<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-2018-0234</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
</tbody>
</table>
| **Complete List of Authors** | Jing Zhou  
Dong Li  
Rui Pan and  
Hansheng Wang |
| **Corresponding Author** | Dong Li |
| **E-mail** | malidong@mail.tsinghua.edu.cn |

Notice: Accepted version subject to English editing.
NETWORK GARCH MODEL

Jing Zhou\textsuperscript{1}, Dong Li\textsuperscript{2}, Rui Pan\textsuperscript{3}, and Hansheng Wang\textsuperscript{4}

\textsuperscript{1}Renmin University of China, \textsuperscript{2}Tsinghua University,
\textsuperscript{3}Central University of Finance and Economics, and \textsuperscript{4}Peking University

Abstract:

The multivariate GARCH (MGARCH) model is popularly used for analyzing financial time series data. However, statistical inference for MGARCH models is quite challenging due to the high dimension issue. To overcome this difficulty, we propose a network GARCH model. The newly proposed model makes use of information derived from an appropriately defined network structure. By doing so, the number of unknown parameters is highly decreased, and the computational complexity is substantially reduced. Strict and weak stationarity of the network GARCH model is rigorously established. In order to estimate the model, a quasi-maximum likelihood estimator (QMLE) is developed, and its asymptotic properties are investigated. Simulation studies are carried out to assess the performance of the QMLE in finite samples and empirical examples are analyzed to illustrate the usefulness of network GARCH models.

\textit{Key words and phrases:} GARCH Model, Multivariate GARCH Model, Network Structure, Quasi-Maximum Likelihood Estimator.
1. Introduction

In the past few decades, financial time series data are becoming increasingly available. A particular focus of this kind of data is on modeling the conditional variance. This issue has given rise to a number of statistical models, among which the ARCH model (Engle, 1982) and the GARCH model (Bollerslev, 1986) have been popularly used. These models have been proved to be important and powerful tools (Tsay, 2003; Fan and Yao, 2017), followed by a large amount of extensions. These extensions include but are not limited to the EGARCH, fGARCH, GARCH-M, GJR-GARCH, IGARCH, NGARCH, QGARCH, TGARCH, and ARFIMA-GARCH. For GARCH models, related literature can be found in Lee and Hansen (1994), Lumsdaine (1996), Hall and Yao (2003), Polonik and Yao (2008), Zhu and Ling (2011) and Conrad and Mammen (2016), etc.

The GARCH models mentioned above are mainly developed for univariate time series data. In the portfolio optimization and risk management of real capital market, cross-sections of hundreds of different stocks are common. This leads to the development of multivariate GARCH (MGARCH) models (Bollerslev and Wooldridge, 1988; Bollerslev, 1990; Engle and Kroner, 1995; Tse and Tsui, 2002), which study dynamic relations among several stocks simultaneously. On one hand, the MGARCH model considers one
Network GARCH Model

particular stock’s historical returns and its conditional variance. On the other hand, it takes other stocks’ information (i.e., correlation or covariance) into account. Conceptually, it is easy to extend univariate GARCH model to the multivariate case. However, statistical inference for MGARCH models is quite challenging. In the MGARCH models, the parameters of interest are all involved in the inverse conditional covariance matrix during the estimation procedure. A particular issue is that the number of parameters increases rapidly with the number of stocks. This makes the resulting estimates highly unstable. Furthermore, the positive definiteness of conditional covariance matrix has to be ensured. To make the model more applicable in practice, additional structures on the covariance matrix should be imposed.

As one can see, the main challenge in MGARCH modeling is to impose a realistic but parsimonious specification to make the covariance matrix positive definite. For instance, many efforts have been done by either imposing certain structures on the conditional covariance matrix (Bollerslev, 1990; Engle and Kroner, 1995; Tse and Tsui, 2002) or applying dimension reduction techniques (Pan et al., 2010; Lam and Yao, 2012; Li et al., 2016). In the first category, the CCC-GARCH models (Bollerslev, 1990; Tse, 2000) and DCC-GARCH models (Tse and Tsui, 2002) are two popular examples.
In the second category, factor modeling has been extensively used by introducing both observed factors (Engle et al., 1990; Bollerslev and Engle, 1993; Tao et al., 2012) and unobserved factors (Pan et al., 2010; Hu and Tsay, 2013; Li et al., 2016). In this article, we develop a new method for MGARCH model, which is illuminated by the emergence of network structure data. We refer to the new model as network GARCH model.

