Conditional Heteroskedasticity with Risk Spillover Through Networks: An Exponential GARCH Approach[*]

Yang Yang[†]

November 2, 2021

Abstract

By introducing both intra-temporal and inter-temporal risk spillover through network, we propose a new multivariate conditional volatility model. For stationary case, the model can capture the dynamic of conditional heteroskedasticity structure when there are long-run stable links among multiple markets, and it is easy to be estimated consistently by QMLE approach. By Monte Carlo simulations, we show good finite sample performance when $n/T \to 0$. When applying the model to monthly stock return innovations of 11 eurozone countries from March 1999 to April 2021, by using geographical and institutional links to capture the network between the countries, the performance of our model dominates single variate GARCH(1,1), EGARCH(1,1) and multivariate GARCH with both constant correlation and dynamic conditional correlation settings by likelihood values and AIC criteria.

1 Introduction

Risk spillover is an important factor in the modern highly integrated global economy. Geographical, political, trade and other relationships among nations and regions are served as bridges to spread risk in local markets from one to others. In existing literature, risk spillovers between financial markets through networks are investigated especially after the global financial crisis. Hong (2001) developed a test for volatility spillover between two time series, and applied it to exchange rates. Hong, Liu and Wang (2009) developed a model to capture the Granger causality correlations of extreme risk spillover between financial markets. Blasques et al. (2016) applied static spatial Durbin model to capture time-varying spillover of sovereign risk in CDS markets in some European countries. Kou, Peng and Zhong (2017) introduced spatial interaction into traditional CAPM and APT models, and applied such models to study the co-movements of Eurozone stock indices and the future contracts on S&P/Case-Shiller Home Price Indices. Richmond (2019) developed a general equilibrium model

1

including trade network and used it to explain the lower interest rates and currency risk premia of central countries in the global trading network.

However, for conditional volatility of assets, network spillover effect had not been investigated seriously. Currently, for multiple markets/assets, there are many different specifications extended from single variate ARCH/GARCH type models. Some of them assume the conditional volatilities of each asset follow single variate ARCH/GARCH model, with some additional assumptions on their conditional covariances, such as multivariate GARCH models developed by Bollerslev (1990) and Engle (2002). These types of models do not consider the spillover at volatility level at all. Some of them assume simultaneous equations structure on conditional variances, such as BEKK and VECH models discussed in Engle and Kroner (1995) and extending literature. Although this type models have some advantages, there are too many parameters included which generate serious identification issue. Asymptotic properties of the estimators for these models are also extremely hard to be discussed. Also, as the spillovers in these specifications are inter-temporal from history without considering intra-temporal interactions, we are not able to use them to analyze the network effects among the markets when we want to focus on some particular links among the assets.

In this paper, we try to introduce the network risk spillover pattern into the conditional heteroskedasticity framework. As directly extending from ARCH/GARCH is hard, we focus on the dynamic of conditional log-volatility, and we try to give a new multivariate exponential GARCH type model with autoregressive effect and with both intra-temporal and inter-temporal network spillover effects. In terms of estimating method, the model can be linearized to a dynamic spatial panel model and easily estimated by likelihood approach. When there are persistent network correlation among multiple countries or regions, our model will be a better choice to capture the dynamic of conditional heteroskedasticity structure of assets associated with the individuals inside the network.

In the following part of this paper, Section 2 introduces the model, including the discussion of alternative model specifications, formation and some properties of the ESPARCH(1,1). Section 3 discusses the QMLE for ESPARCH(1,1) model, including QMLE under different assumptions of residual process, asymptotic property of QMLE estimators, and an LM type of test for Normality of residuals. Section 4 shows the Monte Carlo simulation results of finite sample properties of QMLE and normality test. Finally, Section 5 provides an empirical example focusing on monthly stock returns of 11 eurozone countries from March 1999 to April 2021. A comparison of the performance of our model with some popular traditional conditional heteroskedasticity models is also presented in Section 5. Finally, Section 6 is a brief conclusion of this paper.

2   Model Formation

2.1   Alternative Model Specifications

To capture the conditional heteroskedasticity, the most popular and fundamental model is the univariate ARCH model in Engle (1982). Let $y_t = \sigma_t \varepsilon_t$ where $\varepsilon_t \sim i.i.d$ with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$, the conditional volatility of y_t has a function form

$$\sigma_t^2 = f\left(y_{t-1}, y_{t-2}, \cdots, , \mu\right)$$
$$= f\left(\sigma_{t-1}\varepsilon_{t-1}, \sigma_{t-1}\varepsilon_{t-2}, \cdots, \mu\right)$$

where μ is a constant. When considering risk spillover from multiple markets, let $-i$ denote the markets correlated with market i, the function form of conditional volatility can be modified as

$$\sigma_{i,t}^2 = f_i\left(y_{i,t-1}, y_{i,t-2}, \cdots; y_{-i,t}, y_{-i,t-1}, \cdots, \mu_i\right)$$

for market i.

For the functional choice of f, in existing literature, there are several popular specifications. The simplest model formation is linear, which gives the regular ARCH model:

$$\sigma_t^2 = \mu + \sum_{j=1}^{p} \alpha_j y_{t-j}^2$$

when we only consider one market without any spillover.

However, if we extend the linear specification to a situation with inter- and intra-temporal spillover with other markets, the model will have a complicated structure and be hard to estimate. Let $W_n = (w_{ij,n})_{n \times n}$ be the spatial correlation matrix among n markets, where $w_{ij,n}$ captures the spillover from market i to market j. For regularity, we assume $w_{ij,n} \geq 0$ and $w_{ii,n} = 0$ for every $i, j = 1, \cdots, n$. Considering the effect from one-period time lag, we can write the conditional volatility of market i as

$$\sigma_{i,t}^2 = \mu_i + \lambda \sum_{j=1}^{n} w_{ij,n} y_{j,t}^2 + \gamma y_{i,t-1}^2 + \rho \sum_{j=1}^{n} w_{ij,n} y_{j,t-1}^2$$

without considering effect from external regressors. As $y_{i,t} = \sigma_{i,t}\varepsilon_{i,t}$, the model can be further written as

$$y_{i,t}^2 = \left(\mu_i + \lambda \sum_{j=1}^{n} w_{ij,n} y_{j,t}^2 + \gamma y_{i,t-1}^2 + \rho \sum_{j=1}^{n} w_{ij,n} y_{j,t-1}^2\right) \varepsilon_{i,t}^2$$

Although the linear specification seems simple, as the correlation among $\{y_{i,t}\}$ and $\{\varepsilon_{i,t}\}$ is extremely complex, MLE and GMM approaches are nearly impossible. Consider the simplest case with no time lag effect, the model becomes $y_{i,t}^2 = \mu_i \varepsilon_{i,t}^2 + \lambda \sum_{j=1}^{n} w_{ij,n} y_{j,t}^2 \varepsilon_{i,t}^2$. Denote $y_t^2 = \left(y_{1,t}^2, \cdots, y_{n,t}^2\right)'$, $\mu = (\mu_1, \cdots, \mu_n)'$, $W_n = (w_{ij,n})_{n \times n}$ and $\varepsilon_t^2 = \left(\varepsilon_{1,t}^2, \cdots, \varepsilon_{n,t}^2\right)'$, we can rewrite the model into the following form:

$$y_t^2 = \lambda W_n diag\left(\varepsilon_t^2\right) y_t^2 + diag\left(\mu\right) \varepsilon_t^2$$

where $diag\left(\varepsilon_t^2\right) = \begin{bmatrix} \varepsilon_{1,t}^2 & & \\ & \ddots & \\ & & \varepsilon_{n,.t}^2 \end{bmatrix}$.

Then, we have

$$\left[I_n - \lambda W_n diag\left(\varepsilon_t^2\right)\right] y_t^2 = diag\left(\mu\right) \varepsilon_t^2$$

When the distribution of $\varepsilon_{i,t}$ is continuous with unbounded support on $\mathbb{R}$, such like $N(0,1)$, since all the matrix norms are equivalent, i.e. for any two matrix norm $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$, we have that

$r\left\|A\right\|_{\alpha}\leq\left\|A\right\|_{\beta}\leq s\left\|A\right\|_{\alpha}$ some positive numbers r and s, for all matrices $A\in\mathbb{R}^{n\times n}$ [1], there is a positive possibility that $\left\|\lambda W_n diag\left(\varepsilon_t^2\right)\right\|>1$ for all matrix norms and any real number $\lambda\neq 0$ since they are all equivalent to $\left\|\cdot\right\|_{\infty}$ and $\left\|\lambda W_n diag\left(\varepsilon_t^2\right)\right\|_{\infty}=\sup_i\left|\lambda\right|\sum_{j=1}^{n}\left|w_{ij,n}\varepsilon_j^2\right|$ can be larger than any given positive number with non-zero probability when $W_n\neq 0$. In this scenario, we are not able to write the $\left[I_n-\lambda W_n diag\left(\varepsilon_t^2\right)\right]^{-1}$ into an infinite sum of powers of $\lambda W_n diag\left(\varepsilon_t^2\right)$ and do further discussion on the projection between ε_t^2 and y_t^2. When $\varepsilon_{i,t}$ has a bounded support, for example a continuous uniform distribution, for some λ, $\left\|\lambda W_n diag\left(\varepsilon_t^2\right)\right\|_{\infty}<1$ can be guaranteed since $\lambda W_n diag\left(\varepsilon_t^2\right)$ is a bounded matrix. However, with the following form

$$y_t^2 = \left[I_n-\lambda W_n diag\left(\varepsilon_t^2\right)\right]^{-1}diag\left(\mu\right)\varepsilon_t^2$$
$$= \sum_{l=0}^{\infty}\lambda^l\left[W_n diag\left(\varepsilon_t^2\right)\right]^l diag\left(\mu\right)\varepsilon_t^2$$

, with non-separable terms $\left[W_n diag\left(\varepsilon_t^2\right)\right]^l$, further investigating the relationship between ε_t^2 and y_t^2 is almost impossible. Thus, it is nearly impossible to write down the MLE based on density of $\varepsilon_{i,t}$. For moment condition, as $E\left(y_t^2\right)$ is extremely hard to be calculated, thus GMM is also not a feasible approach.

Due to the problems of linear specifications above, other forms are needed. The most straightforward way is to consider taking log-transformation on square of $y_{i,t}$, instead of using linear specification. Some recent literature had discussed asymptotic properties by comparing different specifications including log-ARCH and exponential GARCH, such like Francq, Wintenberger and Zakoïan (2013). Focusing on conditional log-volatility is easier to build up likelihood based estimators and the conditions for consistency and asymptotic normality are much weaker than directly use conditional volatility. Based on similar idea and consideration, in the remaining parts of this paper, we will focus on the specification with log-transformation to extend the conditional volatility model to a setting with allowing risk-spillovers across regions and markets.

2.2   DGP of ESPARCH(1,1) Model

In Nelson (1991), a general exponential ARCH form is proposed, that the conditional log-volatility of an asset is

$$\ln\left(\sigma_t^2\right)=\alpha_t+\sum_{k=1}^{\infty}\beta_k g\left(\varepsilon_{t-k}\right),\beta_1\equiv 1$$

where $\left\{\alpha_t\right\}_{t=-\infty}^{\infty}$, $\left\{\beta_t\right\}_{t=-\infty}^{\infty}$ are real, non-stochastic, scaler sequences, and g is a function. For multiple assets situation, to capture spill-over effect through networks, we extend the model and have the log-volatility for each asset i as

$$\ln\left(\sigma_t^2\right)=\alpha_t+\sum_{k=1}^{\infty}\beta_k g\left(y_{i,t-k}\right)+\sum_{s=0}^{\infty}\sum_{j=1,j\neq i}^{n}\lambda_k w_{ij,n}g\left(y_{j,t-s}\right)$$

where $g\left(x\right)=\ln x^2$ for $x\neq 0$. Here $w_{ij,n}$ satisfies $w_{ii}=0$ and $w_{ij,n}\geq 0$ for $\forall i\neq j$, which captures

---

[1]See Chapter 5 of Roger A. Horn and Charles R. Johnson (2013).

the network correlation from i to j. Thus, $\lambda_k w_{ij,n} g(y_{j,t-k})$ captures the inter-temporal spillover effect from i to j on conditional volatility $s > 0$ periods ago. When $s = 0$, it captures the intra-temporal spillover effect. In this paper, we will mainly focus a simple situation that $k = 1$ and $s = 0, 1$ which we call a ESPARCH(1,1) model:

$$y_{i,t} = \sigma_{i,t} \varepsilon_{i,t}, \varepsilon_{i,t} \overset{iid}{\sim} (0,1) \tag{1}$$

$$\ln \sigma_{i,t}^2 = \mu_i + \lambda \sum_{j=1}^{n} w_{ij,n} \ln y_{j,t}^2 + \gamma \ln y_{i,t-1}^2 + \rho \sum_{j=1}^{n} w_{ij,n} \ln y_{j,t-1}^2 \tag{2}$$

The order of spatial lag and time lag are both 1, i.e. we only consider risk-spillover through one particular network and only consider dynamic effect from the previous period. In this model, μ_i captures the long-run persistent idiosyncratic risk pattern, the spatial autoregressive term $\lambda \sum_{j=1}^{n} w_{ij,n} \ln y_{j,t}^2$ captures the interaction and intra-temporal spill-over effect, the time series autoregressive $\gamma \ln y_{i,t-1}^2$ captures the dynamic effect from its own previous period, and finally the spatial-temporal autoregressive term $\rho \sum_{j=1}^{n} w_{ij,n} \ln y_{j,t-1}^2$ captures the dynamic spill-over effect from other markets. Since after log-transformation, there is no positive limitation on the LHS of (2), μ_i can take any real numbers.

From (2), for each i, we have

$$\lambda w_{ij,n} = \frac{\partial \ln \sigma_{i,t}^2}{\partial \ln y_{j,t}^2} \approx \frac{\triangle \sigma_{i,t}^2 / \sigma_{i,t}^2}{\triangle y_{j,t}^2 / y_{j,t}^2}$$

$$\gamma = \frac{\partial \ln \sigma_{i,t}^2}{\partial \ln y_{i,t-1}^2} \approx \frac{\triangle \sigma_{i,t}^2 / \sigma_{i,t}^2}{\triangle y_{i,t-1}^2 / y_{i,t-1}^2}$$

$$\rho w_{ij,n} = \frac{\partial \ln \sigma_{i,t}^2}{\partial \ln y_{j,t-1}^2} \approx \frac{\triangle \sigma_{i,t}^2 / \sigma_{i,t}^2}{\triangle y_{j,t-1}^2 / y_{j,t-1}^2}$$

where the change of variable x is small enough, i.e. $\triangle x \approx 0$. Thus, economically, the parameters λ, γ and ρ capture the elasticity of conditional volatility with respect to the volatility of other assets and historical volatility of its own and other assets. In the ESPARCH(1,1) model, all the elasticities are constant, thus it captures the long-run equilibrium reaction of one market to fluctuations of neighborhood markets as well as the information from the past. Additionally, as elasticity is scale-free, we can also avoid the situation that the estimated risk spillover is dominated by extremely volatile neighborhood markets. In this sense, the log-transformed form is better than the linear specification.

