Nonlinear Network Vector Autoregression

October 16, 2020

Chengjie Cai
Department of Mathematics and Statistics
University of Strathclyde
Glasgow, UK

This thesis is submitted to the University of Strathclyde for the degree of Doctor of Philosophy in the Faculty of Science.

Acknowledgements

First of all, I would like to express my sincere thanks to my supervisor, Dr. Jiazhu Pan and Prof. Xuerong Mao for their guidance and support. Dr. Jiazhu Pan spared no effort to guide me how to carry out statistical research and help me overcome the difficulties. Prof. Mao helped us a group of students from China to study in the UK and opened the door to the international community. In any term, this is a remarkable achievement.

Secondly, I would like to thank the China Scholarship Council and University of Strathclyde for their financial support.

I want to thank Dr. Jiafeng Pan. Thanks for her patience and always encouraged me. Her insightful comments and suggestions are very helpful for me in research.

Thanks to all the staff and students at the Department of Mathematics and Statistics in University of Strathclyde for their direct or indirect help in my research.

Thanks to my colleagues and friends, Dr. Siyang Cai, Dr. Yongmei Cai, Dr. Jianqiu Lu and all my friends I met in Glasgow. Because of you, my life in Glasgow was very enjoyable.

Last but not least, I want to thank my parents and my family, thank them for encouraging me, supporting me and giving me patience. My parents have provided me with the best environment and opportunities to promote my development. I am very grateful for their decision to support me abroad many years ago.

Abstract

As we all know, the time series model is one of the most important aspects of modern econometric analysis. The autoregressive model is the theoretical basis of time series.

The classic autoregressive model has two features that can be improved, linear, and one-dimensional. Economic theory shows that many important macroeconomic time series exhibit nonlinear characteristics. If this non-linear feature is ignored, the conclusion is likely to be wrong by only using linear analysis. Therefore, it is necessary to expand the linear model and propose nonlinear methods. Economic theory also shows that there is a mutual influence between individuals, and it is also necessary to consider it as a network.

In this dissertation, we consider three nonlinear autoregressive models for time series with network structure: The Threshold Network autoregressive (TNAR) model, the Threshold Network quantile autoregressive (TNQAR) model and the Markov Switching Network autoregressive (MS-NAR) model.

For the TNAR model, we provide the parameter conditions for the stationary of the time series. Under this parameter condition, the TNAR process can be approximated by the geometrically ergodic process. Under these conditions, we discuss the statistical inference (estimation and test) of the TNAR model and give the asymptotic theory on the inference. The test for nonlinearity is applied. Simulation results and modeling for Twitter data were applied to support our methodology for TNAR models.

For the TNQAR model, we also provide the parameter conditions for the stationary of the time series. Under these parameter conditions, the TNQAR process can be approximated by the geometric traversal process. Under these condition, we discuss the statistical inference (estimation and test) of the TNQAR model and

give the asymptotic theory on the inference. A normality test is applied on the data and a Hill estimator is provided to check whether the conditional distribution of the historical information is a thick tail or not. Simulation results and modeling of hedge fund data were used to support our methodology for TNQAR models.

The Markovian Switching model is provided and its maximum likelihood estimation method is discussed.

Finally, the techniques and the process in collecting Twitter data are presented.

Contents

List of Figures

List of Tables

Abbreviations

ACF	Autocorrelation Function
MLE	Maximum Likelihood Estimation
MSE	Mean Square Error
LM	Lagrange Multiplier
AAE	Average Absolute Error
AR	Autoregressive
ARMA	Autoregressive Moving Average
TAR	Threshold Autoregressive
NAR	Network Vector Autoregressive
MSAR	Markov Switching Autoregressive
GMM	Generalized Method of Moments

Chapter 1

Introduction

1.1 Background

Nowadays, social media produces lots of high-dimensional data with network property. In order to fit the dynamic changes of user behaviour, the Network Vector AutoRegressive (NAR) model [50] is proposed. Social networks are made up of many nodes and network relationships between nodes. For example, the activity of a group of people on Twitter can be described by using a high-dimensional time series. Each person in this group has impact on others. With increasing social media activity, how to analyse such social network is becoming increasingly important in data science. Unlike traditional data, network structure data is no longer independent, but contains information about the relationships between individuals. The relationships are of great commercial value. Modelling such high-dimensional dependent data is a challenge in the research of Big Data. My study aims to develop a nonlinear time series model to describe the dynamics of large-scale social networks. We found the nonlinear phenomenon about a social network. For example, when the activity of a social network increases, it increases gradually and slowly; however when the activity of the social network drops, it drops sharply and quickly. It means the speeds of increase and decrease are different. This is why we want to study the nonlinearity in a social network.

Our network time series problem has a close relationship with the multivariate time series [18] [13] [5]. [32] firstly suggested to model each individual time series separately. This method is simple in both theory and computation. However, the relationships across the different time series were not considered. [4] developed a vector autoregressive (VAR) model. In this model, all the information was taken into account, but the number of parameters that need to be estimated is very large. Especially, when the number of parameters is greater than the number of observations, the parameters cannot be estimated. Therefore, a lot of efforts were then made to reduce the number of parameters. For example, [36] provided a parameter reduction method by factor modelling. [50] considered network structure data and developed a network autoregressive (NAR) model. The NAR model not only considered the relationships across variables, but also reduced the parameter dimension.

However, the NAR model is a linear model. So the existing models are not able to describe the data with nonlinear property. As high dimensional social network has nonlinear property, we therefore propose to extend the NAR model to a series of nonlinear models.

1.2 Overview of the Study

This thesis studies statistical inference of several types of nonlinear high-dimensional time series models, including threshold (TNAR) and Markov-Switching (MSNAR), which are two common non-linear models. At the same time, we also consider how to construct the quantile regression (TQNATR) model to explain the phenomenon of nonlinearity when the distribution of the noise has thick tail, that is, the assumption that the noise follows the normal distribution fails.

In the first chapter, we review the background knowledge of nonlinear time series. The second chapter introduces some theoretical basis of this thesis, including expanding the generalized method of moments (GMM) method and nonlinear test to the high dimension problems. The third chapter introduces the Threshold Network Autoregressive (TNAR) model and discusses the properties of its geometric ergodicity in order to obtain its stationarity. The GMM is applied to obtain parameter estimation of the model and the asymptotic property of the estimation has also been discussed. In addition, the Lagrange Multiplier test is provided to determine the non-linearity of the data.

At the end of the third chapter, three simulations were conducted, where we derived the parameter estimation for the TNAR model using GMM estimators that we have derived. The estimation results were compared to the true parameters and it shows that the estimations were very accurate. And a real-world example using the data extracted from Twitter (the details of data extraction are explained in chapter 6) is presented to support our research results.

In the fourth chapter, we developed the non-linear Quantile Threshold Network Autoregressive (QTNAR) model. Similar to the TNAR model, we discuss the stationary conditions of the QTNAR model, provide its estimation method and explore its asymptotic property. In addition, the Hill Estimator is provided to check whether the data has a thick tail. And simulation results as well as an example of financial markets are provided to support our research results.

The fifth chapter introduces another non-linear time series model, that is, the Markov-Switching Model and the transformation mechanism of the model is decided by an unobservable state variable. Simulation results of the model by using maximum likelihood estimation, as well as real data analysis, have been discussed. In some real data analysis of this thesis, there is no readily available data for us to use, so we need to extract the raw data from the network and process it to obtain the data that we can process and analyze. The sixth chapter introduces the process of extracting data from Twitter website using Python. Chapter 7, the last chapter of the thesis, includes a summary of our findings and some unresolved problems for future works in the field.

Chapter 2

Preliminaries

2.1 Definition of Markov Processes

The Markov process is a kind of stochastic process. Its original model, Markov chain was proposed by Russian mathematician A.A. Markov in 1907. A stochastic process is a mathematical model for a system evolving randomly in time. Depending on the application, time may be modelled as discrete (e.g. 0,1,2,...) or continuous (e.g. the real interval $[0, \infty)$). A stochastic process $\{X(t), t \geq 0\}$ is called a Markov process if the following Markov property is satisfied: for any $t_1 < t_2 < \cdots < t_n < t$, $P\{X(t) \leq x | X(t_n) = x_n, \cdots, X(t_1) = x_1\} = P\{X(t) \leq x | X(t_n) = x_n\}$, which states roughly that "given the present, the future is independent of the past".

2.2 Definition of an aperiodic irreducible Markov chain

First we discuss the definition of irreducibility. We need the following notion.

- For any $i, j \in S$, which S is the state space, we say that the state j is accessible from the state i if $p_{ij} > 0$ for some $n \geq 0$, (notation: $i \rightarrow j$).

The property of accessibility is

- transitive, i.e., $i \rightarrow k$ and $k \rightarrow j$ imply that $i \rightarrow j$.

- Moreover, in case $i \to j$ and $j \to i$, we say that the states i and j communicate, (notation $i \leftrightarrow j$).

The property of communicating is an equivalence relation as

- $i \leftrightarrow i$ (reflexivity),
- $i \leftrightarrow j$ if and only if $j \leftrightarrow i$ (symmetry),
- $i \leftrightarrow k$ and $k \leftrightarrow j$implies $i \leftrightarrow j$ (transitivity).

As a consequence, the state space S can be completely divided into disjoint equivalence classes with respect to the equivalence relation $\leftrightarrow$. The Markov chain $\{X_t\}$ with transition matrix $P = (p_{ij})$ is called irreducible if the state space S consists of only one equivalence class, i.e. $i \leftrightarrow j$ for all $i, j \in S$.

Besides irreducibility, we need the second property of the transition probabilities, i.e. aperiodicity.

The period d_i of the state $i \in S$ is given by $d_i = gcd\{n \geq 1 : p_{ii}^{(n)} > 0$ where "gcs" denotes the greatest common divisor. We define $d_i = \infty$ if $p_{ii}^{(n)} = 0$ for all $n \geq 1$. A state $i \in S$ is said to be aperiodic if $d_i = 1$. The Markov chain $\{X_t\}$ and its transition matrix $P = (p_{ij})$ are called aperiodic if all states of $\{X_t\}$ are aperiodic.

2.3 Introduction to the NAR model

Here we consider a large-scale social network (for example, Facebook or Twitter), which has N nodes, index i from 1 to N. To describe the network structure, an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ is defined, where $a_{ij} = 1$ if there is a social relationship from i to j (e.g. user i follows user j on Twitter), otherwise $a_{ij} = 0$. It can be directed (i.e. $A \neq A^\top$) or undirected (i.e. $A = A^\top$). Any node is not allowed to be self-related, so that $a_{ii} = 0$ (for example, any Twitter user cannot follow himself). Let y_{it} be a continuous response, observed from node i at time t. (e.g. tweet length). $\mathbb{Y}_t = (y_{1t}, y_{2t}, \cdots, y_{Nt})^\top$ is the object we want to study.

Under the network framework, y_{it} may be affected by four different factors. First, y_{it} may be affected by itself, but from the previous point in time, $y_{i(t-1)}$. Second, y_{it} may be affected by its followees, that is, $\{j : a_{ij} = 1\}$. Third, y_{it} may also be affected by a set of node-covariates (Z_i), such as a person's age, gender and location. Finally, unexplained changes should be attributed to an independent random noise. Therefore, [50] propose a network vector autoregressive (NAR) model.

$$y_{it} = \beta_0 + Z_i^\top \gamma + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} y_{j(t-1)} + \beta_2 y_{i(t-1)} + \varepsilon_{it}, i = 1, \cdots, N. \tag{2.1}$$

The NAR model assumes that the response of each node at a given time point is a linear combination of (a) the previous value ($y_{i(t-1)}$), (b) the average value of the connected neighbors $n_i^{-1} \sum_j a_{ij} y_{j(t-1)}$ with $n_i = \sum_j a_{ij}$, called out-degree [47], (c) a set of node-specific covariates Z_i and (d) independent noise. The corresponding coefficients are regarded as: the momentum effect, the network effect and the nodal effect, respectively.

The term $\beta_0 + Z_i^\top \gamma$ constitutes the node intercept of the ith node, where β_0 is the intercept and γ is the corresponding coefficient (i.e. the nodal effect) . We write $\beta_{0i} = \beta_0 + Z_i^\top \gamma$. ε_{it} is the error term, which follows the normal distribution with $E(\varepsilon_{it}) = 0$ and $var(\varepsilon_{it}) = \sigma^2$.

Compared with the usual VAR model which needs to estimate N parameters, the total number of unknown parameters in the NAR model is fixed. Therefore, it is easy to estimate NAR models for large social networks.

2.4 Introduction to the TAR Model

2.4.1 Autoregressive Models

Consider a simple AR(p) model for a time series y_t

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \sigma \epsilon_t. \tag{2.2}$$

where γ_i for $i = 1, 2, \cdots, p$ are autoregressive coefficients, assumed to be constant over time; $\epsilon_t \sim WN(0, 1)$ stands for white-noise error term with constant variance. The AR model can be written in a following vector form:

$$y_t = \mathbf{X_t} \gamma + \sigma \epsilon_\mathbf{t} \tag{2.3}$$

where $\mathbf{X_t} = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ is a row vector of variables and γ is the vector of parameters : $(\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_p)^\top$. $\epsilon_\mathbf{t} \sim WN(0, 1)$ stands for white-noise error term.

2.4.2 Threshold Autoregressive (TAR) Models

The TAR model can be considered as the extension of the autoregressive model, allowing the model parameters to be changed based on a value of the weakly exogenous threshold variable $q_{t-1} = q(y_{t-1}, \cdots, y_{t-p})$. $q(*)$ is an unknown function, which can be defined according to some requirements. Usually y_{t-1} triggers the changes.

Defined in this way, the TAR model can be expressed as follows

$$y_t = \mathbf{X_t}\gamma^{(j)} + \sigma^{(j)}\epsilon_t \qquad \text{if} \qquad r_{j-1} < q_{t-1} < r_j. \tag{2.4}$$

$-\infty = r_0 < r_1 < \cdots < r_k = \infty$ are $k + 1$ non-trivial thresholds that divide the domain of q_t into k different states.

2.4.3 Estimation of the TAR(1) Model

Next, the estimation of a two-regime TAR(1) model is introduced. The TAR(1) model can be written as

$$y_t = (\gamma_0^{(1)} + \gamma_1^{(1)} y_{t-1}) I(q_{t-1} \leq r) + (\gamma_0^{(2)} + \gamma_1^{(2)} y_{t-1}) I(q_{t-1} > r) + \epsilon_t \tag{2.5}$$

where $I(\cdot)$ denotes the indicator function.

Hence, $\mathbf{X_t} = (1, y_{t-1})$ since the order of TAR model is 1. Let $\mathbf{X_t}(r) = (\mathbf{X_t}\mathbf{1}(q_{t-1} \leq r), \mathbf{X_t}\mathbf{1}(q_{t-1} > r))$, where $\mathbf{1}(\cdot) = (I(\cdot), I(\cdot))^\top$, so equation 2.5 can be written as

$$y_t = \mathbf{X_t}(r)\theta + \epsilon_t \tag{2.6}$$

where $\theta = (\gamma^{(1)\top}, \gamma^{(2)\top})^\top$

The least squares (LS) method is used to estimate θ because equation 2.6 is a regression equation (although the parameters are not linear). For a given r, the LS estimate of θ is

$$\hat{\theta}(r) = \left(\sum_{t=1}^{n} \mathbf{X_t}^\top(r)\mathbf{X_t}(r) \right)^{-1} \left(\sum_{t=1}^{n} \mathbf{X_t}(r)^\top y_t \right),$$

with residuals $\hat{e}_t(r) = y_t - \mathbf{X}_t(r)\hat{\theta}(r)$, and residual variance

$$\hat{\sigma}_n^2(r) = \frac{1}{n}\sum_{t=1}^{n}\hat{e}_t(r)^2.$$

2.4.4 Test for Threshold Autoregression

An important question is whether the TAR(1) model, i.e. equation 2.5, is statistically significant for linear AR (1). The null hypothesis is: $H_0 : \gamma^{(1)} = \gamma^{(2)}$. We review the test methodology proposed by [20].