The perspective of network analysis has been successfully incorporated in many fields such as sociology, marketing, organization behavior, and so forth. Empirical findings show that model performance could be significantly improved by incorporating network structure information (Goel and Goldstein, 2014; Nitzan and Libai, 2011; Wei et al., 2014). In real stock market, investors can select different stocks to diversify investment risk. However, those stocks that are heavily held by investors always indicate unique information (Livingston, 2010; Pareek, 2012; Chou and Lee, 2012; Bajo et al., 2017). It is often found that if two stocks have a common shareholder, their financial performance (i.e., stock returns) will be highly correlated. Therefore, the common stock shareholdings can be used to construct an information network. Furthermore, the network structure will be reflected on the conditional covariance matrix. This will lead to a highly parsimonious model and reduction of parameter dimension.
The newly proposed model is distinct from the existing ones in the following perspectives. First of all, different from traditional GARCH model, the network GARCH model makes use of the information from other stocks, by defining an appropriate network structure. Secondly, the network GARCH model has the advantages in processing large number of stocks in terms of volatility forecasting. The number of unknown parameters in the network GARCH model is substantially decreased because of the incorporated network structure. The computational complexity drops from $O(N^2)$ to $O(N)$. Finally, the strictly stationary solution for the new model can be rigorously established under certain conditions. Furthermore, in order to estimate network GARCH model, a quasi-maximum likelihood estimator (QMLE) is proposed and its asymptotics are also investigated.

The rest of the article is organized as follows. Section 2 introduces the network GARCH model. Its strict stationarity solution and asymptotic properties are presented. To demonstrate the finite sample performance of the network GARCH model, extensive numerical studies of both simulated and real datasets are given in Section 3. Lastly, the article is concluded with a brief discussion in Section 4. All detailed techniques are relegated to the supplementary part.
2. The Network GARCH Model

2.1 Model Setup

Let \( i \) be the stock index in the stock market, where \( 1 \leq i \leq N \) and \( N \) is the number of stocks and fixed in our model. For the stock \( i \), we assume a continuous response variable \( y_{it} \in \mathbb{R}^1 \) to be observed for \( t = 1, 2, \cdots, T \).

For example, \( y_{it} \) could be the return of stock \( i \) at time \( t \) in the stock market. Following the idea of Engle (1982) and Bollerslev (1986), a classical GARCH(1,1) model is defined as

\[
y_{it} = \varepsilon_{it} \sqrt{h_{it}}, \quad h_{it} = \omega_0 + \alpha_0 y_{i,t-1}^2 + \beta_0 h_{i,t-1},
\]

(2.1)

where \( \{\varepsilon_{it}\} \) is a sequence of independent and identically distributed (i.i.d.) random variables with \( E(\varepsilon_{it}) = 0 \) and \( \text{var}(\varepsilon_{it}) = 1 \). \( \omega_0 > 0, \alpha_0 \geq 0 \) and \( \beta_0 \geq 0 \) are unknown parameters and are interpreted as volatility parameters. \( h_{it} \) is the conditional variance. Bollerslev (1986) showed that model (2.1) defines a second-order stationary solution if and only if \( \alpha_0 + \beta_0 < 1 \). Further, Nelson (1990) proved that there exists a unique strictly stationary solution to model (2.1) if and only if \( E \log(\beta_0 + \alpha_0 \varepsilon_{it}^2) < 0 \).

To take the network structure into consideration, we first define an adjacency matrix \( A = (a_{ij}) \in \mathbb{R}^{N \times N} \), where \( a_{ij} = 1 \) if stock \( i \) is connected
2.1 Model Setup

with stock \( j \) and \( a_{ij} = 0 \) otherwise. Let \( a_{ii} = 0 \) for \( 1 \leq i \leq N \). In our empirical example, we assume that two stocks are connected with each other (i.e., \( a_{ij} = a_{ji} = 1 \)) if they share at least a common shareholder in their top ten shareholders. In order to model this network dependence structure, we incorporate a new term in model (2.1), which leads to the network GARCH model

\[
y_{it} = \varepsilon_{it} \sqrt{h_{it}}, \quad h_{it} = \omega_0 + \alpha_0 y_{i,t-1}^2 + \lambda_0 \sum d_i^{-1} a_{ij} y_{j,t-1}^2 + \beta_0 h_{i,t-1}, \quad (2.2)
\]

where \( \sum_{j \neq i} \) represents \( \sum_{j=1}^{N} \sum_{j \neq i} \), and \( d_i = \sum_{j=1}^{N} a_{ij} \) is the total number of stocks that \( i \) connects to, which is the out-degree. Its associated coefficient is \( \lambda_0 \). If \( d_i = 0 \) for some \( i \), the stock \( i \) does not have any out-degree and is then regarded as being isolate. In this case, we define \( d_i^{-1} \sum_{j \neq i} a_{ij} y_{j,t-1}^2 = 0 \) with convention. This idea is similar to, but not exactly the same as, that in the network VAR model (Zhu et al. [2016]). Although they take network structure information into model specification, the key difference is that network VAR fits the conditional mean, while our model does the conditional variance. Compared with traditional MGARCH model, we implicitly assume that the focal stock \( i \) is only affected by its direct connected neighbors (i.e., \( a_{ij} = 1 \)). This is typically true in practice, because the activities done by
2.2 Strict Stationarity

those js with $a_{ij} = 0$ can not be observed by $i$. So $\lambda_0$ captures an average influence of other stocks on the focal stock $i$. We then interpret $\lambda_0$ as the network effect. The assumption that a stock can only be averagely affected by all its connected neighbors may have some limitations. This probably makes the proposed network GARCH model only applicable in some specific circumstances. For example, the model is more suitable for the stocks which belong to the same industry. In the next subsection, we intend to derive strict stationarity solution to model (2.2).

