_n)'$, $\log \boldsymbol{\varepsilon}^2 = (\log \varepsilon_1^2, \ldots, \log \varepsilon_n^2)'$, $Z_n = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)'$, $\boldsymbol{1}_n = (1, \ldots, 1)'$ and I_n is a $n \times n$ identity matrix, the model has the following vector form representation,

$$\begin{aligned}\log \boldsymbol{y}^2 &= \log \boldsymbol{h} + \log \boldsymbol{\varepsilon}^2 &(1)\\ \log \boldsymbol{h} &= \lambda W_n \log \boldsymbol{h} + \rho W_n \log \boldsymbol{y}^2 + \alpha \mathbf{1}_n + Z_n\delta, &(2)\end{aligned}$$

where W_n is a spatial weight matrix whose elements are $w_{i,j}$. From (2),

$$\log \boldsymbol{h} = (I_n - \lambda W_n)^{-1}(\rho W_n \log \boldsymbol{y}^2 + \alpha \mathbf{1}_n + Z_n\delta),$$

By substituting (2) into (1),

$$\begin{aligned}\log \boldsymbol{y}^2 &= (I_n - \lambda W_n)^{-1}(\rho W_n \log \boldsymbol{y}^2 + \alpha \mathbf{1}_n + Z_n\delta) + \log \boldsymbol{\varepsilon}^2,\\ (I_n - \lambda W_n)\log \boldsymbol{y}^2 &= \rho W_n \log \boldsymbol{y}^2 + \alpha \mathbf{1}_n + Z_n\delta + (I_n - \lambda W_n)\log \boldsymbol{\varepsilon}^2,\\ \log \boldsymbol{y}^2 &= \lambda W_n \log \boldsymbol{y}^2 + \rho W_n \log \boldsymbol{y}^2 + \alpha \mathbf{1}_n + Z_n\delta + (I_n - \lambda W_n)\log \boldsymbol{\varepsilon}^2. &(3)\end{aligned}$$

This is the SARMA model and the existence condition holds when $|\lambda| + |\rho| < 1$.

3 Estimation

We consider the estimation of the parameters $(\lambda, \rho, \alpha, \delta')'$ and the asymptotic properties of the estimators. Parameters are estimated by a two step procedure. First step is the estimation of $(\lambda, \rho, \delta')'$ by the QML estimation method. The proposed QML estimator are consistent and asymptotically normal. However, $\log \boldsymbol{\varepsilon}^2$ in (3) is not zero mean error terms. Thus, the estimator for α in the first step has bias, therefore we need to estimate α by another method. In second step, α is estimated with consistent estimator derived from the QML based on the likelihood different from the one in the first step.

3.1 First step estimation

Parameters λ, ρ and δ are estimated in first step by the QML estimation method.

To apply the QML estimation method, we need to modify the error term because the mean of $\log \boldsymbol{\varepsilon}^2$ in (3) is not zero as already mentioned. From (3),

$$\begin{aligned}\alpha \mathbf{1}_n + (I_n - \lambda W_n)\log \boldsymbol{\varepsilon}^2 &= \alpha \mathbf{1}_n + (I_n - \lambda W_n)\{\log \boldsymbol{\varepsilon}^2 - E(\log \varepsilon_1^2)\mathbf{1}_n + E(\log \varepsilon_1^2)\mathbf{1}_n\},\\ &= \{\alpha + (1-\lambda)E(\log \varepsilon_1^2)\}\mathbf{1}_n + (I_n - \lambda W_n)\{\log \boldsymbol{\varepsilon}^2 - E(\log \varepsilon_1^2)\mathbf{1}_n\}.\end{aligned}$$

Noting that intercept term has a bias by $(1-\lambda)E(\log \varepsilon_1^2)$.

Denoting $Y_n = \log \boldsymbol{y}^2$, $X_n = [\mathbf{1}_n, \boldsymbol{Z}]$, $V_n = \{\log \boldsymbol{\varepsilon}^2 - E(\log \varepsilon_1^2)\mathbf{1}_n\}$ and $\beta = (\{\alpha + (1-\lambda)E(\log \varepsilon_1^2)\}, \delta')'$, the model has the following representation,

$$Y_n = \lambda W_n Y_n + \rho W_n Y_n + X_n\beta + (I_n - \lambda W_n)V_n, \tag{4}$$

where V_n is already zero mean processes.

Now, let us consider the QML estimation of (4). Regarding v_i's as independent Gaussian variables with mean zero and variance σ^2, the likelihood function of (4) is

$$\log L_n(\psi) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{V_n'(\theta,\beta)V_n(\theta,\beta)}{2\sigma^2} - \log|R_n(\lambda)| + \log|S_n(\theta)|, \tag{5}$$

where $\theta = (\lambda, \rho)'$, $\psi = (\beta', \sigma_2, \theta')'$, $R_n(\lambda) = I_n - \lambda W_n$, $R_n = I_n - \lambda_0 W_n$, $S_n(\theta) = I_n - \lambda W_n - \rho W_n$, $S_n = I_n - \lambda_0 W_n - \rho_0 W_n$ and $V_n(\theta, \beta) = R_n^{-1}(\lambda)[S_n(\theta)Y_n - X_n\beta]$. The QML estimator is the extreme estimator derived form the maximization of (5).

It is convenient to work with the concentrated likelihood by concentrating β and σ^2 out for computation and asymptotic analysis on the estimator. From the first order condition of (5), the concentrated QML estimators of β and σ^2 is

$$\hat{\beta}_n(\theta) = (X_n' R_n'^{-1}(\lambda) R_n^{-1} X_n)^{-1} X_n' R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) Y_n,$$

$$\hat{\sigma}_n^2(\theta) \quad = \quad \frac{\hat{V}_n'(\theta)\hat{V}_n(\theta)}{n},$$

where $\hat{V}_n(\theta) = R_n^{-1}(\lambda)[S_n(\theta)Y_n - X_n\hat{\beta}_n(\theta)]$. The concentrated likelihood function of θ is

$$\log L_n(\theta) \quad = \quad -\frac{n}{2}(\log(2\pi)+1) - \frac{n}{2}\log\hat{\sigma}_n^2(\theta) - \log|R_n(\lambda)| + \log|S_n(\theta)|. \tag{6}$$

The QML estimator $\hat{\theta}_n$ maximizes the concentrated likelihood function (6) and the QML estimators of β and σ^2 are $\hat{\beta}_n(\hat{\theta}_n)$ and $\hat{\sigma}_n^2(\hat{\theta}_n)$, respectively.

For our analysis of the asymptotic properties of first step estimators, we need the following assumptions:

Assumption 1. *The disturbances $\{v_i\}, i = 1, \ldots n$ are i.i.d. across i with zero mean, variance σ_0^2 and $E|v_i|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption 2. *The elements $w_{n,ij}$ of W_n are nonnegative and row normalized and the column sums of W_n are uniformly bounded.*

Assumption 3. *The space Θ is compact, and the true parameter θ_0 lies in its interior.*

Assumption 4. *The matrix $S_n, S_n(\theta), R_n,$ and $R_n(\lambda)$ are uniformly bounded both row and column sums and nonsingular.*

Assumption 5. *The elements of X_n are uniformly bounded constants. The $\lim_{n\to\infty}\frac{1}{n}(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)$ exists and is nonsingular.*

Assumption 6. $0 \leq \underline{c}_y \leq \inf_{\theta\in\Theta}\gamma_{min}(Var(S_n(\theta)Y_n)) \leq \sup_{\theta\in\Theta}\gamma_{max}(Var(S_n(\theta)Y_n)) \leq \bar{c}_y < \infty.$

Assumption 7. $0 \leq \underline{c}_r \leq \inf_{\lambda\in\Lambda}\gamma_{min}(R_n'^{-1}(\lambda)R_n^{-1}(\lambda)) \leq \sup_{\lambda\in\Lambda}\gamma_{max}(R_n'^{-1}(\lambda)R_n^{-1}(\lambda)) \leq \bar{c}_r < \infty$

Assumption 8. $\lim_{n\to\infty}\frac{1}{n}\beta_0'X_n'S_n'^{-1}S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)S_n^{-1}X_n\beta_0 \neq 0,$
where $M_n = I_n - R_n^{-1}X_n(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n'^{-1}$.

To derive the consistency of the QML estimators, we need to show the identification of θ_0. Define $Q_n(\theta) = \max_{\beta,\sigma^2} E(\log L_n(\psi))$. The optimal solutions of this maximization problem are given by

$$\begin{aligned}
\beta_n^*(\theta) &= (X_n'R_n'^{-1}(\lambda)R_n^{-1}X_n)^{-1}X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)S_n(\theta)E(Y_n),\\
\sigma_n^{2*} &= \frac{1}{n}E(V_n'^*(\theta)V_n^*(\theta)),
\end{aligned}$$

where $V_n^*(\theta) = R_n^{-1}(\lambda)[S_n(\theta)Y_n - X_n\beta_n^*(\theta)]$. Therefore,

$$Q_n(\theta) \quad = \quad -\frac{n}{2}(\log(2\pi)+1) - \frac{n}{2}\log\sigma_n^{2*}(\theta) - \log|R_n(\lambda)| + \log|S_n(\theta)|,$$

and identification of θ_0 is based on $\frac{1}{n}Q_n(\theta)$.

Consistency of the QML estimators $\hat{\theta}$ follows from the uniform convergence of $\frac{1}{n}\log L_n(\theta) - \frac{1}{n}Q_n(\theta)$ to zero on Θ and identification of θ_0.

Theorem 1. *Under Assumptions 1-8, θ_0 is globally identifiable and θ_n is a consistent estimator of θ_0.*

To derive the asymptotic distribution of the QMLE $\hat{\psi}_n$, we need to make the Taylor expansion of $\frac{\partial}{\partial\psi}\log L_n(\hat{\psi}) = 0$ at ψ_0. The first-order derivatives of the log-likelihood function at ψ_0 are

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\psi_0)}{\partial\beta} &= \frac{1}{\sigma_0^2\sqrt{n}}X_n'R_n'^{-1}V_n,\\
\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\psi_0)}{\partial\sigma^2} &= \frac{1}{2\sigma_0^4\sqrt{n}}(V_n'V_n - n\sigma_0^2),\\
\frac{1}{\sqrt{n}}\frac{\partial\log L_n(\psi_0)}{\partial\rho} &= \frac{1}{\sigma_0^2\sqrt{n}}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}V_n + \frac{1}{\sigma_0^2\sqrt{n}}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}V_n - \sigma_0^2 tr(S_n^{-1}W_n)),
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial \log L_n(\psi_0)}{\partial \lambda} &= \frac{1}{\sigma_0^2\sqrt{n}}\beta_0' X_n' S_n'^{-1} W_n' R_n'^{-1} V_n + \\
&\quad \frac{1}{\sigma_0^2\sqrt{n}}(V_n'(R_n' S_n'^{-1} W_n' R_n'^{-1} - W_n' R_n'^{-1})V_n - \sigma_0^2 tr(S_n^{-1}W_n) + \sigma_0^2 tr(R_n^{-1}W_n)),
\end{aligned}
$$

where $tr(\cdot)$ denote the trace of a matrix.