Denote the past history from period 0 to period $(t-1)$ as $\mathscr{F}_{t-1}$, we always have

$$E(y_{i,t} | \mathscr{F}_{t-1}) = E(\sigma_{i,t} | \mathscr{F}_{t-1}) E(\varepsilon_{i,t}) = 0$$

By transforming (1), we have

$$\ln y_{i,t}^2 = \ln \sigma_{i,t}^2 + \ln \varepsilon_{i,t}^2 \tag{3}$$

Combine (2) and (3), we have

$$\ln y_{i,t}^2 = \mu_i + \lambda \sum_{j=1}^{n} w_{ij,n} \ln y_{j,t}^2 + \gamma \ln y_{i,t-1}^2 + \rho \sum_{j=1}^{n} w_{ij,n} \ln y_{j,t-1}^2 + \ln \varepsilon_{i,t}^2 \tag{4}$$

Let $logY_t^2 = \left(\ln y_{1,t}^2, \cdots, \ln y_{n,t}^2\right)'$, $\mu = \left(\mu_1, \cdots, \mu_n\right)'$, $log\varepsilon_t^2 = \left(\ln \varepsilon_{1,t}^2, \cdots, \ln \varepsilon_{n,t}^2\right)'$ and $W_n = \left(w_{ij,n}\right)_{n\times n}$, (4) can be transformed into the following vector form:

$$logY_t^2 = \mu + \lambda W_n logY_t^2 + \left(\gamma I_n + \rho W_n\right) logY_{t-1}^2 + log\varepsilon_t^2 \tag{5}$$

When $\left(I_n - \rho W_n\right)^{-1}$ exists, we can rewrite (5) into the following reduced form representation:

$$
\begin{aligned}
logY_t^2 &= \left(I_n - \lambda W_n\right)^{-1} \mu + \left(I_n - \lambda W_n\right)^{-1} \left(\gamma I_n + \rho W_n\right) logY_{t-1}^2 \\
&+ \left(I_n - \lambda W_n\right)^{-1} log\varepsilon_t^2
\end{aligned} \tag{6}
$$

and

$$
\begin{aligned}
log\Sigma_t^2 &= \left(I_n - \lambda W_n\right)^{-1} \mu + \left(I_n - \lambda W_n\right)^{-1} \left(\gamma I_n + \rho W_n\right) logY_{t-1}^2 \\
&+ \left[\left(I_n - \lambda W_n\right)^{-1} - I_n\right] log\varepsilon_t^2
\end{aligned} \tag{7}
$$

since $log\Sigma_t^2 = logY_t^2 - log\varepsilon_t^2$ where $log\Sigma_t^2 = \left(\ln \sigma_{1,t}^2, \cdots, \ln \sigma_{n,t}^2\right)'$. As $log\varepsilon_t^2$ are i.i.d random vectors, $logY_t^2$ follows a VAR model with a specification of neighborhood effects.

By the following decomposition

$$
\begin{aligned}
& \left(I_n - \lambda W_n\right)^{-1} \left(\gamma I_n + \rho W_n\right) \\
=& \sum_{l=0}^{\infty} \left(\lambda W_n\right)^l \left(\gamma I_n + \rho W_n\right) \\
=& \sum_{l=0}^{\infty} \lambda^l W_n^l \left(\gamma I_n + \rho W_n\right) \\
=& \gamma \sum_{l=0}^{\infty} \lambda^l W_n^l + \rho \sum_{l=0}^{\infty} \lambda^l W_n^{l+1} \\
=& \gamma I_n + \left(\rho + \gamma\lambda\right) \sum_{l=1}^{\infty} \lambda^{l-1} W_n^l
\end{aligned}
$$

when $\|\rho W_n\|_\infty < 1$. $\left(I_n - \lambda W_n\right)^{-1} \left(\gamma I_n + \rho W_n\right) logY_{t-1}^2$ captures the combination of time-series momentum effect of the market itself and exponentially decade dynamic risk-spillover effects from neighbor market, neighbor's neighbor, etc. By inverting the intra-temporal spatial correlation operator $\left(I_n - \lambda W_n\right)$, the intra-temporal spillovers from first-order neighbors are transformed to inter-temporal spillovers combined from different orders of neighbors.

2.3 Covariance Stationarity of $\ln y_{i,t}^2$

As this model is to capture the dynamic of conditional log-volatility process, investigating the stationarity of $\ln y_{i,t}^2$ is important. Let $E\left(log\varepsilon_t^2\right) = \omega$ and $\xi_t = log\varepsilon_t^2 - \omega$, when $\left(I_n - \rho W_n\right)^{-1}$ exists, we can rewrite (6) as the following:

6

$$logY_t^2 = (I_n - \lambda W_n)^{-1} (\mu + \omega) + (I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) logY_{t-1}^2$$
$$+ (I_n - \lambda W_n)^{-1} \xi_t \tag{8}$$

By backward induction, given initial observation at $t = 0$, we

$$logY_t^2 = \sum_{s=0}^{t} \left[(I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right]^s (I_n - \lambda W_n)^{-1} (\mu + \omega)$$
$$+ \left[(I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right]^t logY_0^2$$
$$+ \sum_{s=0}^{t} \left[(I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right]^{t-s} (I_n - \lambda W_n)^{-1} \xi_s$$

For any fixed n, to make sure this process is stationary, we need all the eigenvalues of $(I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n)$ lie inside the unit circle to make sure $\lim_{s \to \infty} \left[(I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right]^s = 0$. A necessary condition is $\left\| (I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right\|_\infty < 1$.

For most commonly used row-normalized spatial weighting matrix, i.e. $\sum_{j=1}^{n} w_{ij} = 1$ for $\forall i$, we have $\|\lambda W_n\|_\infty = |\lambda|$. We need $|\lambda| < 1$ to make sure $(I_n - \lambda W_n)^{-1}$ exist. Then, we have

$$\left\| (I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right\|_\infty$$
$$= \left\| \gamma \sum_{l=0}^{\infty} \lambda^l W_n^l + \rho \sum_{l=0}^{\infty} \lambda^l W_n^{l+1} \right\|_\infty$$
$$\leq |\gamma| \sum_{l=0}^{\infty} |\lambda|^l \left\| W_n^l \right\|_\infty + |\rho| \sum_{l=0}^{\infty} |\lambda|^l \left\| W_n^{l+1} \right\|_\infty$$
$$= \frac{|\gamma| + |\rho|}{1 - |\lambda|}$$

Thus, when $\frac{|\gamma| + |\rho|}{1 - |\lambda|} < 1$, i.e. $|\lambda| + |\gamma| + |\rho| < 1$, $\left\{ \ln y_{i,t}^2 \right\}_{t=1,\cdots,T}$ and $\left\{ \ln \sigma_{i,t}^2 \right\}_{t=1,\cdots,T}$ are stationary.

When $\left\| (I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right\|_\infty < 1$ holds, we have

$$E\left[logY_t^2 \right] = \left[I_n - (I_n - \lambda W_n)^{-1} (\gamma I_n + \rho W_n) \right]^{-1}$$
$$\cdot (I_n - \lambda W_n)^{-1} (\mu + \nu)$$
$$= \left[(1 - \gamma) I_n - (\lambda + \rho) W_n \right]^{-1} (\mu + \omega)$$

and

$$E\left[log\Sigma_t^2 \right] = E\left[logY_t^2 \right] - E\left[log\varepsilon_t^2 \right]$$
$$= \left[(1 - \gamma) I_n - (\lambda + \rho) W_n \right]^{-1} (\mu + \omega) - \omega$$

which can be easily computed from (6) and (7). For the covariance matrix of $logY_t^2$ and covariance between $logY_t^2$ and $logY_{t-k}^2$ where k is positive integer and less than t, they are easy to compute

7

by multivariate Yule-Walker equation method as long as $E\xi_t^2 < \infty$. Thus, with proper condition, the ESPARCH(1,1) model is a stable model, and $\left\{\ln Y_t^2\right\}_t$ is covariance stationary. The QMLE estimator discussed in the following session will be limited to stationary situation only.

3 Quasi-maximum Likelihood Estimation

3.1 QMLE for Normal Disturbance Situation

In classical ARCH/GARCH and stochastic volatility literature, we often assume the disturbance terms are standard Normal, especially predicting conditional volatility in the future. When $\varepsilon_{i,t} \overset{iid}{\sim} N(0,1)$, we have $\ln \varepsilon_{i,t}^2$ follows log-χ^2 distribution. The mean and variance of $\ln \varepsilon_{i,t}^2$ are known to be $\psi\left(\frac{1}{2}\right) - \ln\left(\frac{1}{2}\right) \approx -1.2704$ and $\frac{1}{2}\pi^2$ respectively, where $\psi(\cdot)$ is the Digamma function. Thus, $\xi_{i,t}$ is a non-Gaussian white noise with zero mean and $\frac{1}{2}\pi^2$ variance. Let $logY_t^2 = Z_t$ and $\eta = \mu - 1.27l_n$ where $l_n = \left(\underbrace{1,\cdots 1}_{n}\right)^{'}$, then we can rewrite (5) as

$$Z_t = \eta + \lambda W_n Z_t + (\gamma I_n + \rho W_n) Z_{t-1} + \xi_t \tag{9}$$

Similar to stochastic volatility models, we can construct QMLE for our model by using Normal density to approximate the density of ξ_t^2. By (9), let $\theta = (\lambda, \gamma, \rho)^{'}$, then the quasi-density function of ξ_t is

$$f(\xi_t) = \pi^{-3n/2} \exp\left\{-\frac{1}{\pi^2}\xi_t'\xi_t\right\}$$

Then, the conditional quasi-log-density function for $t = 1, \cdots, T$ is

$$q_{n,t}\left(X_t; \theta, \eta | \mathscr{F}_{t-1}\right) = -\frac{3}{2}n\ln(\pi) - \frac{1}{\pi^2}\left[S_n(\lambda) Z_t - (\gamma I_n + \rho W_n) Z_{t-1} - \eta\right]^{'}$$
$$\cdot \left[S_n(\lambda) Z_t - (\gamma I_n + \rho W_n) Z_{t-1} - \eta\right] + \ln|S_n(\lambda)|$$

where $S_n(\lambda) = I_n - \lambda W_n$. Then, given $\{Y_t\}_{t=0}^{T}$, we can estimate θ by maximizing the following quasi-log-likelihood function:

$$Q_{n,T}(\theta, \eta) = -\frac{3}{2}nT\ln(\pi) - \frac{1}{\pi^2}\sum_{t=1}^{T}\left[S_n(\lambda) Z_t - (\gamma I_n + \rho W_n) Z_{t-1} - \eta\right]^{'}$$
$$\cdot \left[S_n(\lambda) Z_t - (\gamma I_n + \rho W_n) Z_{t-1} - \eta\right] + T\ln|S_n(\lambda)| \tag{10}$$

Since the dimension of η is n, which makes the optimization procedure harder to implement with large number of markets, we can concentrate out η to simplify the quasi-log-likelihood function and numerical optimization procedure. From (10), it is easy to get the first order condition of η:

$$\frac{\partial Q_{n,T}(\theta, \eta)}{\partial \eta^{'}} = -\frac{2}{\pi^2}\sum_{t=1}^{T}\left[S_n(\lambda) Z_t - (\gamma I_n + \rho W_n) Z_{t-1} - \eta\right]^{'} \tag{11}$$

When $Q_{n,t}(\theta, \eta)$ is maximized, we have $\frac{\partial Q_{n,T}(\theta,\eta)}{\partial \eta^{'}} = 0$, then

---

[2]See Harvey, Ruiz and Shepherd (1994)

8

$$\hat{\eta} = \frac{1}{T} \sum_{t=1}^{T} \left[S_n \left(\hat{\lambda} \right) Z_t - \left(\hat{\gamma} I_n + \hat{\rho} W_n \right) Z_{t-1} \right] \tag{12}$$

Then, by (12), we can get the concentrated quasi-log-likelihood function as

$$\widetilde{Q}_{n,T} (\theta) = -\frac{3}{2} nT \ln (\pi) - \frac{1}{\pi^2} \sum_{t=1}^{T} \left[S_n (\lambda) \widetilde{Z}_t - (\gamma I_n + \rho W_n) \widetilde{Z}_{t-1} \right]'$$
$$\cdot \left[S_n (\lambda) \widetilde{Z}_t - (\gamma I_n + \rho W_n) \widetilde{Z}_{t-1} \right] + T \ln |S_n (\lambda)| \tag{13}$$

where $\widetilde{Z}_t = Z_t - \frac{1}{T} \sum_{t=1}^{T} Z_t$. By maximizing $\widetilde{Q}_{n,T} (\theta)$ and then back out $\hat{\eta}$ and $\hat{\mu}$, we can get estimate all the parameters.

3.2 QMLE When $\varepsilon_{i,t}$ Is Not Normal

When $\varepsilon_{i,t}$ are not Normal, the properties of $\ln \varepsilon_{i,t}^2$ also changed, especially for its mean and variance, thus the QMLE derived in Section 3.1 would not work. Fortunately, we can use similar technique to construct QMLE for some particular situations. In finance literature, the t- distribution is also widely used as a residual process to capture the heavy-tailed risk. When $\varepsilon_{i,t} \overset{i.i.d}{\sim} \sqrt{\frac{v-2}{v}} t (v)$ for $v \geq 3$ after rescaled variance to 1, we can rewrite it as $\varepsilon_{i,t} = \sqrt{\frac{v-2}{v}} \zeta_{i,t} / \kappa_{i,t}^{\frac{1}{2}}$, thus

$$\ln \varepsilon_{i,t}^2 = \ln \left(\frac{v-2}{v} \right) + \ln \zeta_{i,t}^2 - \ln \kappa_{i,t} \tag{14}$$

where $\zeta_{i,t} \overset{i.i.d}{\sim} N (0,1)$ and $\kappa_{i,t} \overset{i.i.d}{\sim} \chi^2 (v)$ with degree of freedom v. Thus, by properties of log-χ^2 distribution with degree of freedom v[3], we have

$$E \left(\ln \varepsilon_{i,t}^2 \right) = \ln \left(\frac{v-2}{v} \right) + \psi \left(\frac{1}{2} \right) - \ln \left(\frac{1}{2} \right) - \psi \left(\frac{v}{2} \right) + \ln \left(\frac{1}{2} \right)$$
$$= \ln \left(\frac{v-2}{v} \right) - \psi \left(\frac{v}{2} \right) + \psi \left(\frac{1}{2} \right)$$

and

$$var \left(\ln \varepsilon_{i,t}^2 \right) = \frac{1}{2} \pi^2 + \psi' \left(\frac{v}{2} \right)$$

where $\psi (\cdot)$ is the Digamma function and $\psi' (\cdot)$ is the Trigamma function. Thus, with the assumption that $\varepsilon_{i,t} \overset{i.i.d}{\sim} t (v)$, denote $\eta = \mu + \left[-1.27 + \frac{1}{2} \ln \left(\frac{v-2}{v} \right) - \psi \left(\frac{v}{2} \right) + \ln \left(\frac{v}{2} \right) \right] l_n$, we can derive a dynamic spatial panel form with the same form as (9). Thus, the QMLE will have the similar form as (10) despite the variance is no longer $\frac{1}{2} \pi^2$, but $\frac{1}{2} \pi^2 + \psi' \left(\frac{v}{2} \right)$.