2.2 Strict Stationarity

This subsection is concerned with the stationary solution to model (2.2). For simplicity, define $y_t = (y_{1t}, ..., y_{Nt})'$, $h_t = (h_{1t}, ..., h_{Nt})'$, $D = \text{diag}(d_1, ..., d_N)$, $\mathcal{E}_t = \text{diag}(\varepsilon_{1t}^2, ..., \varepsilon_{Nt}^2)$ and $B_t = \beta_0 I_N + \alpha_0 \mathcal{E}_t + \lambda_0 D^{-1} A \mathcal{E}_t$, where $I_N$ stands for an $N \times N$ identity matrix. Then model (2.2) can be re-written in a vector form as

$$h_t = \omega_0 1_N + B_{t-1} h_{t-1}, \quad (2.3)$$
2.2 Strict Stationarity

where \(1_N = (1, \cdots, 1)'\) is a vector with compatible dimension. The top Lyapunov exponent associated with \(h_t\) in (2.3) is defined as

\[
\gamma_0 = \inf \left\{ \frac{1}{n} E \log \| B_n B_{n-1} \cdots B_1 \|_*, \quad n \in \mathbb{N} \right\},
\]

where \(\mathbb{N}\) stands for the set of natural numbers and \(\| \cdot \|_*\) defines the operator norm of \(N \times N\) matrices. By Theorem 3.2 in Bougerol and Picard (1992), \(h_t\) converges almost surely (a.s.) and the process \((y_{it})\) in model (2.2) has the unique strictly stationary solution if and only if \(\gamma_0 < 0\). Under this condition, we have the following expression

\[
h_t = \omega_0 1_N + \omega_0 \sum_{j=1}^{\infty} \left\{ \prod_{i=1}^{j} B_{t-i} \right\} 1_N. \tag{2.4}
\]

However, \(\gamma_0\) is closely related to the distribution of \(\varepsilon_{it}\). Although it can be simulated by Monte Carlo method, it is not easy to be exactly calculated in practice. In what follows, we will give a sufficient condition to ensure \(\gamma_0 < 0\), see the following Theorem \(\square\) and its proof is given in Appendix A.

**Theorem 1.** If \(\alpha_0 + \lambda_0 + \beta_0 < 1\), then there exists a unique strictly stationary solution with finite second moments to model (2.2), i.e., \(E \| y_t \|^2 < \infty\), where \(\| \cdot \|\) is the Euclidean norm. Particularly, under Assumption 1 below, the
2.3 Quasi-Maximum Likelihood Estimator

The long-run variance-covariance matrix of $y_t$ is

$$
\Sigma_y = \text{diag}(\omega_0[(1 - \alpha_0 - \beta_0)I_N - \lambda_0D^{-1}A]^{-1}1_N).
$$

**Remark.** Theorem 1 is constructed with a fixed $N$ which will break if $N$ is diverging. This is because with a diverging $N$, the dimension of $y_t$ is increasing. In this case, it seems how to define a meaningful concept of stationary is highly questionable. As a result, if $N$ is diverging, Theorem 1 not only breaks, but also not fixable. However, as one can see from the following real data analysis, we can indeed cope with a dimension much higher than convention.

2.3 Quasi-Maximum Likelihood Estimator

Assume that the observations $(y_1, \ldots, y_T)$ is from model (2.2) with true value $\theta_0 = (\omega_0, \alpha_0, \lambda_0, \beta_0)' \in \mathbb{R}^4$. Here, $T$ is the sample size. Let $\theta = (\omega, \alpha, \lambda, \beta)' \in \mathbb{R}^4$ be the parameter. The quasi-log-likelihood function (ignoring a constant) is given by

$$
\tilde{L}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{\ell}_t(\theta), \quad \tilde{\ell}_t(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \log \tilde{\sigma}_it(\theta) + \frac{y_{it}^2}{\tilde{\sigma}_it(\theta)} \right\},
$$
2.3 Quasi-Maximum Likelihood Estimator

where \( \tilde{\sigma}^2_{it}(\theta) \) are recursively defined for \( t \geq 1 \) by

\[
\tilde{\sigma}^2_{it}(\theta) = \omega + \alpha y_{i,t-1}^2 + \lambda d_i^{-1} \sum_{j \neq i} a_{ij} y_{j,t-1}^2 + \beta \tilde{\sigma}^2_{i,t-1}(\theta)
\]

with \( \tilde{\sigma}^2_{i0}(\theta) \equiv 0 \).

The QMLE is defined as

\[
\hat{\theta} = (\hat{\omega}, \hat{\alpha}, \hat{\lambda}, \hat{\beta})' = \arg \min_{\theta \in \Theta} \tilde{L}(\theta), \tag{2.5}
\]

where \( \Theta \) is the parameter space.

