DSSR

Discussion Paper No. 82

Spatiotemporal ARCH Models

Takaki Sato
Yasumasa Matsuda

May, 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

Spatiotemporal ARCH Models

Takaki Sato [*] Yasumasa Matsuda [†]

Abstract

This study proposes spatiotemporal extensions of time series autoregressive conditional heteroskedasticity (ARCH) models. We call spatiotemporally extended ARCH models as spatiotemporal ARCH (ST-ARCH) models. ST-ARCH models specify conditional variances given simultaneous observations and past observations, which constitutes a good contrast with time series ARCH models that specify conditional variances given past own observations. We have proposed two types of ST-ARCH models based on cross-sectional correlations between error terms. A spatial weight matrix based on Fama-French 3 factor models are used to quantify the closeness between stock prices. We estimate the parameters in ST-ARCH models by a two-step procedure of the quasi maximum likelihood estimation method. We demonstrate the empirical properties of the models by simulation studies and real data analysis of stock price data in the Japanese market.

Keywords: GARCH model, quasi maximum likelihood, spatial dynamic panel model, spatiotemporal model, spatial weight matrix.

1 Introduction

Volatility which is a conditional variance in a model is one of the most important concepts in financial econometrics. It is used in widely ares such as risk management, option pricing and portfolio selection. The seminal work of Engle (1982) proposes autoregressive conditional heteroscedasticity (ARCH) models and the most important extension of the model is generalized ARCH (GARCH) models proposed by Bollerslev (1986). These univariate volatility models are commonly used to estimate and forecast volatility of financial assets.

Univariate volatility models are generalized to multivariate cases in many ways. The curse of dimensionality becomes a major obstacle for generalization because there are $\frac{n(n+1)}{2}$ quantities in a conditional covariance matrix for a n-dimensional time series. Thus, we attempt to give a conditional covariance matrix some simple structures to reduce the number of parameters. Bollerslev

[*]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

(1990) proposed constant conditional correlation models and Engle and Kroner (1995) proposed BEKK models.

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. 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 spatial extensions 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 contributes to extend spatial ARCH models to spatiotemporal models which we call spatiotemporal ARCH (ST-ARCH) models by using spatial econometrics ideas. ST-ARCH models are the one that describe a conditional variance at an asset given data of simultaneous and past other assets. The model is characterized by a spatial weight matrix which express cross-section correlations between assets and used to reduce the number of parameters. We propose new methods to calculate distance between stock prices from estimation coefficient of Fame-French 3 factor models and make spatial weights based on the distance. We consider two types of correlations between error terms in ST-ARCH models to evaluate covolatiltiy of assets. First one is spatial autoregressive error covariances, i.e. correlations are described with spatial weight matrix and the other is nonparametric error covariances.

The rest of paper proceeds as follows. Section 2 introduces ST-ARCH models. The estimation procedures and their asymptotic properties are described in section 3. Section 4 examines empirical properties of ST-ARCH models by applying the models to simulated data and real data such as stock price in the Japanese market. Section 5 discusses some concluding remarks.

2 Spatiotemporal ARCH models

We propose two types of spatiotemporal ARCH models which have different correlation between error terms as other multivariate volatility models to avoid the cause of dimensionality. Moreover, we explain new methods to make spatial weight matrices for stock data which doesn't include location information.

2.1 Spatiotemporal ARCH models

Let $r_{i,t}$ be the log returns of an asset at time index t. We define spatiotemporal ARCH (ST-ARCH) models by

$$\begin{aligned} r_{i,t} &= \sqrt{h}_{i,t}\varepsilon_{i,t}, \qquad (1) \\ \log h_{i,t} &= c_i + \lambda \sum_{j=1}^{n} w_{i,j} \log y_{j,t} + \gamma \log y_{i,t-1} + \end{aligned}$$

$$\begin{aligned} & \rho \sum_{j=1}^{n} w_{i,j} \log y_{j,t-1} + \boldsymbol{z}_{i,t}\delta, \qquad (2) \\ i &= 1, 2, \ldots, n, \\ t &= 1, 2, \ldots, T, \end{aligned}$$

where $\sqrt{h_i}$ is volatility, $\varepsilon_{i,t}$ is an error term, $\boldsymbol{z}_{i,t}$ is $(k \times 1)$ non-stochastic regressors, c_i is a fixed effect, and $w_{i,j}$ is a spatial weight that quantifies a closeness form point i and j with $w_{i,i} = 0$, and constitutes a spatial weight matrix $W = (w_{i,j})$. For the common parameters $(\lambda, \gamma, \rho, \delta)$ in this model, λ, γ, ρ are the parameters that describe the strength of spatiotemporal dependences of volatility on past observations and simultaneous observations.