Another more general situation is that we do not put any assumptions on the $\varepsilon_{i,t}$. In this case, QMLE is still possible but not all parameters can be identified. In this situation, we are not able to estimate the conditional variance $\sigma_{i,t}^2$ terms based on (7) as μ is not identified due to the unknown expectation of $\ln \varepsilon_{i,n}^2$ after log-transformation. However, we can still capture the network effects of the conditional volatilities. Assume $E \left(\ln \varepsilon_{i,n}^2 \right) = \omega$ and $var \left(\ln \varepsilon_{i,n}^2 \right) = \sigma^2$, denote $log Y_t^2 = Z_t$ and

---

[3] See Abramovitz and Stegun (1970)

$\eta = \mu + \omega l_n$, then we can still get an equation which has the same form as (10). In this situation, as σ^2 is unknown, the QMLE will be

$$H_{n,T}\left(\sigma^2, \theta, \eta\right) = -\frac{nT}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[S_n\left(\lambda\right)Z_t - \left(\gamma I_n + \rho W_n\right)Z_{t-1} - \eta\right]'$$

$$\cdot\left[S_n\left(\lambda\right)Z_t - \left(\gamma I_n + \rho W_n\right)Z_{t-1} - \eta\right] + T\ln\left|S_n\left(\lambda\right)\right| \tag{15}$$

If we believe $\varepsilon_{i,t}$ follows a distribution with skewness or kurtosis different from Normal distribution, it would be a better choice than assuming Normal if the purpose is not to forecast the conditional volatility. Similar to the normal situation, we can concentrate out η by its first order condition, and maximize the following concentrated log-likelihood function

$$\widetilde{H}_{n,T}\left(\sigma^2, \theta\right) = -\frac{nT}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[S_n\left(\lambda\right)\widetilde{Z}_t - \left(\gamma I_n + \rho W_n\right)\widetilde{Z}_{t-1}\right]'$$

$$\cdot\left[S_n\left(\lambda\right)\widetilde{Z}_t - \left(\gamma I_n + \rho W_n\right)\widetilde{Z}_{t-1}\right] + T\ln\left|S_n\left(\lambda\right)\right| \tag{16}$$

where $\widetilde{\eta}\left(\sigma^2, \lambda, \gamma, \rho\right) = \frac{1}{T}\sum_{t=1}^{T}\left[S_n\left(\lambda\right)Z_t - \left(\gamma I_n + \rho W_n\right)Z_{t-1}\right]$, $\widetilde{Z}_t = Z_t - \frac{1}{T}\sum_{t=1}^{T}Z_t$ and $\widetilde{Z}_{t-1} = Z_{t-1} - \frac{1}{T}\sum_{t=1}^{T}Z_{t-1}$.

3.3 Identification and Asymptotic Properties of QMLE

When the distribution of $\varepsilon_{i,t}$ is is unknown, QMLE is based on (16). When $\xi_{i,t}$ are i.i.d residuals with unknown distribution, by denoting $\psi = \left(\sigma^2, \theta'\right)'$, the QMLE $\hat{\psi}_{nT}$ is $\sqrt{nT}-$consistent and asymptotically normal when $n/T \to 0$ in Yu, de Jong and Lee (2008)[4]. The limiting distribution of $\hat{\psi}_{nT}$ is the following

$$\sqrt{nT}\left(\hat{\psi}_{nT} - \psi_0\right) \xrightarrow{d} N\left(0, \Sigma_{\psi_0}^{-1}\left(\Sigma_{\psi_0} + \Omega_{\psi_0}\right)\Sigma_{\psi_0}^{-1}\right)$$

where $\Sigma_{\psi_0} = E\left(\frac{1}{nT}\frac{\partial^2 \widetilde{Q}_{n,T}(\psi_0)}{\partial\psi'\partial\psi}\right)$ and $\Omega_{\psi_0} = \left(\frac{4\mu_4}{\sigma_0^4} - 3\right)\begin{pmatrix} \frac{1}{4\sigma_0^4} & \frac{1}{2\sigma_0^2 n}tr(G_n) & 0_{2\times 2} \\ \frac{1}{2\sigma_0^2 n}tr(G_n) & \frac{1}{n}\sum_{i=1}^{n}G_{n,ii}^2 & \\ 0_{2\times 2} & & 0_{2\times 2} \end{pmatrix}$,

μ_4 is the forth order central moment of $\xi_{i,t}$ which requires $E\left|\xi_{i,t}\right|^{4+\epsilon} < \infty$ for some positive ϵ, and $G_{n,ii}$ is the (i,i) entry of $G_n = W_n S_n^{-1}\left(\psi_0\right)$. For the fixed effects η_i, we have $\sqrt{T}\left(\hat{\eta}_{i,nT} - \eta_{i,0}\right) \xrightarrow{d} N\left(0, \sigma_0^2\right)$ for $i = 1, \cdots, n$ and they are asymptotically independent with each other.

For the situation that n is larger than T, it is still consistent despite the limiting distribution is non-centralized. Under some regularity assumptions, the QMLE should also be $\sqrt{nT}-$consistent and asymptotic normality when $n/T \to \infty$. As this short panel situation is not associated with most applications of conditional heteroskedasticity models and the limiting distribution is quite complicated, we will not show it here.

When $\varepsilon_{i,t}$ is known as standard normal or $t-$distribution, the only difference between quasi-

---

[4]See Section 3 of Yu, de Jong and Lee (2008)

log-likelihood functions is that σ_0^2 is known. The asymptotic distribution of $\hat{\theta}_{nT}$ when $n/T \to 0$ has exactly the same form as the general situation. Additionally, as we can back out $\hat{\mu}_{i,nT}$ from $\hat{\eta}_{i,nT}$ and their difference is just a constant, for $i = 1, \cdots, n$, we have $\sqrt{T}(\hat{\mu}_{i,nT} - \mu_{i,0}) \xrightarrow{d} N(0, \sigma_0^2)$ and they are asymptotically independent with each other. One may concern whether $\xi_{i,t}$ satisfies $E|\xi_{i,t}|^{4+\epsilon} < \infty$ for some positive ϵ. As the existence of positive order moments of log-χ^2 random variables can be derived from its moment generating function, $m(t) = 2^t \Gamma\left(t + \frac{v}{2}\right) / \Gamma\left(\frac{v}{2}\right)$, which is showed in Appendix A.7 in Peter M.Lee (2012). As $\xi_{i,t}$ is $log - \chi^2(1)$ or linear combination of two independent log-χ^2 random variables when $\varepsilon_{i,t}$ is normal or $t-$distributed, the requirement of $(4 + \epsilon)$ moment is satisfied.

In Section 4, the finite sample performance of the QMLE will be investigated by Monte Carlo simulations. Since our model is focused on time-varying conditional variance, a typical application should be the case when T is relatively larger than n. Thus, the Monte Carlo simulations will focus on this scenario.

3.4   A Modified LM Test for Normality

Sometimes, we may be interested in testing whether the true residual process of $\varepsilon_{i,t}$ is standard normal distribution. One potential scenario is forecasting. If the true distribution is normal, then we can estimate all the parameters including the individual fixed effect, and also perform out-of-sample prediction for conditional log-volatility with constructing confidence intervals. Another purpose for the normality test is for testing whether the risk is asymmetric due to leverage effect and extreme tail risks. In our ESPARCH(1,1) model, since the QMLE is based on $\ln \varepsilon_{i,t}^2$ and $\sigma^2 = Var\left(\ln \varepsilon_{i,t}^2\right)$ is one of our estimator, we can construct the normality test based on σ^2 which had been suggested in literature of stochastic volatility models, such like Ruiz (1994). When $\varepsilon_{i,t} \overset{i.i.d}{\sim} N(0,1)$, $Var\left(\ln \varepsilon_{i,t}^2\right) = \frac{1}{2}\pi^2 \approx 4.9348$. Let $H_0 : \sigma^2 = \frac{1}{2}\pi^2$ and $H_1 : \sigma^2 \neq \frac{1}{2}\pi^2$, if we reject H_0, then the true distribution of $\varepsilon_{i,t}$ is $N(0,1)$ can also be rejected.

To test H_0, the most convenient way is to estimate the model under the normal assumption, and then to implement LM test. From (16), based on the concentrated likelihood function, we can get the first order condition for σ^2:

$$\frac{\partial \widetilde{H}_{n,T}(\psi)}{\partial \sigma^2} = -\frac{nT}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{t=1}^{T} \left[S_n(\lambda)\widetilde{Z}_t - (\gamma I_n + \rho W_n)\widetilde{Z}_{t-1} \right]'$$
$$\cdot \left[S_n(\lambda)\widetilde{Z}_t - (\gamma I_n + \rho W_n)\widetilde{Z}_{t-1} \right]$$

Thus, under H_0, given the constrained estimator $\bar{\psi}$ from (13), the score function for σ^2 should be

$$g_{\sigma^2}(\bar{\psi}) = -\frac{nT}{\pi^2} + \frac{2}{\pi^4} \sum_{t=1}^{T} \bar{u}'_{c,t}\bar{u}_{c,t}$$

where $\bar{u}_{c,t} = S_n(\bar{\lambda})\widetilde{Z}_t - (\bar{\gamma}I_n + \bar{\rho}W_n)\widetilde{Z}_{t-1}$.

However, due the concentrated likelihood approach and the existence of the fixed effect terms, the limiting distribution of $\frac{1}{\sqrt{nT}}\frac{\partial \widetilde{H}_{n,T}}{\partial \sigma^2}$ is not centralized around zero, thus traditional LM test statistic does not have the proper limiting distribution. To construct a test statistic with regular limiting distribution, we need to modify the score function as well as the information matrix. Define

$g\left(\bar{\psi}\right)$ and $H\left(\bar{\psi}\right)$ as the following:

$$g_{n,T}\left(\bar{\psi}\right) = \left(g_{\sigma^2}\left(\bar{\psi}\right) + \frac{n}{\pi^2}, 0, 0, 0\right)'$$

and

$$H_{n,T}\left(\bar{\psi}\right) = T\left(\frac{4\mu_4}{\pi^4} - 3\right)\begin{pmatrix} \frac{n}{\pi^4} & \frac{1}{\pi^2}tr(G_n) & 0_{2\times 2} \\ \frac{1}{\pi^2}tr(G_n) & \sum_{i=1}^{n}G_{n,ii}^2 & \\ & 0_{2\times 2} & 0_{2\times 2} \end{pmatrix}$$
$$- E\left(\frac{\partial^2 \widetilde{H}_{n,T}\left(\bar{\psi}\right)}{\partial\psi'\partial\psi}\right) \tag{17}$$

where μ_4 is the forth order central moment of $log - \chi^2\left(1\right)$ distribution and it equals to 170.46 approximately.[5]

By Claim 2 in Yu, de Jong and Lee (2008), we have

$$J_{Norm} = g_{n,T}'\left(\bar{\psi}\right)H_{n,T}^{-1}\left(\bar{\psi}\right)g_{n,T}'\left(\bar{\psi}\right) \xrightarrow{d} \chi^2\left(1\right) \tag{18}$$

when $T/n \to \infty$.

However, we need to be careful that constructing the normality test based on variance of $\ln\varepsilon_{i,n}^2$ is not perfect. Although the only difference among the QMLEs for normal and non-normal situation is the parameter σ^2, we can not rule out the case, that $\varepsilon_{i,t}$ is not normal, but $Var\left(\ln\varepsilon_{i,n}^2\right)$ are close to $\frac{\pi^2}{2}$. Also, in equation (17) , the asymptotic variance matrix contains forth order moment of $\ln\varepsilon_{i,t}^2$. If both the second and forth order moment are close to $log - \chi^2\left(1\right)$ distribution, the test will not have enough power to reject the normality assumption. Fortunately, it is not a serious problem, since the asymptotic variance of QMLE is also only based on second and forth order moments which exactly comes from the terms in $H_{n,T}\left(\bar{\psi}\right)$. If the second order and forth order central moments of $\ln\varepsilon_{i,t}^2$ are very close to $log - \chi^2\left(1\right)$, then the asymptotic variance of the QMLE will also be very close to the case when $\varepsilon_{i,n}$ is normal, thus fail to reject the normality assumption in this situation will not affect further statistical inference too much. In Section 4, we will investigate the finite sample performance of the J_{Norm} statistic, and also provide an example for the lack of power scenario mentioned here.

4 Monte Carlo Simulations for Finite Sample Performance

4.1 Basic Settings

In this section, we try to perform some Monte Carlo simulations to see whether the QMLE under different assumptions has good finite sample properties. By simulating multiple samples from the data generating process (DGP), and then implementing the QMLE procedure, we can evaluate the performance of QMLE for different sample sizes. We will replicate each Monte Carlo simulation exercise by 1,000 times.

---

[5]The approximation values are generated by the following approach implemented by Matlab: generate 1,000 samples with 10 million random draws from the target distributions, and then take the mean of the sample moments as the approximation. For other simulated moments used in this paper, the approach is the same.

To simulate the DGP, we need first to simulate the spatial correlations among regions which satisfy Assumptions 1-8 of Yu, de Jong and Lee (2008). Here we construct the row-stochastic nearest neighbor spatial weight matrix $W_n = (w_{ij,n})$ using LeSage's econometrics toolbox[6]. The procedure is:

1. Generate two random vectors of coordinates as the geographic location for each observation;

2. Find l nearest neighbors for each observation according to their spatial distances and denote the corresponding $w_{ij,n} = 1$, otherwise $w_{ij,n} = 0$;

3. Row-normalize W_n.

In this paper, we consider two situations when $l = 3$ and $l = 6$.

For the fixed effect term μ_i , in the simulation exercises, we assume μ_i is a random draw from $i.i.d$ uniform distribution on $[0, 1]$. For each simulation exercise, we use the same μ_i for all the 1,000 replications. For other parameters, we consider the following two sets: $(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$ and $(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$. For σ^2 in unknown situations, the true values depend on the true distribution of $\varepsilon_{i,n}$, which we will give a detailed illustration in the following sections.

4.2 Performance of QMLE under Normal Assumption

When $\varepsilon_{i,n} \overset{iid}{\sim} N(0, 1)$, the QMLE can be implemented by the procedure described in Section 3.1. The initial value of each parameter in each round of simulation are set to zero. Observation of time 0 is generated from $N(0, 1)$. The finite sample performance of QMLE under different sample sizes are reported in Table 1 and 2. During the concentrated likelihood process, the performance of the QMLE of μ would be coincident with the performance of other parameters which are not reported here for simplicity. In Table 1 and Table 2, the performance of QMLE are evaluated by five indexes: mean, standard deviation, median, upper and lower quantiles. In both of the cases, as T getting larger, the performance are getting better. When T is relatively large (≥ 100), as long as n is much smaller than T, the performance of QMLE will be good. We also tried some other combinations of n and T, like $n = 100, T = 100$ and $n = 150, T = 100$, the performances are similar to the results showed here with $n = 30, T = 100$.