To discuss asymptotic properties of \( \hat{\theta} \), it will be convenient to approximate the sequence \( \{\tilde{\sigma}^2_{it}(\theta)\} \) by an ergodic stationary sequence \( \{\sigma^2_{it}(\theta)\} \), which is defined as

\[
\sigma^2_{it}(\theta) = \omega + \alpha y_{i,t-1}^2 + \lambda d_i^{-1} \sum_{j \neq i} a_{ij} y_{j,t-1}^2 + \beta \sigma^2_{i,t-1}(\theta), \quad \text{for any } t \text{ and each } i. \tag{2.6}
\]

and \( \sigma^2_{i0}(\theta_0) = h_{it} \). Similar to the definitions of \( \tilde{L}(\theta) \) and \( \tilde{\ell}_t(\theta) \), we can define

\[
L(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ell_t(\theta), \quad \ell_t(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \log \sigma^2_{it}(\theta) + \frac{y_{it}^2}{\sigma^2_{it}(\theta)} \right\}. \tag{2.7}
\]
Before stating our main results, we first give two assumptions, which are standard in studying GARCH-type models.

**Assumption 1.** \( \{ \varepsilon_{it} \} \) is i.i.d. across \( i \) and \( t \) with zero mean and unit variance. Further assume \( \varepsilon_{it}^2 \) is non-degenerate.

**Assumption 2.** The parameter space \( \Theta \) is a compact subset of \( \{ \theta : \omega > 0, \alpha > 0, \lambda > 0, \beta > 0, \alpha + \lambda + \beta < 1 \} \) and \( \theta_0 \in \Theta \).

The following theorem states the strong consistency and asymptotic normality of QMLE \( \hat{\theta} \).

**Theorem 2.** If Assumptions 1-2 hold, then \( \hat{\theta} \xrightarrow{a.s.} \theta_0 \) as \( T \to \infty \). Further, if \( \kappa_4 = E \varepsilon_{it}^4 < \infty \) and \( \theta_0 \) is an interior point of \( \Theta \), then as \( T \to \infty \)

\[
\sqrt{NT}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, (\kappa_4 - 1)\Sigma^{-1}),
\]

where

\[
\Sigma = \frac{1}{N} \sum_{i=1}^{N} E \left( \frac{1}{\hat{h}^2_{it}} \frac{\partial \sigma^2_{it}(\theta_0)}{\partial \theta} \frac{\partial \sigma^2_{it}(\theta_0)}{\partial \theta'} \right),
\]

\[
\frac{\partial \sigma^2_{it}(\theta_0)}{\partial \theta} = (1, y_{i,t-1}^2, a_{i}^{-1} \sum_{j \neq i} a_{ij} y_{j,t-1}^2, \sigma^2_{i,t-1})^T + \beta_0 \frac{\partial \sigma^2_{i,t-1}(\theta_0)}{\partial \theta}.
\]
2.3 Quasi-Maximum Likelihood Estimator

The adjacency matrix is responsible for the calculation of parameter derivative. Given an initial guess of \( \theta_0 \), we can repeat the above iteration process until we get convergent \( \theta \). To make statistical inference for \( \theta_0 \), we need to estimate \( \kappa_4 \) and \( \Sigma \). In practice, \( \kappa_4 \) can be consistently estimated as

\[
\hat{\kappa}_4 = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^4, \quad \text{where} \quad \hat{\varepsilon}_{it} = (\hat{h}_{it})^{-1/2} y_{it}.
\]

A consistent estimator of \( \Sigma \) is its sample counterpart, i.e.,

\[
\hat{\Sigma} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \frac{1}{\hat{\sigma}_{it}^2(\hat{\theta})} \frac{\partial \hat{\sigma}_{it}^2(\hat{\theta})}{\partial \theta} \frac{\partial \hat{\sigma}_{it}^2(\hat{\theta})}{\partial \theta'} \right).
\]

Remark. The theories of multivariate GARCH models are well studied in the past. So our work provided little contribution to the parameter consistency theory of a “general” MGARCH model. Our contribution is to propose one very special form of MGARCH models. This model takes an observed network structure into consideration. As a consequence, the number of unknown parameters is substantially reduced. This enables us to cope with a dimension much higher than convention. However, we do pay a price for this benefit, that is, our model can not be used to fit a general multivariate time series if no meaningful network structure is defined.
3. Numerical Studies

3.1 Simulating Data

To demonstrate the finite sample performance of the proposed model, we present in this subsection three simulation examples. These three examples are similar to some extent. The only difference is the generating mechanism of the network structure $A$. Once $A$ is simulated, it is fixed throughout the rest of simulation studies. For a given network structure $A$, the response variable $y_{it}$ is generated according to model (2.2). Then for a given sample size $T$, we first simulate a time series sequence of $y_{it}$ with length of $T_0 + T$. Then the first $T_0$ observations are dropped to eliminate the effect of the initial values. In the simulation study, we set $T_0 = 5000$ for all the three experiments. Finally, a sequence of $\{y_{it}\}$ is simulated.