These involve linear and quadratic function of V_n. The asymptotic distribution of these score functions are derived from the central limit theorems for linear-quadratic forms in Kelejian and Prucha (2001).

The variance matrix of $\frac{1}{\sqrt{n}}\frac{\partial}{\partial\psi}\log L_n(\psi_0)$ is

$$
E\bigg(\frac{1}{\sqrt{n}}\frac{\partial \log L_n(\psi_0)}{\partial \psi}\frac{1}{\sqrt{n}}\frac{\partial \log L_n(\psi_0)}{\partial \psi'}\bigg) = -E\bigg(\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial \psi\psi'}\bigg) + \Omega_{\psi,n},
$$

where $-E\big(\frac{1}{n}\frac{\partial^2}{\partial\psi\psi'}\log L_n(\psi_0)\big)$ is the average Hessian matrix and $\Omega_{\psi,n}$ is a symmetric matrix and both are given in Appendix A. When V_n is normally distributed, $\Omega_{\psi,n} = 0$.

The score function and Hessian matrix have proper asymptotic behavior, therefore we have the following theorem.

Theorem 2. *Under Assumptions 1-8,*

$$
\sqrt{n}\big(\hat{\psi}_n - \psi_0\big) \xrightarrow{d} N\big(0, \Sigma_\psi^{-1} + \Sigma_\psi^{-1}\Omega_\psi\Sigma_\psi^{-1}\big),
$$

where $\Sigma_\psi = -\lim_{n\to\infty} E\big(\frac{1}{n}\frac{\partial^2}{\partial\psi\psi'}\log L_n(\psi_0)\big)$ *and* $\Omega_\psi = \lim_{n\to\infty}\Omega_{\psi,n}$. Σ_ψ *and* Ω_ψ *assume to exist and* $-\Sigma_\psi$ *to be positive definite, sufficiently large* n. *When errors are normally distributed,* $\sqrt{n}\big(\hat{\psi}_n - \psi_0\big) \xrightarrow{d} N\big(0, \Sigma_\psi^{-1}\big)$.

3.2 Second step estimation

Let us consider the estimation of α in the second step. As we mentioned, $\beta_1 = \{\alpha + (1-\lambda)E(\log\varepsilon_1^2)\}$ in the first step has the bias. Therefore, we need to estimate α separately.

We regard ε_i in (3) as independent Gaussian variables, not $\log\varepsilon_i^2$. Then, $\log\varepsilon_i^2$ follows a log chi-squared distribution on 1 degress of freedom (Lee 2012, p379). The probability density function of $\log\varepsilon_i^2$ is given by

$$
f(x) = \frac{1}{\sqrt{2\pi}}\exp\bigg(-\frac{1}{2}\exp(x) + \frac{1}{2}x\bigg). \tag{7}
$$

For notational purposes, we define $\phi = (\lambda, \rho, \alpha, \delta')'$, $Y_n = \log\boldsymbol{y}^2$ and $U = \log\boldsymbol{\varepsilon}^2$. Then, from (3),

$$
\begin{aligned}
U_n(\phi) &= R_n^{-1}(\lambda)(S(\theta)Y - \alpha\mathbf{1}_n - Z_n\delta), \\
&= R_n^{-1}(\lambda)(S(\theta)Y - Z_n\delta) - \frac{\alpha}{1-\lambda}\mathbf{1}_n, \\
&= C - \frac{\alpha}{1-\lambda}\mathbf{1}_n,
\end{aligned}
$$

where $C = R_n^{-1}(\lambda)(S(\theta)Y - Z_n\delta)$.

Therefore, the likelihood function based on the density (7) is

$$
\log L_n(\phi) = \frac{n}{2}\log 2\pi - \sum_{i=1}^{n}\bigg\{-\frac{1}{2}\frac{\exp(C_i)}{\exp(\frac{\alpha}{1-\lambda})} + \frac{1}{2}\bigg(C_i - \frac{\alpha}{1-\lambda}\bigg)\bigg\} - \log|R_n(\lambda)| + \log|S_n(\theta)|,
$$

where C_i is the i-th element of C.

Differentiating it with respect to α, the concentrated QML estimator of α given λ, ρ and δ is

$$
\alpha_n(\lambda,\rho,\delta) = (1-\lambda)\log\bigg(\frac{1}{n}\sum_{i=1}^{n}\exp(C_i)\bigg)
$$

Finally, substituting the proposed QML estimator $(\hat{\lambda}, \hat{\rho}, \hat{\delta}')$ in the first step for $(\lambda, \rho, \delta')'$, we propose

$$\hat{\alpha}_n = (1 - \hat{\lambda}) \log\biggl(\frac{1}{n} \sum_{i=1}^{n} \exp\bigl[\bigl\{ R_n^{-1}(\hat{\lambda})(S(\hat{\theta})Y_n - Z_n\hat{\delta})\bigr\}_i\bigr] \biggr), \tag{8}$$

as an estimator for α.

The estimator $\hat{\alpha}_n$ has consistency.

Theorem 3. *Under Assumptions 1-8, $\hat{\alpha}_n$ is a consistent estimator of α_0.*

4 Empirical analysis

We examine the empirical properties of the S-GARCH model by applying to simulated and land price data in Tokyo areas. Monte Carlo experiments are carried out to investigate the finite sample performance of the proposed estimators.

4.1 Simulation studies

To investigate finite sample properties of the proposed estimators, we use the following data generating process:

$$\begin{aligned} y_i &= \sqrt{h_i}\varepsilon_i, \\ \log h_i &= \lambda \sum_{j=1}^{n} w_{i,j} \log h_j + \rho \sum_{j=1}^{n} w_{i,j} \log y_j^2 + \alpha + x_i\beta, \end{aligned}$$

where x_i's are randomly generated form independent normal distributions and the spatial weights matrix is generated according to Rook contiguity and row normalizing. The error, ε_i, distributions are (i) standard normal distributions, (ii) chi-squared distributions with 3 degrees of freedom and (iii) log normal distributions. Let $\phi = (\lambda, \rho, \alpha, \beta)'$. We choose $\phi_0^1 = (0.9, 0.05, 0.5, 1)'$, $\phi_0^2 = (0.45, 0.45, 0.5, 1)'$, $\phi_0^3 = (0.05, 0.9, 0.5, 1)'$ and $n = 100$ or $n = 400$. Each set of Monte Carlo results is based on 1000 samples and the parameters are estimated by the two step procedure.

The empirical means and square root of mean squared errors (RMSE) for the proposed estimators are reported in Table 1. The results show the estimators in the firs step, $(\hat{\lambda}, \hat{\rho}, \hat{\beta}')'$ are nearly unbiased and not sensitive to the choice of the error distributions. On the other hand, the second step estimator, $\hat{\alpha}$ depends on true parameters and the error distribution. Small λ may be attributed to the poor performance of the estimator because $1 - \hat{\lambda}$ in (8) effects on estimated errors from true value as shown in the proof of Theorem 3. Moreover, as the error distribution is more discrepant from the Gaussian distribution, the estimator has bigger bias and less efficiency. However, the empirical performances of the estimator improve as n becomes larger.

4.2 Land price data analysis

We apply the S-GARCH model to land price data in Tokyo area

Let us introduce land price data used in this section. We use prefectural land price research as land price data. The Japanese Ministry of Land, Infrastructure, Transport, and Tourism publishes land prices on sampling points scattered irregularly all over Japan in the form of price per m^2 in July. We focus on the land prices over Tokyo area (Tokyo, Kanagawa, Saitama, Chiba, Tochigi, Ibaraki, Gunma) observed form 2009 to 2014. The log returns of the land prices are averaged in municipal units. Therefore, our data set consist of 344 discrete unit's average log returns from 2010 to 2014.

Before application of the S-GARCH model, we remove spatial correlations in data with the spatial autoregressive (SAR) model year by year. This modification is similar to that we apply the ARMA model to data before fitting the GARCH model to remove correlation in time series analysis. The SAR model is

$$y_i = \zeta + \kappa \sum_{j=1}^{344} w_{i,j} y_j + u_{i,t}, u_{i,t} \sim i.i.d(0, \tau^2).$$

where $W = (w_{i,j})$ is given the first-order contiguity relation that takes 1 when two units have a common boarder.

We apply the S-GARCH model to the residuals obtained after fitting the SAR model year by year, where the same spatial weight matrix as one in the SAR model was employed. Explanatory variables are intercept term and each unit's area. Areas of observations are included to hold Assumption 8 which is important for the identification uniqueness. Table 2 shows the estimated values of λ, ρ, α and β. Here, α and β is intercept and the coefficient of logarithm of areas, respectively. The standard errors of $\hat{\lambda}$ and $\hat{\rho}$ are derived in Theorem 2 by replacing the population moments with the corresponding sample moments. Figure 1 express the spatial volatility evaluated by

$$\log \hat{\boldsymbol{h}} = (I_n - \hat{\lambda} W_n)^{-1}(\hat{\rho} W_n \log \boldsymbol{y}^2 + \hat{\alpha}\mathbf{1}_n + \boldsymbol{x}\hat{\beta}),$$

where $\boldsymbol{x}$ is the vector of the areas of observations.

we find estimates of spatial correlation of volatility, λ, are significant after the Great East Japan Earthquake in 2011 until 2013 from Table 2. This may show that volatility in land prices have strong correlation when a big event occurs. The effects from simultaneous returns, ρ, are not large and this is similar to empirical results of the time series GARCH model. The sum $\hat{\lambda} + \hat{\rho}$ takes near 1 values between 2011 and 2013. Thus, volatility is persistent to far areas and may generate volatility clustering. From Figure 1, not only the volatility of costal area which hit by the Tsunami but also that of near Fukushima areas is high. This may be caused as the result of Fukushima nuclear accident. Moreover, we find the volatility clustering as explained above. Therefore, volatility in land price takes similar behavior to that of time series financial data like stock price. In addition, we can find volatility in land price has global spillover from figure 2. The model fitting of the S-GARCH model to the data is better than that of the S-ARCH model. The estimated volatility of the S-ARCH model makes small clusters. On the other hand, that of the S-GARCH model generates large clusters. This result shows that the estimated volatility is globally strongly spatially correlated.