4.3 Performance of QMLE under t−distribution Assumption

When $\varepsilon_{i,n} \overset{iid}{\sim} \sqrt{\frac{v-2}{v}} t(v)$ with $v \geq 3$, the third and higher order moments of $\varepsilon_{i,n}$ exist, the QMLE procedure is similar to Normal case despite the assumed known σ^2 is different in the QMLE and concentrated QMLE functions. In this section, we show the simulation results for the case $v = 3$. As

---

[6]See https://www.spatial-econometrics.com

Table 1: Finite Sample Performance for $N(0,1)$ Case When $l = 3$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$			$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$		
		λ	γ	ρ	λ	γ	ρ
	mean	.3859	.1626	-.2743	-.2947	.3467	.1763
n=10	std	.0550	.0560	.0800	.0782	.0550	.1012
	med	.3870	.1615	-.2765	-.2941	.3476	.1790
T=30	$q_{0.25}$	.3473	.1260	-.3300	-.3452	.3091	.1050
	$q_{0.75}$	.4254	.1997	-.2225	.2431	.3865	.2466
	mean	.3964	.1903	-.2923	-.3001	.3849	.1921
n=10	std	.0315	.0295	.0416	.0422	.0291	.0574
	med	.3972	.1906	-.2923	-.2993	.3856	.1924
T=100	$q_{0.25}$	.3751	.1711	-.3205	-.3282	.3648	.1516
	$q_{0.75}$	.4175	.2106	-.2641	-.2709	.4038	.2313
	mean	.3980	.1881	-.2933	-.2979	.3858	.1932
n=30	std	.0200	.0170	.0242	.0266	.0172	.0337
	med	.3988	.1881	-.2933	-.2988	.3859	.1933
T=100	$q_{0.25}$	.3852	.1768	-.3097	-.3157	.3745	.1696
	$q_{0.75}$	.4120	.2002	-.2778	-.2799	.3971	.2160

Table 2: Finite Sample Performance for $N(0,1)$ Case When $l = 6$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$			$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$		
		λ	γ	ρ	λ	γ	ρ
	mean	.3776	.1594	-.2743	-.3131	.34739	.1738
n=10	std	.0735	.0575	.0951	.1254	.0519	.1595
	med	.3829	.1609	-.2801	-.3087	.3473	.1723
T=30	$q_{0.25}$	.3363	.1202	-.3346	-.3970	.3108	.0707
	$q_{0.75}$	.4288	.1985	-.2085	-.2254	.3830	.2823
	mean	.3940	.1896	-.2931	-.3018	.3842	.1910
n=10	std	.0398	.0291	.0526	.0710	.0295	.0903
	med	.3965	.1896	-.2908	-.3020	.3836	.1899
T=100	$q_{0.25}$	.3681	.1698	-.3295	-.3512	.3640	.1273
	$q_{0.75}$	.4197	.2097	-.2560	-.2535	.4039	.2535
	mean	.3965	.1887	-.2932	-.2982	.3848	.1938
n=30	std	.0249	.0177	.0331	.0390	.0167	.0480
	med	.3956	.1890	-.2925	-.2967	.3842	.1921
T=100	$q_{0.25}$	.3799	.1770	-.3164	-.3249	.3742	.1613
	$q_{0.75}$	.4126	.2001	-.2702	-.2724	.3952	.2243

Table 3: Finite Sample Performance for $\sqrt{\frac{1}{3}}t(3)$ Case When $l=3$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$			$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$		
		λ	γ	ρ	λ	γ	ρ
	mean	.3842	.1559	-.2793	-.2907	.3456	.1765
n=10	std	.0562	.0580	.0826	.0852	.0553	.1050
	med	.3837	.1559	-.2804	-.2913	.3445	.1756
T=30	$q_{0.25}$	.3464	.1181	-.3356	-.3478	.3089	.1092
	$q_{0.75}$	.4217	.1970	-.2270	-.2362	.3862	.2487
	mean	.3949	.1877	-.2931	-.2991	.3833	.1935
n=10	std	.0301	.0317	.0442	.0443	.0300	.0598
	med	.3960	.1869	-.2922	-.2989	.3824	.1969
T=100	$q_{0.25}$	.3747	.1677	-.3240	-.3293	.3638	.1561
	$q_{0.75}$	.4144	.2085	-.2629	-.2688	.4038	.2327
	mean	.3976	.1880	-.2924	-.2997	.3858	.1952
n=30	std	.0186	.0183	.0271	.0266	.0173	.0339
	med	.3975	.1881	-.2933	-.3003	.3854	.1934
T=100	$q_{0.25}$	.3850	.1761	-.3091	-.3178	.3741	.1725
	$q_{0.75}$	.4097	.1992	-.2749	-.2816	.3974	.2187

v increases, the t- distribution will approaching to $N(0,1)$, then the performance of QMLE would be closer to the Normal case. With $\sqrt{\frac{v-2}{v}}t(v)$ as the distribution of observation at time 0, and the same zero initial value settings for parameters, Table 3 and 4 show the performance of QMLE with the same sample sizes and true parameters in Table 1 and 2. Similar to the performance for Normal case, as T goes larger, the bias shrinks as long as $n/T \to 0$.

4.4 Performance of QMLE with Unknown Residual Distribution

When the distribution of $\varepsilon_{i,t}$ is unknown, we can still estimate the parameters despite μ using the QMLE method shown in Section 3.3. In this part, we considered three different distributions of $\varepsilon_{i,t}$ to simulate three different situations that standard Normal residuals can not capture. With the same parameter and sample size settings as before, by maximizing the QMLE showed in (14), we investigate the finite sample performance for different situations. The only difference is that

Table 4: Finite Sample Performance for $\sqrt{\frac{1}{3}}t(3)$ Case When $l=6$

		$(\lambda_1,\gamma_1,\rho_1)=(.4,.2,-.3)$			$(\lambda_2,\gamma_2,\rho_2)=(-.3,.4,.2)$		
		λ	γ	ρ	λ	γ	ρ
	mean	.3808	.1590	-.2721	-.3030	.3449	.1622
n=10	std	.0725	.0599	.1025	.1231	.0564	.1644
	med	.3883	.1593	-.2710	-.2954	.3452	.1676
T=30	$q_{0.25}$	.3359	.1176	-.3418	-.3811	.3075	.0650
	$q_{0.75}$	.4302	.2007	-.2015	-.2188	.3826	.2643
	mean	.3943	.1876	-.2939	-.3053	.3850	.1949
n=10	std	.0374	.0310	.0542	.0659	.0297	.0876
	med	.3970	.1870	-.2923	-.3028	.3841	.1978
T=100	$q_{0.25}$	.3701	.1674	-.3303	-.3463	.3649	.1343
	$q_{0.75}$	.4201	.2092	-.2589	-.2618	.4073	.2513
	mean	.3978	.1889	-.2935	-.2977	.3853	.1935
n=30	std	.0259	.0181	.0344	.0382	.0177	.0489
	med	.3977	.1888	-.2926	-.2965	.3854	.1946
T=100	$q_{0.25}$	.3809	.1766	-.3171	-.3221	.3730	.1615
	$q_{0.75}$	.4151	.2016	-.2698	-.2721	.3981	.2250

the expectation and variance of $\ln\varepsilon_{i,t}^2$ are unknown. So, we can still linearize the model by taking logarithm. However, it can not identify μ_i terms and also we need to introduce an additional parameter σ^2 as $Var\left(\ln\varepsilon_{i,t}^2\right)$. Thus, unlike the known situation, the performance of $\hat{\sigma}_{QMLE}^2$ also needs to be evaluated.

The first situation we consider is the fat tail risk. To simulate it, we use $\varepsilon_{i,t}\overset{iid}{\sim}\frac{1}{\sqrt{3}}t(3)$. From the analysis before, the distribution of $\ln\varepsilon_{i,t}^2$ is the difference between independent $\log-\chi^2(1)$ and $\log-\chi^2(3)$, plus $\ln\left(\frac{1}{3}\right)$. Then, the expectation and variance of $\ln\varepsilon_{i,t}^2$ are $\psi\left(\frac{1}{2}\right)-\psi\left(\frac{3}{2}\right)=-2$ and $\frac{1}{2}\pi^2+\psi'\left(\frac{3}{2}\right)\approx 5.8696$ respectively, because $\psi(z+1)-\psi(z)=\frac{1}{z}$ for any positive number z. The finite sample performance of this situation is reported in Table 5 and 6.

Another situation is when a price limit policy or circuit breakers mechanism exists in the market. When the upper bond and lower bond of claim prices or the range of change are limited, the volatility will also be affected. For example, in mainland China, the daily price change of a stock are limited to 10% in both Shanghai and Shenzhen Stock Exchange except for some special occasions. In Japan, South Korea, France, etc, similar policies are implemented. In some markets, the currency prices are also limited in daily exchange. The price limit will also affect the volatility of stock returns, which is investigated in finance literature, e.g. , Subrahmanyam (1994) and Kim and Rhee (1998). Thus, sometimes putting limit on volatility is more reasonable due to the existence of the price limit. To simulate this specific situation, we use $\varepsilon_{i,t}\overset{iid}{\sim}Uniform\left[-\sqrt{3},\sqrt{3}\right]$ which has a bonded support. Then, the density function of $\ln\varepsilon_{i,t}^2$ is

$$f(x)=\frac{\sqrt{3}}{6}\exp\left(\frac{1}{2}x\right),x\in(-\infty,\ln 3]$$

where we can derive that $E\left(\ln \varepsilon_{i,t}^2\right) = \ln 3 - 2$ and $Var\left(\ln \varepsilon_{i,t}^2\right) = 4$. The finite sample performance of this situation is reported in Table 7 and 8.

Finally, we consider the situation when the risk is asymmetric, i.e. the return is skewed. Due to leverage effect and risk of extreme bad events, i.e. "black swan" incidents, many asset prices and returns would be skewed. Finance literature had investigated the skewness of US stock returns, e.g., Harvey and Siddique (2000) and Chang, Christofferen and Jacobs (2013). To capture the skewness, we consider to use extreme value distribution for $\varepsilon_{i,t}$ which has the density function:

$$f\left(x\right) = \exp\left\{\frac{\pi}{\sqrt{6}}\left(x - \gamma_{EM}\right) - \exp\left\{\frac{\pi}{\sqrt{6}}\left(x - \gamma_{EM}\right)\right\}\right\}, x \in \mathbb{R} \tag{19}$$

where γ_{EM} is the Euler-Mascheroni constant and $\gamma_{EM} = \lim_{n\to\infty}\left(-\ln n + \sum_{k=1}^n \frac{1}{k}\right) \approx 0.5772$. The skewness of this distribution is $-\frac{12\sqrt{6}}{\pi^3}\zeta\left(3\right) \approx -1.14$ where $\zeta\left(\cdot\right)$ is the Riemann zeta function. The distribution is derived from extreme value theory, which is widely used to capture extreme risk in financial system, e.g. McNeil and Frey (2000) and Poon, Rockinger and Tawn (2004). The extreme value distribution we use here has a fatter and longer tail on the left-hand side, capturing the larger impact of bad events to investors. By using simulation, the mean and variance of $\ln \varepsilon_{i,t}^2$ are approximately equal to -1.37 and 4.89 respectively. Table 9 and 10 show the performance of QMLE in this scenario.

By the results showed in Table 5-10, we can see when the distribution of $\varepsilon_{i,t}$ is unknown, the finite sample performance is similar to the case when we know the true distribution of $\varepsilon_{i,t}$ is normal or Student's t. As T increases, the bias is shrinking as long as T/n also getting smaller. The levels of biases for λ, γ and ρ are also similar to results showed in Table 1-4 when the sample size is the same. Also, for σ^2 , the variance of $\ln \varepsilon_{i,t}^2$, the QMLE estimator has good performance for the all the three situations. Combining the results from Section 4.1 to Section 4.4, we can see that using the density of normal distribution to build up QMLE is a good way to estimate the ESPARCH(1,1) model. For variant situations with considering different types of risks, the finite sample estimators perform well for small samples.

Table 5: Finite Sample Performance for Unknown $\sqrt{\frac{1}{3}}t(3)$ Case When $l=3$

$$\sigma^2 \approx 5.87$$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$				$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$			
		λ	γ	ρ	σ^2	λ	γ	ρ	σ^2
	mean	.3924	.1562	-.2755	5.6420	-.3064	.3480	.1891	5.5664
n=10	std	.0605	.0574	.0812	.7220	.0902	.0542	.1136	.7160
	med	.3976	.1571	-.2782	5.5892	-.3031	.3485	.1872	5.5077
T=30	$q_{0.25}$	.3533	.1167	-.3311	5.1516	-.3640	.3126	.1097	5.0692
	$q_{0.75}$	.4335	.1978	-.2202	6.0903	-.2468	.3867	.2630	5.9948
	mean	.3969	.1886	-.2922	5.8217	-.2879	.3839	.1950	5.7795
n=10	std	.0314	.0317	.0419	.4168	.0462	.0292	.0581	.3947
	med	.3965	.1877	-.2925	5.7900	-.2980	.3849	.1960	5.7729
T=100	$q_{0.25}$	.3766	.1684	-.3192	5.5310	-.3291	.3634	.1561	5.5018
	$q_{0.75}$	.4183	.2109	-.2651	6.0963	-.2653	.4044	.2330	6.0410
	mean	.3969	.1882	-.2943	5.7962	-.3002	.3852	.1946	5.7881
n=30	std	.0197	.0175	.0264	.0243	.0259	.0170	.0324	.2296
	med	.4003	.1880	-.2950	5.7873	-.3008	.3851	.1954	5.7925
T=100	$q_{0.25}$	.3866	.1763	-.3135	5.6314	-.3167	.3738	.1724	5.6381
	$q_{0.75}$	.4131	.1999	-.2760	5.9518	-.2827	.3960	.2165	5.9366

Table 6: Finite Sample Performance for Unknown $\sqrt{\frac{1}{3}}t(3)$ Case When $l=6$

$$\sigma^2 \approx 5.87$$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$				$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$			
		λ	γ	ρ	σ^2	λ	γ	ρ	σ^2
	mean	.3876	.1567	-.2720	5.5864	-.3169	.3452	.1763	5.5510
n=10	std	.0734	.0595	.1024	.7206	.1326	.0565	.1636	.6988
	med	.3931	.1561	-.2722	5.5210	-3121	.3438	.1764	5.5165
T=30	$q_{0.25}$	.3432	.1157	-.3398	5.0804	-.3993	.3067	.0674	5.0433
	$q_{0.75}$	.4419	.1958	-.2008	5.9998	-.2342	.3839	.2778	6.0157
	mean	.3942	.1862	-.2919	5.7910	-.3042	.3836	.1884	5.7818
n=10	std	.0381	.0304	.0566	.4171	.0713	.0299	.0893	.4049
	med	.3958	.1861	-.2918	5.7610	-.3009	.3844	.1901	5.7850
T=100	$q_{0.25}$	.3705	.1661	-.3306	5.5106	-.3525	.3637	.1327	5.4903
	$q_{0.75}$	.4199	.2068	-.2522	6.0679	-.2558	.4030	.2454	6.0632
	mean	.3976	.1875	-.2931	5.7896	-.3008	.3859	.1940	5.8086
n=30	std	.0252	.0188	.0345	.2380	.0384	.0172	.0508	.2372
	med	.3982	.1874	-.2951	5.7881	-.3019	.3855	.1942	5.8020
T=100	$q_{0.25}$	.3820	.1751	-.3168	5.6316	-.3268	.3749	.1631	5.6449
	$q_{0.75}$	.4143	.2008	-.2712	5.9355	-.2755	.3974	.2251	5.9666

Table 7: Finite Sample Performance for Unknown $U\left[-\sqrt{3}, \sqrt{3}\right]$ Case When $l = 3$