If the errors are i.i.d., a standard F statistic can be used, i.e.

$$F_n = n\left(\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2}{\hat{\sigma}_n^2}\right),$$

where

$$\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \mathbf{X}_t\tilde{\gamma})^2,$$

and

$$\tilde{\gamma} = \left(\sum_{t=1}^{n}\mathbf{X_t}^\top\mathbf{X_t}\right)^{-1}\left(\sum_{t=1}^{n}\mathbf{X_t}^\top y_t\right)$$

is the ordinary least squares (OLS) estimate of γ under the null hypothesis that $\gamma^{(1)} = \gamma^{(2)}$. Then the F statistic for the threshold autoregression can be derived that

$$F_n = \sup_{r\in\Gamma} F_n(r)$$

where $\Gamma = [\underline{r}, \overline{r}]$ and

$$F_n(r) = n\left(\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2(r)}{\hat{\sigma}_n^2(r)}\right)$$

is the pointwise F-statistic.

2.5 Generalized Method of Moments

According to the Chapter 14 in Time Series Analysis by [18], the aim of GMM is to minimalize the equation of

$$Q(\beta) = m(\beta)^\top W^{-1} m(\beta)$$

and the estimate is

$$\hat{\beta} = \arg\min(m(\beta)^\top W^{-1} m(\beta)). \tag{2.7}$$

The choice of the weight matrix is the core issue of the moment estimation method. [22] proposed the best weight matrix

$$W = (1/T)\sum_{t=1}^{T}[m(\beta_0)][m(\beta_0)]^\top, \tag{2.8}$$

if the random errors were serially uncorrelated.

If the random error is serially correlated, [33] proposed the estimate of W:

$$W = \Gamma_0 + \sum_{v=1}^{q}\{1 - [v/(q+1)]\}(\Gamma_v + \Gamma_v^\top),$$

where

$$\Gamma_v = (1/T)\sum_{t=v+1}^{T}[m(\beta)][m(\beta)]^\top.$$

The GMM estimator is asymptotically effective in large sample and invalid in small sample. So parameter estimation can apply GMM only in large sample.

The procedure of GMM estimation is as follows. An initial estimate β is obtained by OLS. Then the estimate of β is used in 2.8 to obtain an estimate of W. Final, apply 2.7 to obtain the estimate of GMM.

2.6 Nonlinearity Test

The basic assumption of the classical linear regression model is that the regression variables, disturbance terms, and parameters are linear. In most cases, this assumption is difficult to find theoretical foundation. Linearity is just an approximation and simplification.

If the correct regression model is a nonlinear model and replaced with a linear model, the model setting error will be generated and the estimator will be biased and non-uniform. One of the reasons for the autocorrelation, heteroscedasticity, and non-normal errors is the linear setting of this nonlinear model.

Therefore, nonlinearity test should be carried out to decide whether fitting a linear model to data is appropriate. The nonlinearity test can be carried out using the Lagrange Multiplier (LM) test [39] or equivalent [38]'s score test. The LM statistic has the same asymptotic distribution as the likelihood ratio test and the Wald test, but only the estimator of the null hypothesis is considered for the LM test. Usually, if the null hypothesis model is linear, its estimator is relatively easy.

2.7 Threshold Effect Test

In empirical research, for a time series type of economic variable, whether to establish a linear time series model or a nonlinear time series model need to be judged. This problem cannot be visually identified, but it can be transformed into a statistical hypothesis test problem. Regarding to the test of the threshold effect, many available test methods have been developed. At present, the two most widely used tests include [44]'s F test, which is based on the principle of arranged autoregression and the SupWald test proposed by [20].

2.8 Quantile Regression

According to the idea of quantile regression proposed by [28], [29] further proposed a quantile autoregressive model (QAR). In this model, the autoregressive coefficients are variable at different quantile points, which can describe the different behavioral features of time series at different quantile points, and provide complete information of the entire conditional distribution of time series. For the time series $\{y_t\}$, the p-order linear quantile autoregressive model can be expressed as:

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1} + \cdots + \theta_p(\tau)y_{t-p}, \tag{2.9}$$

where $\mathcal{F}_t$ is the σ-field generated by $\{y_s, s \leq t\}$ and $\tau \in (0,1)$ is the quantile. $Q_{y_t}(\tau|\mathcal{F}_{t-1})$ is the conditional quantile of y_t at the τ quantile. $\theta_0(\tau), \theta_1(\tau), \cdots, \theta_p(\tau)$ represents the autoregressive coefficient at the τ quantile.

Equation 2.9 can be simplified as

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \mathbf{x}_t^\top \theta(\tau).$$

where $\mathbf{x}_t = (1, y_{t-1}, \cdots, y_{t-p})^\top$ and $\theta(\tau) = (\theta_0(\tau), \theta_1(\tau), \cdots, \theta_p(\tau))$.

The parameter estimates for this model can be obtained by optimizing the equation

$$\hat{\theta}(\tau) = \arg\ \min_{\theta} \sum_{t=1}^{T} \rho_{\tau}(y_t - Q_{y_t}(\tau|\mathcal{F}_{t-1}))$$

where $\rho_{\tau}(u) = u(\tau - I(u < 0))$ is the loss function in [28].

Chapter 3

Threshold Network Autoregressive Model

3.1 Background

In the field of classical econometrics, the linear model has an important position and is the base of other econometric models. The linear model is relatively simple in terms of model setting, and its parameter estimation. The model prediction methods for the linear model are also relatively mature, and the results of the linear model are also easy to explain and understand in economic theory. The linear autoregressive model is an important time series linear model and is usually used to describe the linear dynamic of the adjustment mechanism for the economic variable. It is the base of time series analysis and is important in time series analysis. However, economic theory also shows that many important macroeconomic time series exhibit nonlinear characteristics. Many empirical studies also support the conclusion that "a large number of macroeconomic sequences have the characteristics of nonlinear dynamic adjustment", such as interest rate [2], inflation rate [11], etc. It is clearly not appropriate to still use the linear autoregressive model to model these economic variables that exhibit nonlinear dynamic mechanisms.

In order to adapt to the rapid development of economic theory, the nonlinear methodology has also been rapidly developed. In the development of nonlinear time series analysis, one of the focuses is on various nonlinear parametric models. Threshold autoregressive (TAR) built an important theory for studying the nonlinear dynamic behaviour of time series. It was first proposed by [41] and discussed in detail by [43]and [42].

TAR is an approach to model the data using a multi-regime segmented local linear autoregressive model. According to the threshold value, the time series is divided into multiple regimes, and establishes different linear autoregressive models for each regime.

3.2 The Model and Its Stationarity

3.2.1 Model and Notations

Our TNAR model is based on the NAR model. The model in [50] shows as follows

$$y_{it} = \beta_0 + Z_i^\top \gamma + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} y_{j(t-1)} + \beta_2 y_{i(t-1)} + \varepsilon_{it}, i = 1, \cdots, N. \tag{3.1}$$

We considered a large-scale social network with N nodes indexed by $1 \leq i \leq N$ and y_{it} is the response of ith node at time point t. The response y_{it} is a linear combination of: (a) $y_{i(t-1)}$, (b) $n_i^{-1} \sum_{j=1}^{N} a_{ij} y_{j(t-1)}$ with $n_i = \sum_j a_{ij}$, (c) node-specific covariates Z_i and (d) an independent noise.

$Z_i = (Z_{i1}, \cdots, Z_{ip})^\top$ is a p-dimensional node-specific random vector for each node i and a_{ij} describes the relationship between ith node and jth node by the following rule that $a_{ij} = 1$ if the ith node follows the jth node, otherwise $a_{ij} = 0$. Then the adjacency matrix A can be defined by $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ and $\mathbb{Y}_t = (y_{1t}, \cdots, y_{Nt})^\top \in \mathbb{R}^N$ constitutes an ultra-high dimensional vector.

This is a linear model which can capture linear dynamic structure of a network. However, empirical study on our data from Twitter showed the social networks have nonlinear dynamic structures. To incorporate the nonlinear property, we propose the following threshold network autoregressive ($TNAR$) model,

$$\begin{aligned} y_{it} = \quad & \beta_0 + Z_i^\top \gamma + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} y_{j(t-1)} + [\beta_2^{(1)} I_{\{y_{i(t-1)} \geq r\}} \\ & + \beta_2^{(2)} I_{\{y_i(t-1) < r\}}] y_{i(t-1)} + \varepsilon_{it}, i = 1, \cdots, N, \end{aligned} \tag{3.2}$$

where $I_{\{y_{i(t-1)} \geq r\}}$ is an indicator function which takes 1 when $y_{i(t-1)} < r$ and 0 otherwise. For simplicity in the exposition the following sections are restricted to studying nonlinear processes with at most two regimes. Notes, however, that the methodology introduced here can be easily extended to more regimes.

Define $\mathbb{Z} = (Z_1, \cdots, Z_N)^\top \in \mathbb{R}^{N\times p}$ and $\mathcal{B}_0 = \beta_0\mathbf{1} + \mathbb{Z}\gamma \in \mathbb{R}^N$, where $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_p)^\top$ is the corresponding coefficient. The above equation (3.2) can be rewritten in vector-matrix form as follows

$$\mathbb{Y}_t = \beta_0\mathbf{1} + \beta_1 W\mathbb{Y}_{t-1} + \beta_2\mathbb{Y}_{t-1} + \beta_3 J_{t-1}\mathbb{Y}_{t-1} + \mathbb{Z}\gamma + \varepsilon_t, \tag{3.3}$$

where $\beta_2 = \beta_2^{(1)}, \beta_3 = \beta_2^{(2)} - \beta_2^{(1)}, \mathbf{1} = (1, 1, \cdots, 1)^\top, W = diag\{n_1^{-1}, \cdots, n_N^{-1}\}A$, is the row-normalized adjacency matrix and $J_{t-1} = \mathrm{diag}\{I_{\{y_{1(t-1)}<r\}}, I_{\{y_{2(t-1)}<r\}}, \cdots, I_{\{y_{N(t-1)}<r\}}\}$. We have the final *TNAR* model:

$$\mathbb{Y}_t = \mathcal{B}_0 + \mathbb{G}_{t-1}\mathbb{Y}_{t-1} + \varepsilon_t, \tag{3.4}$$

where $\mathbb{G}_{t-1} = \beta_1 W + \beta_2 I + \beta_3 J_{t-1}$.

3.2.2 Strict Stationarity: Type I

In order to do statistical inference for the *TNAR* model, we first need to show when the model has a unique stationary solution. According to whether N is fixed or $N \to \infty$, we can define two different types of stationary with type I (N is fixed) and type II ($N \to \infty$).

For Type I stationarity, the common method to show whether nonlinear time series is stationary is to represent time series as a Markov chain and show that the Markov chain is ergodic. We will use the following Lemma 3.2.1 that shows the Markov chain is ergodic under this sufficient condition and Lemma 3.2.2 is also used in the proof of Theorem 3.2.1.

Lemma 3.2.1. *Assume that $\{\mathbb{Y}_t\}$ is an aperiodic ϕ-irreducible Markov chain and let g be a nonnegative measurable function. Then $\{\mathbb{Y}_t\}$ is geometrically ergodic if there exists a small set C and constants $\lambda_1 > 0, \lambda_2 > 0, 0 < \lambda < 1$ such that*

$$\begin{aligned} (i) \quad & E\{g(\mathbb{Y}_t)|\mathbb{Y}_{t-1} = Y\} \leq \lambda g(Y) - \lambda_1, \ \textit{for any } Y \notin C; \\ (ii) \quad & E\{g(\mathbb{Y}_t)|\mathbb{Y}_{t-1} = Y\} \leq \lambda_2, \ \textit{for any } Y \in C \end{aligned}$$

This lemma is called drift-criteria for the geometric ergodicity of a Markov chain, which comes from [45] (see also [35]). In order to get the ergodicity of the TNAR model, we need an assumption on error terms:

Assumption 3.2.1. *$\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})^\top$ are independent and distributed with positive density functions and finite fourth moments.*

The following Lemma 3.2.2, which can be easily obtained by the usual method in Markov chian theory, is also used in proof of Theorem 3.2.1.

Lemma 3.2.2. *Under assumption 3.2.1,$\{\mathbb{Y}_t\}$ is an aperiodic ϕ-irreducible Markov chain and every bounded compact set with positive Lebesgue measure is a small set.*

Hence, we are now ready to state the following theorem that the *TNAR* model is ergodic and hence stationary when $|\beta_1| + \max\{|\beta_2|, |\beta_2 + \beta_3|\} < 1$.

Theorem 3.2.1. *If $\rho = |\beta_1| + max\{|\beta_2|, |\beta_2 + \beta_3|\} < 1$, then the stochastic process $\{\mathbb{Y}_t\}$ defined by model (3.3) is geometrically ergodic, and hence it has a unique stationary distribution as*

$$\mathbb{Y}_t = \sum_{j=0}^{\infty} \Pi_j(\mathcal{B}_0 + \varepsilon_{t-j}), \tag{3.5}$$

where $\Pi_j = \prod_{i=1}^{j} \mathbb{G}_{t-i}$ and $\Pi_0 = I_N$

Proof. First, we prove that $\mathbb{Y}_t$ has the following form of solution. According to the difinition of equation 3.4, that is, $\mathbb{G}_{t-1} = \beta_1 W + \beta_2 I + \beta_3 J_{t-1}$,

$$\begin{aligned}
\mathbb{Y}_t &= \mathcal{B}_0 + \mathbb{G}_{t-1}\mathbb{Y}_{t-1} + \varepsilon_t \\
&= \mathcal{B}_0 + \mathbb{G}_{t-1}(\mathcal{B}_0 + \mathbb{G}_{t-2}\mathbb{Y}_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \mathcal{B}_0 + \mathbb{G}_{t-1}\mathcal{B}_0 + \mathbb{G}_{t-1}\mathbb{G}_{t-2}\mathbb{Y}_{t-2} + \mathbb{G}_{t-1}\varepsilon_{t-1} + \varepsilon_t \\
&\cdots \\
&= \sum_{j=1}^{\infty}(\prod_{i=1}^{j} \mathbb{G}_{t-i}\mathcal{B}_0 + \prod_{i=1}^{j} \mathbb{G}_{t-i}\varepsilon_{t-j}) + \mathcal{B}_0 + \varepsilon_t \\
&= \sum_{j=0}^{\infty} \Pi_j(\mathcal{B}_0 + \varepsilon_{t-j}),
\end{aligned}$$

where $\Pi_j = \mathbb{G}_{t-1}\mathbb{G}_{t-2}\cdots\mathbb{G}_{t-j}$ for $j > 0$ and $\Pi_0 = I_N$.