We need to consider cross-sectional correlations between error terms to calculate covolatility which becomes zero when error terms are uncorrelated. Here, we consider two types of correlation among error terms. First one is spatial autoregressive cross-sectional correlations defined by,

$$\log \boldsymbol{\varepsilon}_t^2 = \alpha W \log \boldsymbol{\varepsilon}_t^2 + \log \boldsymbol{V}_t^2, \qquad (3)$$

where $\log \boldsymbol{\varepsilon}_t^2 = (\log \varepsilon_{1,t}^2, \ldots, \log \varepsilon_{n,t}^2)'$, $log \boldsymbol{V}_t^2 = (\log v_{1,t}^2, \ldots, \log v_{n,t}^2)'$ are $n \times 1$ column vectors and $v_{i,t}$ is independent and identically distributed (i.i.d.) across i and t with mean zeros and variance σ^2. The benefit of using spatial correlations is to reduce the number of parameters from $\frac{n(n+1)}{2}$ to 1. The other one is nonparametric cross-sectional correlations, hence we assume $\log \boldsymbol{\varepsilon}_t$ follows i.i.d. across t with mean $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Omega}$ which have no restriction. Although the number of parameters is huge when we assume nonparametric cross-sectional correlations, we can capture correlations in more flexible conditions. We call ST-ARCH models which have the former error terms ST-ARCH models with spatial autoregressive error covariance (ST-ARCH-SCOV)models and models which have the later error terms ST-ARCH models with nonparametric error covariance (ST-ARCH-NCOV)models.

The ST-ARCH models are different from the multivariate time series GARCH model in the following two points. Firstly, ST-ARCH models describe logged spatiotemporal volatility defined by linear combinations of past observations and simultaneous observations, which is an analogy of time series GARCH models. Although the definitions of volatility are different, it will be shown in later section that they have some common features such as volatility clustering. Secondly, the log transformation of $h_{i,t}$ is used to ensure the existence of observations. If we defined non logged volatility, it would be difficult to guarantee the existence of observations unlike that for time series ARCH models that can be proved by Markov process theories (Fan and Yao (2003)). The log transformation of volatility makes it much easier to prove the existence in the following way. Substituting $\log h_{i,t}$ in the log squared equation of (2) with (1), we have

$$\log y_{i,t}^2 \quad = \quad c_i + \lambda \sum_{j=1}^{n} w_{i,j} \log y_{j,t} + \gamma \log y_{i,t-1} + \rho \sum_{j=1}^{n} w_{i,j} \log y_{j,t-1} +$$

$$z_{i,t}\delta + \log \varepsilon_{i,t}^2,$$

which is a spatial dynamic panel model whose existence conditions have been well established.

2.2 Spatial weight matrices

Spatial weight matrices are nonnegative matrices and predetermined based on the spatial configuration of observations in sample such as first-order contiguity relation or inverse distance between observations. Spatial weight matrices is used to reduce estimation parameters, which have critical role in spatial analysis because dependence relations between a set of n observations have $n^2 - n$ relations and this leads the over-parameterization problem. However, stock prices doesn't include location information, hence we define an economic distance between stock prices to make spatial weight matrices.

We define an economic distance from Fama-French three factor models. Fama-French models are defined by

$$r_{i,t} = \alpha_i + r_{M,t}\beta_{i,M} + SMB_t\beta_{i,SMB} + HML_t\beta_{i,HML} + e_{i,t},$$

where $e_{i,t}$ is an i.i.d. random variables across i ant t with zero mean and variance, $r_{M,t}$ is an expected excess return of the market portfolio, SMB_t is Small Minus Big, i.e., the return of a portfolio of small stocks in excess of the return on a portfolio of large stocks, and HML_t is High Minus Low, i.e., the return of a portfolio of stocks with a high book-to-market ratio in excess of the return on a portfolio of stocks with a low book-to-market ratio (Bodie, et al (2013)). The parameters, $\beta_{i,M}, \beta_{i,SMB}, \beta_{i,HML}$, are risk premiums derived from exposure to each risk sources, therefore if beta's of two assets take similar values then two assets may behave similarly. We define a distance between two assets, $d_{i,j}$, by

$$d_{i,j} = \frac{< \beta_i, \beta_j >}{||\beta_i||||\beta_j||},$$

where $\beta_i = (\beta_{i,M}, \beta_{i,SMB}, \beta_{i,HML})'$, $< \beta_i, \beta_j >$ is a real inner-product of column vectors and $||\beta_i||$ is the Euclidean norm of a column vector. After that we calculate the (i, j) element of a spatial weight matrix, W, based on the above distance from

$$w_{i,j} = \frac{d_{i,j}}{\sum_{j=1}^{n} d_{i,j}}.$$

3 Estimation

We shall propose the estimation method of parameters in ST-ARCH models. We have proposed two types of ST-ARCH models and parameters in each model are estimated by a two step procedure. First step is the estimation of parameters except for individual effects by quasi-maximum likelihood (QML) estimation methods because the estimators for individual effects are biased. In second step, individual effects are estimated by QML methods which based on the likelihood different from the one in the first step.