**Example 1.** (Random Distributed Network Structure) We present a simple network structure whose in-degree (i.e., $q_i = \sum_{j=1}^{N} a_{ji}$) follows a random distribution. This means there are no influential nodes (i.e., with a relatively large in-degree) in the network. We first generate $N$ i.i.d. random variables according to a uniform distribution between 0 and 5. Denote these variables by $U_i$ with $1 \leq i \leq N$. For each node $i$, we randomly select a sample size of $[U_i]$ from $S_F = \{1, 2, \cdots, N\}$ without replacement, where $[U_i]$
3.1 Simulating Data

stands for the smallest integer no less than $U_i$. Denote the sample by $S_i$. Define $a_{ij} = 1$ if $j \in S_i$ and $a_{ij} = 0$ otherwise. This leads to the adjacency matrix $A$. The histogram of in-degree and visualization of this network structure are shown Figure 1. From this figure we can see that distribution of in-degree is almost random and there are no influential nodes.

Example 2. (Power-Law Distributed Network Structure) We next consider the power-law distributed network structure ([Clauset et al.] 2009). This network structure reflects a general phenomenon in reality that the majority of nodes have very few connections while a small amount have a gigantic number of connections. In such networks, there always exist some influential nodes with very large in-degree, such as celebrities. To mimic this network structure, we follow [Clauset et al.] (2009) and generate $A$ as follows. First, we simulate for each node its out-degree in the same way as in Example 1. Next, we generate another $N$ i.i.d. random variables (e.g., denoted as $r_i, i = 1, \cdots, N$) according to the discrete power-law distribution, i.e., $P(r_i) = ck^{-\alpha}$ for a normalizing constant $c$ and the exponent parameter $s = 2.5$. A smaller $s$ value implies a heavier distribution tail. We then normalize each $r_i$ to its corresponding probability $p_i = r_i/\sum_{i=1}^{N} r_i$. For each node $i$, we select a sample size of $[U_i]$ according to the probability of $p_i$ from $S_F = \{1,2,\cdots,N\}$ without replacement. Denote the sample
3.1 Simulating Data

by $S_i$. Define $a_{ij} = 1$ if $j \in S_i$ and $a_{ij} = 0$ otherwise. The histogram of in-degree and visualization of this network structure are shown Figure 2. We can see that there is at least one node with very large degree, which means this could be an influential node.

Example 3. (Stochastic Block Network Structure) The stochastic block model (Nowicki and Snijders, 2011) is another popularly studied network topology in previous literature. For example, in the stock market, different stocks may belong to different industries. The performance of a stock is very likely to be influenced by its neighbors who are in the same industry block. Following Nowicki and Snijders (2011), let $K = N/10$ be the total number of blocks in this case. For simplicity, we randomly assign a block label ($k = 1, 2, \cdots, K$) to each node with equal probability. Next, set $P(a_{ij} = 1) = 0.5$ if $i$ and $j$ belong to the same block and $P(a_{ij} = 1) = 0.001/N$ otherwise. This means nodes in the same block are more likely to be connected with each other compared with the nodes from different blocks. The histogram of in-degree and visualization of this network structure are shown Figure 3. From Figure 3 we can clearly see there are 5 blocks in this simulated network structure.
3.2 Simulation Results

For each example, various combinations of network size (i.e., \(N = 50, 100, 200\)) and sample size (i.e., \(T = 100, 200, 400\)) are investigated. To make the simulation results more stable, we consider different setup of the true parameters \(\theta_0 = (\omega_0, \alpha_0, \lambda_0, \beta_0)'\). In the first example, \(\theta_0\) is fixed as \((0.005, 0.1, 0.1, 0.7)'\). In the second example, \(\theta_0\) is fixed as \((0.01, 0.1, 0.2, 0.6)'\). Lastly, in the block case, \(\theta_0\) is fixed as \((0.02, 0.1, 0.3, 0.5)'\). Each simulation is randomly replicated for \(M = 1000\) times. Let \(\hat{\theta}^{(m)} = (\hat{\theta}_1^{(m)}, \hat{\theta}_2^{(m)}, \hat{\theta}_3^{(m)}, \hat{\theta}_4^{(m)})'\) be the estimators obtained in the \(m\)th \((1 \leq m \leq M)\) replication. We consider two measures to evaluate the finite-sample performance of the proposed method. First, for a given parameter \(\theta_k\) with \(1 \leq k \leq 4\), the root-mean-square error is evaluated by \(\text{RMSE}_k = \{M^{-1} \sum_{m=1}^{M} (\hat{\theta}_k^{(m)} - \theta_k)^2\}^{1/2}\). Second, for each \(1 \leq k \leq 4\), a 95\% confidence interval is constructed for \(\theta_k\) as \(\text{CI}_k^{(m)} = (\hat{\theta}_k^{(m)} - z_{0.975} \hat{\text{SE}}_k^{(m)}, \hat{\theta}_k^{(m)} + z_{0.975} \hat{\text{SE}}_k^{(m)})\), where \(\hat{\text{SE}}_k^{(m)}\) is the square root of the \(j\)th diagonal element of \(\hat{\Sigma}_T\) and \(z_{\alpha}\) is the \(\alpha\)th quantile of a standard normal distribution. Then, the coverage probability is computed as \(\text{CP}_k = M^{-1} \sum_{m=1}^{M} I(\theta_k \in \text{CI}_k^{(m)})\), where \(I(\cdot)\) is the indicator function. Lastly, the network density (i.e., \(\{N(N - 1)\}^{-1} \sum_{i,j} a_{ij}\)) is also reported.