Second, we apply Lemma 3.2.1 to prove the model (3.3) is geometrically ergodic. Define a norm by

$$\|X\|^2 = \sum_{i=1}^{N} x_i^2 \quad for \quad X = (x_1, \cdots, x_N)^\top \in \mathbb{R}^N. \tag{3.6}$$

Let $g(Y) = \|Y\|$, where $Y = (y_1, \cdots, y_N) \in \mathbb{R}^N$. We have

$$\begin{aligned} E\{g(\mathbb{Y}_t) \mid \mathbb{Y}_{t-1} = Y\} &= E\{\|\mathcal{B}_0 + GY + \varepsilon_t\|\} \\ &\leq E\{\|\mathcal{B}_0\| + \|GY\| + \|\varepsilon_t\|\} \\ &= E\|\mathcal{B}_0\| + E\|GY\| + E\|\varepsilon_t\| \\ &\leq E\|\mathcal{B}_0\| + (|\beta_1| + \max\{|\beta_2|, |\beta_2 + \beta_3|\})\|Y\| + E\|\varepsilon_t\| \end{aligned}$$

since

$$\begin{aligned} E\|GY\| &= E\|\left(\beta_1 W + \beta_2 I + \beta_3 J_{t-1}\right) Y\| \\ &\leq E\|\beta_1 WY\| + E\|\beta_2 IY + \beta_3 J_{t-1} Y\| \\ &\leq |\beta_1| E\|Y\| + \max\{|\beta_2|, |\beta_2 + \beta_3|\} E\|Y\| \\ &= (|\beta_1| + \max\{|\beta_2|, |\beta_2 + \beta_3|\})\|Y\| \end{aligned}$$

and note that

$$\begin{aligned} \|WY\|^2 &= (WY)^T (WY) \\ &\leq Y^T (W^T W) Y \\ &\leq \rho(W^T W) Y^T Y \end{aligned}$$

which implies

$$\|WY\| \leq \|Y\|.$$

This is because $W^T W$ is a symmetric matrix and its maximum eigenvalue $\rho(W^T W) \leq \rho(W^T)\rho(W) \leq \rho(W) = 1$. The spectral radius of W in Appendix A.1 are applied here. Notice we have $E\|\varepsilon_t\| = \sqrt{\sum_{i=1}^N E(\varepsilon_{it}^2)} = \sqrt{N\sigma^2} = \sqrt{N}\sigma < \infty$. Let $\rho = |\beta_1| + \max\{|\beta_2|, |\beta_2 + \beta_3| < 1$ and take λ and M such that $0 < \rho < \lambda < 1$ and

$$M > \frac{E\|\mathcal{B}_0\| + \sqrt{N}\sigma}{\lambda - \rho}, \tag{3.7}$$

where $E\|\mathcal{B}_0\| < \infty$. Denote $C = \{Y : \|Y\| \leq M\}$. By Lemma 3.2.2, C is a small set.

When $\|Y\| > M$, that is, $Y \notin C$, we have

$$E\{g(\mathbb{Y}_t) | \mathbb{Y}_{t-1} = Y\} \leq \lambda g(Y) - \lambda_1, \tag{3.8}$$

where $\lambda_1 = (\lambda - \rho)M - E\|\mathcal{B}_0\| - \sqrt{N}\sigma > 0$.

For any $Y \in C$,

$$E\{g(\mathbb{Y}_t)|\mathbb{Y}_{t-1} = Y\} \leq \lambda_2 \tag{3.9}$$

where $\lambda_2 = E\|\mathcal{B}_0\| + \rho M + \sqrt{N}\sigma > 0$.

By Lemma 3.2.1, $\{\mathbb{Y}_t\}$ is geometrically ergodic, and hence it has a unique stationary distribution. $\square$

3.2.3 Strict Stationarity: Type II

Next we show the Type II stationarity. [50] provided a reasonable definition of type II stationarity.

Definition 3.2.1. *Let $\mathbb{Y}_t \in \mathbb{R}^N$ be a N-dimensional time series with $N \to \infty$. Define $\mathcal{W} = \{\omega \in \mathbb{R}^\infty : \sum |\omega_i| < \infty\}$, where $\omega = (\omega_i \in \mathbb{R}^1 : 1 \leq i \leq \infty)^\top \in \mathbb{R}^\infty$. For each $\omega \in \mathcal{W}$, let $\mathbf{w}_N = (\omega_1, \cdots, \omega_N)^\top \in \mathbb{R}^N$ be the truncated N-dimensional vector. $\{\mathbb{Y}_t\}$ is then said to be strictly stationary, if it satisfies the following conditions: for any $\omega \in \mathcal{W}$, (1) $Y_t^\omega = \lim_{N\to\infty} \mathbf{w}_N^\top \mathbb{Y}_t$ exists in the almost sure sense; and (2) $\{Y_t^\omega\}$ is strictly stationary.*

We then have the following theorem for the TNAR model.

Theorem 3.2.2. *Assume the same conditions as in Theorem 3.2.1 with $N \to \infty$. Then TNAR model has a unique strictly type II stationary solution with finite first order moment.*

Proof. To prove the existence of a stationary solution, it is sufficient to show that $\{\mathbb{Y}_t\}$ is strictly stationary according to Definition 3.2.1.

Define $|M|_e$ as $|M|_e = (|m_{ij}|) \in \mathbb{R}^{n\times p}$ for any arbitrary matrix $M = (m_{ij}) \in \mathbb{R}^{n\times p}$. Moreover, for matrices $M_1 = (m_{ij}^{(1)}) \in \mathbb{R}^{n\times p}$ and $M_2 = (m_{ij}^{(2)}) \in \mathbb{R}^{n\times p}$, define $M1 \preceq M2$ as $m_{ij}^{(1)} \leq m_{ij}^{(2)}$ for $1 \leq i \leq n$ and $1 \leq j \leq p$.

Let β^{max} denote $max(|\beta_2|, |\beta_2 + \beta_3|)$ and define $G^{max} = |\beta_1|W + \beta^{max}I$. Then we have

$$|G|_e \preceq |G^{max}|_e \tag{3.10}$$

Now we want to prove

$$E|\mathbf{w_N}^\top \mathbb{Y}_\mathbf{t}| < \infty,$$

which implies that $\lim_{N\to\infty} \mathbf{w_N}^\top \mathbb{Y}_\mathbf{t}$ exists. We have

$$|\mathbf{w_N}^\top \mathbb{Y}_\mathbf{t}| \leq \sum_\mathbf{i} |\omega_\mathbf{i}| \sum_\mathbf{i} |\mathbf{y_{it}}| \tag{3.11}$$

and

$$E|\mathcal{B}_0 + \varepsilon_{t-j}|_e \preceq C\mathbf{1}, \tag{3.12}$$

where $C = |\beta_0| + E|Z_i^\top \gamma| + E|\varepsilon_{it}|$ and

$$|G^{max}|_e^j = (|\beta_1| W + \beta^{max} I)^j \mathbf{1}. \tag{3.13}$$

Hence, we have

$$\begin{aligned} E|\mathbf{w_N}^\top \mathbb{Y}_\mathbf{t}| &= E|\mathbf{w_N}^\top \sum_{j=1}^{\infty} (\prod_{i=1}^{j} \mathbb{G}_{t-i})(\mathcal{B}_0 + \varepsilon_{t-j})| \\ \text{by (3.11)} \quad &\leq \sum |\omega_i| \sum_{j=0}^{\infty} E|(\prod_{i=1}^{j} \mathbb{G}_{t-i})(\mathcal{B}_0 + \varepsilon_{t-j})| \\ &\leq \sum |\omega_i| \sum_{j=0}^{\infty} (\prod_{i=1}^{j} |\mathbb{G}_{t-i}|) E|(\mathcal{B}_0 + \varepsilon_{t-j})| \\ \text{by (3.10 and 3.12)} \quad &\leq \sum |\omega_i| \sum_{j=0}^{\infty} |G^{max}|_e^j C\mathbf{1} \\ \text{by (3.13)} \quad &= C \sum |\omega_i| \sum_{j=0}^{\infty} (|\beta_1| + \beta^{max})^j < \infty, \end{aligned}$$

which implies that $\lim_{N\to\infty} \mathbf{w}_N^\top \mathbb{Y}_t$ exists with probability one. Let $Y_t^\omega = \lim_{N\to\infty} \mathbf{w}_N^\top \mathbb{Y}_t$, and it is obvious that $\{Y_t^\omega\}$ is strictly stationary. Hence, $\{\mathbb{Y}_t\}$ is strictly stationary according to Definition 3.2.1.

Next, we verify the uniqueness of the strictly stationary solution. Assume that $\{\widetilde{\mathbb{Y}}_t\}$ is another strictly stationary solution to the *TNAR* model with finite first order moment. Therefore, $E|\widetilde{\mathbb{Y}}_t|_e \preceq C_1\mathbf{1}$ for some constant C_1.

Then we have

$$E|\mathbf{w}_N^\top \mathbb{Y}_t - \mathbf{w}_N^\top \widetilde{\mathbb{Y}}_t| = E\left|\sum_{j=m}^{\infty} \mathbf{w_N}^\top (\prod_{\mathbf{i=1}}^{\mathbf{j}} \mathbb{G}_\mathbf{t-i})(\mathcal{B}_\mathbf{0} + \varepsilon_\mathbf{t-j}) - \mathbf{w_N}^\top (\prod_{\mathbf{i=1}}^{\mathbf{m}} \mathbb{G}_\mathbf{t-i}) \widetilde{\mathbb{Y}}_\mathbf{t-m}\right|$$

$$\begin{aligned}
&\leq \sum_i |\omega_i| E \left| \sum_{j=m}^{\infty} (\prod_{i=1}^{j} \mathbb{G}_{t-i})(\mathcal{B}_0 + \varepsilon_{t-j}) - \prod_{i=1}^{m} \mathbb{G}_{t-i} \widetilde{\mathbb{Y}}_{t-m} \right| \\
&\leq \sum_i |\omega_i| \left(\sum_{j=m}^{\infty} |G^{max}|^j E|\mathcal{B}_0 + \varepsilon_{t-j}| + |G^{max}|_e^m E|\widetilde{\mathbb{Y}}_{t-m}| \right) \\
&\leq \sum_i |\omega_i| \left(C \sum_{j=m}^{\infty} (|\beta_1| + \beta^{max})^j + C_1(|\beta_1| + \beta^{max})^m \right)
\end{aligned}$$

for any N and weight ωv. Consequently, by the arbitrary specification of m, we have $\mathbb{Y}^\omega = \widetilde{\mathbb{Y}}_t^\omega$ with probability one. This completes the proof of Theorem 3.2.2. □

3.3 GMM Parameter Estimation

Let $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^\top \in \mathbb{R}^3$ and $\theta = (\theta_j^\top) = (\beta^\top, \gamma^\top)^\top \in \mathbb{R}^{p+4}$. In order to estimate the unknown parameter θ, we rewrite the *TNAR* model (3.2) as

$$\begin{aligned}
y_{it} &= \beta_0 + Z_i^\top \gamma + \beta_1 w_i^\top \mathbb{Y}_{t-1} + \beta_2 y_{i(t-1)} + \beta_3 I_{\{y_i(t-1)<r\}} y_{i(t-1)} + \varepsilon_{it} \\
&= X_{i(t-1)}^\top \theta + \varepsilon_{it} \qquad (3.14)
\end{aligned}$$

where $X_{i(t-1)} = (1, w_i^\top \mathbb{Y}_{t-1}, y_{i(t-1)}, I_{\{y_i(t-1)<r\}} y_{i(t-1)}, Z_i^\top)^\top \in \mathbb{R}^{p+4}$ and $w_i = (a_{ij}/n_i : 1 \leq j \leq N)^\top \in \mathbb{R}^N$ is the ith row vector of W. We denote $\mathbb{X}_t = (X_{1t}, X_{2t}, \cdots, X_{Nt})^\top = (\mathbf{1}, W\mathbb{Y}_t, \mathbb{Y}_t, J_t\mathbb{Y}_t, \mathbb{Z}) \in \mathbb{R}^{N\times(p+4)}$. Then model (3.2) can be rewritten in vector form as $\mathbb{Y}_t = \mathbb{X}_{t-1}^\top \theta + \varepsilon_t$. The parameters in the *TNAR* model can be estimated by the *generalized method of moments* (*GMM*) [22].

3.3.1 General Idea of GMM Method

GMM was originally proposed by [22]. It has become an important unified framework for estimation and inference in econometrics. It is a powerful moment-based parameter estimation method. To obtain explicit expressions of unknown parameters from a set of moment conditions, GMM performs parameter estimation by minimizing the weighted distance between the overall moment and the sample moment. The key advantage of GMM is that it only needs to specify some specific moment conditions instead of all densities. The most famous implementation of the GMM method is the Hansen two-step algorithm, which is an iterative regression process proposed by [22] in his original GMM paper. In this section, we will outline the process of parameter estimation of TNAR model using GMM according to the Chapter 14 in Time Series Analysis by [18] (p. 410-413).

$\mathbb{Y}_t$ is a $N \times 1$ vector and θ is an unknown $(p+4) \times 1$ vector of coefficients, $h(\theta, \mathbb{Y}_t)$ can be viewed as an $(p+4) \times 1$ vectored-valued function. Let θ_0 denote the true value and suppose we have $E\{h(\theta_0, \mathbb{Y}_t)\} = 0$. Let $\mathfrak{Y}_{T,N} = (\mathbb{Y}_T^\top, \mathbb{Y}_{T-1}^\top, \cdots, \mathbb{Y}_1^\top)^\top$ be a $(TN \times 1)$ vector and let the vector-valued function $g(\theta, \mathfrak{Y}_{T,N}) \in \mathbb{R}^{(p+4)\times 1}$ denote the sample average of $h(\theta, \mathbb{Y}_t)$

$$g(\theta, \mathfrak{Y}_{T,N}) = (1/T) \sum_{t=1}^{T} h(\theta, \mathbb{Y}_t). \tag{3.15}$$

GMM estimates $\hat{\theta}$ for real parameters θ is the minimum orthogonality condition:

$$Q(\theta, \mathfrak{Y}_{T,N}) = g(\theta, \mathfrak{Y}_{T,N})^\top W g(\theta, \mathfrak{Y}_{T,N}),$$

where W is an weighting matrix. Here, we set $W = \mathrm{I}$ (Identity matrix) to assign the same weight to all moment conditions. We can solve the GMM estimation θ starting from the well-known least squares (LS) problem:

$$\theta = \arg\min_{\theta} g(\theta, \mathfrak{Y}_{T,N})^\top g(\theta, \mathfrak{Y}_{T,N}).$$

By setting the different moment condition $h(\theta)$, we will provide two special cases to show how to use GMM.