3.1 Estimation for ST-ARCH-SCOV models

The parameters of ST-ARCH-SCOV models are $\lambda, \gamma, \rho, \delta, \alpha$ and individual effects, and the parameters except for individual effects are estimated in first step by the QML method. To apply the QML method, we need to demean the error term.

$$\begin{aligned}
\boldsymbol{c} + (I-\alpha W)^{-1} \log \boldsymbol{V}_t^2 &= \boldsymbol{c} + (I-\alpha W)^{-1}(\log \boldsymbol{V}_t^2 - E(\log v_{1,1}^2)\mathbf{1} + \\
&\quad E(\log v_{1,1}^2)\mathbf{1}), \\
&= \{\boldsymbol{c} + (1-\lambda)E(\log v_{1,1}^2)\mathbf{1}\} + \\
&\quad (I-\alpha W)^{-1}(\log \boldsymbol{V}_t^2 - E(\log v_{1,1}^2)\mathbf{1}),
\end{aligned}$$

where $\boldsymbol{c} = (c_1, \ldots, c_n)'$, $\mathbf{1} = (1, \ldots, 1)'$ and I is an $n \times n$ identity matrix. We see the individual effects have the bias by $(1-\lambda)E(\log v_{1,1}^2)$. Denoting $Y_t = (\log y_{1,t}^2, \ldots, \log y_{n,t}^2)'$, $A = \boldsymbol{c} + (1-\lambda)E(\log v_{1,1}^2)\mathbf{1}$, $\xi = \log \boldsymbol{V}_t^2 - E(\log v_{1,1}^2)\mathbf{1}$ and $Z_t = (\boldsymbol{z}'_{1,t}, \ldots, \boldsymbol{z}'_{n,t})'$, we have the following modified representation,

$$Y_t = A + \lambda W Y_t + \gamma Y_{t-1} + \rho W Y_{t-1} + Z_t \delta + (I - \alpha W)^{-1}\xi, \tag{4}$$

where ξ is zero mean processes.

Now, let us consider the QML estimation in (4). The Gaussian likelihood function of (4) by regarding the elements of xi as independent Gaussian noises with mean zero and variance σ^2 is

$$\log L(\psi) = -\frac{nT}{2}\log(2\pi\sigma^2) - \frac{\sum_{t=1}^{T} V_t'(\psi) V_t(\psi)}{2\sigma^2} + T\log|R(\alpha)| + T\log|S(\lambda)|,$$

where $\theta = (\lambda, \gamma, \rho, \alpha, \delta',)'$, $\psi = (\theta, A', \sigma^2)'$, $R(\alpha) = I - \alpha W$, $S(\lambda) = I - \lambda W$, and $V_t(\psi) = R^{-1}(\alpha)[S_n(\theta)Y_t - (\gamma I + \rho W)Y_{t-1} - Z_t\delta - A]$.

Concentrating out A and $\sigma^2$by,

$$\begin{aligned}
\hat{A}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} V(\theta), \\
\hat{\sigma}^2(\theta) &= \frac{1}{T}\sum_{t=1}^{T} V(\theta)' R'^{-1}(\alpha) R^{-1}(\alpha) V(\theta),
\end{aligned}$$

where $V(\theta) = S_n(\theta)Y_t - (\gamma I + \rho W)Y_{t-1} - Z_t\delta$, we obtain the concentrated likelihood function of θ,

$$\begin{aligned}
\log L_n(\theta) &= -\frac{nT}{2}(\log(2\pi) + 1) - \frac{nT}{2}\log\hat{\sigma}^2(\theta) + T\log|R_n(\lambda)| + \\
&\quad T\log|S_n(\theta)|.
\end{aligned} \tag{5}$$

The QML estimator fo θ is $\hat{\theta}$ that maximizes the concentrated likelihood function (5), and σ^2 are estimated by $\hat{\sigma}^2(\hat{\theta})$.

We move to the estimation of individual effects. As the estimated individual effects in the first step is biased, we need to estimate individual effects separately in the second step. In the second step, we employ the quasi likelihood by regarding ε_i, not $\log \varepsilon_i^2$, as Gaussian and assume errors are independent because individual effects are firm-specific characteristics.

The probability density function of $\log \varepsilon^2$ for a standard normal variable ε is shown by Lee (2012, p. 379) as

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\exp(x) + \frac{1}{2}x\right), \tag{6}$$

we obtain the quasi likelihood function as

$$\log L_n(\phi) = \frac{nT}{2}\log 2\pi - \sum_{t=1}^{T}\sum_{i=1}^{n}\left\{-\frac{1}{2}\frac{\exp(V_{i,t}(\phi))}{\exp(c_i)} + \frac{1}{2}(V_{i,t}(\phi) - c_i)\right\} + \log|S(\lambda)|,$$

where $\phi = (\lambda, \gamma, \rho, \delta), V(\phi) = S(\lambda)Y_t - (\gamma I + \rho W)Y_{t-1} - Z_t\delta, V_{i,t}$ is the i-th element of V_t. Differentiating the likelihood with respect to c_i and solving the equation by setting it to be 0, we have the relation

$$c_i(\phi) = \log\left(\frac{1}{n}\sum_{t=1}^{T}\exp(V_{i,t}(\phi))\right).$$

Replacing ϕ with the QML estimator $\hat{\phi}(\hat{\lambda}, \hat{\gamma}, \hat{\rho}, \hat{\delta}')$ in the first step, we finally obtain

$$\hat{c}_i(\hat{\phi}) = \log\left(\frac{1}{n}\sum_{t=1}^{T}\exp(V_{i,t}(\hat{\phi}))\right). \tag{7}$$

as estimators for individual effects.