		$\sigma^2 = 4$							
		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$				$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$			
		λ	γ	ρ	σ^2	λ	γ	ρ	σ^2
	mean	.3936	.1567	-.2711	3.8002	-.3011	.3479	.1823	3.8274
n=10	std	.0581	.0549	.0787	.6160	.0849	.0555	.1075	.5956
	med	.3953	.1546	-.2735	3.7561	-.3028	.3500	.1826	3.7999
T=30	$q_{0.25}$	.3571	.1185	-.3251	3.3890	-.3601	.3118	.1098	3.4089
	$q_{0.75}$	.4329	.1937	-.2216	4.2081	-.2435	.3826	.2550	4.1882
	mean	.3964	.1875	-.2932	3.9538	-.3018	.3842	.1947	3.9556
n=10	std	.0332	.0313	.0440	.3643	.0487	.0300	.0593	.3423
	med	.3974	.1883	-.2969	3.9323	-.3020	.3828	.1952	3.9394
T=100	$q_{0.25}$	.3741	.1647	-.3233	3.6912	.3636	.3636	.1561	3.7163
	$q_{0.75}$	.4181	.2090	-.2645	4.1842	.4026	.4026	.2364	4.1643
	mean	.3995	.1881	-.2929	3.9458	-.2993	.3856	.1928	3.9502
n=30	std	.0191	.0183	.0262	.2024	.0283	.0283	.0347	.2023
	med	.3990	.1891	-.2940	3.9448	-.2992	.3867	.1920	3.9525
T=100	$q_{0.25}$	.3867	.1764	-.3102	3.8022	-.3183	.3733	.1709	3.8148
	$q_{0.75}$	.4122	.2000	-.2761	4.0868	-.2803	.3974	.2160	4.0788

Table 8: Finite Sample Performance for Unknown $U\left[-\sqrt{3}, \sqrt{3}\right]$ Case When $l = 6$

		$\sigma^2 = 4$							
		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$				$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$			
		λ	γ	ρ	σ^2	λ	γ	ρ	σ^2
	mean	.3814	.1594	-.2790	3.8574	-.3117	.3484	.1760	3.8180
n=10	std	.0719	.0556	.1034	.6281	.1183	.0524	.1597	.6444
	med	.3899	.1587	-.2808	3.8148	-.3060	.3480	.1773	3.7475
T=30	$q_{0.25}$	.3335	.1205	-.3533	3.4135	-.3874	.3112	.0661	3.3790
	$q_{0.75}$	.4346	.1969	-.2060	4.2552	-.2295	.3841	.2860	4.2158
	mean	.3959	.1878	-.2924	3.9546	-.3038	.3834	.1931	3.9595
n=10	std	.0411	.0313	.0578	.3677	.0737	.0298	.0915	.3587
	med	.3971	.1883	-.2918	3.9139	-.3046	.3843	.1929	3.9338
T=100	$q_{0.25}$	.3698	.1659	-.3312	3.6984	-.3504	.3644	.1381	3.7223
	$q_{0.75}$	.4230	.2086	-.2542	4.1892	-.2538	.4043	.2567	4.1855
	mean	.3978	.1868	-.2939	3.9517	-.3025	.3850	.1935	3.9478
n=30	std	.0283	.0176	.0342	.2062	.0387	.0169	.0543	.2057
	med	.3991	.1868	-.2936	3.9506	-.3024	.3852	.1950	3.9508
T=100	$q_{0.25}$	.3803	.1753	-.3154	3.8131	-.3260	.3737	.1634	3.8078
	$q_{0.75}$	.4156	.1984	-.2721	4.0904	-.2768	.3968	.2302	4.0800

Table 9: Finite Sample Performance for Unknown EV Case When $l = 3$

		$\sigma^2 \approx 4.89$							
		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$				$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$			
		λ	γ	ρ	σ^2	λ	γ	ρ	σ^2
	mean	.3872	.1606	-.2763	4.7104	-.3087	.3459	.1821	4.6739
n=10	std	.0605	.0566	.0811	.7101	.0853	.0564	.1112	.6906
	med	.3907	.1596	-.2783	4.6445	-.3088	.3485	.1833	4.6321
T=30	$q_{0.25}$	.3513	.1217	-.3318	4.2034	-.3658	.3114	.1092	4.1850
	$q_{0.75}$	.4277	.1992	-.2222	5.1490	-.2518	.3836	.2555	5.1090
	mean	.3980	.1880	-.2910	4.8315	-.3018	.3850	.1954	4.8250
n=10	std	.0300	.0317	.0436	.3669	.0464	.0295	.0600	.3790
	med	.3993	.1881	-.2932	4.8117	-.3012	.3854	.1966	4.8067
T=100	$q_{0.25}$	.3786	.1675	-.3223	4.5670	-.3337	.3660	.1531	4.5616
	$q_{0.75}$	.4188	.2084	-.2618	5.0707	-.2711	.4036	.2350	5.0661
	mean	.4000	.1878	-.2922	4.8423	-.2996	.3852	.1925	4.8449
n=30	std	.0198	.0183	.0268	.2120	.0268	.0167	.0381	.2178
	med	.4006	.1876	-.2934	4.8439	-.2992	.3851	.1923	4.8370
T=100	$q_{0.25}$	.3873	.1750	-.3101	4.6938	-.3178	.3743	.1701	4.6940
	$q_{0.75}$	.4136	.2000	-.2740	4.9930	-.2824	.3956	.2175	4.9946

Table 10: Finite Sample Performance for Unknown EV Case When $l = 6$

		$\sigma^2 \approx 4.89$							
		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$				$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$			
		λ	γ	ρ	σ^2	λ	γ	ρ	σ^2
	mean	.3816	.1588	-.2765	4.6974	-.3134	.3463	.1677	4.6793
n=10	std	.0717	.0578	.1087	.6845	.1315	.0551	.1694	.6756
	med	.3845	.1597	-.2740	4.6340	-.3082	.3481	.1720	4.6314
T=30	$q_{0.25}$	.3369	.1181	-.3508	4.2146	-.3946	.3090	.0579	4.1874
	$q_{0.75}$	.4321	.1990	-.2060	5.1251	-.2251	.3846	.2799	5.1076
	mean	.3944	.1876	-.2913	4.8291	-.3080	.3840	.1885	4.8254
n=10	std	.0403	.0313	.0560	.3919	.0716	.0291	.0982	.3744
	med	.3950	.1858	-.2907	4.8117	-.3034	.3844	.1929	4.8080
T=100	$q_{0.25}$	.3676	.1662	-.3265	4.5680	-.3542	.3653	.1255	4.5538
	$q_{0.75}$	.4229	.2076	-.2560	5.0849	-.2590	.4021	.2552	5.0638
	mean	.3978	.1877	-.2944	4.8379	-.3030	.3851	.1919	4.8369
n=30	std	.0245	.0176	.0341	.2082	.0395	.0170	.0577	.2044
	med	.3984	.1872	-.2944	4.8317	-.3032	.3853	.1958	4.8380
T=100	$q_{0.25}$	.3810	.1759	-.3168	4.6916	-.3305	.3740	.1627	4.6806
	$q_{0.75}$	.4154	.1994	-.2725	4.9859	-.2773	.3970	.2266	4.9771

4.5 Performance of Normality Test

In this section, we investigate the finite sample performance of the J_{Norm} statistic for normality which we derived in Section 3.4. To simulate the test size and test power for different situations, we simulate the samples by setting $\varepsilon_{i,t}$ follows the distributions used in the previous sessions: $N(0,1)$, $\sqrt{\frac{1}{3}}t(3)$, $U\left[-\sqrt{3}, \sqrt{3}\right]$ and the extreme value distribution with density function showed by (19). With 1,000 times replications, we can get the empirical probability of rejections by using the asymptotic critical value at $\alpha = 0.05$. The results of the simulated size and power using the asymptotic critical value are showed in Table 11, 12 and 13. Additionally, simulated critical values for $p = 0.1$, $p = 0.05$ and $p = 0.01$ are reported in Table 14 and Table 15.

From Table 11, 14 and 15, we can see that the size and critical values of the J_{Norm} in general are approaching the asymptotic values, but for different adjacent correlations and different parameter settings, the finite sample performance varies a lot. For both of our two parameter settings, the critical values are closer to asymptotic values for $l = 6$ situation, which may indicate our test may have better performance for more densely connected networks than sparse ones. For sparse situation when $l = 3$, i.e. each individual only has three neighbors, using asymptotic critical values will over-reject H_0 by nearly 5% more chance even when we have 200 periods observations. By contrast, for the more dense case $l = 6$, the over-reject rate will be less than 2%.

In general, for finite sample application, the J_{Norm} test statistic is applicable but we need to be cautious. If the network or spatial correlation considered is sparse, it will have a larger chance to make Type I error when sample size is small. In this situation, it may be better to simulate the distribution of J_{Norm} statistic for the sample size you considered, rather than using the asymptotic value of $\chi^2(1)$.

For test power, the results are more interesting. From Table 12 and 13, when the distributions of $\varepsilon_{i,t}$ are $\sqrt{\frac{1}{3}}t(3)$ and $U\left[-\sqrt{3}, \sqrt{3}\right]$, the test power for finite sample is good. For both $n = 10$ and $n = 30$, as T increases from 30 to 200, the test power by using asymptotic critical value converges to 1 fast for both parameter settings. However, for extreme value distribution, the test power is pretty small. In our simulation exercises, even the largest test power is around 10%, which is not of strong power. In fact, for the parameter settings and adjacent matrix we use here, even with $n = 50$ and $T = 1000$, the test power will be less than 30%. Obviously, our modified LM test is powerless in this particular situation.

The reason for a lack of power issue is exactly the limitation we mentioned in Section 3.4. Although the difference between the standard normal distribution and the extreme value distribution we considered in Section 4.4 is quite obvious, after taking square and natural logarithm, their variance and fourth order moments are both very close. By our simulation, when $\varepsilon_{i,n}$ have the density function as in (19), the variance and fourth order central moment are approximately 4.89 and 170.72 which are very close to what we have for $log - \chi^2(1)$. As our test statistic is constructed

Table 11: Test Size of J_{Norm} $(\chi^2_{0.95}(1) = 3.8415)$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$		$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$	
n	T	$l = 3$	$l = 6$	$l = 3$	$l = 6$
	30	.113	.073	.065	.048
10	100	.099	.069	.059	.045
	200	.092	.058	.054	.052
	30	.084	.043	.064	.041
30	100	.092	.065	.061	.052
	200	.077	.061	.057	.051

by only second and fourth order moments, the difference is so small to be distinguished with finite samples. In fact, the distribution of $\ln \varepsilon^2_{i,t}$ after re-centralization at zero is very close for these two distributions, not only for the two particular moments. Although the density function of standard normal and EV are quit different which are showed in Figure 1, after the log-square transformation and re-centralization, the distributions are very similar despite tails.

In Figure 2, we show the simulated CDFs of the two processes which are estimated by kernel density method based on 100000 random samples. The only significant difference between these two distributions is the magnitude of left tail probability where the values of two random variables are both less than negative 3 times the variance. In fact, with smaller sample size, similar to our J_{Norm} statistic, non-parametric Kolmogorov-Smirnov (K-S) test also faces lack of power issue even if it use more information from the sample than our test. We simulated 10000 times for the test power of K-S statistic for three different sample size, and report the results in Table 16. When the sample size is 1000, the test power of K-S is less than 10% when $\alpha = 0.05$, which is very similar to the power of J_{Norm}. Even increasing the sample size to 5000, we still have less than 50% of chance to correctly distinguish the two distributions. Since our QMLE is based on $\ln \varepsilon^2_{i,t}$, fail to reject the normality of $\varepsilon_{i,t}$ will not affect our statistical inference seriously, we don not have to worry the lack of power issue in this situation which we discussed before in Section 3.4.

Table 12: Test Power of J_{Norm} When $l = 3$ $(\chi^2_{0.95}(1) = 3.8415)$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$			$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$		
n	T	$\sqrt{\frac{1}{3}}t(3)$	$U\left[-\sqrt{3}, \sqrt{3}\right]$	EV	$\sqrt{\frac{1}{3}}t(3)$	$U\left[-\sqrt{3}, \sqrt{3}\right]$	EV
	30	.225	.414	.093	.198	.404	.055
10	100	.668	.823	.097	.625	.777	.062
	200	.950	1	.112	.928	.956	.087
	30	.673	.767	.085	.553	.744	.075
30	100	.991	.997	.082	.987	.994	.060
	200	1	1	.094	1	1	.076

Table 13: Test Power of J_{Norm} When $l = 6$ $(\chi^2_{0.95}(1) = 3.8415)$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$			$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$		
n	T	$\sqrt{\frac{1}{3}}t(3)$	$U\left[-\sqrt{3}, \sqrt{3}\right]$	EV	$\sqrt{\frac{1}{3}}t(3)$	$U\left[-\sqrt{3}, \sqrt{3}\right]$	EV
	30	.213	.399	.058	.185	.367	.063
10	100	.645	.800	.064	.605	.754	.062
	200	.926	.959	.058	.911	.959	.049
	30	.588	.747	.064	.540	.723	.055
30	100	.992	.991	.065	.991	.994	.066
	200	1	1	.068	1	1	.065

Table 14: Simulated Critical Values of J_{Norm} When $l = 3$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$			$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$		
n	T	.1	.05	.01	.1	.05	.01
	30	4.2272	6.2467	14.6318	3.1822	4.3007	6.1742
10	100	3.8308	5.4743	9.8324	2.9027	4.0807	6.8274
	200	3.5140	5.3132	8.9933	2.7984	4.0245	6.3566
	30	3.4224	5.3055	9.1624	3.0261	4.2566	6.7994
30	100	3.5865	5.3535	9.5370	3.0883	4.1514	7.5637
	200	3.8205	5.1297	8.3532	3.0586	4.0961	6.6775

Table 15: Simulated Critical Values of J_{Norm} When $l = 6$

		$(\lambda_1, \gamma_1, \rho_1) = (.4, .2, -.3)$			$(\lambda_2, \gamma_2, \rho_2) = (-.3, .4, .2)$		
n	T	.1	.05	.01	.1	.05	.01
	30	3.2527	4.6289	8.2464	2.7346	3.7093	6.2319
10	100	3.0831	4.4957	7.5095	2.7555	3.6425	6.7524
	200	3.1494	4.4324	7.1762	2.8433	4.1054	6.6510
	30	2.7133	3.6259	6.8676	2.5756	3.4757	6.3536
30	100	3.0378	4.6709	8.2586	2.8496	3.8673	7.5614
	200	2.9709	4.2578	7.2924	3.0428	3.8904	7.2347