3.3.2 Special Case 1: Ordinary Least Squares

The critical assumption needed to justify OLS regression is that the regression residual ε_t is uncorrelated with the explanatory variables:

$$h(\theta, \mathbb{Y}_t) = \mathbb{X}_{t-1}(\mathbb{Y}_t - \mathbb{X}_{t-1}^\top \theta). \tag{3.16}$$

Hence, the *GMM* estimate of θ is the solution of the following equation:

$$g(\theta, \mathfrak{Y}_{N,T}) = (1/NT) \sum_{t=2}^{T} h(\theta, \mathbb{Y}_t) = (1/NT) \sum_{t=2}^{T} \mathbb{X}_{t-1}(\mathbb{Y}_t - \mathbb{X}_{t-1}^\top \theta) = \mathbf{0}. \tag{3.17}$$

Hence, the ordinary least squares type estimator can be obtained as

$$\hat{\theta}_{OLS} = \left(\sum_{t=1}^{T} \mathbb{X}_{t-1} \mathbb{X}_{t-1}^\top \right)^{-1} \sum_{t=1}^{T} \mathbb{X}_{t-1} \mathbb{Y}_t. \tag{3.18}$$

3.3.3 Special Case 2

Another assumption is that the regression residual ε_t is uncorrelated with the previous value of $\mathbb{Y}_t$. Take

$$h(\theta, \mathbb{Y}_t) = \mathbb{G}_{t-1}(\mathbb{Y}_t - \mathbb{X}_{t-1}^\top \theta) \tag{3.19}$$

where $\mathbb{G}_{t-1} = (\mathbb{Y}_{t-1}, \mathbb{Y}_{t-2}, \cdots, \mathbb{Y}_{t-p-4})^\top$. The *GMM* estimate of θ is the solution of the following equation:

$$g(\theta, \mathfrak{Y}_{N,T}) = (1/NT) \sum_{t=1}^{T} \mathbb{G}_{t-1}(\mathbb{Y}_t - \mathbb{X}_{t-1}^\top \theta) = \mathbf{0}. \tag{3.20}$$

Hence, the estimator of the second case can be obtained as

$$\hat{\theta}_{N,T} = \left(\sum_{t=1}^{T} \mathbb{G}_{t-1} \mathbb{X}_{t-1}^\top \right)^{-1} \sum_{t=1}^{T} \mathbb{G}_{t-1} \mathbb{Y}_t. \tag{3.21}$$

3.4 Asymptotic Distribution of GMM Estimator

Given certain moment conditions, the corresponding estimator can be obtained. However, how to select the best moment conditions is a problem. In our opinions, the smaller the asymptotic variance, the better the moment condition. Therefore, the next thing is to find the asymptotic distribution of the GMM estimator when $min\{N, T\} \to \infty$. We have the following proposition.

Proposition 3.4.1. *Let $g(\theta; \mathfrak{Y}_{N,T})$ be differentiable in θ for all $\mathfrak{Y}_{N,T}$, and $\hat{\theta}_{N,T}$ be the GMM estimator. Let $\{\hat{S}_{N,T}\}_{T=1}^{\infty}$ be a sequence of positive definite $r \times r$ matrices such that $\hat{S}_{N,T} \xrightarrow{P} S$. Suppose, further, that the following hold:*

(a) $\hat{\theta}_{N,T} \xrightarrow{P} \theta_0$;

(b) $\sqrt{NT} g(\theta_0; \mathfrak{Y}_{N,T}) \xrightarrow{d} N(0, S)$;

(c) *for any sequence* $\{\theta_{N,T}^*\}_{N,T=1}^{\infty}$ *satisfying* $\theta_{N,T}^* \xrightarrow{P} \theta_0$

$$\lim \left\{ \left. \frac{\partial g(\theta; \mathfrak{Y}_{N,T})}{\partial \theta^\top} \right|_{\theta = \theta_{N,T}^*} \right\} = \lim \left\{ \left. \frac{\partial g(\theta; \mathfrak{Y}_{N,T})}{\partial \theta^\top} \right|_{\theta = \theta_0} \right\} = D^\top$$

in probability, then

$$\sqrt{NT}(\hat{\theta}_{N,T} - \theta_0) \xrightarrow{d} N(0, V) \tag{3.22}$$

where $V = \{DS^{-1}D^\top\}^{-1}$

Proof. Let $g_i(\theta;\mathfrak{Y}_{N,T})$ denote the ith element of $g(\theta;\mathfrak{Y}_{N,T})$, so that $g_i : R^a \to R^1$. By the mean-value theorem,

$$g_i(\hat{\theta}_{N,T}) = g_i(\theta_0;\mathfrak{Y}_{N,T}) + [d_i(\theta^*_{i,N,T};\mathfrak{Y}_{N,T})]^\top(\hat{\theta}_{N,T} - \theta_0) \tag{3.23}$$

where

$$d_i(\theta^*_{i,N,T};\mathfrak{Y}_{N,T}) = \left.\frac{\partial g_i(\theta;\mathfrak{Y}_{N,T})}{\partial\theta^\top}\right|_{\theta=\theta^*_{i,N,T}}$$

for some $\theta^*_{i,N,T}$ between θ_0 and $\hat{\theta}_{N,T}$. Define

$$D^\top_{N,T} = \begin{bmatrix} [d_1(\theta^*_{1,N,T};\mathfrak{Y}_{N,T})]^\top \\ [d_2(\theta^*_{1,N,T};\mathfrak{Y}_{N,T})]^\top \\ \vdots \\ [d_r(\theta^*_{1,N,T};\mathfrak{Y}_{N,T})]^\top \end{bmatrix}. \tag{3.24}$$

Stacking the equations in an $(r \times 1)$ vector produces

$$g(\hat{\theta}_{N,T};\mathfrak{Y}_{N,T}) = g(\theta_0;\mathfrak{Y}_{N,T}) + D^\top_{N,T}(\hat{\theta}_{N,T} - \theta_0). \tag{3.25}$$

If both sides of (3.25) are multiplied by the matrix

$$\left\{\left.\frac{\partial g(\theta;\mathfrak{Y}_{N,T})}{\partial\theta^\top}\right|_{\theta=\hat{\theta}_{N,T}}\right\}^\top \times \hat{S}^{-1}_{N,T}, \tag{3.26}$$

the result is

$$\begin{aligned}
&\left\{\left.\frac{\partial g(\theta;\mathfrak{Y}_{N,T})}{\partial\theta^\top}\right|_{\theta=\hat{\theta}_{N,T}}\right\}^\top \times \hat{S}^{-1}_{N,T} \times [g(\hat{\theta}_{N,T};\mathfrak{Y}_{N,T})] \\
&= \left\{\left.\frac{\partial g(\theta;\mathfrak{Y}_{N,T})}{\partial\theta^\top}\right|_{\theta=\hat{\theta}_{N,T}}\right\}^\top \times \hat{S}^{-1}_{N,T} \times [g(\theta_0;\mathfrak{Y}_{N,T})] \\
&+ \left\{\left.\frac{\partial g(\theta;\mathfrak{Y}_{N,T})}{\partial\theta^\top}\right|_{\theta=\hat{\theta}_{N,T}}\right\}^\top \times \hat{S}^{-1}_{N,T} \times D^\top_{N,T}(\hat{\theta}_{N,T} - \theta_0).
\end{aligned} \tag{3.27}$$

(14.1.22) in [18] shows the left side of (3.27) equals zero, so

$$\hat{\theta}_{N,T} - \theta_0 = -\left[\left\{\left.\frac{\partial g(\theta;\mathfrak{Y}_{N,T})}{\partial\theta^\top}\right|_{\theta=\hat{\theta}_{N,T}}\right\}^\top \times \hat{S}^{-1}_{N,T} \times D^\top_{N,T}\right]^{-1}$$

$$\times \left\{ \left. \frac{\partial g(\theta; \mathfrak{Y}_{N,T})}{\partial \theta^\top} \right|_{\theta=\hat{\theta}_{N,T}} \right\}^\top \times \hat{S}_{N,T}^{-1} \times [g(\theta_0; \mathfrak{Y}_{N,T})].$$

Now, $\theta^*_{i,N,T}$ in (3.23) is between θ_0 and $\hat{\theta}_{N,T}$, so that $\theta^*_{i,N,T} \xrightarrow{P} \theta_0$ for each i. Each row of $D^\top_{N,T}$ converges in probability to the corresponding row of $D^\top$

$$\sqrt{NT}(\hat{\theta}_{N,T} - \theta_0) = -\left\{DS^{-1}D^\top\right\}^{-1} \times \left\{DS^{-1}\sqrt{NT}g(\theta_0; \mathfrak{Y}_{N,T})\right\}(1+o_p(1)) \quad (3.28)$$

Define

$$C = -\left\{DS^{-1}D^\top\right\}^{-1} \times DS^{-1},$$

then (3.28) becomes

$$\sqrt{NT}(\hat{\theta}_{N,T} - \theta_0) = C\sqrt{NT}g(\theta_0; \mathfrak{Y}_{N,T})(1 + o_p(1)). \quad (3.29)$$

Since

$$\sqrt{NT}g(\theta_0; \mathfrak{Y}_{N,T}) \xrightarrow{d} N(0, S), \quad (3.30)$$

therefore

$$\sqrt{NT}(\hat{\theta}_{N,T} - \theta_0) \xrightarrow{d} N(0, V), \quad (3.31)$$

where

$$V = CSC^\top = \left\{DS^{-1}D^\top\right\}^{-1} \times DS^{-1} \times S \times S^{-1}D^\top\{DS^{-1}D^\top\}^{-1} = \left\{DS^{-1}D^\top\right\}^{-1}. \quad (3.32)$$

□

By Proposition 3.4.1, we know that $\hat{\theta}_{N,T}$ is $\sqrt{NT}$-consistent with asymptotic variance V. This is of great importance for performance analysis and comparison of large sample size scenarios. Then we apply the proposition into two special cases.

Before proving the asymptotic properties of the estimated parameters, we need to follow the assumptions as in [50].

- **Assumption 2.** (Node assumption) Assume $\rho < 1$, where ρ is defined in Theorem 3.2.1. Moreover, assume that the Z_i's are independent and identically distributed random vectors, with mean 0 and covariance $\Sigma_z \in \mathbb{R}^{p\times p}$. In addition, its fourth-order moment is finite. ϵ_{it} has the same assumption for every i and t. In addition, we need $\{Z_i\}$ and $\{\epsilon_{it}\}$ to be independent of each other.

- **Assumption 3.** (Network structure)

  A3.1) (Connectivity) Think of W as a probability transfer matrix whose state space is defined as all nodes in the network. We assume that this Markov chain is irreducible and aperiodic. Also define $\pi = (\pi_i) \in \mathbb{R}^N$ to be the stationary distribution of this Markov chain, which has the following properties, a) $\pi_i \geq 0$, b) $\sum_i \pi_i = 1$ and c) $\pi = W^\top \pi$. Further, $\sum_{i=1}^N \pi_i^2 \to 0$ as $N \to \infty$.

  A3.2) (Sparsity) Assume $|\lambda_1(W + W^\top)| = O(\log(N))$, where $\lambda_1(\cdot)$ is the largest eigenvalue of a matrix.

- **Assumption 4.** (Law of Large Numbers) Denote Σ_Y is the covariance matrix of $\mathbb{Y}_t$. Define $Q = (I - G)^{-1}(I - G^\top)^{-1}$ and $G = \beta_1 W + \beta_2 I$. Assume that the following limits exist: $k_1 = \lim\limits_{N\to\infty} N^{-1} tr(\Sigma_Y)$, $k_2 = \lim\limits_{N\to\infty} N^{-1} tr(W\Sigma_Y)$, $k_3 = \lim\limits_{N\to\infty} N^{-1} tr\{(I - G)^{-1}\}$, $k_4 = \lim\limits_{N\to\infty} N^{-1} tr\{Q\}$. Here k_1, k_2, k_3 and k_4 are fixed constants. In addition, different from the Zhu's assumptions, the following limits are also assumed to exist: $c_\beta^- = \lim\limits_{N\to\infty} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T Y_{i(t-1)}^-$, $\Sigma_2^- = \lim\limits_{N\to\infty} (NT)^{-1} \sum_{t=1}^T J_{t-1} \mathbb{Y}_{t-1}^\top W \mathbb{Y}_{t-1}$, $\Sigma_3^- = \lim\limits_{N\to\infty} (NT)^{-1} \sum_{t=1}^T J_{t-1} \mathbb{Y}_{t-1}^\top \mathbb{Y}_{t-1}$ and $\Sigma_4^- = \lim\limits_{N\to\infty} (NT)^{-1} \sum_{t=1}^T J_{t-1} \mathbb{Y}_{t-1}^\top \mathbb{Z}$.

Remark 2. The above assumptions can be explained as follows:

Assumption 2 is the basic assumption of the node covariates Z_i and the noise term ε_{it}, so that the law of large numbers and the central limit theorem can be used. In fact, Assumption 2 can be relaxed to be weak dependent as long as the law of large numbers and the central limit theorem hold.

Assumption 3 is about the network structure. (A3.1) ensures that all nodes can reach each other within a limited number of steps, such that the network is irreducible. (A3.2) ensures that the network is sufficiently sparse, so that the divergence rate of λ_1 can be controlled by $log(N)$.

Assumption 4 is used to apply the law of large numbers, from which the asymptotic covariance matrix is derived. Consider for example the first condition in Assumption 4, that is $k_1 = \lim\limits_{N\to\infty} N^{-1} tr(\Sigma_Y) = \lim\limits_{N\to\infty} N^{-1} \sum_{i=1}^N var^*(Y_{it})$, where $var^*(\cdot) = var(\cdot|\mathbb{Z})$ and $\mathbb{Z}$ is about the nodal information. With the help of Assumption 4, the following limits can be verified to exist: $k_5 = \lim\limits_{N\to\infty} N^{-1} tr(WQW^\top)$,

$k_6 = \lim_{N\to\infty} N^{-1}tr(W\Sigma_Y W^\top)$, $k_7 = \lim_{N\to\infty} N^{-1}tr\{WQ\}$, $k_8 = \lim_{N\to\infty} N^{-1}tr\{W(I - G)^{-1}\}$.

Now under these assumptions, we get asymptotic normality for the following two special cases when we apply the Proposition 3.4.1.

3.4.1 Special Case 1: Ordinary Least Squares

As a result, an ordinary least squares type estimator can be obtained by Corollary 3.4.1. Before that, Lemma 3.4.1 is also needed.

Lemma 3.4.1. *Assuming the stationarity condition $\rho < 1$, and the Assumption 2-4 hold. Set $\hat{S} = (NT)^{-1}\sum_{t=1}^{T}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^\top$, $Y_{i(t-1)}^- = I_{\{Y_{i(t-1)}<r\}}Y_{i(t-1)}$ and $\mathbb{Y}_{t-1}^- = (I_{\{Y_{1(t-1)}<r\}}Y_{1(t-1)}), \cdots, I_{\{Y_{N(t-1)}<r\}}Y_{N(t-1)})^\top$. Then we have*

$$\hat{S} \xrightarrow{P} S = \begin{pmatrix} 1 & c_\beta & c_\beta & c_\beta^- & 0_p^\top \\ & \Sigma_1 & \Sigma_2 & \Sigma_2^- & k_8\Sigma_z\gamma \\ & & \Sigma_3 & \Sigma_3^- & k_3\Sigma_z\gamma \\ & & & \Sigma_3^- & \Sigma_4^- \\ & & & & \Sigma_z \end{pmatrix},$$

as $\min\{N, T\} \to \infty$. Here $c_\beta = \beta_0(1-\beta_1-\beta_2)^{-1}, \Sigma_1 = c_\beta^2 + k_5\gamma^\top\Sigma_z\gamma + k_6, \Sigma_2 = c_\beta^2 + k_7\gamma^\top\Sigma_z\gamma + k_2, \Sigma_3 = c_\beta^2 + k_4\gamma^\top\Sigma_z\gamma + k_1$ and $0_p = (0, \cdots, 0)^\top$ is a vector with p dimension.