3.2 Estimation for ST-ARCH-NCOV models

The parameters in ST-ARCH-NCOV models are estimated by a tow step procedure in order to avoid the bias of individual effect, c_i, similarly. Since the error term $\log \varepsilon_{i,t}^2$ is not a zero mean process, the QML estimation would not work.

First step is the estimation of $\gamma, \rho, \lambda, \Omega$. We slight modify the original form as

$$\begin{aligned}\log y_{i,t}^2 \quad &= \quad \{c_i + E(log\varepsilon_{i,t}^2)\} + \lambda\sum_{j=1}^{N} w_{ij}\log y_{i,t}^2 + \gamma \log y_{i,t-1}^2 + \\ &\qquad \rho\sum_{j=1}^{N} w_{ij}\log y_{i,t-1}^2 + z_{i,t}\delta + \{\log \varepsilon_{i,t}^2 - E(\log \varepsilon_{i,t}^2)\}.\end{aligned}$$

Define three symbols for simplicity by $Y_{i,t} = \log y_{i,t}^2$, $A_i = c_i + E(\log \varepsilon_{i,t}^2)$ and $v_{i,t} = \log \varepsilon_{i,t}^2 - E(\log \varepsilon_{it}^2)$. Then we have

$$Y_{i,t} = A_i + \lambda \sum_{j=1}^{N} w_{ij} Y_{i,t} + \gamma Y_{i,t-1} + \rho \sum_{j=1}^{N} w_{ij} Y_{i,t-1} + z_{i,t}\delta + v_{i,t}, \tag{8}$$

Let us estimate $\gamma, \rho, \lambda, \Omega$ in (8) by the QML estimation which be obtained by regarding $v_{i,t}$s as Gaussian variables with mean 0. We have the quasi log-likelihood function of $\psi_{np} = (\gamma, \rho, \lambda, \Omega, A')'$ by

$$\begin{aligned} \log L(\psi_{np}) &= -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log(\Omega) + T\log|I - \lambda W| \\ &\quad - \sum_{t=1}^{T} \frac{V_t'(\psi_{np})\Omega^{-1}V_t(\psi_{np})}{2}, \end{aligned} \tag{9}$$

where $V_t(\psi_{np}) = Y_t - A - \lambda W Y_t - \gamma Y_{t-1} - \rho W Y_{t-1} - Z_t\delta$.

Concentrating out A and Ω by

$$\begin{aligned} \hat{A}(\theta_{np}) &= \frac{1}{T}\sum_{t=1}^{T} V(\theta_{np}), \\ \hat{\Omega}(\theta_{np}) &= \frac{1}{T}\sum_{t=1}^{T} [V(\theta_{np}) - \hat{A}(\theta_{np})][V(\theta_{np}) - \hat{A}(\theta_{np})]', \end{aligned} \tag{10}$$

where $\theta_{np} = (\lambda, \gamma, \rho, \delta')'$, $V_t(\theta_{np}) = Y_t - \lambda W Y_t - \gamma Y_{t-1} - \rho W Y_{t-1} - Z_t\delta$. We have the concentrated log-likelihood function by

$$\log L(\theta_{np}) = -\frac{nT}{2}\log(2\pi + 1) - \frac{T}{2}\log(\hat{\Omega}(\theta_{np})) + T\log|I - \lambda W|. \tag{11}$$

Maximizing this with respect to λ, γ, ρ and δ, we have the estimator $\hat{\lambda}, \hat{\gamma}$ $\hat{\rho}$ and $\hat{\delta}$. Moreover, substituting $\hat{\lambda}, \hat{\gamma}$ and $\hat{\rho}$ into (10), we have the estimator $\hat{\Omega}(\hat{\theta}_{np})$.

Second step is the estimation of individual effects. We estimate it in the same manner as previous section, i.e., maximizing the quasi log-likelihood derived by regarding $\varepsilon_{i,t}$ in the error terms standard Gaussian variables.

3.3 Asymptotic results

This section considers consistency and asymptotic normality of estimators in a special case. Consistency and asymptotic normality of estimators in first step are proved by the results of Yu and Lee (2008). Consistency and asymptotic normality of estimators in second step are proved independently of estimators in first step. These asymptotic properties are proofed when the covariance matrix of error terms in ST-ARCH models is homoskedastic and has no cross sectional correlation, $\Omega = \sigma^2 \boldsymbol{I}$.

We will make use of the following set of assumptions.

Assumption 1. *The* $\varepsilon_{i,t}, i = 1, \ldots, n, t = 1, \ldots, T$ *in* $\boldsymbol{\varepsilon}_t$ *are i.i.d with mean zero and variance* σ^2. *Its moment* $E(|log\varepsilon_{i,t}|^{4+\gamma})$ *for some* $\gamma \geq 0$ *exists.*

Assumption 2. $\boldsymbol{W}$ *is a constant spatial weights matrix and its diagonal elements satisfy* $w_{i,i} = 0$ *for* $i = 1, \ldots, N$.