5 Conclusion

We have proposed a spatial generalized autoregressive conditional heteroskedasticity (S-GARCH) model as extension of a spatial autoregressive conditional heteroskedasticity (S-ARCH) model (Sato and Matsuda (2017)) to evaluate spatial volatility. The S-GARCH models can be expressed in the form of a spatial autoregressive moving average (SARMA) model and we propose the two step estimation procedure to estimate the parameters in the model. The quasi maximum likelihood (QML) estimators in each step have desired asymptotic properties. Finite sample performances of the estimators are reasonably good from Monte Carlo experiments. We find volatility in land prices is similar behavior to that of time series data from real data analysis.

For future research, we describe possible extensions. We used the first-order contiguity relations to make the spatial weight matrix. The choice of spatial weight matrices is important matter in empirical analysis. Thus, applying other spatial weight matrices can improve our volatility analysis using the S-GARCH model. Moreover, spatiotemporal extension of the S-GARCH model which considers effects from both space and time would make it possible to analyze volatility structures in more detail.

Table 1: The empirical means and square root of mean squared errors (RMSE) of the estimators.

	normal				chi(3)				log normal			
	n=100		n=400		n=100		n=400		n=100		n=400	
ϕ	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.9	0.029	0.082	0.007	0.029	0.032	0.078	0.009	0.030	0.031	0.080	0.009	0.029
0.05	-0.039	0.069	-0.009	0.026	-0.040	0.066	-0.010	0.027	-0.038	0.068	-0.011	0.026
0.5	0.039	0.378	0.006	0.105	0.015	0.310	-0.003	0.100	-0.037	0.310	-0.018	0.101
1.0	0.021	0.188	0.009	0.089	0.023	0.176	0.004	0.082	0.020	0.173	0.004	0.077
0.45	-0.060	0.238	-0.015	0.098	-0.065	0.243	-0.015	0.103	-0.053	0.224	-0.016	0.097
0.45	-0.001	0.155	0.002	0.072	0.002	0.159	0.002	0.075	0.002	0.151	0.003	0.073
0.5	-0.014	0.292	-0.002	0.092	-0.054	0.313	-0.007	0.113	-0.277	0.595	-0.086	0.255
1.0	0.034	0.232	0.012	0.113	0.044	0.229	0.018	0.109	0.041	0.216	0.007	0.103
0.05	-0.027	0.139	-0.011	0.080	-0.023	0.141	-0.011	0.079	-0.029	0.132	-0.013	0.079
0.9	-0.011	0.108	0.002	0.069	-0.017	0.115	0.002	0.068	-0.006	0.108	0.003	0.068
0.5	-0.431	0.829	-0.100	0.240	-0.627	1.089	-0.114	0.295	-1.009	1.660	-0.313	0.630
1.0	0.013	0.236	0.007	0.114	0.022	0.228	0.006	0.109	0.007	0.215	0.004	0.105

Note: $\phi = (\lambda, \rho, \alpha, \beta)'$

Table 2: Estimated values and standard errors of λ and ρ and estimated values of α and β in the S-ARCH model and the S-GARCH model applied to the residuals by fitting the SAR model to log returns of land priced data year by year.

	S-ARCH					S-GARCH				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
$\hat{\lambda}$						0.772	0.845	0.874	0.893	0.601
se(λ)						0.206	0.139	0.128	0.100	0.415
$\hat{\rho}$	0.240	0.244	0.274	0.279	0.184	0.110	0.076	0.059	0.060	0.104
se(ρ)	0.083	0.081	0.082	0.083	0.084	0.077	0.055	0.048	0.045	0.086
$\hat{\alpha}$	0.569	-0.518	-0.606	-0.193	-0.804	0.162	-0.121	-0.130	-0.021	-0.412
$\hat{\beta}$	-0.022	0.212	0.232	0.109	0.225	-0.001	0.052	0.049	0.025	0.120
AIC	1538.7	1481.7	1549.8	1573.8	1537.7	1536.6	1475.3	1547.9	1570.4	1537.9

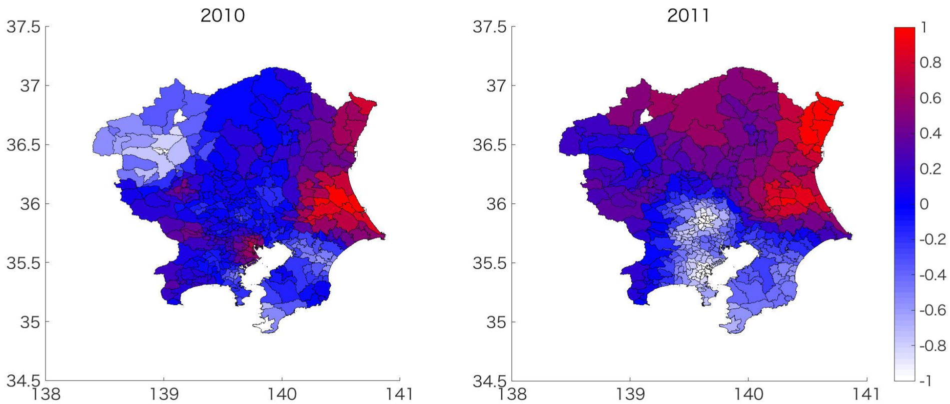

Figure 1: The identified volatility in 2010 and 2011. Notice that the great earth quake occurred in 2011.

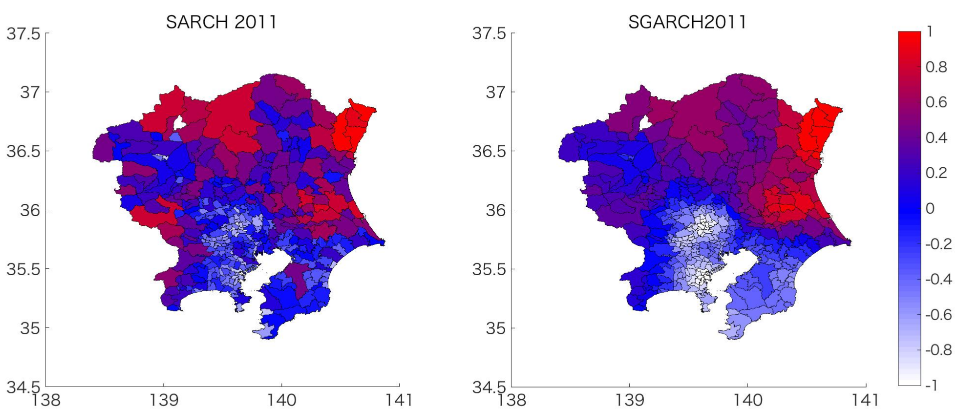

Figure 2: A comparison between the estimated volatility of the S-ARCH model and that of the S-GARCH model.

A Hessian, average Hessian and symmetric matrix $\Omega_{\psi,n}$

The Hessian matrix $H_n(\psi) \equiv \frac{\partial^2}{\partial\psi\partial\psi'} \log L_n(\psi)$ has the elements:

$$H_{\beta\beta'} = -\frac{1}{\sigma^2} X_n' R_n'^{-1}(\lambda) R_n^{-1}(\lambda) X_n,$$

$$
\begin{aligned}
H_{\beta\sigma^2} &= -\frac{1}{\sigma^4}X_n'R_n'^{-1}(\lambda)V(\theta),\\
H_{\beta\rho} &= -\frac{1}{\sigma^2}X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)W_nY_n,\\
H_{\beta\lambda} &= \frac{1}{\sigma^2}X_n'R_n'^{-1}(\lambda)(W_n'R_n'^{-1}(\lambda)V_n(\theta)+R_n^{-1}(\lambda)W_nV_n(\theta)-R_n^{-1}(\lambda)W_nY_n),\\
H_{\sigma^2\sigma^2} &= \frac{n}{2\sigma^4}-\frac{V_n'(\theta)V_n(\theta)}{\sigma^6},\\
H_{\sigma^2\rho} &= -\frac{1}{\sigma^4}Y_n'W_n'R_n'^{-1}(\lambda)V(\theta),\\
H_{\sigma^2\lambda} &= \frac{1}{\sigma^4}(V_n'(\theta)-Y_n')W_n'R_n'^{-1}(\lambda)V_n(\theta),\\
H_{\rho\rho} &= -\frac{1}{\sigma^2}Y_n'W_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)W_nY_n-tr(S_n^{-1}(\theta)W_nS_n^{-1}(\theta)W_n),\\
H_{\rho\lambda} &= \frac{1}{\sigma^2}Y_n'W_n'R_n'^{-1}(\lambda)(W_n'R_n'^{-1}(\lambda)V(\theta)+R_n^{-1}(\lambda)W_nV_n(\theta)-R_n^{-1}(\lambda)W_nY_n)-tr(S_n^{-1}(\theta)W_nS_n^{-1}(\theta)W_n),\\
H_{\lambda\lambda} &= \frac{1}{\sigma^2}(Y_n'-V_n'(\theta))W_n'R_n'^{-1}(\lambda)(2W_n'R_n'^{-1}(\lambda)V_n(\theta)+R_n^{-1}(\lambda)W_nV_n(\theta)-R_n^{-1}(\lambda)W_nY_n)\\
&\quad +tr(R_n^{-1}(\lambda)W_nR_n^{-1}(\lambda)W_n)-tr(S_n^{-1}(\theta)W_nS_n^{-1}(\theta)W_n).
\end{aligned}
$$

The average Hessian matrix $\Sigma_{\psi,n}\equiv -E\big(\frac{1}{n}\frac{\partial^2}{\partial\psi\psi'}\log L_n(\psi_0)\big)$ has the elements:

$$
\begin{aligned}
\Sigma_{\beta\beta'} &= \frac{1}{n\sigma_0^2}X_n'R_n'^{-1}R_n^{-1}X_n,\\
\Sigma_{\beta\sigma^2} &= 0,\\
\Sigma_{\beta\rho} &= \frac{1}{n\sigma_0^2}X_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}X_n\beta_0,\\
\Sigma_{\beta\lambda} &= \frac{1}{n\sigma_0^2}X_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}X_n\beta_0,\\
\Sigma_{\sigma^2\sigma^2} &= \frac{1}{2\sigma_0^4},\\
\Sigma_{\sigma^2\rho} &= \frac{1}{n\sigma_0^2}tr(W_nS_n^{-1}),\\
\Sigma_{\sigma^2\lambda} &= \frac{1}{n\sigma_0^2}tr(W_nS_n^{-1}-W_nR_n^{-1}),\\
\Sigma_{\rho\rho} &= \frac{1}{n\sigma_0^2}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}X_n\beta_0+\frac{1}{n}tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n+S_n^{-1}W_nS_n^{-1}W_n),\\
\Sigma_{\rho\lambda} &= \frac{1}{n\sigma_0^2}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}X_n\beta_0+\frac{1}{n}tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n+S_n^{-1}W_nS_n^{-1}W_n)\\
&\quad -\frac{1}{n}tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n+S_n^{-1}W_nR_n^{-1}W_n),\\
\Sigma_{\lambda\lambda} &= \frac{1}{n\sigma_0^2}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}X_n\beta_0+\frac{1}{n}tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n+S_n^{-1}W_nS_n^{-1}W_n)\\
&\quad -\frac{2}{n}tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n+S_n^{-1}W_nR_n^{-1}W_n)+\frac{1}{n}tr(R_n^{-1}W_nR_n^{-1}W_n+W_n'R_n'^{-1}R_n^{-1}W_n).
\end{aligned}
$$