Detailed simulation results are summarized in Tables 1-3. For the first example in Table I, we can see that the estimators are consistent, with the
3.3 Real Data Analysis

RMSE values decrease towards 0 as \( T \to \infty \). Take \( \lambda_0 \) (i.e., the estimated network effect) with \( N = 100 \) for an example, the RMSE value drops from 1.18% to 0.58% as \( T \) increases from 100 to 400. Furthermore, the reported coverage probabilities (i.e., CP) for each parameter (\( \theta_k \)) are all fairly close to their nominal level 95%. This suggests that the estimated standard error (i.e., \( \hat{SE} \)) can approximate the true SE well. Quantitatively similar results are obtained for Example 2 in Table 2 and Example 3 from Table 3. All these findings confirm the fact that the proposed estimator \( \hat{\theta} \) is indeed consistent and asymptotically normal.

3.3 Real Data Analysis

In this subsection, we use the proposed model to analyze some real stock market datasets. The stock data is from Chinese A share market, which are traded in Shanghai Stock Exchange and Shenzhen Stock Exchange during the year of 2014. Specifically, the response variable \( y_{it} \) is daily log return. According to the industry classification criteria provided by the China Securities Regulatory Commission, each of the stocks is divided into one of the 18 different categories. For illustration purpose, we only select four categories to assess the performance of the proposed model. They are the Mining Industry with 68 stocks, Real Estate with 127 stocks, Wholesale
3.3 Real Data Analysis

and Retail with 139 stocks, and Manufacturing with 1,515 stocks. For each category, the network structure is constructed according to common shared ownership information. Specifically, we try three different choices of adjacency matrix to compare the estimation results. The first one is that we collect the top ten shareholders’ information for each stock. The network structure (i.e., adjacency matrix) is constructed as follows. For any two arbitrary stocks $i$ and $j$, $a_{ij} = 1$ if they share at least one common shareholder, otherwise $a_{ij} = 0$. The second one is similar as the first one except that for stocks $i$ and $j$, $a_{ij} = 1$ if they share at least two common shareholders, otherwise $a_{ij} = 0$. Lastly, we collect the top five shareholders’ information for each stock. For stocks $i$ and $j$, $a_{ij} = 1$ if they share at least one common shareholder, otherwise $a_{ij} = 0$.

To give a descriptive analysis about these four categories, the daily averaged stock return for each industry is plotted in Figure 4. We also display their network structure in Figure 5. It is clear to see that the network structure for the 4 industries is different from each other. Then we apply the network GARCH model to each dataset. Estimation results are given in Table 4.

From Table 4 we can find the results are consistent across different choices of adjacency matrix. Nearly all estimates are statistically signifi-
significant at 1% or 5% level, except for the second case with the category of manufacturing and retail. This is understandable because network density in case two is too sparse. This leads to an insignificant effect of network structure. Taking Mining Industry in case one as an example, the estimated network effect $\lambda_0$ (0.0571) suggests that the return of a stock is positively related with the performance of its connected neighbors. The estimated $\alpha_0$ (0.1933) confirms that a stock with higher (lower) return in the past is likely to exhibit a higher (lower) performance in the future. Finally, the estimated $\beta_0$ (0.5614) is very strong compared with the other two effects, indicating the variance of a stock could be very large in real practice. Similar results could be found in the other three categories.

4. Conclusion

We propose a network GARCH model which takes network structure information into consideration. To capture the impact of connected neighbors, we introduce a network structure term in the traditional GARCH (1,1) model. The new model can substantially decrease the computational complexity from $O(N^2)$ to $O(N)$. The resulting estimators enjoy asymptotic properties and these findings are also confirmed by extensive numerical studies. We further illustrate our model by a real dataset from Chinese stock
market. Significant network structure term is detected.

To conclude this article, we discuss here a number of interesting topics for future study. First, the network structure term in model (2.2) is added on $y_{j,t-1}^2$. However, the adjacency matrix can also be added on $h_{j,t-1}$. This makes the resulting model considerably more complicated. The associated theoretical development would be more challenging. Nevertheless, the research effort to link $h_{j,t-1}$ should be a separate research topic in the future. Second, the parameter $\theta$ are assumed to be the same across different stocks. However, stock heterogeneity problems may exist in reality. That means different stocks may have a different reaction to the effect. Then investigating a vary-coefficient network GARCH model is another interesting problem worthwhile pursuing. Lastly, the network structure discussed in the paper is simple and straightforward. The network term is only driven by one parameter $\lambda$. This may limit our model in some specific context. However, the empirical results show that the model performs very well when the stocks come from the same category for instance. In the future, we could consider more flexible network structures, such as sub-block structure.