Assumption 3. *The matrix* $(\mathbf{I} - \lambda\mathbf{W})$ *is invertible for all* $\lambda \in \Lambda$. *Furthermore,* Λ *is compact and* λ_0 *is in the interior of* Λ.

Assumption 4. $\mathbf{W}$ *is uniformly bounded in row and column sums. Also,* $(\mathbf{I} - \lambda\mathbf{W})^{-1}$ *is uniformly bounded in* $\lambda \in \Lambda$.

Assumption 5. $\sum_{h=1}^{\infty} abs(A^h)$ *is uniformly bounded, where* $A = \{(\mathbf{I} - \lambda_0\mathbf{W})^{-1}\}(\gamma_0\mathbf{I} + \rho_0\mathbf{W})$.

Assumption 6. N *is a non-decreasing function of* T *and* T *goes to infinity.*

Assumption 7. $\frac{1}{nT}\sum_{t=1}^{T}(\tilde{D}_t, G\tilde{D}_t\zeta_0)'(\tilde{D}_t, G\tilde{D}_t\zeta_0))$ *is nonsingular where* $G = W(I - \lambda_0 W)^{-1}$, $\zeta = (\gamma, \rho, \delta)'$, $\tilde{D}_t = (\tilde{Y}_{t-1}, W\tilde{Y}_{t-1}, \tilde{Z}_t)$, *and* $\tilde{Y}_t$ *and* $\tilde{Z}_t$ *are demeaning of* Y *and* Z.

First, we consider consistency and asymptotic normality of estimators in first step.

Theorem 1. *Under assumptions 1-7, estimators in first step have consistency and asymptotic normality.*

Proof. The assumptions hold the assumptions in Yu and Lee (2008). Therefore, estimators in first step have consistency and asymptotic normality from the theorem 1 and 3 in Yu and Lee (2008). □

Secondly, we consider consistency and asymptotic normality of estimators in second step.

Theorem 2. *Under the assumptions 1-7,* $\hat{c}_i$ *converges to* $c_{i,0}$ *in probability.*

Proof. The consistency of $\hat{c}_i$ will follow from the convergence in probability to zero of $(\exp(\hat{c}_i) - \exp(c_{i,0}))$.

Let $\log y_{i,t}^2$ be $z_{i,t}$ for simplicity.

$$\exp(\hat{c}_i) = \frac{1}{T}\sum_{t=1}^{T}\exp\left(z_{i,t} - \hat{\lambda}\sum_{j=1}^{N} z_{j,t} - \hat{\gamma}z_{i,t-1} - \hat{\rho}\sum_{j=1}^{N} z_{j,t-1}\right).$$

Moreover,

$$\begin{aligned} & \exp\left(z_{i,t} - \hat{\lambda}\sum_{j=1}^{N} z_{j,t} - \hat{\gamma}z_{i,t-1} - \hat{\rho}\sum_{j=1}^{N} z_{j,t-1}\right) \\ = \quad & \exp\left(c_{i,0} + \log\varepsilon_{i,t}^2 + (\lambda_0 - \hat{\lambda})\sum_{j=1}^{N} z_{j,t} + (\gamma_0 - \hat{\gamma})z_{i,t-1} + (\rho_0 - \hat{\rho})\sum_{j=1}^{N} z_{j,t-1}\right), \end{aligned}$$

$$
\begin{aligned}
&= \exp(c_{i,0} + \log \varepsilon_{i,t}^2 + A), \\
&\quad \left(A = (\lambda_0 - \hat{\lambda}) \sum_{j=1}^{N} z_{j,t} + (\gamma_0 - \hat{\gamma}) z_{i,t-1} + (\rho_0 - \hat{\rho}) \sum_{j=1}^{N} z_{j,t-1} \right) \\
&= \exp(c_{i,0} + \log \varepsilon_{i,t}^2) + \exp(c_{i,0} + \log \varepsilon_{i,t}^2)\{\exp(A) - 1\}.
\end{aligned}
$$

Using Theorem 1 of Yu and Lee (2008), we have

$$
\begin{aligned}
A &= (\lambda_0 - \hat{\lambda}) \sum_{j=1}^{N} z_{j,t} + (\gamma_0 - \hat{\gamma}) z_{i,t-1} + (\rho_0 - \hat{\rho}) \sum_{j=1}^{N} z_{j,t-1}, \\
&= o_p(1) O_p(1), \\
&= o_p(1).
\end{aligned}
$$

Since $\exp(A) \xrightarrow{p} 1$,, $\exp(A) - 1 \xrightarrow{p} 0$. Thus,

$$
\begin{aligned}
\exp(c_{i,0} + \log \varepsilon_{i,t}^2)\{\exp(A) - 1\} &= O_p(1) o_p(1), \\
&= o_p(1).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\exp(c_i) &= \frac{1}{T} \sum_{t=1}^{T} \{\exp(c_{i,0} + \log \varepsilon_{i,t}^2) + o_p(1)\}, \\
&= \exp(c_{i,0}) \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{i,t}^2 + o_p(1).
\end{aligned}
$$