The symmetric matrix $\Omega_{\psi,n}$ has the elemetns:

$$
\begin{aligned}
\Omega_{\beta\beta'} &= 0,\\
\Omega_{\beta\sigma^2} &= \frac{\mu_3}{2n\sigma_0^6}X_n'R_n'^{-1}\mathbf{1}_n,\\
\Omega_{\beta\rho} &= \frac{\mu_3}{n\sigma_0^4}\sum_i^n\{(R_n^{-1}X_n)_i\}'(R_n^{-1}W_nS_n^{-1}R_n)_{ii},
\end{aligned}
$$

$$
\begin{aligned}
\Omega_{\beta\lambda} &= \frac{\mu_3}{n\sigma_0^4}\sum_i^n\{(R_n^{-1}X_n)_i\}'(R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)_{ii},\\
\Omega_{\sigma^2\sigma^2} &= \frac{\mu_4-3\sigma_0^4}{4\sigma_0^8},\\
\Omega_{\sigma^2\rho} &= \frac{\mu_3}{2n\sigma_0^6}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}\mathbf{1}_n + \frac{\mu_4-3\sigma_0^4}{2n\sigma_0^6}tr(S_n^{-1}W_n),\\
\Omega_{\sigma^2\lambda} &= \frac{\mu_3}{2n\sigma_0^6}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}\mathbf{1}_n + \frac{\mu_4-3\sigma_0^4}{2n\sigma_0^6}tr(S_n^{-1}W_n - R_n^{-1}W_n),\\
\Omega_{\rho\rho} &= \frac{2\mu_3}{n\sigma_0^4}\sum_{i=1}^n(R_n^{-1}W_nS_n^{-1}X_n\beta_0)_i(R_n^{-1}W_nS_n^{-1}R_n)_{ii} + \frac{\mu_4-3\sigma_0^4}{n\sigma_0^4}\sum_{i=1}^n\{(R_n^{-1}W_nS_n^{-1}R_n)_{ii}\}^2,\\
\Omega_{\rho\lambda} &= \frac{\mu_3}{n\sigma_0^4}\sum_{i=1}^n(R_n^{-1}W_nS_n^{-1}X_n\beta_0)_i(2R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)_{ii}\\
&\quad+\frac{\mu_4-3\sigma_0^4}{n\sigma_0^4}\sum_{i=1}^n(R_n^{-1}W_nS_n^{-1}R_n)_{ii}(R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)_{ii},\\
\Omega_{\lambda\lambda} &= \frac{2\mu_3}{n\sigma_0^4}\sum_{i=1}^n(R_n^{-1}W_nS_n^{-1}X_n\beta_0)_i(R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)_{ii}\\
&\quad+\frac{\mu_4-3\sigma_0^4}{n\sigma_0^4}\sum_{i=1}^n\{(R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)_{ii}\}^2,
\end{aligned}
$$

where μ_3 and μ_4 are the third and fourth moments of v_is, respectively, $(R_n^{-1}X_n)_i$ is the i-th row of $(R_n^{-1}X_n)$, $(R_n^{-1}W_nS_n^{-1}X_n\beta_0)_i$ is the i-th element of $(R_n^{-1}W_nS_n^{-1}X_n\beta_0)$ and $(R_n^{-1}W_nS_n^{-1}R_n)_{ii}$, $(R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)_{ii}$ and $(2R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)_{ii}$ are the (i,j)th element of $(R_n^{-1}W_nS_n^{-1}R_n)$, $(R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)$ and $(2R_n^{-1}W_nS_n^{-1}R_n - R_n^{-1}W_n)$, respectively.

B Some useful Lemmas

Lemma B.1 (Proposition 8.4.13, Bernstein (2009))**.** *Let A and B be matrices. We use $\gamma_{\max}$ and $\gamma_{\min}$ to denote the largest and smallest eigenvalues of a matrix. If A is symmetric and B is positive semi definite, then*

$$\gamma_{min}(A)tr(B) \leq tr(AB) \leq \gamma_{max}(A)tr(B).$$

Lemma B.2 (Lee, 2002, p.256; Lee, 2004, p1918)**.** *Let $\{A_n\}$ and $\{B_n\}$ be two two sequences of $n\times n$ matrices that are uniformly bounded in both row and column sums and the elements of an $n\times n$ matrix $\{C_n\}$ be $O(1)$ uniformly. Then*

1. *the sequence $\{A_nB_n\}$ are uniformly bounded in both row and column sums,*
2. *the elements of C_nB_n have the uniform order $O(1)$, and*
3. *the elements of A_n are uniformly bounded and $tr(A_n) = O(n)$.*

Lemma B.3 (Lee, 2004, p1918)**.** *The elements, the v_i's of V_n are assumed to be i.i.d. with zero mean and a finite variance and the fourth moment of the $v's$ is assumed to exist. Suppose that A_n is a square matrix with tis column sums being uniformly bounded and elements of the $n\times K$ matrix Z_n are uniformly bounded. Let $\{B_n\}$ be uniformly bounded either in row or column sums and their elements $b_{n,ij}$ have $O(1)$ uniformly in i and j. Then*

1. *$\frac{1}{\sqrt{n}}Z_n'A_nV_n = O_p(1)$ and*
2. *$\frac{1}{n}E(V_n'B_nV_n) = O(1)$ and $\frac{1}{n}[V_n'B_nV_n - E(V_n'B_nV_n)] = o_p(1)$.*

C Proofs of Theorems 1-3

C.1 Proof of Theorem1

The consistency of $\hat{\theta}$ will follow from the uniform convergence of $\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))$ to zero on Θ and the uniqueness identification condition that, for any $\epsilon > 0, \limsup_{n\to\infty} \max_{\theta \in N_\epsilon^c(\theta_0)} \frac{1}{n}(Q_n(\theta) - Q_n(\theta_0)) < 0$, where $N_\epsilon^c(\theta_0)$ is the complement of an open neighborhood of θ_0 in Θ of diameter ϵ (Theorem 3.4 of white (1994)).

C.1.1 Proof of the uniform convergence of $\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))$

First, we shall prove the uniform convergence of $\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))$ to zero on Θ. The proof follows from:

(a) $\inf_{\theta\in\Theta} \sigma_n^{*2}(\theta)$ is bounded away from zero,

(b) $\sup_{\theta\in\Theta} |\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta)| = o_p(1)$,

(c) $\sup_{\theta\in\Theta} |\frac{1}{n}(\log L_n(\theta) - Q_n(\theta))| = o_p(1)$.

Proof of (a) By the definition of $V_n^*(\theta)$,

$$\begin{aligned}
V_n^*(\theta) &= R_n^{-1}(\lambda)(S_n(\theta)Y_n - X_n\beta_n^*(\theta)), \\
&= R_n^{-1}(\lambda)S_n(\theta)Y_n - R_n^{-1}X_n(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)S_n(\theta)E(Y_n), \\
&= R_n^{-1}(\lambda)S_n(\theta)Y_n - P_nR_n^{-1}(\lambda)S_n(\theta)E(Y_n), \\
&= M_nR_n^{-1}(\lambda)S_n(\theta)Y_n + P_nR_n^{-1}(\lambda)S_n(\theta)(Y_n - E(Y_n)),
\end{aligned}$$

where, $P_n = R_n^{-1}X_n(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n'^{-1}$ and $M_n = I_n - P_n$.

From the orthogonality between the two symmetric idempotent matrices M_n and P_n, we have,

$$\begin{aligned}
\sigma_n^{*2}(\theta) &= \frac{1}{n}E(V_n'^*(\theta)V_n^*(\theta)), \\
&= \frac{1}{n}E[Y_n'S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)Y_n + (Y_n - E(Y_n))'S_n'(\theta)R_n'^{-1}(\lambda)P_nR_n^{-1}(\lambda)S_n(\theta)(Y_n - E(Y_n))], \\
&= \frac{1}{n}E(Y_n')S_n'(\theta)R_n'^{-1}(\lambda)M_nR_n^{-1}(\lambda)S_n(\theta)E(Y_n) + \frac{1}{n}tr(R_n'^{-1}(\lambda)R_n^{-1}(\lambda)Var(S_n(\theta)Y_n)).
\end{aligned}$$

The matrix M_n is positive semi definite because M_n is a symmetric idempotent matrix (Lemma 14.2.14 of Harville (1997)). Thus, the first term is nonnegative uniformly in $\theta \in \Theta$.

Because the matrix $Var(S_n(\theta)Y_n)$ is symmetric and $\gamma_{min}Var(S_n(\theta)Y_n) > 0$ from the assumption, the matrix is positive semi definite (Theorem 3.25 of Schott (2005)). By Lemma A.1, the second term is

$$\begin{aligned}
\frac{1}{n}tr(R_n'^{-1}(\lambda)R_n^{-1}(\lambda)Var(S_n(\theta)Y_n)) &\geq \frac{1}{n}\gamma_{min}(R_n'^{-1}(\lambda)R_n^{-1}(\lambda))tr(Var(S_n(\theta)Y_n)), \\
&\geq \frac{1}{n}\underline{c}_r\underline{c}_y, \\
&> 0, \text{uniformly in } \theta \in \Theta.
\end{aligned}$$

It follow that $\inf_{\theta\in\Theta} \sigma_n^{*2}(\theta)$ is bounded away from zero.