Proof. Recall that $\hat{S}$ is a symmetric matrix, so we only calculate the upper triangle of $\hat{S}$.

$$\begin{aligned} \hat{S} &= (NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^\top \\ &= \begin{pmatrix} 1 & S_{12} & S_{13} & S_{14} & S_{15} \\ & S_{22} & S_{23} & S_{24} & S_{25} \\ & & S_{33} & S_{34} & S_{35} \\ & & & S_{44} & S_{45} \\ & & & & S_{55} \end{pmatrix}, \end{aligned}$$

where $S_{12} = (NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} w_i^\top\mathbb{Y}_{t-1}, S_{13} = (NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{i(t-1)}$, $S_{14} = (NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{i(t-1)}^-, S_{15} = N^{-1}\sum_{i=1}^{N} Z_i^\top, S_{22} =$

$(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}(w_i^\top \mathbb{Y}_{t-1})^2$,
$S_{23}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}w_i^\top \mathbb{Y}_{t-1}Y_{i(t-1)}, S_{24}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}w_i^\top \mathbb{Y}_{t-1}Y_{i(t-1)}^-$,
$S_{25}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}w_i^\top \mathbb{Y}_{t-1}Z_i^\top, S_{33}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}^2$,
$S_{34}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}Y_{i(t-1)}^-, S_{35}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}Z_i^\top$,
$S_{44}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{i(t-1)}^-)^2, S_{45}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}^-Z_i^\top$,
$S_{55}=N^{-1}\sum_{i=1}^{N}Z_iZ_i^\top$.

From the [50]'s work, $S_{12}\xrightarrow{P}c_\beta, S_{13}\xrightarrow{P}c_\beta, S_{13}\xrightarrow{P}0_p^\top, S_{22}\xrightarrow{P}\Sigma_1, S_{23}\xrightarrow{P}\Sigma_2, S_{25}\xrightarrow{P}k_8\gamma^\top\Sigma_z, S_{33}\xrightarrow{P}\Sigma_3, S_{35}\xrightarrow{P}k_3\gamma^\top\Sigma_z$ and $S_{55}\xrightarrow{P}\Sigma_z$.

Next, we prove the convergence of the remaining five elements in $\hat{S}$ one by one.

Step 1.1 Convergence of S_{14}. Note that

$$S_{14}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}^-\xrightarrow{P}c_\beta^-.$$

Step 1.2 Convergence of S_{24}. Note that

$$S_{24}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}w_i^\top\mathbb{Y}_{t-1}Y_{i(t-1)}^-=(NT)^{-1}\sum_{t=1}^{T}J_{t-1}\mathbb{Y}_{t-1}^\top W\mathbb{Y}_{t-1}\xrightarrow{P}\Sigma_2^-.$$

Step 1.3 Convergence of S_{34}. Note that

$$S_{34}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}Y_{i(t-1)}^-=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}^2I_{\{Y_{i(t-1)}<r\}}\xrightarrow{P}\Sigma_3^-.$$

Step 1.4 Convergence of S_{44}. Note that

$$S_{44}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}^-Y_{i(t-1)}^-=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}^2I_{\{Y_{i(t-1)}<r\}}=S_{34}\xrightarrow{P}\Sigma_3^-.$$

Step 1.5 Convergence of S_{45}. Note that

$$S_{45}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}Y_{i(t-1)}^-Z_i^\top\xrightarrow{P}\Sigma_4^-.$$

This completes the proof. □

Corollary 3.4.1. *(Ordinary Least Squares) Under the assumption of Theorem 3.2.2*

$$\sqrt{NT}(\hat{\theta}_{OLS}-\theta_0)\xrightarrow{d}N(0,S^{-1}) \tag{3.33}$$

where

$$S=\lim_{T\to\infty,N\to\infty}(NT)^{-1}\sum_{t=1}^{T}\mathbb{X}_{t-1}\mathbb{X}_{t-1}^\top. \tag{3.34}$$

Proof. First of all, the GMM estimator is a consistent estimator of θ_0, which meets the first condition of proposition 3.4.1.

Secondly, we need to prove the second condition of proposition 3.4.1

$$\sqrt{NT}g(\theta_0, \mathfrak{Y}_{N,T}) \xrightarrow{d} N(0, S)$$

We know

$$g(\theta, \mathfrak{Y}_{N,T}) = (1/NT)\sum_{t=1}^{T} X_{t-1}(Y_t - X_{t-1}^\top \theta) \tag{3.35}$$

To prove the result, it suffices to show that

$$\sqrt{NT}\eta^\top g(\theta_0, \mathfrak{Y}_{N,T}) = (NT)^{-\frac{1}{2}}\sum_{t=1}^{T} \eta^\top X_{t-1}\varepsilon_t \xrightarrow{d} N(0, \eta^\top S\eta) \tag{3.36}$$

for any η, where σ^2 is set to be 1 in this step for simplicity. Denote $\xi_{Nt} = (NT)^{-\frac{1}{2}}\eta^\top X_{t-1}\varepsilon_t$, $S_{Nt} = \sum_{s=1}^{t}\xi_{Ns}$ and $\mathcal{F}_{Nt} = \sigma\{\varepsilon_{is}, 1 \le i \le N, -\infty < s \le t\}$ so $\{S_{Nt}, \mathcal{F}_{Nt}, -\infty < t \le T, N \ge 1\}$ is a martingale array.

For any $\delta > 0$, as $N \to \infty$,

$$\sum_{t=1}^{T_N} E\{\xi_{Nt}^2 I(|\xi_{Nt}| > \delta)|\mathcal{F}_{N,t-1}\} \le \sum_{t=1}^{T_N} E\left(\frac{\xi_{Nt}^4}{\delta^2}|\mathcal{F}_{N,t=1}\right) \tag{3.37}$$

$$\le \quad \frac{C}{(NT_N)^2\delta^2}\sum_{t=1}^{T_N}(\eta^\top X_{t-1}X_{t-1}^\top \eta)^2 \xrightarrow{P} 0 \tag{3.38}$$

where $\xi_{Nt}^4 = (NT_N)^{-2}(\eta^\top X_{t-1}\varepsilon_t\varepsilon_t^\top X_{t-1}^\top \eta)^2$ and $C = E(\varepsilon_t\varepsilon_t^\top \varepsilon_t\varepsilon_t^\top)$ is some constant.

$$\sum_{t=1}^{T_N} E(\xi_{Nt}^2|\mathcal{F}_{N,t-1}) = \frac{1}{NT_N}\sum_{t=1}^{T_N}\eta^\top X_{t-1}X_{t-1}^\top \eta \xrightarrow{P} \eta^\top S\eta \tag{3.39}$$

where $\hat{S} = \frac{1}{NT}\sum_{t=1}^{T} X_{t-1}X_{t-1}^\top \xrightarrow{p} S$ because of Lemma 3.4.1 . According to the central limit theorem for martingale difference sequences, we have $S_{NT_N} = \sqrt{NT}\eta^\top g(\theta_0, \mathfrak{Y}_{N,T}) \xrightarrow{d} N(0, \eta^\top S\eta)$.

Thirdly,

$$\frac{\partial g(\theta; \mathfrak{Y}_{N,T})}{\partial \theta^\top} = \frac{1}{NT}\sum_{t=1}^{T} X_{t-1}X_{t-1}^\top = \hat{S} \xrightarrow{P} S,$$

which meets the third condition of proposition 3.4.1.

Hence, $V = \{DS^{-1}D^\top\}^{-1} = S^{-1}$. This completes the proof. □

By Corollary 3.4.1, $\hat{\theta}_{OLS}$ is $\sqrt{NT}$ consistent with asymptotic variance S^{-1}. The asymptotic distribution for estimate can be used to construct confidence intervals.

Below we prove a lemma which shows the convergence in the above (3.34) is true and then gives the asymptotic distribution of parameter estimates.

3.4.2 Special Case 2

In this example, we want to show how to use proposition 3.4.1 to prove the new GMM estimator that we defined with asymptotic distribution. Due to time and space constraints, the proof of the existence of the limits of $\hat{S} = \frac{1}{NT}\sum_{t=1}^{T}\mathbb{G}_{t-1}\mathbb{G}_{t-1}^{\top}$ and $\frac{1}{NT}\sum_{t=1}^{T}\mathbb{G}_{t-1}X_{t-1}^{\top}$. is skipped. We assume that the two limits exist.

As a result, a GMM estimator can be obtained by Corollary 3.4.2.

Corollary 3.4.2. *Under assumption of Theorem 3.2.2,*

$$\sqrt{NT}(\hat{\theta}_{N,T} - \theta_0) \xrightarrow{d} N(0, V), \tag{3.40}$$

where

$$V = \lim_{T\to\infty N\to\infty} NT\left(\sum_{t=p+4+1}^{T}\mathbb{G}_{t-1}\mathbb{X}_{t-1}^{\top}\right)^{-1}\left(\sum_{t=p+4+1}^{T}\mathbb{G}_{t-1}\mathbb{G}_{t-1}^{\top}\right)\left(\sum_{t=p+4+1}^{T}\mathbb{G}_{t-1}\mathbb{X}_{t-1}^{\top}\right)^{-1}.$$

Proof. First of all, the same as OLS, the GMM estimator is a consistent estimator of θ_0, which meet the first condition of proposition 3.4.1.

Secondly, we need to prove the second condition of proposition 3.4.1

$$\sqrt{NT}g(\theta_0, \mathfrak{Y}_{N,T}) \xrightarrow{d} N(0, S).$$

We know

$$g(\theta, \mathfrak{Y}_{N,T}) = (1/NT)\sum_{t=1}^{T}\mathbb{G}_{t-1}(Y_t - X_{t-1}^{\top}\theta). \tag{3.41}$$

To prove the result, it suffices to show that

$$\sqrt{NT}\eta^{\top}g(\theta_0, \mathfrak{Y}_{N,T}) = (NT)^{-\frac{1}{2}}\sum_{t=1}^{T}\eta^{\top}\mathbb{G}_{t-1}\varepsilon_t \xrightarrow{d} N(0, \eta^{\top}S\eta) \tag{3.42}$$

for any η.

Denote $\xi_{Nt} = (NT)^{-\frac{1}{2}}\eta^{\top}\mathbb{G}_{t-1}\varepsilon_t, S_{Nt} = \sum_{s=1}^{t}\xi_{Ns}$ and $\mathcal{F}_{Nt} = \sigma\{\varepsilon_{is}, 1 \leq i \leq N, -\infty < s \leq t\}$, so $\{S_{Nt}, \mathcal{F}_{Nt}, -\infty < t \leq T, N \geq 1\}$ is a martingale array.

For any $\delta > 0$, as $N \to \infty$,

$$\sum_{t=1}^{T_N} E\{\xi_{Nt}^2 I(|\xi_{Nt}| > \delta)|\mathcal{F}_{N,t-1}\} \leq \sum_{t=1}^{T_N} E\left(\frac{\xi_{Nt}^4}{\delta^2}|\mathcal{F}_{N,t-1}\right) \tag{3.43}$$

$$\leq \frac{C}{(NT_N)^2\delta^2}\sum_{t=1}^{T_N}(\eta^\top \mathbb{G}_{t-1}\mathbb{G}_{t-1}^\top \eta)^2 \xrightarrow{p} 0 \tag{3.44}$$

where $\xi_{Nt}^4 = (NT)^{-2}(\eta^\top \mathbb{G}_{t-1}\mathbb{G}_{t-1}^\top \eta)^2$ and $C = E(\varepsilon_t\varepsilon_t^\top \varepsilon_t\varepsilon_t^\top)$ is some constant.

$$\sum_{t=1}^{T_N} E(\xi_{Nt}^2|\mathcal{F}_{N,t-1}) = \frac{1}{NT_N}\sum_{t=1}^{T_N}\eta^\top \mathbb{G}_{t-1}\mathbb{G}_{t-1}^\top \eta \xrightarrow{p} \eta^\top S\eta \tag{3.45}$$

where $\hat{S} = \frac{1}{NT}\sum_{t=1}^{T}\mathbb{G}_{t-1}\mathbb{G}_{t-1}^\top \xrightarrow{p} S$. According to the central limit theorem for martingale difference sequences, we have $S_{NT_N} = \sqrt{NT}\eta^\top g(\theta_0, \mathfrak{Y}_{N,T}) \xrightarrow{d} N(0, \eta^\top S\eta)$.

Thirdly,

$$\frac{\partial g(\theta; \mathfrak{Y}_{N,T})}{\partial \theta^\top} = \frac{1}{NT}\sum_{t=1}^{T}\mathbb{G}_{t-1}X_{t-1}^\top,$$

which meets the third condition of proposition 3.4.1 and

$$D^\top = \lim_{T\to\infty N\to\infty}\frac{1}{NT}\sum_{t=1}^{T}\mathbb{G}_{t-1}X_{t-1}^\top.$$

Hence,

$$\begin{aligned} V &= \{DS^{-1}D^\top\}^{-1} \\ &= \lim_{T\to\infty N\to\infty} NT\left(\sum_{t=2}^{T}\mathbb{G}_{t-1}\mathbb{X}_{t-1}^\top\right)^{-1}\left(\sum_{t=2}^{T}\mathbb{G}_{t-1}\mathbb{G}_{t-1}^\top\right)\left(\sum_{t=2}^{T}\mathbb{G}_{t-1}\mathbb{X}_{t-1}^\top\right)^{-1}. \end{aligned}$$

This completes the proof.

□

With Corollary 3.4.1 and Corollary 3.4.2, we will compare the performance of the asymptotic variance of the GMM estimators with the Monte Carlo simulation results in Section 3.6.

3.5 Lagrange Multiplier (LM) Test

In this section, we will discuss the nonlinear test for the threshold effect. It is important to determine whether the threshold effect is statistically significant. In empirical studies, for a time series type economic variable, whether a linear time series model or a non-linear time series model should be established requires the use of standardized measurement methods. This judgment cannot be made intuitively, but it can be transformed into a statistical hypothesis test to identify the TNAR model. We will focus on the LM Test [6, 25] with the threshold autoregressive model as the alternative hypothesis.

Now we develop such a test under the proposed $TNAR$ model

$$\mathbb{Y}_t = \mathcal{B}_0 + G\mathbb{Y}_{t-1} + \varepsilon_t, \tag{3.46}$$

where $G = \beta_1 W + \beta_2 I + \beta_3 J_{t-1}$. The matrices G and J_{t-1} contain a threshold parameter r. We assume that the threshold parameter r belongs to a known bounded subset $\tilde{R}$ of R, usually a finite interval.

Given observations, $\mathbb{Y}_1, \cdots, \mathbb{Y}_T$, consider the null hypothesis is $H_0 : \beta_3 = 0$, which means that the model has one regime (linear NAR model) and the alternative one is $H_1 : \beta_3 \neq 0$. Let $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma) \in \mathbb{R}^{p+4}$. Under H_0, there is no nuisance parameter r. The conditional log likelihood is

$$\mathbf{L}(\theta) = \sum_t l_t = \sum_{t=1}^{T} \left(-\frac{1}{2}\mathsf{log}|\Sigma_0| - \frac{1}{2}\varepsilon_t^\top \Sigma_0^{-1} \varepsilon_t \right),$$

where

$$\Sigma_0 = \begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{pmatrix}.$$

Suppose that $\hat{\theta}_0$ is the maximum likelihood estimate of θ under the null hypothesis H_0. Then the LM test depends on the score function at $\hat{\theta}_0$, where

$$LM = \mathbf{D}(\hat{\theta}_0)^\top \mathbf{I}(\hat{\theta}_0)^{-1} \mathbf{D}(\hat{\theta}_0) \to \chi^2_k. \tag{3.47}$$

Here $\mathbf{D}(\hat{\theta}_0)$ and $\mathbf{I}(\hat{\theta}_0)$ are the score function and Fisher information matrix, respectively. k is the k equations in the null hypothesis, that is, k constraints on the parameter θ.