By the law of large numbers (Brockwell and Davis (1991), p206),

$$
\exp(\hat{c}_i) \xrightarrow{p} \exp(c_{i,0}) + o_p(1).
$$

Therefore, $\exp(\hat{c}_i) - \exp(c_{i,0}) \xrightarrow{p} o_p(1)$. □

Lemma 1. *Under the assumptions 1-7,*

$$
\sqrt{T} \left\{ \exp\left((\lambda_0 - \hat{\lambda}) \sum_{j=1}^{N} \log y_{j,t}^2 + (\gamma_0 - \hat{\gamma}) \log y_{i,t-1}^2 + (\rho_0 - \hat{\rho}) \sum_{j=1}^{N} \log y_{j,t-1}^2 \right) - 1 \right\} = o_p(1).
$$

Proof. Let $(\lambda_0 - \hat{\lambda}) \sum_{j=1}^{N} \log y_{j,t}^2 + (\gamma_0 - \hat{\gamma}) \log y_{i,t-1}^2 + (\rho_0 - \hat{\rho}) \sum_{j=1}^{N} \log y_{j,t-1}^2$ be A for simplicity.

Denote $\theta = (\lambda, \gamma, \rho)'$. At the true value, $\theta_0 = (\lambda_0, \gamma_0, \rho_0)'$. Moreover, $\hat{\theta}_{NT}$ is an estimator of θ. Then, $\hat{\theta}_{NT} - \theta_0 = O_p\left(max\left(\sqrt{\frac{1}{NT}}, \frac{1}{T} \right) \right)$ by theorem 3 of Yu and Lee (2008).

Thus,

$$\begin{aligned}
A &= (\lambda_0 - \hat{\lambda})\sum_{j=1}^{N} \log y_{j,t}^2 + (\gamma_0 - \hat{\gamma})\log y_{i,t-1}^2 + (\rho_0 - \hat{\rho})\sum_{j=1}^{N}\log y_{j,t-1}^2, \\
&= O_p\left(max\left(\sqrt{\frac{1}{NT}}, \frac{1}{T}\right)\right)O_p(1), \\
&= O_p\left(max\left(\sqrt{\frac{1}{NT}}, \frac{1}{T}\right)\right).
\end{aligned}$$

Therefore, $\sqrt{T}|A| = o_p(1)$.

From the Taylor's theorem,

$$\begin{aligned}
\sqrt{T}|\exp(A) - 1| &= \sqrt{T}|1 + \exp(b)A - 1|, \\
&= \exp(b)\sqrt{T}|A|.
\end{aligned}$$

where b is some real number between 0 and A.

Hence,

$$\begin{aligned}
\sqrt{T}|\exp(A) - 1| &= O_p(1)o_p(1), \\
&= o_p(1).
\end{aligned}$$

□

Theorem 3. *Under the assumptions in 1-7 and there exist the variance of* $\log \varepsilon_{i,t}^2$*,* σ_0^2*,* $\sqrt{T}(\hat{c}_i - c_0) \xrightarrow{d} N(0, \sigma_0^2)$*.*

Proof. Let $\log y_{i,t}^2$ be $z_{i,t}$ for simplicity.

$$\begin{aligned}
\exp(\hat{c}_i) &= \frac{1}{T}\sum_{t=1}^{T}\exp\left(z_{i,t} - \hat{\lambda}\sum_{j=1}^{N} z_{j,t} - \hat{\gamma}z_{i,t-1} - \hat{\rho}\sum_{j=1}^{N} z_{j,t-1}\right), \\
&= \frac{1}{T}\sum_{t=1}^{T}\exp\Bigg(c_{i,0} + \log\varepsilon_{i,t}^2 + (\lambda_0 - \hat{\lambda})\sum_{j=1}^{N} z_{j,t} + (\gamma_0 - \hat{\gamma})z_{i,t-1} + \\
&\quad (\rho_0 - \hat{\rho})\sum_{j=1}^{N} z_{j,t-1}\Bigg), \\
&= \frac{1}{T}\sum_{t=1}^{T}\exp(c_{i,0} + \log\varepsilon_{i,t}^2 + A), \\
&\quad \left(A = (\lambda_0 - \hat{\lambda})\sum_{j=1}^{N} z_{j,t} + (\gamma_0 - \hat{\gamma})z_{i,t-1} + (\rho_0 - \hat{\rho})\sum_{j=1}^{N} z_{j,t-1}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\exp(c_{i,0} + \log\varepsilon_{i,t}^2) + \frac{1}{T}\sum_{t=1}^{T}\exp(c_{i,0} + \log\varepsilon_{i,t}^2)\{\exp(A) - 1\}.
\end{aligned}$$