Proof of (b) Noting that

$$\begin{aligned}
\hat{V}_n(\theta) &= R_n^{-1}(\lambda)(S_n(\theta)Y_n - X_n\hat{\beta}_n(\theta)), \\
&= R_n^{-1}(\lambda)S_n(\theta)Y_n - R_n^{-1}X_n(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)S_n(\theta)Y_n, \\
&= M_nR_n^{-1}(\lambda)S_n(\theta)Y_n.
\end{aligned}$$

Hence,

$$\hat{\sigma}_n^2(\theta) = \frac{1}{n}\hat{V}'_n(\theta)\hat{V}_n(\theta),$$

$$= \frac{1}{n} Y_n' S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) Y_n.$$

It follows that

$$\begin{aligned}
\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta) &= \frac{1}{n} Y_n' S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) Y_n - \frac{1}{n} E\big(Y_n' S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) Y_n\big) \\
&\quad - \frac{1}{n} E\big((Y_n - E(Y_n))' S_n'(\theta) R_n'^{-1}(\lambda) P_n R_n^{-1}(\lambda) S_n(\theta)(Y_n - E(Y_n))\big), \\
&= (Q_1 - EQ_1) - EQ_2,
\end{aligned}$$

where, $Q_1 = \frac{1}{n} Y_n' S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) Y_n$ and $EQ_2 = \frac{1}{n} E\big((Y_n - E(Y_n))' S_n'(\theta) R_n'^{-1}(\lambda) P_n R_n^{-1}(\lambda) S_n(\theta)(Y_n - E(Y_n))\big)$.

To show the result, it sufficient to show $Q_1 - EQ_1 \xrightarrow{p} 0$ and $EQ_2 \longrightarrow 0$, uniformly in $\theta \in \Theta$.

First, we show that $Q_1 - EQ_1 \xrightarrow{p} 0$ uniformly in $\theta \in \Theta$. By Theorem 1 of Andrews (1992), the uniform convergence of $Q_1 - EQ_1$ to zero in probability follows from the pointwise convergence for each $\theta \in \Theta$ and stochastic equicontinuity of Q_1, i.e., for any $\epsilon > 0$, there exists a positive number δ such that $\limsup_{n\to\infty} P(\sup_{\theta\in\Theta} \sup_{\theta' \in B(\theta,\delta)} > \epsilon) < \epsilon$, where $B(\theta, \delta)$ denote a closed ball in Θ of radius $\delta \geq 0$ centered at θ.

First of all, the pointwise convergence of $Q_1 - EQ_1$ will be shown. We have, by the identity: $Y_n = S_n^{-1} X_n \beta_0 + S_n^{-1} R_n V_n$,

$$\begin{aligned}
Q_1 &= \frac{1}{n}(S_n^{-1} X_n \beta_0 + S_n^{-1} R_n V_n)' S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta)(S_n^{-1} X_n \beta_0 + S_n^{-1} R_n V_n), \\
&= \frac{1}{n}(\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0 + 2\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n \\
&\quad + V_n' R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n), \\
&= Q_{1,1}(\theta) + 2Q_{1,2}(\theta) + Q_{1,3}(\theta),
\end{aligned}$$

where $Q_{1,1}(\theta) = \frac{1}{n}(\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0)$,
$Q_{1,2}(\theta) = \frac{1}{n}(\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n)$ and
$Q_{1,3}(\theta) = \frac{1}{n}(V_n' R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n V_n)$. The two terms $Q_{1,2}(\theta)$ and $Q_{1,3}(\theta)$ are stochastic.

For the second term, the column sums of $S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$ are uniformly bounded from assumption 3 and Lemma 2 and $E(Q_{1,2}(\theta)) = 0$. Thus, the pointwise convergence of $Q_{1,2}(\theta) - E(Q_{1,2})(\theta)$ follow from Lemma 3. Similarly, the column sums of $R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$ are uniformly bounded and the pointwise convergence of $Q_{1,3}(\theta) - E(Q_{1,3})(\theta)$ follows from Lemma 3. Therefore, $Q_1 - EQ_1 \xrightarrow{p} 0$, for each $\theta \in \Theta$.

Next, we show that Q_1 is stochastic equicontinuous. We have by the mean value theorem:

$$\begin{aligned}
Q_{1,\ell}(\theta_1) - Q_{1,\ell}(\theta_2) &= \frac{\partial}{\partial\theta'} Q_{1,\ell}(\bar{\theta})(\theta_2 - \theta_1), \\
&\leq \sup_{\theta\in\Theta} \left| \frac{\partial}{\partial\theta'} Q_{1,\ell}(\theta) \right| (\theta_2 - \theta_1),
\end{aligned}$$

where $\ell = 1, 2, 3$ and $\bar{\theta}$ lies between θ_1 and θ_2. For stochastic equicontinuous, it suffices to show that $\sup_{\theta\in\Theta} \big|\frac{\partial}{\partial\theta'} Q_{1,\ell}(\theta)\big| = O_p(1)$ by Theorem 21.10 of Davidson (1994). Let Π_1 be $S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1}$, Π_2 be $\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$ and Π_3 be $R_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n$. The partial derivatives $\frac{\partial}{\partial\theta'} \Pi_{1,\ell}$ take simple form and consequently $\frac{\partial}{\partial\theta'} \Pi_{1,\ell}$ are also uniformly bounded in both row and column sums. For $Q_{1,1}$, for any θ, the elements of $\beta_0' X_n \frac{\partial}{\partial\theta'} S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1}$ and $X_n \beta_0$ are uniformly bounded. Thus, there exists constants c_1 and c_2 such that $|\{\beta_0' X_n' (\frac{\partial}{\partial\theta'} S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1})\}_i| \leq c_1$ and $|(X_n \beta_0)_i| \leq c_2$ where $\{\beta_0' X_n' (\frac{\partial}{\partial\theta'} S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1})\}_i$ and $(X_n \beta_0)_i$ are the i-th elements of each vector. It follows that $\big|\frac{\partial}{\partial\theta'} Q_{1,1}\big| \leq c_1 c_2 = O(1)$. For $Q_{1,2}$, for any θ, $\big|\frac{\partial}{\partial\theta'} \Pi_{1,2,i}\big| \leq c_3$ where $\frac{\partial}{\partial\theta'} \Pi_{1,2,i}$ is the i-th element of $\frac{\partial}{\partial\theta'} \Pi_{1,2}$. Therefore, from Lemma B.3, $P\big(\big|\frac{\partial}{\partial\theta'} Q_{1,2}\big| > M\big) \leq P\big(\big|\frac{1}{n}\sum_{i=1}^n c_3 v_i\big| > M\big) = O\big(n^{-\frac{1}{2}}\big)$. For $Q_{1,3}$, for any θ, $\big|\frac{\partial}{\partial\theta'} \Pi_{1,3,ij}\big| \leq c_4$ where where $\frac{\partial}{\partial\theta'} \Pi_{1,3,ij}$ is the

(i,j)th element of $\frac{\partial}{\partial\theta'}\Pi_{1,3}$. Thus, from Lemma B.3, $P\big(\big|\frac{\partial}{\partial\theta'}Q_{1,3}\big| > M\big)\leq P\big(\big|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}c_4 v_i v_j\big| > M\big)= O(1)$. Thus, $\sup_{\theta\in\Theta}\big|\frac{\partial}{\partial\theta'}Q_{1,\ell}(\theta)\big| = O_p(1)$ It follow that Q_1 is stochastic equicontinuous. Hence, by Theorem 1 of Andrews (1992), $Q_1 - EQ_1 \xrightarrow{p} 0$ uniformly in $\theta\in\Theta$.

Secondly, we show that $EQ_2 \longrightarrow 0$, uniformly in $\theta\in\Theta$. There exist $\underline{c}_x$ such that $0<\underline{c}_x\leq\inf_{\lambda\in\Lambda}\gamma_{min}\big(\frac{1}{n}X_n'R_n'^{-1}R_n^{-1}X\big)$ from assumption. By Assumption, Lemma 1 and 2 and theorem 3.4 of Schott (2005), We have,

$$
\begin{aligned}
EQ_2 &= \frac{1}{n}E\big((Y_n-E(Y_n))'S_n'(\theta)R_n'^{-1}(\lambda)P_nR_n^{-1}(\lambda)S_n(\theta)(Y_n-E(Y_n))\big),\\
&= \frac{1}{n}tr(R_n'^{-1}(\lambda)P_nR_n^{-1}(\lambda)Var(S_n(\theta)Y_n)),\\
&= \frac{1}{n}tr(R_n'^{-1}(\lambda)R_n^{-1}X_n(X_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n)^{-1}X_n'R_n'^{-1}R_n^{-1}(\lambda)Var(S_n(\theta)Y_n)),\\
&\leq \frac{1}{n}\gamma_{min}^{-1}(X_n'R_n'^{-1}R_n^{-1}X)\gamma_{max}^2(R_n'^{-1}(\lambda)R_n^{-1}(\lambda))\gamma_{max}(Var(S_n(\theta)Y))tr(X_n'X_n)),\\
&= \frac{1}{n}\gamma_{min}^{-1}\bigg(\frac{X_n'R_n'^{-1}R_n^{-1}X}{n}\bigg)\gamma_{max}^2(R_n'^{-1}(\lambda)R_n^{-1}(\lambda))\gamma_{max}(Var(S_n(\theta)Y))\frac{1}{n}tr(X_n'X_n)),\\
&\leq \frac{1}{n}\underline{c}_x^{-1}\bar{c}_r^2\bar{c}_y\frac{1}{n}tr(X_n'X_n)),\\
&= O\big(n^{-1}\big)
\end{aligned}
$$

Hence, $EQ_2\longrightarrow 0$, uniformly in $\theta\in\Theta$.

Therefore, $\sup_{\theta\in\Theta}|\hat{\sigma}_n^2(\theta)-\sigma_n^{*2}(\theta)| = o_p(1)$, completing the proof of (b).

Proof of (C) We show that $\sup_{\theta\in\Theta}\big|\frac{1}{n}(\log L_n(\theta)-Q_n(\theta))\big| = o_p(1)$. Note that

$$\frac{1}{n}(\log L_n(\theta)-Q_n(\theta)) = -\frac{1}{2}(\log\hat{\sigma}_n^2(\theta)-\log\sigma_n^{*2}(\theta)).$$

By the Taylor expansion,

$$\big|\log\hat{\sigma}_n^2(\theta)-\log\sigma_n^{*2}(\theta)\big| = \frac{1}{\tilde{\sigma}_n^2(\theta)}\big|\hat{\sigma}_n^2(\theta)-\sigma_n^{*2}(\theta)\big|,$$

where $\tilde{\sigma}_n^2(\theta)$ lies between $\hat{\sigma}_n^2(\theta)$ and $\sigma_n^{*2}(\theta)$. From the proof (a) and (b), it follow that $\hat{\sigma}_n^2(\theta)$ is uniformly bounded away from zero on Θ. Moreover, $\tilde{\sigma}_n^2(\theta)$ is also uniformly bounded away from zero on Θ because $\tilde{\sigma}_n^2(\theta)$ exists between $\hat{\sigma}_n^2(\theta)$ and $\sigma_n^{*2}(\theta)$ and thereby $\frac{1}{\tilde{\sigma}_n^2(\theta)}$ is uniformly bounded. As $\hat{\sigma}_n^2(\theta)-\sigma_n^{*2}(\theta)$ coverges in probability to zero uniformly on Θ, $|\log\hat{\sigma}_n^2(\theta)-\log\sigma_n^{*2}(\theta)| = o_p(1)$ uniformly on Θ.