Figure 1: Density Function of $N(0,1)$ and EV
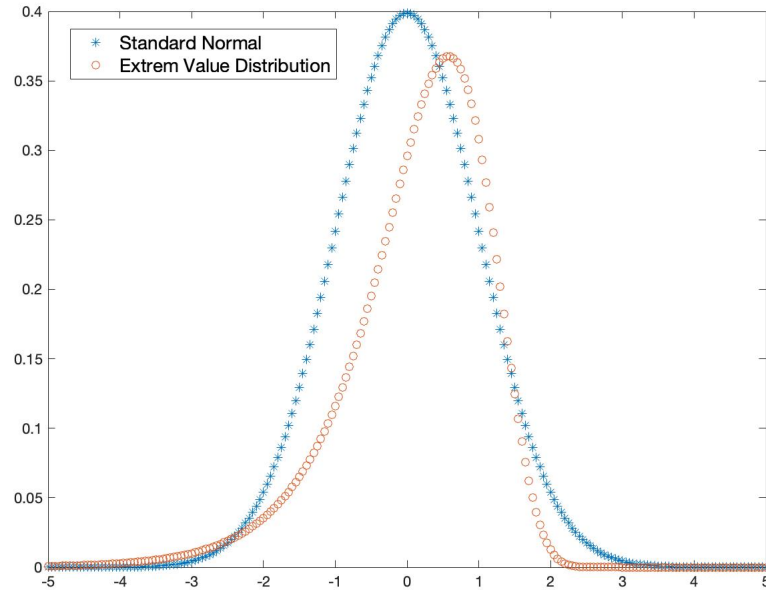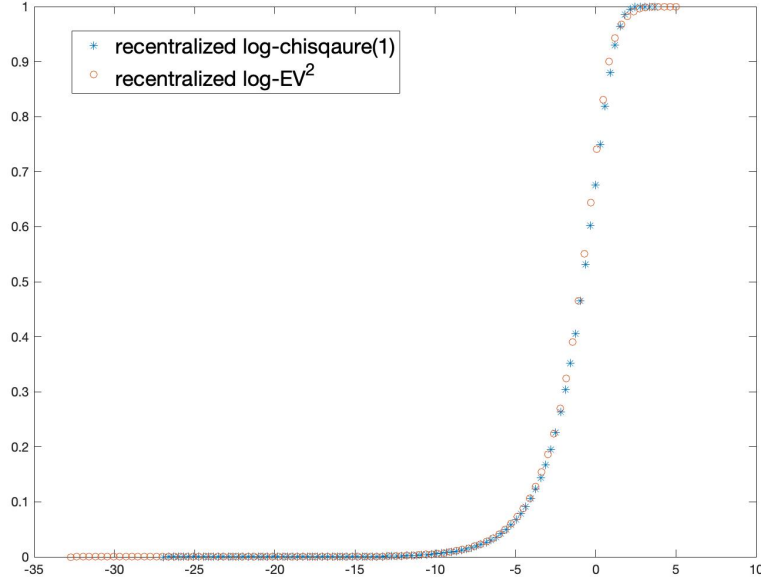

Table 16: Simulated Test Power of Two Sample K-S Test For Re-centralized $\ln \varepsilon_{i,t}^2$

	$\alpha = .1$	$\alpha = .05$	$\alpha = .01$
$n = 1000$	.1394	.0751	.0169
$n = 5000$	.5477	.4196	.2026
$n = 10000$	.9013	.7926	.5566

Figure 2: Comparison of Kernel CDF of 100,000 Random Sample of Re-centralized $\ln \varepsilon_{i,t}^2$

5 Application: Risk Spillover Among Eurozone Countries

5.1 Data Description

When considering the links among economies, the risk spillover from one country to another through geographical links is important and persistent. Historically, due to limitations of transportations, political and economical incidents that happens in one country will only affect its neighboring countries through trades and migrations. Although the economy are globally linked in modern society, most trade and economic integration organizations are still based on regional corporations among nearby countries, such like NAFTA and EU. In the long run, asset returns associated with one country will fluctuate due to risk in neighboring countries. Thus, their conditional volatility of those assets will be spatially correlated. In this section, we are going to investigate the risk spillovers among some eurozone economies by focusing on the stock markets.

Eurozone, officially called the euro area, is a monetary union of 19 member states of the European Union that adopted the euro as their primary currency and sole legal tender. The beginning of the eurozone is January 1, 1999, with 11 original countries: Belgium, Germany, Spain, France, Italy, Netherland, Portugal, Austria, Finland, Luxembourg and Ireland. In this paper, we will focus on the economic links among these 11 countries.

The assets we will investigate are the prices of common stock shares of companies associated with a country traded on national or foreign stock exchanges. OECD provides monthly share price index of each country which is constructed by averaging the prices of common stock shares and

then calculating the relative prices comparing to the average price in 2015[7]. Using this national index portfolio constructed by OECD, we can better capture the whole picture of a country's economic and financial market, especially for the small local companies which are not included in more widely used stock indexes suck like SP500, FTSE100, etc. Since small local companies depend on the domestic market more than large companies in general, portfolios including stocks of these small companies can reflect more domestic risk of economy, it would be better when considering the country-level risk spillovers.

For each country i, we can construct the monthly return rate (%) of this national portfolio return of month t by

$$MR_{i,t} = (P_{i,t} - P_{i,t-1}) / P_{i,t-1} * 100$$

where $P_{i,t}$ is the relative share price of country i in month t comparing to price in 2015. Then, we can further construct the monthly stock return innovation of country i by

$$RI_{i,t} = MR_{i,t} - MR_{i,t-1}$$

If no information updated, at equilibrium, the expectation of aggregate stock return in an economy should be a constant at equilibrium. Thus, in the long-run, the expectation of national stock return innovation should be zero. Table 17 reports summary statistics of the monthly stock return innovations PRI_i we constructed for the 11 original countries in eurozone, from March 1999 to April 2021. Table 18 shows the correlation coefficients between the PRI_is of these 11 countries for the 266 sample months. From Table 17, for all the 11 eurozone countries, the expectation of their monthly stock return innovation in the sample period is close to zero, but very volatile. Comparing the standard deviation and extreme values of monthly return innovations, the financial markets in south European countries, such as Portugal, Italy and Spain, are relatively more volatile than other countries. This indicates that the financial risk may link to geographical factors. In Table 18, it is obvious that the links among financial markets of each eurozone country are strong. For any two countries in our sample, the correlation coefficients among their monthly stock return innovation are larger than 0.4. However, their correlations seem to have no geographical links which conflicts with Table 17. The correlations among three south European countries are not very strong. In contrast, although Finland and Ireland are not adjacent to other eurozone countries, their financial conditions seem to have stronger correlations with others, even stronger than many neighborhood countries.

To further investigate the properties of the stock return innovation processes, in Table 19, we report the results of ADF test of unit root, KPSS test for stationarity (intercept only) and Engle's ARCH test for conditional heteroskedasticity. To including momentum and seasonal effect, we use the 12 periods lags. On one hand, for all the countries, the existence of unit root is rejected. On the other hand, despite for Ireland, Engle's test rejects the homoskedasticity assumption. Thus, conditional volatility models are needed to capture the dynamic of $RI_{i,t}$'s. In Figure 2, we can also see the heteroskedasticity across time directly from the plots of $RI_{i,t}^2$ processes. It seems like

---

26

Table 17: Summary Statistics of Stock Return Innovations (%)

	Mean	S.D.	Minimum	Maximum
Belgium	.0204	6.0182	-23.7896	28.2239
Germany	.0006	6.0780	-26.9865	21.5954
Spain	.0178	7.3668	-31.3490	30.7723
France	.0054	6.0733	-22.8660	23.5420
Italy	.0395	7.3879	-32.5638	28.9739
Netherland	.0295	5.5467	-23.1061	19.4025
Portugal	-.0034	9.0223	-26.9166	51.9512
Austria	.0073	5.4106	-24.6367	27.3145
Finland	.0157	5.6129	-23.7090	23.3261
Luxembourg	.0260	5.7835	-22.2640	27.0218
Ireland	.0181	6.637	-27.5134	31.5424

Table 18: Sample Correlation Coefficients Between Stock Return Innovations

	BEL	DEU	ESP	FRA	ITA	NLD	PRT	AUT	FIN	LUX	IRL
BEL	1										
DEU	.82	1									
ESP	.71	.61	1								
FRA	.78	.67	.55	1							
ITA	.78	.69	.58	.70	1						
NLD	.77	.72	.64	.63	.64	1					
PRT	.66	.58	.43	.57	.52	.49	1				
AUT	.80	.72	.54	.74	.69	.67	.83	1			
FIN	.93	.87	.75	.79	.76	.78	.69	.84	1		
LUX	.92	.80	.71	.77	.79	.75	.73	.87	.94	1	
IRL	.87	.86	.70	.74	.70	.79	.61	.76	.91	.86	1

GARCH type models should be suitable. However, from the sample autocorrelation and partial correlation functions shown in Figure 4 and Figure 5, the order of autoregression of $RI_{i,t}^2$ processes are not clear. For most countries, both ACF and PACF of $RI_{i,t}^2$ cut edge for more than one-month lags and show strong positive correlation with only one month lag, which is not typical for an autoregressive process. For Spain, the ACF and PACF are even more wired, with irregular correlations with lagged terms. The wired shape of ACFs and PACFs can be evidence of risk-spillovers, which can not be explained by any time-series autoregressive terms.

Table 19: ADF, KPSS and Engle's ARCH Test Results (Lag=12)

	BEL	DEU	ESP	FRA	ITA	NLD	RPT	AUT	FIN	LUX	IRL
ADF	-7.52	-7.39	-6.79	-6.88	-7.07	-7.88	-7.98	-7.12	-7.01	-7.35	-7.30
$p-$value	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
KPSS	.0257	.0247	.0224	.0266	.0318	.0321	.0237	.0248	.0272	.0261	.0245
$p-$value[8]	>.10	>.10	>.10	>.10	>.10	>.10	>.10	>.10	>.10	>.10	>.10
Engles's	28.84	26.19	40.78	38.40	39.10	19.94	44.73	46.55	34.21	34.72	21.03
$p-$value	.00	.01	.00	.00	.00	.07	.00	.00	.00	.00	.15

5.2 Empirical Specification and Results

To capture the risk spillovers between the monthly stock return innovations, the major specification we consider is the ESPARCH(1,1) model proposed in Section 2.2:

$$RI_{i,t} = \sigma_{i,t}\varepsilon_{i,t}, \varepsilon_{i,t} \overset{iid}{\sim}(0,1)$$

$$\ln \sigma_{i,t}^2 = \mu_i + \lambda \sum_{j=1}^{n} w_{ij,n} \ln RI_{j,t}^2 + \gamma \ln RI_{i,t-1}^2 + \rho \sum_{j=1}^{n} w_{ij,n} \ln RI_{j,t-1}^2 \tag{20}$$

where i indicates the 11 original countries of eurozone and March 1999 is treated as the original time period ($t = 0$).

For the network between the countries, we consider four different specifications. We denote the network weighting matrices as $W_{adjacent}$, $W_{distinv}$, $W_{dist2inv}$ and W_{EU}. These four specifications satisfy the standard assumption of spatial weighting matrix: the diagonal elements are all zeros. However, the interactions between countries were constructed differently. The first specification, $W_{adjacent}$, is constructed based on the adjacent relationship between countries. If country i and j share a boarder on the land, then $w_{ij,adjacent} = 1$. Otherwise $w_{ij,adjacent}$ will be assigned

---

[8]For KPSS test p-valus, Matlab reports the p-value larger than .10 as .10, not the exact p-value
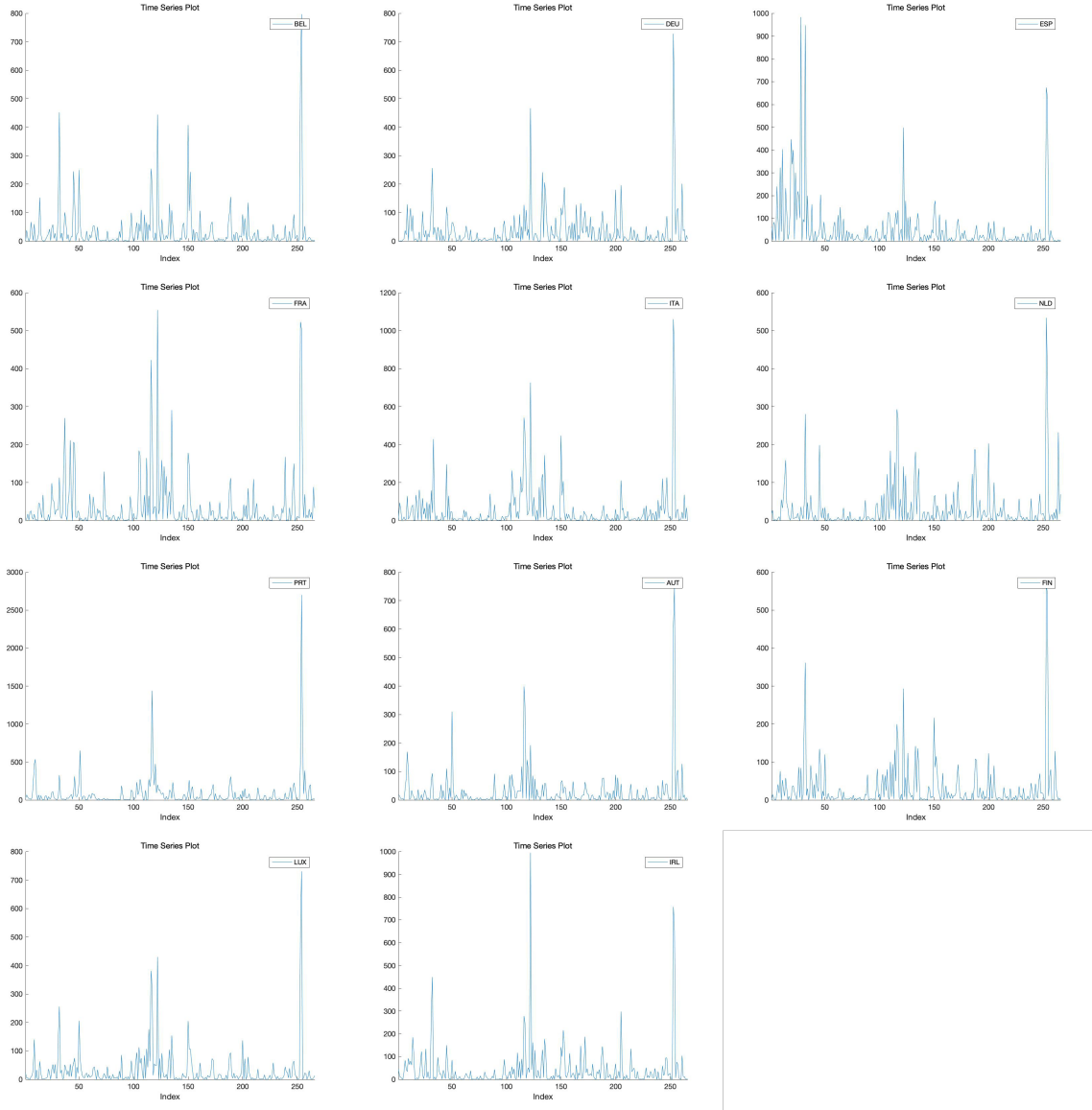
Figure 3: Plot of $RI^2_{i,t}$ Processes