**Supplementary Materials**

All technical details can be found in the supplementary materials.
Acknowledgements

We express our sincere thanks to the Co-editor, an Associate Editor and an anonymous referee for their constructive suggestions. Zhou’s research is supported in part by the NSFC (No.71702185). Li’s research is supported in part by the NSFC (No.11571348, No.11771239). Pan’s research is supported in part by the NSFC (No.11601539, No.11631003, No.71771224). The Fundamental Research Funds for the Central Universities (QL18010). Wang’s research is partially supported by the NSFC (No. 71532001, No. 11525101, No. 71332006). China’s National Key Research Special Program (No. 2016YFC0207700).

References


REFERENCES


REFERENCES

*Marketing Science* 33, 82-93.

Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors.


Wei, Y., Yildirim, P., Christophe, V. D. B. and Dellarocas, C. (2014). Credit scoring with social


Figure 1: Left panel: in-degree distribution for Example 1 with $N = 50$. Right panel: network visualization. Dot stands for node and line stands for edge. The deeper the color and the larger the dot means a larger in-degree.

Table 1: Simulation results for Example 1 with 1000 replications. The RMSE values ($\times 10^{-2}$) are reported for each estimate. The corresponding CP (in %) is given in parentheses. Network density (ND) is also reported.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\omega_0$</th>
<th>$\alpha_0$</th>
<th>$\lambda_0$</th>
<th>$\beta_0$</th>
<th>ND(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.14 (94.9)</td>
<td>1.43 (94.9)</td>
<td>1.66 (95.1)</td>
<td>4.23 (93.7)</td>
<td>6.41</td>
</tr>
<tr>
<td>200</td>
<td>0.08 (95.7)</td>
<td>0.99 (94.9)</td>
<td>1.12 (94.7)</td>
<td>2.79 (94.8)</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.06 (95.2)</td>
<td>0.70 (95.3)</td>
<td>0.79 (94.8)</td>
<td>1.87 (95.2)</td>
<td></td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.11 (95.0)</td>
<td>1.00 (95.4)</td>
<td>1.18 (94.5)</td>
<td>3.02 (94.6)</td>
<td>3.20</td>
</tr>
<tr>
<td>200</td>
<td>0.07 (95.0)</td>
<td>0.71 (94.6)</td>
<td>0.78 (96.5)</td>
<td>2.04 (94.1)</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.05 (95.2)</td>
<td>0.49 (95.5)</td>
<td>0.58 (94.2)</td>
<td>1.38 (94.7)</td>
<td></td>
</tr>
<tr>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.08 (94.0)</td>
<td>0.72 (94.2)</td>
<td>0.81 (94.6)</td>
<td>2.25 (92.0)</td>
<td>1.52</td>
</tr>
<tr>
<td>200</td>
<td>0.05 (93.0)</td>
<td>0.52 (94.4)</td>
<td>0.55 (95.3)</td>
<td>1.47 (94.2)</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.03 (95.0)</td>
<td>0.36 (93.9)</td>
<td>0.40 (94.6)</td>
<td>1.00 (93.3)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Left panel: in-degree distribution for Example 2 with \( N = 50 \). Right panel: network visualization. Dot stands for node and line stands for edge. The deeper the color and the larger the dot means a larger in-degree.

Table 2: Simulation results for Example 2 with 1000 replications. The RMSE values (\( \times 10^{-2} \)) are reported for each estimate. The corresponding CP (in %) is given in parentheses. Network density (ND) is also reported.

<table>
<thead>
<tr>
<th>( T )</th>
<th>( \omega_0 )</th>
<th>( \alpha_0 )</th>
<th>( \lambda_0 )</th>
<th>( \beta_0 )</th>
<th>ND(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 50 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.47 (91.7)</td>
<td>1.56 (94.3)</td>
<td>2.00 (92.2)</td>
<td>3.93 (88.7)</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.16 (93.5)</td>
<td>1.04 (93.6)</td>
<td>1.27 (93.6)</td>
<td>2.35 (91.9)</td>
<td>5.51</td>
</tr>
<tr>
<td>400</td>
<td>0.11 (93.5)</td>
<td>0.73 (94.3)</td>
<td>0.85 (94.0)</td>
<td>1.61 (92.0)</td>
<td></td>
</tr>
<tr>
<td>( N = 100 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.20 (94.5)</td>
<td>1.04 (95.4)</td>
<td>1.44 (94.3)</td>
<td>2.82 (93.9)</td>
<td>3.09</td>
</tr>
<tr>
<td>200</td>
<td>0.13 (94.0)</td>
<td>0.74 (93.8)</td>
<td>0.95 (95.9)</td>
<td>1.97 (93.1)</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.09 (95.0)</td>
<td>0.50 (95.4)</td>
<td>0.70 (95.5)</td>
<td>1.33 (93.6)</td>
<td></td>
</tr>
<tr>
<td>( N = 200 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.14 (93.3)</td>
<td>0.73 (94.8)</td>
<td>1.01 (94.4)</td>
<td>2.13 (93.5)</td>
<td>1.50</td>
</tr>
<tr>
<td>200</td>
<td>0.09 (93.6)</td>
<td>0.54 (94.6)</td>
<td>0.72 (94.0)</td>
<td>1.40 (93.5)</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.06 (95.3)</td>
<td>0.37 (94.7)</td>
<td>0.50 (94.7)</td>
<td>0.95 (94.6)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3: Left panel: in-degree distribution for Example 3 with $N = 50$. Right panel: network visualization with $K = 5$. The dot stands for node and the line stands for edge.