Consequently, $\sup_{\theta\in\Theta}\big|\frac{1}{n}(\log L_n(\theta)-Q_n(\theta))\big| = o_p(1)$.

C.1.2 Proof of the identification uniqueness condition

Secondly, we shall prove the identification uniqueness condition. The proof follow from:

(i) $\frac{1}{n}Q_n(\theta)$ is uniformly equicontinuous on Θ.

(ii) Show some properties of an auxiliary model.

(iii) Show that the identification uniqueness condition holds.

Proof of (i) We show that $\frac{1}{n}Q_n(\theta) = \frac{1}{2}(\log 2\pi+1)-\frac{1}{2}\log\sigma_n^{*2}(\theta)-\frac{1}{n}\log|R_n(\lambda)|+\frac{1}{n}\log|S_n(\theta)|$ is uniformly equicontinuous on Θ. It is sufficient to show that partial derivatives of each term are uniformly bounded. The uniform continuity of $\log\sigma_n^{*2}(\theta)$ on Θ follows because $\frac{1}{\sigma_n^{*2}(\theta)}$ is uniformly bounded since $\sigma_n^{*2}(\theta)$ is uniformly bounded away form zero on Θ. For $\frac{1}{n}\log|R_n(\lambda)|$, $\frac{d}{d\lambda}\frac{1}{n}\log|R_n(\lambda)| = \frac{1}{n}tr(R_n^{-1}(\lambda)W_n)$. From assumption and Lemma 2, the elements of $R_n^{-1}(\lambda)W_n$ are uniformly bounded. Thus, $\frac{1}{n}tr(R_n^{-1}(\lambda)W_n) = O(1)$ from Lemma 2. Similarly, $\frac{\partial}{\partial\theta}\frac{1}{n}\log|S_n(\theta)| = O(1)$. Hence, $\frac{1}{n}Q_n(\theta)$ is uniformly equicontinuous on Θ.

Proof of (ii) It is useful to establish an auxiliary process:

$$Y_n = \lambda W_n Y_n + \rho W_n Y_n + R_n(\lambda) V_n,$$

where $V_n \sim N(0, \sigma_0^2 I_n)$. The log-likelihood function of the above auxiliary process is given by

$$\log L_{p,n}(\theta, \sigma^2) \;=\; -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2(\theta)) - \log|R_n(\lambda)| + \log|S_n(\theta)| - \frac{1}{2\sigma^2} Y_n' S_n'(\theta) R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) Y_n.$$

Let E_p be the expectation under this auxiliary process. Define $Q_{p,n}(\theta) = \max_{\sigma^2} E_p(\log L_{p,n}(\theta))$. The optimal solutions of this maximization problem is

$$\begin{aligned}
\sigma_n^2(\theta) &= \frac{1}{n} E_p(Y_n' S_n'(\theta) R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) Y_n), \\
&= \frac{\sigma^2}{n} tr(R_n S_n^{-1} S_n'(\theta) R_n'^{-1}(\lambda) R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} R_n).
\end{aligned}$$

Hence,

$$Q_{p,n}(\theta) = -\frac{n}{2}\log(2\pi + 1) + \frac{n}{2}\log\sigma_n^2(\theta) - \log|R_n(\lambda)| + \log|S_n(\theta)|.$$

By Shannon-Kolmogorov Information Inequality (Ferguson (1996), p113), $Q_{p,n}(\theta) \leq Q_{p,n}(\theta_0)$ for all $\theta \in \Theta$. This implies that $\frac{1}{n}(Q_{p,n}(\theta) - Q_{p,n}(\theta_0) \leq 0$ for all $\theta \in \Theta$.

Proof of (iii) We show that the identification uniqueness condition holds by contradiction.

$$\begin{aligned}
\frac{1}{n}(Q_n(\theta) - Q_n(\theta_0)) &= -\frac{1}{2}\log\sigma_n^{*2}(\theta) - \log|R_n(\lambda)| + \log|S_n(\theta)| - \left(-\frac{1}{2}\log\sigma_0^2 - \log|R_n| + \log|S_n|\right) \\
&= \left(-\frac{1}{2}(\log\sigma_n^2(\theta) - \log\sigma_0^2) - \frac{1}{n}(\log|R_n(\lambda)| - \log|R_n|) + \frac{1}{n}(\log|S_n(\theta)| - \log|S_n|)\right) \\
&\quad -\frac{1}{n}(\log\sigma_n^{*2}(\theta) - \log\sigma_n^2(\theta)), \\
&= \frac{1}{n}\big(Q_{p,n}(\theta) - Q_{p,n}(\theta_0)\big) - \frac{1}{2}(\log\sigma_n^{*2}(\theta) - \log\sigma_n^2(\theta)).
\end{aligned}$$

Moreover,

$$\sigma_n^{*2}(\theta) - \sigma_n^2(\theta) = \frac{1}{n}\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0.$$

M_n is positive semi definite and thereby $\sigma_n^{*2}(\theta) - \sigma_n^2(\theta) \geq 0$. This implies $-\frac{1}{2}(\log\sigma_n^{*2}(\theta) - \log\sigma_n^2(\theta)) \leq 0$.

Now, suppose that the identification uniqueness condition does not hold. Then, there exists an $\epsilon > 0$ and a sequence $\{\theta_n\}$ in $N_\epsilon^c(\theta_0)$ such that $\lim_{n\to\infty}\frac{1}{n}\big(Q_n(\theta) - Q_n(\theta_0)\big) = 0$. By the compactness of $N_\epsilon^c(\theta_0)$, there exists a convergent subsequence $\{\theta_{n_m}\}$ of $\{\theta_n\}$ with the limit θ_+ of θ_{n_m} being in $N_\epsilon^c(\theta_0)$. This implies that $\theta_+ \neq \theta_0$. As $\frac{1}{n}Q_n(\theta)$ is uniformly equicontinuous, $\lim_{n_m\to\infty}\frac{1}{n_m}\big(Q_{n_m}(\theta_+) - Q_{n_m}(\theta_0)\big) = 0$. Because $\frac{1}{n}\big(Q_{p,n}(\theta) - Q_{p,n}(\theta_0)\big) \leq 0$ and $-\frac{1}{2}\big(\log\sigma_n^{*2}(\theta) - \log\sigma_n^2(\theta)\big) \leq 0$, this is possible only if $\lim_{n_m\to\infty}\frac{1}{n_m}\big(Q_{n_m}(\theta_+) - Q_{n_m}(\theta_0)\big) = 0$ and $-\frac{1}{2}\big(\log\sigma_n^{*2}(\theta) - \log\sigma_n^2(\theta)\big) \leq 0$. However, $\lim_{n\to\infty}\frac{1}{n}\beta_0' X_n' S_n'^{-1} S_n'(\theta) R_n'^{-1}(\lambda) M_n R_n^{-1}(\lambda) S_n(\theta) S_n^{-1} X_n \beta_0 \neq 0$ from the assumption in Theorem 3.1 . Thus, $-\frac{1}{2}\big(\log\sigma_n^{*2}(\theta) - \log\sigma_n^2(\theta)\big) < 0$ and consequently $\lim_{n_m\to\infty}\frac{1}{n_m}\big(Q_{n_m}(\theta_+) - Q_{n_m}(\theta_0)\big) \neq 0$. This is a contradiction. Therefore, the identification uniqueness condition must hold.

The consistency of $\hat{\theta}$ follow form uniform convergence and the identification uniqueness condition. This completes the proof of the theorem. □

C.2 Proof of Theorem 2

We have by the Taylor expansion,

$$0 \;=\; \frac{1}{\sqrt{n}}\frac{\partial \log L_n(\hat{\psi}_n)}{\partial \psi},$$

$$= \frac{1}{\sqrt{n}}\frac{\partial \log L_n(\psi_0)}{\partial \psi} + \left(\frac{1}{n}\frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial \psi \partial \psi'}\right)\sqrt{n}(\hat{\psi}_n - \psi_0),$$

where $\bar{\psi}_n$ lies between $\hat{\psi}_n$ and ψ_0. Thus, the asymptotic normality of $\hat{\psi}_n$ follows if

(a) $\frac{1}{\sqrt{n}}\frac{\partial \log L_n(\psi_0)}{\partial \psi} \xrightarrow{D} N\big(0, \lim_{n\to\infty}\Gamma(\psi_0)\big)$,

(b) $\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} - E\big(\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'}\big) \xrightarrow{p} 0$, and

(c) $\frac{1}{n}\frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial \psi \partial \psi'} - \frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} \xrightarrow{p} 0$.

Proof of (a) The asymptotic normality of $\frac{1}{\sqrt{n}}\frac{\partial \log L_n(\psi_0)}{\partial \psi}$ follows from the central limit theorems for linear-quadratic forms in Kelejian and Prucha (2001). We need to check that the score vector holds Assumptions in Kelejian and Prucha (2001). To check assumptions for asymptotic normality, it is sufficient to show some matrices hold desired boundaly conditions. From assumptions of this paper and Lemma A.2, $(R_n'S_n'^{-1}W_n'R_n'^{-1} - W_n'R_n'^{-1})$ and $R_n'S_n'^{-1}W_n'R_n'^{-1}$ are uniformly bounded in column sums, and the elements of $X_n'S_n'^{-1}W_n'R_n'^{-1}$ are uniformly bounded. Thus, each score function holds the assumptions and the asymptotic normality of each score function follows. Finally, the Cramér-Wold devise (Proposition 6.3.1 of Brockwell and Davis (1991)) leads to the joint asymptotic normality.