Figure 4: Sample Autocorrelation Function of $RI^2_{i,t}$ Processes
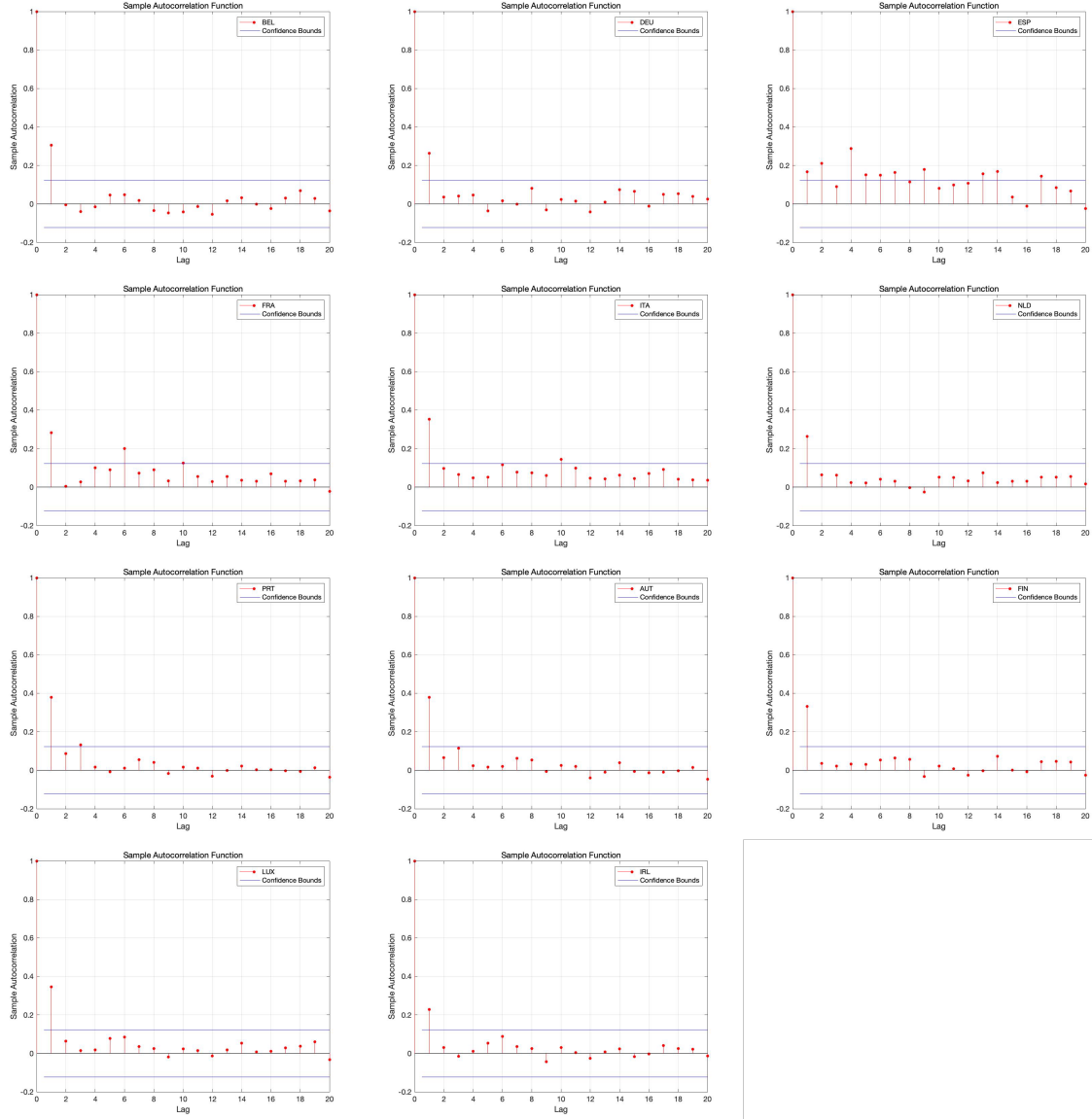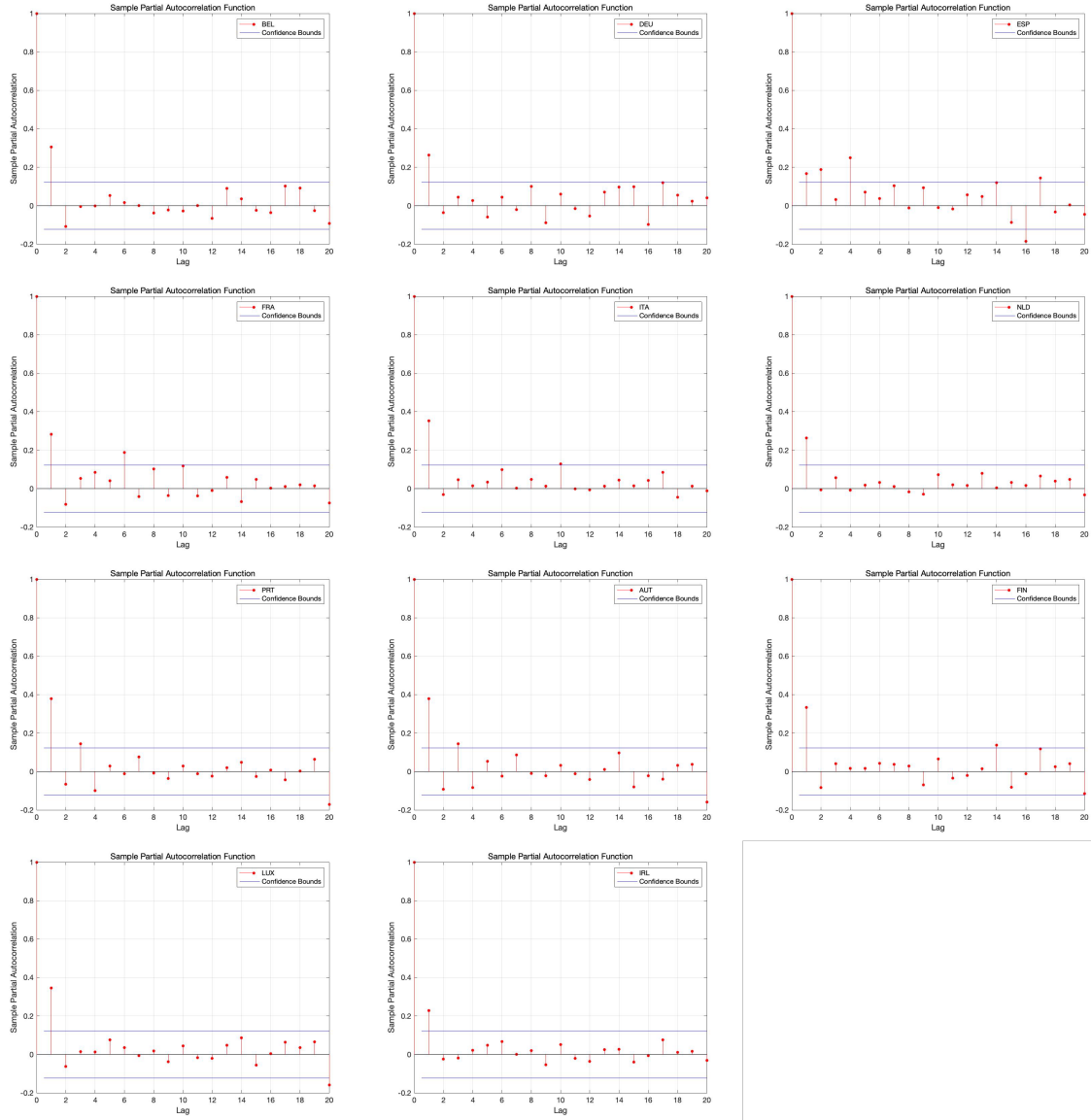
Figure 5: Sample Partial Autocorrelation Function of $RI_{i,t}^2$ Processes

as zero. For Finland and Ireland, as they do not share a boarder with any other original euro-zone countries, all the elements associated with these two countries will be zero. Then, despite the two zero rows, we do row normalization for all non-zero rows, i.e. normalize $w_{ij,adjacent}$ as $w_{ij,adjacent} / \left(\sum_{j=1}^{11} w_{ij,adjacent} \right)$, to make sure the neighborhood effect is normalized. Thus, in these network specifications, risk spillovers are assumed to happen through direct land neighborhood relationships, and isolated countries are assumed not sharing or receiving any risk spillovers.

The second and third specifications, $W_{distinv}$ and $W_{dist2inv}$, are constructed by the distances among the capital cities of each country. Using Google Map, we can measure the great-circle distance between any two points on the earth, which is the shortest distance between two points on the earth's surface[9]. Denote d_{ij} be great-circle distance between the marks of capital cities of country i and j on the Google Maps, each element of $W_{distinv}$ is normalized inverse of the distance, i.e. $w_{ij,distinv} = \frac{1}{d_{ij}} / \left(\sum_{j=1}^{11} \frac{1}{d_{ij}} \right)$. For $W_{dist2inv}$, each element $w_{ij,dist2inv} = \frac{1}{d_{ij}^2} / \left(\sum_{j=1}^{11} \frac{1}{d_{ij}^2} \right)$ which is the normalized inverse of the square of distance. Thus, both of these two specifications use the distance between capitals to approximately measure the geographical distance between countries, and assume the risk spillovers are decreasing with the distance between countries. The difference between these two specifications is how the spillover effect is decreasing. Obviously, $W_{distinv}$ puts less weights on the nearby countries and more weights on faraway countries, which assume the spillover decreasing slower than $W_{dist2inv}$ specification.

Finally, the last specification W_{EU} has the simplest form, where $w_{ij,EU} = \frac{1}{10}$ for any $i \neq j$. In this specification, we employ the linear-in-mean type of network formation inside the eurozone, and using equal weights for the link between any two countries. As eurozone countries share the same central bank and have some common monetary policies, any country inside is affected by other member countries no matter how far away they are. By using these specification, we assume the risk spillover are transmitted through institutional links instead of geographical links although the eurozone itself is based on the geographical concept of Europe.

In Table 20, we implement the normality test constructed in Section 3. 4 for each specification and report the test results. For $W_{adjecent}$, as normality is not rejected, we implement the QMLE under the normality assumption in Section 3.1. For the other three specifications, the normality is rejected, so we implement the QMLE with unknown distribution in Section 3.3. The estimation results are reported in Table 21. The value of quasi log-likelihood function and the AIC value are also shown as well as asymptotic standard deviation of each parameter. We also report McFadden's pseudo-R^2 as a comparision between the intercept only model. Since our QMLE approach is based on linearization, we can also compare models on another dimension, which is Efron's pseudo R^2 for fitness of $\ln y_{i,t}^2$ processes. As Efron's pseudo-R^2 is calculated by the same way as regular R^2 for linear regression, i.e. sum of square of estimated residual divided by sum of square of independent variables, we can use it to compare the models on how much variance of $\ln y_{i,t}^2$ can be explained by the models. It is a convenient measure and only works for linearized models.

For all the four different specifications, λ and ρ are significant both economically and statistically, which indicates strong spillover effects across the stock returns of the 11 eurozone economies. By

---

[9]Here we use an online website using Google Map API which provides the measure of great-circle distance between any two cities on the earth: https://www.distancefromto.net

introducing risk spillovers, all of our four specifications fit the data much better then intercept only model which is showed by the relatively large value of McFadden's pseudo R^2. As the magnitude of λ is much larger than ρ, intra-temporal interactions among the markets affect the stock return innovations much more than dynamic spillovers from the history. Also, we should notice that both λ and ρ are much larger than γwhich captures the dynamic effect from own history. The larger volatility elasticity to the historical volatility of neighboring countries, may indicate that investors care more about what happened in neighboring economists than domestic, which is quite an interesting observation. A potential reason for the larger intra-temporal elasticity to neighbors would be that we focus on monthly stock returns, the strong intra-temporal spillover effects contain the interactions during the trading time and absorb dynamic spillovers at daily and even shorter frequency. Due to current Internet technology, serious incidents that happen in a country will be known globally within a few minutes or hours, so it is really hard to separate the dynamic and instantaneous reactions unless focusing on high frequency trading data. In general, the results showed in Table 20 indicate strong spillover patterns in conditional volatility.

Another interesting question is how does the risk spread from one country to the whole eurozone. Clearly, the quasi-log-likelihood function and Efron's pseudo R^2 of $W_{adjacent}$ are much smaller than the other three specifications. As $W_{adjacent}$ only captures the neighborhood relationship of countries sharing land borders, it is the most narrow channel of risk spillovers among the four specifications. In real world, especially for the financial markets and modern economic interactions, it is clearly too narrow thus not a good enough approximation to capture the networks of risk spillovers. Especially, in this specification, Finland and Ireland are treated as isolated countries. However from Table 18, we can clearly see that their stock return innovations are closely correlated with other eurozone countries. For the other three specifications, we can see the results of $W_{distinv}$ and W_{EU} are pretty close, and their quasi-log-likelihood function are larger than $W_{dist2inv}$. It indicates that the channel of risk spillover inside the eurozone is relevant with geographical links, but not decreasing with the distance among countries too much. When putting more weights on faraway countries, the estimated intra-temporal risk spillover λ is getting larger, from zero weights in $W_{adjacent}$ to all equal weights in W_{EU}. The results also indicate that the institutional links among eurozone countries seem to play a more important role than geographical distance during risk spreading across countries, as W_{EU} gives us the best model with respect to both AIC criteria and Efron's pseudo R^2. However, it is only marginally better than $W_{distinv}$ for the sample period we consider. Thus, without additional information, we can not tell whether distance is important or not in the risk-spreading network. Without additional information and data, we can only say these two specifications both capture the long-term persistent links at similar level.

Table 20: J_{Norm} Test Results for Different Spatial Correlations

	J_{Norm} Statistic	p-Value
$W_{adjacent}$	.1581	.6909
$W_{distinv}$	17.7100	.0000
$W_{dist2inv}$	10.6971	.0011
W_{EU}	20.7423	.0000

Table 21: Major Specification for Different Spatial Correlations

	$W_{adjacent}$	$W_{distinv}$	$W_{dist2inv}$	W_{EU}
σ^2		4.1053***	4.2818***	4.0382***
		(.0351)	(.0375)	(.0346)
λ	.3005***	.5455***	.4634***	.5752***
	(.0353)	(.0399)	(.0457)	(.0353)
γ	.0808***	.0479***	.0565***	.0456***
	(.0198)	(.0185)	(.0194)	(.0175)
ρ	.1219***	.0925**	.0942**	.0750*
	(.0464)	(.0542)	(.0586)	(.0495)
quasi-LogLike	-6467.5	-6272.5	-6330.5	-6249.2
AIC	12963.0	12575.0	12691.0	12528.4
McFadden R^2	.3266	.3469	.3409	.3493
Efron R^2	.1357	.2659	.2343	.2777

5.3 Comparison With Traditional Conditional Heteroskedasticity Models

To evaluate the performance of our model, we also need comparison with existing conditional heteroskedasticity models. As benchmarks, we try to compare our model with two most popular single-variate conditional heteroskedasticity models: GARCH(1,1) purposed in Bollerslev (1986) and EGARCH(1,1) in Nelson (1991). For the monthly stock return innovation of each country i, assume $RI_{i,t} = \sigma_{i,t}\varepsilon_{i,t}$ where $\varepsilon_{i,t} \overset{i.i.d}{\sim} N(0,1)$. If it follows GARCH(1,1) process, we should have

$$\sigma_{i,t}^2 = \alpha_i + \beta_i \sigma_{i,t-1}^2 + \gamma_i RI_{i,t-1}^2$$

If it follows EGARCH(1,1), we should have

$$\ln \sigma_{i,t}^2 = \kappa_i + \theta_i \ln \sigma_{i,t-1}^2 + \xi_i \varepsilon_{i,t-1}^2 + \iota_i \left(|\varepsilon_{i,t}| - E|\varepsilon_{i,t}| \right)$$

By estimating each country's monthly stock return innovation process in our sample, the performance of fitting GARCH(1,1) and EGARCH(1,1) are reported in Table 21 and Table 22. Both the models are estimated by maximum likelihood approach.