Table 3: Simulation results for Example 3 with 1000 replications. The RMSE values ($\times 10^{-2}$) are reported for each estimate. The corresponding CP (in %) is given in parentheses. Network density (ND) is also reported.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\omega_0$</th>
<th>$\alpha_0$</th>
<th>$\lambda_0$</th>
<th>$\beta_0$</th>
<th>ND(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.45 (95.4)</td>
<td>1.61 (95.1)</td>
<td>3.50 (95.1)</td>
<td>5.42 (94.4)</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.29 (95.3)</td>
<td>1.13 (94.5)</td>
<td>2.37 (95.6)</td>
<td>3.64 (94.4)</td>
<td>9.43</td>
</tr>
<tr>
<td>400</td>
<td>0.21 (95.4)</td>
<td>0.79 (95.5)</td>
<td>1.75 (95.4)</td>
<td>2.61 (94.4)</td>
<td></td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.30 (95.9)</td>
<td>1.15 (94.8)</td>
<td>2.53 (93.8)</td>
<td>3.79 (94.2)</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.20 (95.3)</td>
<td>0.79 (95.2)</td>
<td>1.73 (95.1)</td>
<td>2.57 (94.7)</td>
<td>4.65</td>
</tr>
<tr>
<td>400</td>
<td>0.14 (95.4)</td>
<td>0.55 (95.2)</td>
<td>1.20 (94.6)</td>
<td>1.71 (95.8)</td>
<td></td>
</tr>
<tr>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.20 (94.3)</td>
<td>0.79 (95.6)</td>
<td>1.79 (93.8)</td>
<td>2.75 (93.3)</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.14 (94.2)</td>
<td>0.58 (95.4)</td>
<td>1.22 (95.7)</td>
<td>1.86 (94.6)</td>
<td>2.41</td>
</tr>
<tr>
<td>400</td>
<td>0.09 (94.7)</td>
<td>0.40 (95.3)</td>
<td>0.87 (94.7)</td>
<td>1.27 (95.4)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4: Average stock return for the 4 categories in 2014.
Figure 5: Network structure for the 4 categories.
<table>
<thead>
<tr>
<th>Category</th>
<th>Parameter</th>
<th>Estimate</th>
<th>p-value</th>
<th>Estimate</th>
<th>p-value</th>
<th>Estimate</th>
<th>p-value</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining Industry</td>
<td>$\omega_0$</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0183</td>
<td>&lt;.001</td>
<td>0.0015</td>
<td>&lt;.001</td>
<td>0.5514</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\alpha_0$</td>
<td>0.0571</td>
<td>&lt;.001</td>
<td>0.0712</td>
<td>&lt;.001</td>
<td>0.0047</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\lambda_0$</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>0.0169</td>
<td>&lt;.001</td>
<td>0.0268</td>
<td>&lt;.001</td>
<td>0.0068</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Real Estate</td>
<td>$\omega_0$</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0186</td>
<td>&lt;.001</td>
<td>0.0015</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\alpha_0$</td>
<td>0.0311</td>
<td>&lt;.001</td>
<td>0.0498</td>
<td>&lt;.001</td>
<td>0.0068</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\lambda_0$</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>0.0015</td>
<td>&lt;.001</td>
<td>0.0010</td>
<td>&lt;.001</td>
<td>0.0010</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wholesale and Retail</td>
<td>$\omega_0$</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0185</td>
<td>&lt;.001</td>
<td>0.0015</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\alpha_0$</td>
<td>0.0447</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\lambda_0$</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>0.0015</td>
<td>&lt;.001</td>
<td>0.0010</td>
<td>&lt;.001</td>
<td>0.0010</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>$\omega_0$</td>
<td>$0.5 \times 10^{-1}$</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\alpha_0$</td>
<td>0.0141</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\lambda_0$</td>
<td>0.0016</td>
<td>&lt;.001</td>
<td>0.0027</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>0.0010</td>
<td>&lt;.001</td>
<td>0.0010</td>
<td>&lt;.001</td>
<td>0.0010</td>
<td>&lt;.001</td>
<td>0.0001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>