Since our model has extra parameters , [7] shows the expression (3.47) need to be modified. $\mathbf{D}(\theta)$ and $\mathbf{I}(\theta)$ are written into the block vectors or the block matrices according to the dimensions of the parameter to be estimated β_3 and the extra parameters $(\beta_0, \beta_1, \beta_2, \gamma)$, i.e.

$$\mathbf{D}(\theta) = \begin{bmatrix} \mathbf{D_1}(\theta) \\ \mathbf{D_2}(\theta) \end{bmatrix}, \ \mathbf{I}(\theta) = \begin{bmatrix} \mathbf{I_{11}}(\theta) & \mathbf{I_{12}}(\theta) \\ \mathbf{I_{21}}(\theta) & \mathbf{I_{22}}(\theta) \end{bmatrix}.$$

By routine operations, it is easy to obtain the scoring vector, $(\frac{\partial \mathbf{L}(\theta)}{\partial \beta_0}, \frac{\partial \mathbf{L}(\theta)}{\partial \beta_1}, \frac{\partial \mathbf{L}(\theta)}{\partial \beta_2}, \frac{\partial \mathbf{L}(\theta)}{\partial \beta_3}, \frac{\partial \mathbf{L}(\theta)}{\partial \gamma})$, and the expectation of the second derivatives with respect to $\beta_0, \beta_1, \beta_2, \beta_3$ and γ. The detailed calculation of this part is in Appendix A.2.

The Lagrange-multiplier test statistic is

$$\mathbf{LM} = \sup_{r \in \tilde{R}} \mathbf{D_1}(\hat{\theta}_0)^\top [\mathbf{I_{11}}(\hat{\theta}_0) - \mathbf{I_{12}}(\hat{\theta}_0)\mathbf{I_{22}}(\hat{\theta}_0)^{-1}\mathbf{I_{21}}(\hat{\theta}_0)]^{-1}\mathbf{D_1}(\hat{\theta}_0), \tag{3.48}$$

where

$$\mathbf{D_1}(\hat{\theta}_0) = \left.\frac{\partial \mathbf{L}(\theta)}{\partial \beta_3}\right|_{\hat{\theta}_0} = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \sum_{i=1}^{N} Y_{(t-1)i} \hat{\varepsilon}_{ti} I_{\{Y_{(t-1)i<r}\}},$$

$$\mathbf{I_{11}}(\hat{\theta}_0) = E\left(-\frac{\partial^2 L(\theta)}{\partial \beta_3^2}\right)_{\hat{\theta}_0} = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \sum_{i=1}^{N} Y_{(t-1)i}^2 I_{\{Y_{(t-1)i<r}\}},$$

$$\begin{aligned}\mathbf{I_{21}}(\hat{\theta}_0) = \mathbf{I_{12}}(\hat{\theta}_0)^\top &= E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_0} \quad -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_1} \quad -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_2} \quad -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \gamma_1} \quad \cdots \quad -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \gamma_p}\right)_{\hat{\theta}_0} \\ &= \begin{pmatrix} \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \sum_{i=1}^{N} Y_{(t-1)i} I_{\{Y_{(t-1)i<r}\}} \\ \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} W_{ij} Y_{(t-1)j}\right) Y_{(t-1)i} I_{\{Y_{(t-1)i<r}\}} \\ \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \sum_{i=1}^{N} Y_{(t-1)i}^2 I_{\{Y_{(t-1)i<r}\}} \\ \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \sum_{i=1}^{N} Y_{(t-1)i}^2 I_{\{Y_{(t-1)i<r}\}} Z_{1i} \\ \vdots \\ \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \sum_{i=1}^{N} Y_{(t-1)i}^2 I_{\{Y_{(t-1)i<r}\}} Z_{pi} \end{pmatrix},\end{aligned}$$

$$\mathbf{I_{22}}(\hat{\theta}_0) = E\begin{pmatrix} -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0^2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \beta_2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \gamma_1} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \gamma_p} \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1^2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \beta_2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \gamma_1} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \gamma_p} \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2^2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \gamma_1} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \gamma_p} \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \beta_2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1^2} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \gamma_p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_p \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_p \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_p \partial \beta_2} & -\frac{\partial^2 L(\theta)}{\partial \gamma_p \partial \gamma_1} & \cdots & -\frac{\partial^2 L(\theta)}{\partial \gamma_p^2} \end{pmatrix}_{\hat{\theta}_0}.$$

Here $\hat{\varepsilon}_{ti}$ and $\hat{\sigma}^2$ are the residual and residual variance under the null hypothesis, obtained by OLS estimate from [50]. By the first Theorem in [48], LM is asymptotically distributed as χ_k^2. Therefore, for a given confidence level α, it is not difficult to determine the rejection domain of H_0.

3.6 Numerical Studies

3.6.1 Simulation Models

To demonstrate the performance of the proposed methodology, we use the same three examples as $\mathbb{X}$ [50]. The main difference among the three examples is the generation mechanism of the adjacency matrix A and the selection of $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^\top \in \mathbb{R}^4$ in (3.3). Except for this, they are similar. For each example, the random error ε_{it} is generated from the standard normal distribution $N(0,1)$, and the covariate $Z_i = (Z_{i1}, Z_{i2}, \cdots, Z_{i5}) \in \mathbb{R}^5$ is generated by multivariate normal distribution with mean value $\mathbf{0}$ and covariance $\Sigma_z = (\sigma_{j_1j_2})$, where $\sigma_{j_1j_2} = 0.5^{|j_1-j_2|}$. For each example, γ is fixed at $\gamma = (-0.5, 0.3, 0.8, 0, 0)^\top$. To obtain $\mathbb{Y}_t$, starting from any value, we generate 1000 times by (3.4) and obtain the 1000th value as our initial value $\mathbb{Y}_0$ to remove the effect of the initial value. Then $\mathbb{Y}_t$ can be generated according to (3.4).

Example 1. (Dyad Independence Model) A Dyad Independence Model was introduced in [24] with Dyad defined as $D_{ij} = (a_{ij}, a_{ji})$ for $1 \le i < j \le N$. It is assumed that different $D_{ij}s$ are independent by Dyad independence. We let $P(D_{ij} = (1,1)) = 20N^{-1}$ to make sure the network sparsity. Moreover, we let $P(D_{ij} = (1,0)) = P(D_{ij} = (0,1)) = 0.5N^{-0.8}$. The result is that the expected number of connected dyads is $O(N^{0.2})$. Thereby, $P(D_{ij} = (0,0))$ will be $1 - 20N^{-1} - N^{-0.8} \to 1$ as $N \to \infty$. For this example, we fix T=10, 30, 100 and $\beta = (0.3, 0.0, 0.5, 0.1)^\top$.

Example 2. (Stochastic Block Model) Next we consider the stochastic block model [34, 46], which is a popular network structure. According to [46], the block network structure are randomly assigned for each node a block label $(k = 1, \cdots, K)$, where $K \in \{5, 10, 20\}$. We let $P(a_{ij} = 1) = 0.3N^{-0.3}$ if i and j stay in the same block and $P(a_{ij} = 1) = 0.3N^{-1}$ otherwise. This means that the nodes are more likely to be connected in the same block than those from different blocks. For this example, we fix T=30 and $\beta = (0.0, 0.1, -0.2, 0.1)^\top$.

Example 3. (Power-Law Distribution Model) According to [3], there is a common network phenomenon that most nodes have very small links but there are a small amount of nodes that have a large number of links. Such phenomenon can be described by the power-law distribution. According to [9], we simulate A as follows. First, the in-degree $d_i = \sum_j a_{ji}$ are generated for each nodes according to the discrete power-law distribution as $P(d_i = k) = ck^{-\alpha}$, where c is a normalizing constant and the exponent parameter $\alpha \in \{1.2, 2.0, 3.0\}$. Finally, we randomly choose d_i nodes as the ith node's followers. For this example, we fix T=30 and $\beta = (0.3, -0.1, 0.5, 0.1)^\top$.

3.6.2 Performance Measurements and Simulation Results

We consider the different size (N=100, 500 or 1000) for each simulation example. The experiment will be randomly repeated R=1000 times. Let $\hat{\theta}^{(r)} = (\hat{\theta}_j^{(r)})^\top = (\hat{\beta}_0^{(r)}, \hat{\beta}_1^{(r)}, \hat{\beta}_2^{(r)}, \hat{\beta}_3^{(r)}, \hat{\gamma}^{(r)\top})^\top$ be the estimator of the rth replication. We utilize the following measures to assess these performances. For each parameter θ_j with $1 \le j \le p+4$, the root mean square error (RMSE) is obtained by $\text{RMSE}_j = \{R^{-1}\sum_{r=1}^R(\hat{\theta}_j^{(r)} - \theta_j)^2\}^{1/2}$. Then, for each parameter θ_j with $1 \le j \le p+4$, the 95% confidence interval is defined by $\text{CI}_j^{(r)} = (\hat{\theta}_j^{(r)} - z_{0.975}\widehat{SE}_j^{(r)}, \hat{\theta}_j^{(r)} + z_{0.975}\widehat{SE}_j^{(r)})$, where $\widehat{SE}_j^{(r)}$ is root square of the jth diagonal element of $(\sum_t \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1})^{-1}\hat{\sigma}^2$ with $\hat{\sigma}^2 = (NT)^{-1}\sum_{i,t}(y_{it} - X_{it}^\top\hat{\theta}^{(r)})^2$, and z_α is the αth quantile of a standard normal distribution. The coverage probability (CP) is computed as $\text{CP}_j = R^{-1}\sum_{r=1}^R I(\theta_j \in CI_j^{(r)})$ where $I(\cdot)$ is the indicator function. Lastly, the total number of observed edges (TNOE) (i.e., $\sum_{ij} a_{ij}$) and the network density (ND) is given by $\{N(N-1)\}^{-1}\sum_{ij} a_{ij}$.

These detailed results are summarized in Table 3.1, Table 3.2 and Table 3.3. For fixed T, the RMSE decreases towards 0 as N increases. For example, the RMSE value of β_3 with $T = 30$ drops from 2.6% to 0.8% as N increases from 100 to 1000 in Example 1. The network is sparse as N increases (ND drops from 22.7% to 2.4% for Dyad Independence Model with N increases from 100 to 1000). The coverage probabilities for each parameter (i.e., θ_j) are stable at the normal level 95%, which means the theoretical results is reasonable and the proposed estimator

$\hat{\theta}$ is indeed consistent and asymptotically normal.

Table 3.1: Simulation results for Example 1 with 1000 replications

	$T = 10$			$T = 30$			$T = 100$		
	$N = 100$	$N = 500$	$N = 1000$	$N = 100$	$N = 500$	$N = 1000$	$N = 100$	$N = 500$	$N = 1000$
β_0	6.4 (94.3)	2.8(94.7)	1.9(95.9)	3.9(94.4)	1.8(94.1)	1.3(94.4)	1.9(94.4)	0.9(94.9)	0.6(93.9)
β_1	7.0 (96.0)	2.7(96.0)	2.1(94.5)	8.8(93.7)	2.9(95.7)	2.0(93.6)	2.3(94.4)	0.9(95.2)	0.7(94.9)
β_2	3.7 (93.6)	1.5(95.8)	1.1(95.0)	2.1(93.9)	0.9(94.7)	0.6(95.9)	1.1(94.7)	0.5(94.7)	0.3(95.4)
β_3	4.5 (94.0)	2.0(94.7)	1.3(95.1)	2.6(94.2)	1.1(94.6)	0.8(94.0)	1.4(94.1)	0.6(95.0)	0.4(94.2)
γ_1	4.9 (93.8)	2.1(95.8)	1.5(94.9)	2.7(94.7)	1.2(94.5)	0.9(95.0)	1.5 (93.9)	0.7 (94.3)	0.5 (94.3)
γ_2	4.8 (93.8)	2.0(96.3)	1.4(94.4)	2.5(95.2)	1.2(94.9)	0.8(95.8)	1.4 (95.1)	0.7 (93.8)	0.4 (94.5)
γ_3	6.5 (95.5)	2.8(95.5)	1.9(95.6)	3.7(93.0)	1.6(95.2)	1.1(94.6)	2.0 (95.7)	0.9 (94.9)	0.6 (96.2)
γ_4	4.9 (95.0)	1.9(95.5)	1.3(94.6)	2.0(95.5)	1.0(94.8)	0.7(94.6)	1.5 (96.0)	0.6 (95.7)	0.4 (95.1)
γ_5	4.2 (95.2)	1.6(94.6)	1.1(95.0)	2.1(95.5)	1.0(95.9)	0.7(92.9)	1.4 (93.8)	0.6 (93.5)	0.3 (94.7)
TNOE	2065	10715	21913	2065	10715	21913	2065	10715	21913
ND(%)	20.9	4.3	2.2	20.9	4.3	2.2	20.9	4.3	2.2

Table 3.2: Simulation results for Example 2 with 1000 replications

	$K=5$			$K=10$			$K=20$		
	$N=100$	$N=500$	$N=1000$	$N=100$	$N=500$	$N=1000$	$N=100$	$N=500$	$N=1000$
β_0	2.9(95.3)	1.3(96.1)	0.9(95.8)	3.0(94.8)	1.4(94.9)	1.0 (95.3)	3.0(95.5)	1.3 (95.4)	0.9 (94.8)
β_1	1.8(95.4)	1.2(94.6)	1.1(94.0)	1.8(95.3)	0.8(95.2)	0.8 (95.1)	1.7(95.1)	0.8 (95.1)	0.6 (94.4)
β_2	3.0(95.2)	1.2(95.5)	0.9(95.7)	3.1(95.4)	1.3(95.4)	0.9 (95.7)	3.0(95.8)	1.3 (95.4)	1.0 (94.4)
β_3	4.5(94.7)	2.0(96.4)	1.4(96.2)	4.5(95.7)	2.0(96.0)	1.5 (95.9)	4.7(95.8)	2.0 (95.7)	1.5 (95.3)
γ_1	2.2(93.9)	1.0(95.8)	0.7 (95.3)	2.2(94.4)	1.0 (94.8)	0.7 (95.0)	2.2(94.4)	1.0(94.8)	0.7 (95.0)
γ_2	2.4(93.8)	1.1(94.1)	0.7 (96.2)	2.6(95.4)	1.1 (93.4)	0.8 (94.5)	2.6(95.0)	1.1(95.3)	0.8 (93.9)
γ_3	3.0(95.5)	1.2(93.9)	0.8 (94.5)	3.0(95.8)	1.2 (95.5)	0.8 (95.1)	2.6(95.9)	1.2(95.9)	0.8 (95.7)
γ_4	2.9(94.8)	1.1(95.5)	0.7 (95.0)	2.4(94.0)	1.0 (94.9)	0.7 (94.9)	2.5(94.9)	1.0(94.3)	0.7 (95.2)
γ_5	2.2(94.5)	0.9(95.1)	0.6 (95.5)	2.3(94.8)	0.9 (95.9)	0.7 (94.9)	1.9(95.3)	0.9(95.0)	0.7 (95.3)
TNOE	261	2431	7946	184	1287	4220	165	945	2308
ND(%)	2.6	1.0	0.8	1.9	0.5	0.4	1.7	0.4	0.2