Proof of (b) Let $D_{\psi\psi}$ be $\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'} - E\big(\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial \psi \partial \psi'}\big)$. Then, $D_{\psi\psi}$ has the elements:

$$\begin{aligned}
D_{\beta\beta'} &= 0, \\
D_{\beta\sigma^2} &= -\frac{1}{n\sigma_0^4}X_n'R_n'^{-1}V_n, \\
D_{\beta\rho} &= -\frac{1}{n\sigma_0^2}X_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_nV_n, \\
D_{\beta\lambda} &= \frac{1}{n\sigma_0^2}X_n'(R_n'^{-1}W_n'R_n'^{-1} + R_n'^{-1}R_n^{-1}W - R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n)V_n, \\
D_{\sigma^2\sigma^2} &= \frac{1}{\sigma_0^4} - \frac{1}{n\sigma_0^6}V_n'V_n, \\
D_{\sigma^2\rho} &= -\frac{1}{n\sigma_0^4}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}V_n - \frac{1}{n\sigma_0^4}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}V_n - \sigma_0^2 tr(S_n'^{-1}W_n')), \\
D_{\sigma^2\lambda} &= -\frac{1}{n\sigma_0^4}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}V_n + \frac{1}{n\sigma_0^4}(V_n'W_n'R_n'^{-1}V_n - \sigma_0^2 tr(W_n'R_n'^{-1})) \\
&\quad -\frac{1}{n\sigma_0^4}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}V_n - \sigma_0^2 tr(S_n'^{-1}W_n')), \\
D_{\rho\rho} &= -\frac{2}{n\sigma_0}\beta_0'X_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_nV_n \\
&\quad -\frac{1}{n\sigma_0^2}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_nV_n - \sigma_0^2 tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n)), \\
D_{\rho\lambda} &= \frac{1}{n\sigma_0^2}\beta_0'X_n'(S_n'^{-1}W_n'R_n'^{-1}W_n'R_n'^{-1} + S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n - 2S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n)V_n \\
&\quad +\frac{1}{n\sigma_0^2}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}W_n'R_n'^{-1}V_n - \sigma_0^2 tr(S_n'^{-1}W_n'R_n'^{-1}W_n')) \\
&\quad +\frac{1}{n\sigma_0^2}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nV_n - \sigma_0^2 tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n)) \\
&\quad -\frac{1}{n\sigma_0^2}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_nV_n - \sigma_0^2 tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n)), \\
D_{\lambda\lambda} &= \frac{1}{n\sigma_0^2}\beta_0'X_n'(2S_n'^{-1}W_n'R_n'^{-1}W_n'R_n'^{-1} + S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n - 2S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n'^{-1}R_n
\end{aligned}$$

$$
\begin{aligned}
&-2R_n'^{-1}W_n'R_n'^{-1}W_n'R_n'^{-1} - R_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n + 2R_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n \\
&+2R_n'^{-1}W_n'R_n'^{-1}W_n'R_n^{-1} + R_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n - R_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n)V_n \\
&+\frac{2}{n\sigma_0^2}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}W_n'R_n'^{-1}V_n - \sigma_0^2 tr(S_n'^{-1}W_n'R_n'^{-1}W_n')) \\
&+\frac{1}{n\sigma_0^2}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nV_n - \sigma_0^2 tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_n)) \\
&-\frac{1}{n\sigma_0^2}(V_n'R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_nV_n - \sigma_0^2 tr(R_n'S_n'^{-1}W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n)) \\
&-\frac{2}{n\sigma_0^2}(V_n'W_n'R_n'^{-1}W_n'R_n'^{-1}V_n - \sigma_0^2 tr(W_n'R_n'^{-1}W_n'R_n'^{-1})) \\
&+\frac{1}{n\sigma_0^2}(V_n'W_n'R_n'^{-1}R_n^{-1}W_nV_n - \sigma_0^2 tr(W_n'R_n'^{-1}R_n^{-1}W_n)) \\
&+\frac{1}{n\sigma_0^2}(V_n'W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_nV_n - \sigma_0^2 tr(W_n'R_n'^{-1}R_n^{-1}W_nS_n^{-1}R_n)).
\end{aligned}
$$

Thus, the elements of $D_{\psi\psi}$ are decomposed into sums of the forms: $\frac{1}{n}X_n'A_n(\theta)V_n, \frac{1}{n}\beta_0'X_n'A_n(\theta)V_n$, $\frac{1}{n}(V_n'A_n(\theta)V_n - E(V_n'A_n(\theta)V_n))$ and $\frac{1}{\sigma_0^4} - \frac{1}{n\sigma_0^6}V_n'V_n$, where a matrix $A_n(\theta)$ is uniformly bounded in both row and column sums. From Lemma A.3, $\frac{1}{n}X_n'A_n(\theta)V_n, \frac{1}{n}\beta_0'X_n'A_n(\theta)V_n$ and $\frac{1}{n}(V_n'A_n(\theta)V_n - E(V_n'A_n(\theta)V_n))$ are convergence to zero in probability. Moreover, $\frac{1}{\sigma_0^4} - \frac{1}{n\sigma_0^6}V_n'V_n \xrightarrow{p} 0$ because $\frac{1}{n}V_nV_n \xrightarrow{p} \sigma_0^2$ by the law of large numbers. Therefore, it follow that $\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial\psi\partial\psi'} - E\big(\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial\psi\partial\psi'}\big) \xrightarrow{p} 0$.

Proof of (c) From Lemma B.2 and B.3, it is easy to show that $\frac{1}{n}\frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial\psi\partial\psi'} = O_p(1)$ and $\frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial\psi\partial\psi'} = O_p(1)$. Here, $\bar{\sigma}^{-r} = \sigma_0^{-r} + o_p(1), r = 2, 4, 6$ because $\bar{\sigma}^2 \xrightarrow{p} \sigma_0^2$ and σ^r appears in $H_n(\psi) \equiv \frac{\partial^2}{\partial\psi\partial\psi'}\log L_n(\psi)$ multiplicatively, thus it results in an asymptotically negligible error to replace $\bar{\sigma}^2$ by σ_0^2. The elements of the Hessian matrix, $H_n(\psi) \equiv \frac{\partial^2}{\partial\psi\partial\psi'}\log L_n(\psi)$, are decomposed into sums of terms of the forms: $X_n'A_n(\theta)X_n$, $X_n'A_n(\theta)Y_n$, $X_n'A_n(\theta)V(\theta)$, $Y_n'A_n(\theta)Y_n$, $\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}V_n'(\theta)V_n(\theta), Y_n'A_n(\theta)V_n(\theta), V_n'(\theta)A_n(\theta)V_n(\theta)$ and $tr(A_n(\theta))$, where a matrix $A_n(\theta)$ is uniformly bounded in both row and column sums. Therefore, it is sufficient to show that the difference between each term at $\bar{\psi}$ and ψ_0 converges to zero in probability and moreover this can be easily shown. We show some examples corresponding each term of the Hessian matrix.

Noting that

$$
\begin{aligned}
R_n^{-1}(\lambda) - R_n^{-1} &= R_n^{-1}(\lambda)(R_n - R_n(\lambda))R_n^{-1}, \\
&= (\lambda_0 - \lambda)R_n^{-1}(\lambda)W_nR_n^{-1}.
\end{aligned}
$$

For $X_n'A_n(\theta)X_n$,

$$
\begin{aligned}
\frac{1}{n}X_n'R_n'^{-1}(\bar{\lambda})R_n^{-1}(\bar{\lambda})X_n - \frac{1}{n}X_n'R_n'^{-1}R_n^{-1}X_n &= \frac{1}{n}X_n'(R_n'^{-1}(\bar{\lambda}) - R_n'^{-1} + R_n'^{-1})R_n^{-1}(\bar{\lambda})X_n - \frac{1}{n}X_n'R_n'^{-1}R_n^{-1}X_n, \\
&= \frac{1}{n}X_n'(R_n'^{-1}(\bar{\lambda}) - R_n'^{-1})R_n^{-1}(\bar{\lambda})X_n + \\
&\quad \frac{1}{n}X_n'R_n'^{-1}R_n^{-1}(\bar{\lambda})X_n - \frac{1}{n}X_n'R_n'^{-1}R_n^{-1}X_n, \\
&= (\lambda_0 - \bar{\lambda})\frac{1}{n}X_nR_n'^{-1}(\lambda)W_n'R_n^{-1}R_n^{-1}(\bar{\lambda})X_n \\
&\quad +(\lambda_0 - \bar{\lambda})\frac{1}{n}X_n'R_n'^{-1}R_n^{-1}(\lambda)W_nR_n^{-1}X_n, \\
&= o_p(1)O(1) + o_p(1)O(1), \\
&= o_p(1).
\end{aligned}
$$

Moreover, the convergence of $X_n'A_n(\theta)Y_n$ is shown similarly.

Noting that

$$\begin{aligned} V_n(\theta) &= R_n^{-1}(\lambda)R_n(\lambda)V_n(\theta), \\ &= R_n^{-1}(\lambda)(S(\theta)Y_n - X_n\beta), \\ &= R_n^{-1}(\lambda)((\lambda_0-\lambda)W_nY_n + (\rho_0-\rho)W_nY_n + X_n(\beta_0-\beta) + R_nV_n). \end{aligned}$$

Thus, for $X_n'A_n(\theta)V(\theta)$,

$$\begin{aligned} \frac{1}{n}X_n'R_n'^{-1}(\bar\lambda)V_n(\bar\theta) - \frac{1}{n}X_n'R_n'^{-1}V_n &= \big((\lambda_0-\bar\lambda)+(\rho_0-\bar\rho)\big)\frac{1}{n}X_n'R_n'^{-1}(\bar\lambda)W_nY_n + \frac{1}{n}X_n'R_n'^{-1}(\bar\lambda)X_n(\beta_0-\beta) \\ &\quad + \frac{1}{n}X_n'R_n'^{-1}(\bar\lambda)R_nV_n - \frac{1}{n}X_n'R_n'^{-1}V_n, \\ &= o_p(1)O_p(1) + O_p(1)o_p(1) + o_p(1) + o_p(1), \\ &= o_p(1), \end{aligned}$$

where the convergence of last two terms follow from Lemma B.3.

Here,

$$\begin{aligned} \frac{1}{n}V_n'(\bar\theta)V_n(\bar\theta) &= \big((\lambda_0-\bar\lambda)+(\rho_0-\bar\rho)\big)^2\frac{1}{n}Y_n'W_n'R_n'^{-1}(\bar\lambda))R_n^{-1}(\bar\lambda)W_nY_n \\ &\quad +(\beta_0-\beta)'\frac{1}{n}X_n'R_n'^{-1}(\bar\lambda)R_n^{-1}(\bar\lambda)X_n(\beta_0-\beta) + \frac{1}{n}V_n'R_n'R_n'^{-1}(\bar\lambda)R_n^{-1}(\bar\lambda)R_nV_n \\ &\quad +\frac{2}{n}\big((\lambda_0-\bar\lambda)+(\rho_0-\bar\rho)\big)Y_n'W_n'R_n'^{-1}(\bar\lambda)R_n^{-1}(\bar\lambda)X_n(\beta_0-\beta) \\ &\quad +\frac{2}{n}\big((\lambda_0-\bar\lambda)+(\rho_0-\bar\rho)\big)Y_n'W_n'R_n'^{-1}(\bar\lambda)R_n^{-1}(\bar\lambda)R_nV_n + (\beta_0-\beta)'\frac{2}{n}X_n'R_n'^{-1}(\bar\lambda)R_n^{-1}(\bar\lambda)R_nV_n, \\ &= o_p(1)O_p(1) + o_p(1)O(1)o_p(1) + \sigma_0^2 + o_p(1)O_p(1)o_p(1) + o_p(1)O_p(1) + o_p(1)o_p(1), \\ &= \sigma_0^2 + o_p(1). \end{aligned}$$

It follows that $\frac{1}{2\sigma_0^4} - \frac{1}{n\sigma_0^6}V_n'(\theta)V_n(\theta) = o_p(1)$.