Moreover,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\exp(c_{i,0}+\log\varepsilon_{i,t}^2) &= \frac{1}{T}\sum_{t=1}^{T}\exp(c_{i,0})+\frac{1}{T}\sum_{t=1}^{T}\exp(c_{i,0})(\varepsilon_{i,t}^2-1),\\ &= \exp(c_{i,0})+\frac{\sigma_0\exp(c_{i,0})}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{i,t}^2-1}{\sigma_0}\right).\end{aligned}$$

Thus,

$$\begin{aligned}\sqrt{T}(\exp(\hat{c}_i)-\exp(c_{i,0})) &= \frac{\sigma_0\exp(c_{i,0})}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_{i,t}^2-1}{\sigma_0}\right)+\\ &\quad \frac{1}{T}\sum_{t=1}^{T}\sqrt{T}[\exp(c_{i,0}+\log\varepsilon_{i,t}^2)\{\exp(A)-1\}]. \qquad (12)\end{aligned}$$

From the central limit theorem, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\varepsilon_{i,t}^2-1}{\sigma_0}\right)\xrightarrow{d}N(0,1)$. Furthermore, the second term of (12) converges in probability to zero by lemma 1.

Therefore,

$$\sqrt{T}(\exp(\hat{c}_i)-\exp(c_{i,0}))\xrightarrow{d}N(0,\{\sigma_0\exp(c_{i,0})\}^2)$$

The asymptotic variance $\frac{\{\sigma_0\exp(c_{i,0})\}^2}{T}\to 0$ as $T\to\infty$ and the natural logarithm function is differentiable at $\exp(c_{i,0})$. Using proposition 6.4.1. in (Brockwell and Davis 1991), we have

$$\hat{c}_i\xrightarrow{d}N\left(c_{i,0},\frac{\sigma_0^2}{T}\right)$$

□

4 Empirical analysis

We examine empirical properties of ST-ARCH models by applying to simulation data, stock price data in the Japanese market. Monte Carlo experiments are carried out to investigate finite sample performances of the estimators, and stock price data analysis is employed to demonstrate practical performances of ST-ARCH models.

4.1 Simulation studies

To investigate finite sample properties of the two stage estimators, we simulate data by ST-ARCH models where $z_{i,t}$'s are randomly generated from independent normal distributions, individual effects are randomly generated from independent normal distributions which variance 4 and the spatial weights matrix is generated according to Rook contiguity and row normalizing. For ST-ARCH-SCOV models, we consider the two cases of error terms, $\varepsilon_{i,t}$: (i) independent standard normal distributions, (ii) spatially correlated error terms defined by (3) where $v_{i,t}$ follow independent stand normal distributions. The parameters $(\lambda, \gamma, \rho, \alpha, \delta)' = (0.4, 0.2, 0.2, 0.2, 2)$. For ST-ARCH-NCOV models, we consider the two cases of error terms, $\varepsilon_{i,t}$: (i) independent standard normal distributions, (ii) nonparametric covariance matrices in which error terms have correlations. Nonparametric covariance is made by test matrices in MATLAB. The parameters $(\lambda, \gamma, \rho, \delta)' = (0.4, 0.2, 0.2, 0.2, 2)$. The sample size n is 49 and time period T is 100 in both models Each set of Monte Carlo results is based on 1000 repetitions of the two step estimation.

The empirical means and square root of mean squared errors (RMSE) for the two stage estimators are reported in Table 1. The results show the estimators in the firs step, $(\hat{\lambda}, \hat{\gamma}, \hat{\rho}, \hat{\delta}')'$ are nearly unbiased and not sensitive to the choice of the error distributions. On the other hand, the second step estimators for individual effects, c, have biases when error terms have cross-sectional correlations

Table 1: The empirical means and root mean squared errors (RMSE) of the estimators of ST-ARCH-SCOV models and ST-ARCH-NCOV models. We consider the two cases of error terms for both models, respectively.

	ST-ARCH-SCOV models				ST-ARCH-NCOV models			
	Case (i)		Case (ii)		Case (i)		Case (ii)	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
λ	0.000	0.024	0.000	0.026	-0.001	0.035	0.000	0.026
γ	-0.006	0.013	-0.007	0.013	-0.008	0.018	-0.005	0.012
ρ	-0.003	0.019	0.001	0.019	-0.003	0.027	-0.002	0.019
α	0.000	0.032	0.000	0.034				
δ	-0.008	0.031	-0.002	0.033	-0.003	0.045	-0.001	0.026
c_1	0.007	0.187	-0.343	0.414	0.008	0.187	-0.021	0.147

4.2 Stock price data analysis

We apply ST-ARCH models to stock price data in Japanese financial market. We shall examine the characteristics of volatility in stock price data.

Let us introduce data used in this section. stock price data and trading volume data of each stock are collected by Yahoo finance. Stock price data is transformed into logged returns and trading volume data is transformed into logarithm of volumes of each stock and used as explanatory variables, Z_t. Our data set is comprised of the 30 companies in TOPIX core 30. We used the spatial weight matrix based on financial distance discussed in section 2. The data of $r_{M,t}, SMB_t$ and HML_t is collected from Financial Data Solutions, Inc.