Table 3.3: Simulation results for Example 3 with 1000 replications

	α=1.2			α=2.0			α=3.0		
	$N = 100$	$N = 500$	$N = 1000$	$N = 100$	$N = 500$	$N = 1000$	$N = 100$	$N = 500$	$N = 1000$
β_0	4.2(93.8)	2.3(94.4)	4.0 (95.2)	3.1(94.5)	1.3 (95.2)	1.1 (94.5)	3.0 (96.7)	1.4 (95.1)	1.0 (94.6)
β_1	9.4(94.8)	7.7(93.9)	7.6 (95.2)	1.7(95.0)	0.6 (95.7)	0.6 (95.2)	1.1 (95.0)	0.4 (95.2)	0.3 (94.5)
β_2	2.1(94.8)	0.9(93.8)	0.6 (95.6)	2.0(93.7)	0.9 (95.8)	0.6 (94.5)	2.0 (94.3)	0.9 (94.7)	0.6 (93.0)
β_3	2.7(95.3)	1.1(94.7)	0.8 (94.8)	2.6(95.3)	1.2 (94.7)	0.8 (94.4)	2.6 (96.1)	1.1 (95.8)	0.8 (94.1)
γ_1	2.9(94.5)	1.2(95.0)	0.8 (95.7)	2.8(94.9)	1.2 (95.4)	0.9 (94.7)	2.8 (95.0)	1.3 (94.0)	0.9 (94.4)
γ_2	2.9(94.9)	1.2(95.4)	0.8 (95.1)	2.8(94.0)	1.2 (95.4)	0.8 (94.5)	2.7 (95.7)	1.2 (94.9)	0.8 (94.4)
γ_3	3.7(95.2)	1.6(95.7)	1.1 (95.5)	3.8(94.9)	1.6 (95.4)	1.1 (95.2)	4.0 (93.6)	1.6 (93.5)	1.1 (94.5)
γ_4	2.4(95.4)	1.1(94.8)	0.7 (94.9)	2.5(94.6)	1.1 (94.5)	0.8 (93.6)	2.5 (93.9)	1.0 (94.2)	0.8 (95.5)
γ_5	2.2(95.9)	0.9(94.6)	0.7 (95.4)	2.4(94.2)	1.0 (95.4)	0.7 (94.8)	2.2 (94.7)	1.0 (94.5)	0.7 (96.1)
TNOE	3535	80066	287465	262	3218	6064	222	1085	2161
ND(%)	35.7	32.1	28.8	2.7	1.3	0.6	2.2	0.4	0.2

3.7 A Twitter Dataset

The data are collected from Twitter(www.twitter.com) and the details of data extraction are provided in Chapter 6. Our dataset contains weekly tweets length for a total of N=9908 active followers of Strathclyde official Twitter account. These Twitter users are observed for a total of $T = 8$ continuous weeks. The response (y_{it}) considered here is the number of characters contained in the post by node i in week t. Moreover, we used the in-degrees and out-degrees as two time-invariant nodal covariates. We provide the histogram of the in-degrees and out-degrees in Figure 3.1. It can be seen that the distribution of in-degrees is much more skewed than that of out-degree. Their median values are 12 and 20 respectively. The network structure A is defined to be the followee-follower relationship. The resulting network density is around 4.0%. The histogram of responses is plotted in Figure 3.2. The response distribution is normal with the mean value 6.78.

The simple linear regression is conducted for each node with y_{it} as the response and $y_{i(t-1)}$ as the only covariate. As a consequence, R-squares can be computed for each node. This leads to a total of N=9908 R-square values, whose median level is about 7.3%. This suggests the existence of the momentum impact. Next, we compute residuals from this model for each node. These residuals are treated as the responses and regressed against $\sum_j w_{ij} y_{j(t-1)}$ (i.e., the network impact). this leads to another N=9908 R-squares values, whose median is around 20%. This suggests that, even after controlling the momentum impact, the network effect exists.

We applied the LM test to our Twitter data and LM value is 3041 (p-value<0.01). Therefore, it is significant to reject the linear assumption of these data. It suggests that it is more appropriate to use our non-linear TNAR model to provide the estimation results. The detailed estimation results are given in Table 3.4. The estimated network effect (0.17) suggests that the activeness of a node is positively related to its connected neighbours. The estimated momentum effect (0.62) confirms that a node with higher (lower) activeness level in the past is likely to exhibit higher (lower) activeness in the future. The nodal effect indicates that uses who have more followers and followees tend to be more active.

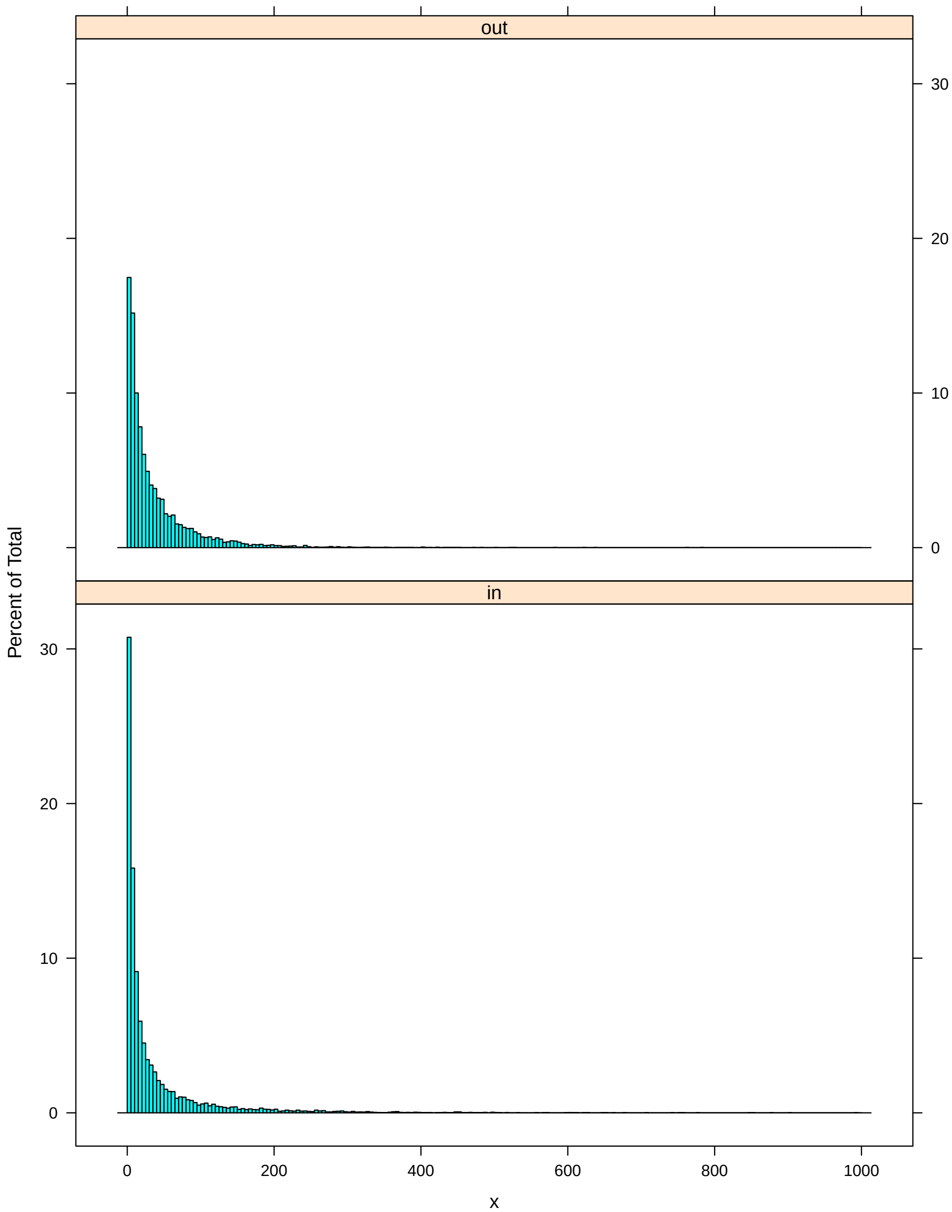


Figure 3.1: The histogram of in-degrees (below) and out-degrees (upper).

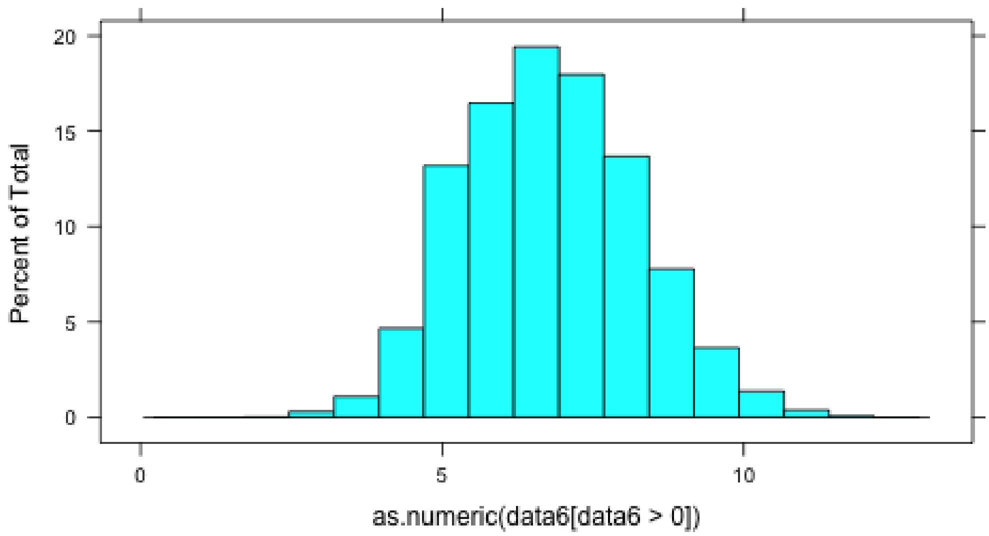


Figure 3.2: The histogram of the response.

Table 3.4: The detailed NAR analysis results for the Twitter Data Set

Regression coefficient	Estimate	SE($\times 10^2$)	p-value
$\hat{\beta}_0$	1.2442	3.78	< 0.001
$\hat{\beta}_1$	0.1715	0.65	< 0.001
$\hat{\beta}_2$	0.6247	0.30	< 0.001
$\hat{\beta}_3$	-0.2184	0.72	< 0.001
$\hat{\gamma}_1$	0.0018	0.02	< 0.001
$\hat{\gamma}_2$	0.0015	0.02	< 0.001

3.8 Conclusion

To summarise, in the Chapter 6, we have collected data from Twitter to check the validation of the model in [50]. We carried out LM test and rejected the linear assumption of their model. The $TNAR$ model is proposed to describe the nonlinear property of high dimensional data. The strictly stationary condition of $TNAR$ model and the parameter estimation method by GMM have been provided. The asymptotic properties of GMM is also investigated. The numerical studies (simulation and real data analysis) have been carried out.

Chapter 4

Threshold Network Quantile Autoregression

As we all know, the time series model is one of the most important contents in modern econometric analysis. The autoregressive (AR) model is the theoretical base of all time series models. The traditional model describes the process where the conditional distribution of the explanatory variable is affected by its lagged variable. Ordinary least squares (OLS) is an important method for estimating model parameters. If random disturbances of the AR model follow the normal distribution, then the estimator has consistency and asymptotic efficiency. However, in practice, the errors usually do not follow normal distributions, for example, data with biased, peak or heavy tailed distribution, or data with significant heteroscedasticity, outliers, etc. In these situations, the OLS estimator has a large deviation, and the robustness of the OLS estimator is poor. The linear quantile regression model proposed by [28] can solve the above problem, which becomes increasingly popular in the field of time series econometrics in recent years. [29] further proposed a quantile autoregressive model for a conditional quantile function, which does not assume an independent and identically distributed (i.i.d) underlying process.

4.1 Introduction

For non-linear modelling, there are a few very important types of nonlinear model such as threshold autoregressive (TAR) models, smooth transition (STAR) models, and Markov switching models. Among them, the threshold autoregressive model proposed by [42] has an important influence. This captures the asymmetrical character of the time series when they are at different stages. A comprehensive summary of TAR models in theory and the fields of econometrics and economics can be found in [21].

However, the above research work is based on the framework of mean value, which only describes the dynamics of the conditional mean process of the response variable. The proposed linear quantile model by [28] can reveal the impact of the explanatory variables on the response variables at each quantile point. [29] developed a quantile autoregressive (QAR) model. In this model, the parameter can be varying at different quantiles.

However, the quantile autoregressive model can only be applied in the univariate case. To the authors' knowledge, the method in the existing literature cannot directly be applied to high dimensional data since the total number of parameters is quite large.

In this chapter, we first propose the threshold network quantile autoregressive (TNQAR) model.

The rest of the chapter is arranged as follows. In Section 2 we introduce the threshold network quantile autoregressive model, where the stationary condition are established. Details about parameter estimation method are given in Section 3, where the asymptotic properties are also given. Simulation studies and a real data analysis are conducted in Section 4. Lastly, a brief conclusion is discussed in Section 5.

4.2 Threshold Network Quantile Autoregression (TNQAR)

In this section, we first give a brief overview of network quantile regression (NQAR). Then we propose our TNQAR model.

4.2.1 Network Quantile Autoregression

NQAR is developed by [52] under the framework of quantile regression. Let $U_{it}(1 \leq i \leq N, 1 \leq t \leq T)$ be a sequence of *iid* random variables, which follows the standard uniform distribution. Assume that a q-dimensional random nodal covariate vector $Z_i \in \mathbb{R}^q$ belongs to the ith node. The network relationship is defined by $A = (a_{ij} \in \mathbb{R}^{N\times N})$ as the adjacency matrix, in which $a_{ij} = 1$ if the ith node follows the jth node, otherwise $a_{ij} = 0$ and the nodes cannot be self-related (i.e. $a_{ii} = 0$). The NQAR can be written as

$$Y_{it} = \beta_0(U_{it}) + \sum_{l=1}^{q} Z_{il}\gamma_l(U_{it}) + \beta_1(U_{it})n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{i(t-1)} + \beta_2(U_{it})Y_{i(t-1)} \triangleq g_\theta(U_{it}),$$

where the β_j's $(0 \leq j \leq 3)$ and the γ_l's $(1 \leq l \leq q)$ are unknown coefficient functions from [0,1] to $\mathbb{R}^1$ and $n_i = \sum_{j\neq i} a_{ij}$ is the out-degree for the ith node.

4.2.2 Model and Notations

Inspired by [52] which proposes the network quantile autoregression (NQAR) model to describe the dynamic behaviour in a high dimensional system, we consider a nonlinear quantile regression for analysing high dimensional data with network structure and propose the TNQAR model as

$$\begin{aligned} Y_{it} =& \beta_0(U_{it}) + \sum_{l=1}^{q} Z_{il}\gamma_l(U_{it}) + \beta_1(U_{it})n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{i(t-1)} + \beta_2(U_{it})Y_{i(t-1)} + \\ & \beta_3(U_{it})\mathbf{1}_{\{Y_{i(t-1)}>r\}}Y_{i(t-1)} \triangleq g_\theta(U_{it}), \end{aligned} \tag{4.1}$$

Assuming the right side of (4.1) is monotonically increasing in U_{it}, the conditional quantile function of Y_{it} given $(Z_i, \mathbb{Y}_{t-1})$ as

$$Q_{Y_{it}}(\tau|Z_i, \mathbb{Y}_{t-1}) = \beta_0(\tau) + \sum_{l=1}^{q} Z_{il}\gamma_l(\tau) + \beta_1(\tau)n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{i(t-1)} + \beta_2(\tau)Y_{i(t-1)} + \beta_3(\tau)\mathbf{1}_{\{Y_{i(t-1)}>r(\tau)\}}Y_{i(t-1)}.$$

In the above equation, the quantile autoregressive coefficients are functions of τ and vary over the quantiles.