Before next proof, we show an example. $Y_n'S_n(\theta)V_n = \beta'X_n'S_n^{-1}S(\theta)V_n + V_n'R_n'S_n^{-1}S_n(\theta)V_n$ and

$$\begin{aligned} \frac{1}{n}V_n'R_n'S_n^{-1}S_n(\theta)V_n - \frac{1}{n}V_n'R_n'S_n^{-1}S_nV_n &= \big((\lambda_0-\lambda)+(\rho_0-\rho)\big)\frac{1}{n}V_n'R_n'S_n^{-1}V_n, \\ &= o_p(1)O_p(1), \\ &= o_p(1). \end{aligned}$$

It follows that $\frac{1}{n}Y_n'S_n(\theta)V_n - \frac{1}{n}Y_n'S_nV_n = o_p(1)$ and similarly $\frac{1}{n}Y_n'A_n(\theta)V_n - \frac{1}{n}Y_n'A_nV_n = o_p(1)$ and $\frac{1}{n}Y_n'A_n(\theta)Y_n - \frac{1}{n}Y_n'A_nY_n = o_p(1)$ where A_n is $A_n(\theta)$ at true value θ_0.

Now, for $Y_n'A_n(\theta)V_n(\theta)$,

$$\begin{aligned} \frac{1}{n}Y_n'W_n'R_n'^{-1}(\lambda)V_n(\theta) - \frac{1}{n}Y_n'W_n'R_n'^{-1}V_n &= \big((\lambda_0-\bar\lambda)+(\rho_0-\bar\rho)\big)\frac{1}{n}Y_n'W_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)W_nY_n \\ &\quad +\frac{1}{n}Y_n'W_n'R_n'^{-1}(\lambda)R_n^{-1}(\lambda)X_n(\beta_0-\bar\beta) \\ &\quad +\frac{1}{n}Y_n'W_n'R_n'^{-1}(\bar\lambda)R_n^{-1}(\bar\lambda)R_nV - \frac{1}{n}Y_n'W_n'R_n'^{-1}V_n \\ &= o_p(1)O_p(1) + O_p(1)o_p(1) + o_p(1) \\ &= o_p(1). \end{aligned}$$

Moreover, the convergence of $V_n(\theta)'A_n(\theta)V_n(\theta)$ is also shown similary.

Finally, for $tr(A_n(\theta))$, by the Taylor expansion,

$$\frac{1}{n}tr(R_n^{-1}(\lambda)W_nR_n^{-1}(\lambda)W_n) - \frac{1}{n}tr(R_n^{-1}W_nR_n^{-1}W_n) = \frac{d}{d\lambda}tr(R_n^{-1}(\tilde\lambda)W_nR_n^{-1}(\tilde\lambda)W_n)(\bar\lambda-\lambda_0),$$

$$
\begin{aligned}
&= O(1)o_p(1), \\
&= o_p(1),
\end{aligned}
$$

where $\tilde{\lambda}$ lies between $\bar{\lambda}$ and λ_0.

The convergence of the other elements of the Hessian matrix are shown similarly, hence $\frac{1}{n}\frac{\partial^2 \log L_n(\bar{\psi}_n)}{\partial\psi\partial\psi'} - \frac{1}{n}\frac{\partial^2 \log L_n(\psi_0)}{\partial\psi\partial\psi'} \xrightarrow{p} 0$.

This completes the proof of the theorem. □

C.3 Proof of Theorem 3

The estimator for α is

$$
\hat{\alpha}_n = (1-\hat{\lambda})\log\left(\frac{1}{n}\sum_{i=1}^{n}\exp\{(R_n^{-1}(\hat{\lambda})[S(\hat{\theta})Y_n - Z_n\hat{\delta}])_i\}\right),
$$

Here,

$$
\begin{aligned}
S(\hat{\theta})Y_n - Z_n\hat{\delta} &= Y_n - \hat{\lambda}W_nY_n - \hat{\rho}W_nY_n - Z_n\hat{\delta}, \\
&= (\lambda_0 - \hat{\lambda})W_nY_n + (\rho_0 - \hat{\rho})W_nY_n + Z_n(\delta_0 - \hat{\delta}) + \alpha_0\mathbf{1}_n + R_nV_n, \\
&= D + \alpha_0\mathbf{1}_n + R_nV_n,
\end{aligned}
$$

where $D = (\lambda_0 - \hat{\lambda})W_nY_n + (\rho_0 - \hat{\rho})W_nY_n + Z_n(\delta_0 - \hat{\delta})$.

Because $R_n^{-1}(\hat{\lambda})(S(\hat{\theta})Y_n - Z_n\hat{\delta}) = \frac{\alpha_0}{1-\hat{\lambda}}\mathbf{1}_n + R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n$,

$$
\frac{1}{n}\sum_{i=1}^{n}\exp\{(R_n^{-1}(\hat{\lambda})[S(\hat{\theta})Y_n - Z_n\hat{\delta}])_i\} = \exp\left(\frac{\alpha}{1-\lambda}\right)\frac{1}{n}\sum_{i=1}^{n}\exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n)_i\}.
$$

Thus,

$$
\hat{\alpha} - \alpha_0 = (1-\hat{\lambda})\log\left(\frac{1}{n}\sum_{i=1}^{n}\exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n)_i\}\right). \tag{9}
$$

To prove consistency, it is sufficient that the right side of (9) converges to zero in probability.

By the Taylor expansion,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n)_i\} &= 1 + \frac{1}{n}\sum_{i=1}^{n}\exp(b_i)\big\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n)_i\big\} \\
&= 1 + \frac{1}{n}\boldsymbol{b}'(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n),
\end{aligned}
$$

where b_i lies between 0 and $(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n)_i$, and $\boldsymbol{b} = (b_1, \ldots, b_n)'$.

From Assumptions, Theorem 1 and Lemma B.2 and B.3,

$$
\begin{aligned}
\frac{1}{n}\boldsymbol{b}'(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n) &= (\lambda_0 - \hat{\lambda})\frac{1}{n}\boldsymbol{b}'R_n^{-1}(\hat{\lambda})W_nY_n + (\rho_0 - \hat{\rho})\frac{1}{n}\boldsymbol{b}'R_n^{-1}(\hat{\lambda})W_nY_n \\
&\quad + \frac{1}{n}\boldsymbol{b}'R_n^{-1}(\hat{\lambda})Z_n(\delta_0 - \hat{\delta}) + \frac{1}{n}\boldsymbol{b}'R_n^{-1}(\hat{\lambda})R_nV_n, \\
&= o_p(1)O_p(1) + o_p(1)O_p(1) + O(1)o_p(1) + o_p(1), \\
&= o_p(1).
\end{aligned}
$$

Thus, $\frac{1}{n}\sum_{i=1}^{n}\exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n)_i\} \xrightarrow{p} 1$ and
$(1-\hat{\lambda})\log\big(\frac{1}{n}\sum_{i=1}^{n}\exp\{(R_n^{-1}(\hat{\lambda})D + R_n^{-1}(\hat{\lambda})R_nV_n)_i\}\big) \xrightarrow{p} 0$. □

References

[1] Andrews, D. W. (1992). Generic uniform convergence. Econometric theory, 8(2), 241-257.

[2] Bernstein, D. S. (2009). Matrix mathematics: Theory, facts, and formulas. 2nd edition. Princeton: Princeton University Press.

[3] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

[4] Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. The review of economics and statistics, 498-505.

[5] Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. Journal of political Economy, 96(1), 116-131.

[6] Borovkova, S. & Lopuhaa, R., (2012). Spatial GARCH: A Spatial Approach to Multivariate Volatility Modeling. Available at SSRN: https://ssrn.com/abstract=2176781.

[7] Brockwell, P. J., & Davis, R. A. (1991). Time series: theory and methods. Springer-Verlag.

[8] Caporin, M., and Paruolo, P., 2009, Structured multivariate volatility models, Available at SSRN: http://ssrn.com/abstract=1318639.

[9] Davidson, J. (1994). Stochastic limit theory. Oxford: Oxford University Press.

[10] Doan, O., & Tapnar, S. (2013). GMM estimation of spatial autoregressive models with moving average disturbances. Regional Science and Urban Economics, 43(6), 903-926.

[11] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50, 987-1007.

[12] Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. Econometric theory, 11(1), 122-150.

[13] Fan, J., & Yao, Q. (2003). Nonlinear time series: nonparametric and parametric methods. Springer Science & Business Media.

[14] Ferguson, T. S. (1996). A course in large sample theory. Chapman & Hall.

[15] Harville, D. A. (1997). Matrix algebra from a statistician's perspective. New York: Springer.

[16] Kelejian, H. H., & Prucha, I. R. (1997). Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. International regional science review, 20(1-2), 103-111.

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

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

[19] Kelejian, H. H., & Robinson, D. P. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in regional science, 72(3), 297-312.

[20] Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric theory, 18(2), 252-277.

[21] Lee, L. F. (2004). Asymptotic Distributions of QuasiMaximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica, 72(6), 1899-1925.

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

[23] Lee, P. M. (2012). Bayesian statistics: an introduction. John Wiley & Sons.

[24] Robinson, P. M. (2009). Largesample inference on spatial dependence. The Econometrics Journal, 12(s1).

[25] Sato, T. and Matsuda, Y. (2017). Spatial Autoregressive Conditional Heteroskedasticity Models. J.Japan Statist. Soc., Vol. 47, 2

[26] Schott, J. R. (2005). Matrix analysis for statistics. 2nd edition. John Wiley & Sons.

[27] Su, L., & Yang, Z. (2015). QML estimation of dynamic panel data models with spatial errors. Journal of Econometrics, 185(1), 230-258.

[28] White, H. (1994). Estimation, inference and specification analysis. New York: Cambridge university press.

[29] Yan, J. (2007). Spatial stochastic volatility for lattice data. Journal of agricultural, biological, and environmental statistics, 12(1), 25.

[30] Yang Z. (2015). Unified M-Estimation of Fixed-Effects Spatial Dynamic Models with Short Panels, Working Papers 14-2015, Singapore Management University, School of Economics.

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