The sample period under consideration is 2013/01/04-2016/12/30 and we apply models in two periods to compare estimation and prediction results between a short period and a long period. Period 1 starts on January 4, 2016 and ends on December 30, 2016 and Period 2 starts on January 4, 2013 and ends on December 30, 2016. In each period, last 30 days are used for prediction.

We compared three models which are ST-ARCH-SCOV models, ST-ARCH-NCOV models and GARCH models based on AIC and prediction errors (PE). Prediction errors are calculated by

$$
\begin{aligned}
E(\log y_{i,t}^2 - \log \hat{y}_{i,t}^2)^2 &= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}(\log y_{i,t}^2 - \log \hat{y}_{i,t}^2)^2, \\
\log \hat{y}_{i,t}^2 &= \log \hat{\sigma}^2 + E(\log \varepsilon_{i,t}^2),
\end{aligned}
$$

where $\log \hat{y}_{i,t}^2$'s are prediction values given past information and those of ST-ARCH models are calculated easily from estimates in the first step. Let $\log \hat{y}_t^2 = (\log \hat{y}_{1,t}^2, \ldots, \log \hat{y}_{n,t}^2)'$ and $\log \hat{y}_t^2 = (I - \hat{\lambda}W)^{-1}(\hat{A}_i + (\hat{\gamma}I + \hat{\rho}W)\log y_{t-1} + Z_{t-1}\hat{\delta})$. Moreover, those of GARCH models are calculated by regarding $E(\log \varepsilon_{i,t}^2)$ as -1.27 which is the expectation values when errors are Gaussian noises.

We find from Table 2 that estimated values of spatial correlation, $\hat{\lambda}$ is bigger than other parameters. Simultaneous information may plays an important role for estimation and prediction of volatility. Table 3 shows model fits of ST-ARCH models is better than those of GARCH models in the short period, whereas this relations are opposite in the long period. This results may show that spatial correlations change through time and spatial weight matrices in ST-ARCH models which describe constant spatial correlations can't capture cross-sectional correlations in stock prices in the long period. This suggests the need to consider time-varying spatial weight matrices. Predictions errors of ST-ARCH models are smaller than those of GARCH models in both periods. This is the advantage point of ST-ARCH models and shows the possibility of a wide rage of financial applications.

Table 2: Estimated values and their t values for estimators in ST-ARCH models which are ST-ARCH-SCOV models and ST-ARCH-NCOV models applied to log returns of stock price data in Japanese financial market.

	Period 1: 2016/01/04 $\sim$ 2016/12/30		Period 2: 2013/01/04 $\sim$ 2016/12/30	
	ST-ARCH-SCOV	ST-ARCH-NCOV	ST-ARCH-SCOV	ST-ARCH-SCOV
λ	0.678	0.936	0.661	0.876
t(λ)	42.123	147.745	68.683	319.830
γ	-0.028	-0.025	0.016	0.0163
t(γ)	-1.727	-0.221	2.575	3.905
ρ	0.059	0.026	0.050	0.003
t(ρ)	2.115	0.248	7.051	0.897
α	0.092		0.021	
t(α)	2.213		4.365	
δ	0.311	0.137	0.189	0.125
t(δ)	6.639	4.890	73.872	72.278

Table 3: AIC and prediction errors (PE) of ST-ARCH models which are ST-ARCH-SCOV (SCOV) models and ST-ARCH-NCOV (NCOV) models and GARCH models applied to log returns of stock price data in Japanese financial market.

	Period 1: 2016/01/04 $\sim$ 2016/12/30			Period 2: 2013/01/04 $\sim$ 2016/12/30		
	SCOV	NCOV	GARCH	SCOV	NCOV	GARCH
AIC	27312	27357	28637	125170	124687	115357
PE	5.641	5.843	16.777	5.158	5.209	13.255

5 Conclusion

We have proposed spatiotemporal autoregressive conditional heteroskedasticity (ST-ARCH) models as spatiotemporal extensions of a spatial autoregressive conditional heteroskedasticity (S-ARCH) model by Sato and Matsuda (2017). We consider two types of correlations between error terms in ST-ARCH models to evaluate covolatiltiy of assets. First one is spatial autoregressive error covariances and the other is nonparametric error covariances. By re-expressing ST-ARCH as spatial dynamic panel models, we employ spatial econometrics methodology to estimate the parameters by the two step procedures. Applications to stock price data in Japanese financial market demonstrate that prediction errors of ST-ARCH models are smaller than those of GARCH models.

Finally let us introduce possible extensions of ST-ARCH models. We employed time-invariant spatial weight matrices. However, empirical analysis shows that spatial correlations change through time and spatial weight matrices in ST-ARCH models which describe constant spatial correlations can't capture cross-sectional correlations in stock prices in the long period. This suggests the need to consider spatiotemporal volatility models which include time-varying spatial weight matrices.

References

[1] Bodie, Z., Kane, A., & Marcus, A. (2013) Investment, 10th ed., Irwin/McGraw-Hill, Boston.

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

[3] 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.

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

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

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

[7] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.

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

[9] Fan, J., and Yao, Q. (2003). Nonlinear Time Series, Springer-Verlag, New York.

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

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

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

[13] 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.