The results in Table 22 and 23 show us the drawbacks when we ignore the risk spillovers through networks and only focusing on individual assets. Although for each country, the GARCH(1,1) and EGARCH(1,1) can capture some dynamic of conditional volatilities, combining all 11 countries as a whole, the results are not as good as our Log-SPARCH(1,1) model. Comparing to the results in Table 21, the sum of log-likelihood is much smaller than all the specifications in Section 5.2. Even comparing to $W_{adjecent}$, the worst network approximation, the sum of log-likelihood is around 40% less. The performance of EGARCH(1,1) is slightly better than GARCH, perhaps due the flexibility of asymmetric risks and leverage effect. However, it is still much worse than results of ESPARCH(1,1). Considering that they have more parameters combined, the performance is even relatively worse shown by the sum of AIC of each individual model.

After comparing with single variate conditional volatility models, we can also compare the performance of our model with other multivariate conditional volatility models. To capture the stable long-run relationships between different assets' volatility, one popular model is multivariate GARCH model with constant correlation (CCC) developed in Bollerslev (1990). For $y_t = (y_{1,t}, \cdots, y_{n,t})'$, the DGP of CCC model is

$$E(y_t|\mathscr{F}_{t-1}) = 0$$

$$Var(y_t|\mathscr{F}_{t-1}) = \Omega_t$$

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \sigma_{i,t-1}^2 + \beta_i y_{i,t-1}^2$$

$$\sigma_{ij,t} = \rho_{ij}\sigma_{i,t}\sigma_{j,t}$$

where $\Gamma = \{\rho_{ij}\}$ is the correlation matrix of y_t, $\sigma_{i,t}$ and $\sigma_{ij,t}$ are the variance and covariance terms of Ω_t. In Table 24, we report the maximum likelihood estimation results of the constant correlation multi-variate GARCH model.

Another popular multivariate GARCH model is dynamic conditional correlation specification

(DCC) suggested in Engle (2002) and then further simplified by Aielli (2013). It is a more flexible model allowing the covariance terms be changing over time. The conditional covariance matrix of y_t is assumed as $Var\left(y_t|\mathscr{F}_{t-1}\right) = D_t^{1/2} R_t D_t^{1/2}$, where $R_t \equiv [\rho_{ij,t}]$ is the conditional correlation matrix and $D_t \equiv diag\left(h_{1,t}, \cdots, h_{n,t}\right)$ is a diagonal matrix with the asset conditional variances as diagonal elements. By construction, R_t is the conditional covariance matrix of the standardized return innovations, $\epsilon_t \equiv [\epsilon_{1,t}, \cdots, \epsilon_{n,t}]'$, where $\epsilon_{i,t} \equiv y_{i,t}/\sqrt{h_{i,t}}$. In DCC model, the diagonal elements of D_t are modeled as univariate GARCH models:

$$h_{i,t} = \omega_i + \alpha_i h_{i,t-1} + \beta_i y_{i,t-1}^2$$

which is similar to CCC model. The conditional correlation matrix is then modeled as a function of the past standardized $RI_{i,t}$, which is

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}$$

where

$$Q_t = (1 - \lambda_1 - \lambda_2) S + \lambda_1 \epsilon_{t-1}' \epsilon_{t-1} + \lambda_2 Q_{t-1}$$

and $Q_t \equiv [q_{ij,t}]_{n\times n}$, $Q_t^* \equiv diag\left(q_{11,t}, \cdots, q_{nn,t}\right)$, $S \equiv [s_{ij}]_{n\times n}$, and λ_1 and λ_2 are nonnegative scalars which satisfy $\lambda_1 + \lambda_2 < 1$. In Table 25, we report the maximum likelihood estimation results of the dynamic conditional correlation multi-variate GARCH model. For convenience, we will not report $\hat{S}$ since it contains too many terms.

Comparing the results in Table 21 and Table 24, our ESPARCH(1,1) model also outperforms the multivariate GARCH model with constant correlations when explaining the conditional volatility of monthly stock return innovations of the 11 eurozone countries. With fewer parameters, the estimated log-likelihood function is much larger than multivariate GARCH, even for the worst $W_{adjacent}$. The reason is simple. Although multivariate GARCH model is more flexible with allowing different individual level autoregressive correlations and the correlations among different assets, the interactions of conditional variance terms are totally ignored. Comparing to results in Table 25, the conclusion is similar. With a good approximation of stable networks among the countries, both intra- and inter-temporal correlations can be captured by ESPARCH(1,1) specification. Thus, from our example, when the long-run network structure among markets is clear and easy to be described by some proper approximations, our ESPARCH(1,1) model is a better choice to capture the dynamic structure of conditional volatility than existing models without considering network correlations, even if our model does not directly consider correlations among the returns and is less flexible on the autoregressive effects.

Table 22: GARCH (1,1) Estimation Results For Each Country

	α_i	β_i	γ_i	LogLike	AIC
BEL	25.3197***	.0000	.2908***	-840.8	1687.6
	(5.7665)	(.1874)	(.0935)		
DEU	19.7952**	.3082	.1468*	-851.2	1708.4
	(9.8641)	(.3172)	(.0834)		
ESP	4.3624**	.7305***	.1850**	-878.7	1763.4
	(1.9513)	(.0916)	(.0729)		
FRA	2.7939**	.7684***	.1668***	-839.7	1685.4
	(1.3259)	(.0625)	(.0558)		
ITA	1.8181**	.8336***	.1431***	-885.0	1775.9
	(.8809)	(.0296)	(.0360)		
NLD	2.2968**	.8009***	.1366***	-821.7	1649.4
	(1.0861)	(.0562)	(.0452)		
PRT	26.4918***	.0369	.6853***	-916.6	1839.1
	(4.2140)	(.0761)	(.1365)		
AUT	7.2264***	.5003***	.2550***	-806.2	1618.3
	(2.1463)	(.1191)	(.0874)		
FIN	2.9993**	.7306***	.1889***	-822.2	1650.3
	(1.4738)	(.0731)	(.0636)		
LUX	11.8708***	.2592*	.4108***	-822.4	1650.7
	(2.748)	(.1511)	(.1320)		
IRL	2.3281	.8089***	.1538***	-869.3	1744.6
	(1.5525)	(.0645)	(.0506)		

Sum of Log-Likelihood: -9353.6

Sum of AIC: 18773.1

Table 23: EGARCH (1,1) Estimation Results For Each Country

	κ_i	θ_i	ξ_i	ι_i	LogLike	AIC
BEL	.5279**	.8434***	.3247***	-.2848***	-833.0	1674.0
	(.2225)	(.0631)	(.0990)	(.0783)		
DEU	.2719	.9221***	.0953	-.2037**	-850.1	1708.1
	(.2927)	(.0806)	(.0991)	(.0933)		
ESP	.2372**	.9381***	.2902***	-.0926	-877.1	1762.1
	(.1070)	(.0285)	(.0917)	(.0666)		
FRA	.1232	.9571***	.1128	-.3317***	-831.7	1674.3
	(.1012)	(.0276)	(.0752)	(.0863)		
ITA	.1788**	.9544***	.2198***	-.1228	-881.5	1770.9
	(.0731)	(.0189)	(.0610)	(.0791)		
NLD	.3060**	.9104***	.2434***	-.1735**	-817.6	1643.1
	(.1237)	(.0376)	(.0776)	(.0799)		
PRT	2.1029***	.4793***	1.0000***	-.9783	-914.0	1835.9
	(.4145)	(.1036)	(.1624)	(.1010)		
AUT	.2377	.9193***	.2316***	-.3082***	-798.4	1604.7
	(.1576)	(.0490)	(.0857)	(.0808)		
FIN	.1619	.9413***	.1513	-.3627***	-812.6	1633.2
	(.1281)	(.0380)	(.1009)	(.0984)		
LUX	.5101*	.8400***	.2834***	-.2872***	-817.0	1641.9
	(.2949)	(.0880)	(.1046)	(.1044)		
IRL	.1514	.9546***	.1943**	-.2791***	-859.6	1727.1
	(.1222)	(.0327)	(.0919)	(.0984)		

Sum of Log-Likelihood: -9293.7

Sum of AIC: 18675.3

Table 24: Multivariate GARCH with CCC Estimation Results

	ω_i	α_i	β_i
BEL	1.8419*	.8590***	.0837***
	(1.0211)	(.0471)	(..0213)
DEU	1.9730**	.8648***	.0773***
	(.8499)	(.0386)	(.0227)
ESP	1.4508*	.8648***	.0989***
	(.7677)	(.0416)	(.0316)
FRA	3.2654**	.7889***	.1077***
	(1.3643)	(.0650)	(.0320)
ITA	3.1068**	.8282***	.1014***
	(1.2426)	(.0425)	(.0292)
NLD	1.6480**	.8597***	.0810***
	(.6819)	(.0353)	(.0235)
PRT	18.198***	.1977*	.5970***
	(5.4876)	(.1109)	(.0940)
AUT	2.8688***	.7700***	.1123***
	(.9141)	(.0443)	(.0249)
FIN	1.6850***	.8528***	.0818***
	(.5045)	(.0277)	(.0169)
LUX	2.8512***	.7822***	.1118***
	(.9429)	(.0481)	(.0245)
IRL	2.4934**	.8290***	.1051***
	(1.0300)	(.0394)	(.0260)

Log-Likelihood: -7417.0

AIC: 14900.0

Table 25: Multivariate GARCH with DCC Estimation Results

	ω_i	α_i	β_i	λ_1	λ_2
				.1048***	.6851***
				(.0148)	(.0420)
BEL	7.0076***	.6799***	.2305***		
	(2.5125)	(.0731)	(.0387)		
DEU	2.6344***	.8319***	.1490***		
	(1.003)	(.0357)	(.0299)		
ESP	2.9656**	.8120***	.1630***		
	(1.3026)	(.0462)	(.0386)		
FRA	4.9371***	.7154***	.2145***		
	(1.6703)	(.0633)	(.0510)		
ITA	4.8557***	.7679***	.1986***		
	(1.6906)	(.0449)	(.0414)		
NLD	2.3643**	.8320***	.1199***		
	(.9429)	(.0389)	(.0265)		
PRT	16.0833***	.2639***	.7869***		
	(4.4069)	(.0706)	(.1080)		
AUT	3.9910***	.7311***	.1920***		
	(1.1043)	(.0397)	(.0277)		
FIN	3.4676***	.7843***	.1671***		
	(.8504)	(.0332)	(.0242)		
LUX	5.3657***	.6950***	.2202***		
	(1.2887)	(.0467)	(.0311)		
IRL	4.2313***	.7766***	.1784***		
	(1.3581)	(.0371)	(.0299)		

Log-Likelihood: -7304.7

AIC: 14921.4

6   Conclusion

With introducing spatial autoregressive terms into exponential ARCH type of models, we give a new model specification, ESPARCH(1,1), to capture the volatility spillover through network. By this extension, the new model can capture both intra-temporal interaction among markets and dynamic spillover from history. Due to the log-volatility structure, it is easier to discuss the stationarity and develop likelihood based estimation method. Since it can be transformed into a linear dynamic spatial panel, the QMLE has good asymptotic properties. In finite sample, simulation results show that the QMLE works well when the sample period is larger than number of individuals. In addition, to test normality of residuals, we develop a modified LM test and simulate its performance. When we apply it to monthly stock return innovations of euro-zone countries, we identify strong interactions among capital markets of 11 original euro-zone countries, through geographical and institutional links. Also, our model has better performance than existing ARCH/GARCH type of models.

References

[1] Alexander J. McNeil, Rüdiger Frey, "Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach", *Journal of Empirical Finance*, Volume 7, Issues 3–4, November 2000, 271-300

[2] Andrew Harvey, Esther Ruiz, Neil Shephard, "Multivariate Stochastic Variance Models", *The Review of Economic Studies*, Volume 61, Issue 2, April 1994, 247-264

[3] Avanidhar Subrahmanyam, "Circuit Breakers and Market Volatility: A Theoretical Perspective", *The Journal of Finance*, Volume 49, Issue 1, March 1994, 237-254

[4] Bo Young Chang, Peter Christoffersen, Kris Jacobs, "Market Skewness Risk and The Cross Section of Stock Returns", *Journal of Financial Economics*, Volume 107, Issue 1, January 2013, 46-68

[5] Campbell R. Harvey, Akhtar Siddique, "Conditional Skewness in Asset Pricing Tests", *The Journal of Finance*, Volume 55, Issue 3, June 2000, Pages 1263-1295

[6] Christian Francqac, Olivier Wintenbergerb, Jean-Michel Zakoïan, "GARCH models without positivity constraints: Exponential or log GARCH?", *Journal of Econometrics*, Volume 177, Issue 1, November 2013, 34-46

[7] Daniel B. Nelson, "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, Volume 59, No. 2, March 1991, 347-370

[8] Esther Ruiz, "Quasi-maximum Likelihood Estimation of Stochastic Volatility Models", *Journal of Econometrics*, Volume 63, Issue 1, July 1994, 289-306

[9] Francisco Blasques, Siem Jan Koopman, Andre Lucas, Julia Schaumburg, "Spillover Dynamics of Systemic Risk Measurement Using Spatial Financial Time Series Models", *Journal of Econometrics*, Volume 195, Issue 2, December 2016, 211-223

[10] Gian Piero Aielli, "Dynamic Conditional Correlation: On Properties and Estimation", *Journal of Business & Economic Statistics*, Volume 31, Issue 3, July 2013, 282-299

[11] James LeSage, R. Kelley Pace, "Introduction to Spatial Econometrics", January 2009, CRC Press/Taylor & Francis Group

[12] Jihai Yu, Robert de Jong, Lung-fei Lee, "Quasi-maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects When both n and T are Large", *Journal of Econometrics*, Volume 146, Issue 1, September 2008 , 118-134

[13] Kenneth A. Kim, S. Ghon Rhee, "Price Limit Performance: Evidence from the Tokyo Stock Exchange", *The Journal of Finance*, Volume 52, Issue 2, June 1997, 885-901

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

[15] Robert Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, Volume 50, Issue 4, July 1982, 987-1008

[16] Robert Engle, "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business & Economic Statistics*, Volume 20, Issue 3, January 2002, 339-350

[17] Robert Engle and Kenneth Kroner, "Multivariate Simultaneous Generalized Arch ", *Econometric Theory*, Volume 11, No. 1, March 1995, 122-150

[18] Robert Richmond, "Trade Network Centrality and Currency Risk Premia", *The Journal of Finance*, Volume 74, Issue 3, June 2019, 1083-1581

[19] Roger Horn and Charles Johnson, "Matrix Analysis", Second Edition, Cambridge University Press, 2013, ISBN 978-0-521-83940-2

[20] Ser-Huang Poon, Michael Rockinger and Jonathan Tawn, "Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications", *The Review of Financial Studies*, Volume 17, Issue 2, April 2004, 581–610

[21] Steven Kou, Xianhua Peng, Haowen Zhong, "Asset Pricing with Spatial Interactions", *Management Science*, Volume 64, No. 5, February 2017, 2083-2101

[22] Tim Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics,* Volume 31, Issue 3, April 1986, 307-327

[23] Tim Bollerslev, "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model", *The Review of Economics and Statistics*, Volume 72, No. 3, August 1990, 498-505

[24] Yongmiao Hong, "A Test for Volatility Spillover with Application to Exchange Rates", *Journal of Econometrics*, Volume 103, Issues 1-2, July 2001, 183-224

[25] Yongmiao Hong, Yanhui Liu, Shouyang Wang, "Ganger Causality in Risk and Detection of Extreme Risk Spillovers between Financial Markets", *Journal of Econometrics*, Volume 150, Issue 2, June 2009, 271-287