Denote $\mathbb{Y}_t = (Y_{1t}, Y_{2t}, \cdots, Y_{Nt})^\top \in \mathbb{R}^N, \mathbb{Z} = (Z_1, Z_2, \cdots, Z_N)^\top \in \mathbb{R}^{N\times q}$. Let $\mathcal{B}_{0t} = \beta_0(U_{it}) + \sum_l Z_{il}\gamma_l(U_{it}, 1 \le i \le N)^\top \in \mathbb{R}^N, \mathcal{B}_{kt} = diag\{\beta_k(U_{it}), 1 \le i \le N\} \in \mathbb{R}^{N\times N}$ for $k = 1, 2, 3$. $J_{t-1} = diag\{\mathbf{1}_{\{Y_{i(t-1)}>r(U_{it})\}}, 1 \le i \le N\} \in \mathbb{R}^{N\times N}$. $\Gamma = E(\mathcal{B}_{0t}) = c_0\mathbf{1}_N \in \mathbb{R}^N$, where $c_0 = b_0 + c_Z, b_0 = \int_0^1 \beta_0(u)du$ and $c_Z = E(Z_1)^\top\tilde{\gamma}$ with $\tilde{\gamma} = (\int_0^1 \gamma_l(u)du, 1 \le l \le q)^\top$. Then the TNQAR model (4.1) can be re-written in vector form as

$$\mathbb{Y}_t = \Gamma + G_t\mathbb{Y}_{t-1} + V_t \tag{4.2}$$

where $G_t = \mathcal{B}_{1t}W + \mathcal{B}_{2t} + \mathcal{B}_{3t}J_{t-1} \in \mathbb{R}^{N\times N}, W = (w_{ij}) = (n_i^{-1}a_{ij}) \in \mathbb{R}^{N\times N}$ is the row-normalized adjacency matrix and $V_t = \mathcal{B}_{0t} - \Gamma \in \mathbb{R}^N$.

4.2.3 Stationarity

For convenience, set $b_k = E\{\beta_k(U_{it})\}$ for $k = 1, 2, 3$, then we have the following theorem.

Theorem 4.2.1. *Assuming $|b_1| + max\{|b_2|, |b_2 + b_3|\} < 1$, the stochastic process $\{\mathbb{Y}_t\}$ is geometrically ergodic. Then there exists a unique stationary solution of the TNQAR model (4.1) as*

$$\mathbb{Y}_t = \sum_{l=0}^{\infty} \Pi_l\Gamma + \sum_{l=0}^{\infty} \Pi_l V_{t-l}$$

where $\Pi_l = \prod_{k=1}^{l} G_{t-k+1}$ for $l \ge 1$ and $\Pi_0 = I_N$.

Proof. First of all, by iteration, we can get the solution of the TNQAR model (4.1) as

$$\mathbb{Y}_t = \sum_{l=0}^{L-1} \Pi_l\Gamma + \Pi_L\mathbb{Y}_{t-L} + \sum_{l=0}^{L-1} \Pi_l V_{t-l} = \sum_{l=0}^{\infty} \Pi_l\Gamma + \sum_{l=0}^{\infty} \Pi_l V_{t-l}, \tag{4.3}$$

where $\Pi_l = \prod_{k=1}^{l} G_{t-k+1}$ for $l \ge 1$ and $\Pi_0 = I_N$.

As in the previous chapter, we utilize the previous lemma 3.2.1 to show the time series is stationary.

Define a norm by $\|X\|^2 = \sum_{i=1}^{N} x_i^2 \quad \text{for} \quad X = (x_1, \cdots, x_N)^\top \in \mathbb{R}^N$ and $g(Y) = \|Y\|$.

$$\begin{aligned} E\{g(\mathbb{Y}_t)|\mathbb{Y}_{t-1} = Y\} &= E(\|\Gamma + G_tY + V_t\|) \\ &\leq E\|\Gamma\| + E\|G_tY\| + E\|V_t\| \\ &\leq E\|\Gamma\| + E\|V_t\| + (|b_1| + max\{|b_2|, |b_2 + b_3|\})\|Y\| \end{aligned}$$

since

$$\begin{aligned} E\|G_tY\| &= E\|(\mathcal{B}_{1t}W + \mathcal{B}_{2t} + \mathcal{B}_{3t}J_{t-1})Y\| \\ &\leq E\|\mathcal{B}_{1t}WY\| + E\|\mathcal{B}_{2t}Y + \mathcal{B}_{3t}J_{t-1}Y\| \\ &\leq (|b_1| + max\{|b_2|, |b_2 + b_3|\})\|Y\| \end{aligned}$$

and

$$\begin{aligned} \|WY\| &\leq \rho(W^\top W)\|Y\| \\ &\leq \|Y\| \qquad (\text{since} \qquad \rho(W^\top W) \leq \rho(W^\top)\rho(W) = 1) \end{aligned}$$

Let $\rho = |b_1| + max\{|b_2|, |b_2 + b_3|\} < 1$ and take λ and M such that $0 < \rho < \lambda < 1$ and

$$M > \frac{E\|\Gamma\| + E\|V_t\|}{\lambda - \rho}.$$

Denote $C = \{Y : \|Y\| \leq M\}$ and easy to know that C is small set. When $\|Y\| > M$ (i.e. $Y \notin C$), we have

$$E\{g(\mathbb{Y}_t)|\mathbb{Y}_{t-1} = Y\} \leq \lambda g(Y) - \lambda_1,$$

where $\lambda_1 = (\lambda - \rho)M - (E\|\Gamma\| + E\|V_t\|) > 0$. And when $Y \in C$,

$$E\{g(\mathbb{Y}_t)|\mathbb{Y}_{t-1} = Y\} \leq \lambda_2,$$

where $\lambda_2 = \rho M + E\|\Gamma\| + E\|V_t\| > 0$. By lemma 3.2.1, $\{\mathbb{Y}_t\}$ is geometrically ergodic and there is a unique stationary distribution. $\square$

4.3 Statistical Inference

4.3.1 Parameter Estimation and Asymptotic Property

In this part, we introduce an estimation method to the parameters of the TNQAR model 4.1. The conditional quantile of y_{it} is modeled by $Q_{y_{it}}(\tau|Z_i, \mathbb{Y}_{t-1}) = g(\tau, r(\tau))$. $g(\cdot,\cdot)$ is a piecewise linear process, defined by $g(\tau, r(\tau)) = X_{it}^\top(r(\tau))\theta(\tau)$, where $\theta(\tau) = [\beta_0(\tau), \gamma^\top(\tau), \beta_1(\tau), \beta_2(\tau), \beta_3(\tau)]^\top \in \mathbb{R}^{q+4}$ and

$$X_{it}(r(\cdot)) = (1, Z_i^\top, n_i^{-1}\sum_{j=1}^{N} a_{ij}y_{j(t-1)}, y_{i(t-1)}, \mathbf{1}_{\{y_{i(t-1)}>r(\cdot)\}}y_{i(t-1)}) \in \mathbb{R}^{q+4},$$

with $\mathbf{1}\{\cdot\}$ being the indicator function. Note that $V_{it\tau} = y_{it} - g(\tau, r(\tau))$.

For models with a known threshold parameter $r_0 = r_0(\tau)$ i.e. different τ, the corresponding threshold value is given, the classical quantile autoregression estimation process can be applied. The estimation method is similar as the classic network QAR [52]. Then the parameter vector $\theta_{r_0}(\tau)$ can be estimated by

$$\hat{\theta}_{r_0}(\tau) = \arg\min_{\theta} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau\{y_{it} - X_{it}(r_0)^\top\theta\}, \tag{4.4}$$

where $\rho_\tau(u) = u\{\tau - \mathbf{1}(u<0)\}$ is the loss function for quantile regression in [28].

[52] demonstrates the asymptotic property of the standard linear network quantile autoregressive estimator. [14] proved the asymptotic property of the one-dimensional threshold quantile autoregressive estimator. In order to develop [14] to high-dimensional situations and develop [52] to regime switching framework, we need the following assumptions.

Assumptions:

A1: $\tau \in B \subset (0,1)$ and B is a compact set. $r(\tau)$ lies in a compact set $\mathcal{G} \subset \mathbb{R}$ for every $\tau \in B$ and $\theta(\tau) \in \Theta$, with Θ compact and convex;

A2: Let $F_{it}(\cdot|\mathcal{F}_t) = F_{it}(\cdot)$ denote the conditional distribution function of y_{it} given $\mathcal{F}_t$, $F_{it}(\cdot)$ has a continuous density of $f_{it}(\cdot)$ with $0 < f_{it}(u) < \infty$ on $U = \{u : 0 < F_t(u) < 1\}$ and f_{it} is uniformly integrable on U.

A3: Define

$$\hat{\Omega}_0(r, r^*) = (NT)^{-1}\sum_{it} X_{it}(r)X_{it}^\top(r^*)$$

and

$$\hat{\Omega}_1(\tau,r) = (NT)^{-1}\sum_{it} f_{it}(X_{it}^\top\theta(\tau))X_{it}(r)X_{it}^\top(r)$$

converge almost surely to $\Omega_0(r,r^*) = E[X_{it}(r)X_{it}^\top(r^*)]$ and $\Omega_1(\tau,r) = E[f_{it}(X_{it}^\top\theta(\tau))X_{it}(r)X_{it}^\top(r)]$ respectively, for any given $r(\tau)\in\mathcal{G}$ and all $\tau\in B$. Define $\Sigma_\theta(\tau,r) = \Omega_1^{-1}(\tau,r)\Omega_0(r,r)\Omega_1(\tau,r)^{-1}$. A1 imposes that $r(\cdot)$ lies on a compact set. This assumption was used by [14]. A2 is common in the QR literature. A3 guarantees the consistency of the variance parameter estimators $\Omega_0(r,r)$ and $\Omega_1(\tau,r)$ in the parameter space $\mathcal{G}\times B$.

A4: For all $\tau\in B$, $(\theta_0(\tau),r_0(\tau)) = \arg\min_{(\theta,r)}E[\sum_i\rho_\tau(V_{it\tau})]$ exists and is unique.

A4 guarantees that for each threshold value and quantile, this TNQAR problem has a unique solution.

These assumptions A1-A4 are common in quantile regression and regime switching literature. These assumptions are used in [14].

(C1) (Nodal Assumption) Assume that the Z_i's are independent and identically distributed random vectors with a mean of zero and a covariance of $\Sigma_z\in\mathbb{R}^{p\times p}$. In addition, its fourth-order moment is finite. The same settings are used for $V_{it\tau}$ across both $1\le i\le N$ and $0\le t\le T$. Also we need $\{Z_i\}$ and $\{V_{it\tau}\}$ are independent of each other.

(C2) (Connectivity) Think of W as a transfer matrix of a Markov chain whose state space is the nodes in the network. We assume that this Markov chain is irreducible and aperiodic. In addition, we define π as the stationary distribution of this Markov chain. And $\sum_{i=1}^N\pi_i^2$ converges to zero as N tends to infinity.

(C3) (Sparsity) Define $W^* = W + W^\top$ is a symmetric matrix, assuming $\lambda_1(W^*) = O(\log N)$

(C4) (Monotonicity) It is assumed that $X_{it}^\top\theta(\tau)$ $(1\le i\le N, 1\le t\le T)$ is a monotone increasing function with the respect to $\tau\in(0,1)$.

These assumptions follow the assumption from [50], which has a detailed description.

Therefore, lemma 4.3.1 shows that when r_0 is known, the asymptotic distribution of the estimator $\hat{\theta}_{r_0}(\tau)$ follows a normal distribution. This is similar to Corollary 4.1 in [52].

Lemma 4.3.1. *Given assumptions C1-C4 and A1-A4, and for r_0 known with $\tau \in B$ fixed,*

$$\sqrt{NT}\{\hat{\theta}_{r_0}(\tau) - \theta_{r_0}(\tau)\} \xrightarrow{d} N(0, \tau(1-\tau)\Sigma_\theta(\tau, r_0)),$$

as $min\{N, T\} \to \infty$.

Proof. The proof of this result uses standard NQAR asymptotic theory. This proof is a simple extension of Appendix B.2 in [52] for the asymptotic normality in Network Quantile Autoregressive models. □

The next thing to discuss is when the threshold value $r(\tau)$ is unknown and needs to be estimated. The estimator of TNQAR model and the threshold parameter are given by

$$(\hat{\theta}(\tau), \hat{r}(\tau)) = \arg\min_{(\theta,r)} \sum_{it} \rho_\tau(y_{it} - X_{it}^\top(r)\theta).$$

The above estimator problem can be translated into a two-stages method. For fixed τ, consider the r value of a set of grids on the real value line and for each r, the estimator can be obtained by model 4.4, and save $\hat{\theta}_r(\tau)$. Next, minimize

$$\hat{r}(\tau) = \arg\min_r \sum_{it} \rho_\tau(y_{it} - X_{it}^\top(r)\hat{\theta}_r(\tau)).$$

The following lemma gives the consistency of the estimator of $(\theta(\tau), r(\tau))$.

Lemma 4.3.2. *Given the assumptions A1-A4 and C1-C4, fixed $\tau \in B$,*

$$(\hat{\theta}(\tau), \hat{r}(\tau)) = (\theta_0(\tau), r_0(\tau)) + o_p(1)$$

Proof. Define $\mathbb{E}_T = (NT)^{-1}\sum_{it}\delta_{x_{it}}$, where δ_x assigns mass 1 at x and zero elsewhere, such that for any class $\mathcal{F}$ of the measurable function $f : \chi \to \mathbb{R}$. This is the same as maximizing $M_T(\mu) = \mathbb{E}_T m_\mu$, where $m_\mu(x_{it}) = -(\rho_\tau(y_{it} - X_{it}(r)^\top\theta(\tau)) - \rho_\tau(y_{it} - X_{it}(r_0)^\top\theta_0(\tau)))$.

We need to establish the conditions of the argmax theorem, which will give the consistency of the estimator.

For each $\tau \in B$, $(\hat{\theta}(\tau), \hat{r}(\tau))$ minimize $Q_T(\theta, r) = \mathbb{E}_T(\rho_\tau(y_{it} - X_{it}(r)^\top\theta(\tau)) - \rho_\tau(y_{it} - X_{it}(r_0)^\top\theta_0(\tau)))$. Define $Q(\theta, r) = E(\rho_\tau(y_{it} - X_{it}(r)^\top\theta(\tau)) - \rho_\tau(y_{it} - X_{it}(r_0)^\top\theta_0(\tau)))$. By assuming that A2 and A4, $Q(\theta, r)$ is minimized only at $(\theta_0(\tau), r_0(\tau))$ for each $\tau \in B$.

We will use these notations $\mu = (\theta, r)$ where $\theta = (\alpha, \beta)$. Fix a compact set $K \subset \mathcal{G} \times \Theta$, $\mathcal{F}_K = \{m_\mu : \mu \in K\}$ is Glivenko-Cantelli, since our $m_\mu(x_{it})$ and $m_\mu(x)$ in the literature [14] can be regarded as the same and they have the same properties. The existency of a solution $(\hat{\theta}, \hat{r})$ is also provided by [14].

Therefore, by argmax theorem from Lemma 1 in [8] , $(\hat{\theta}(\tau), \hat{r}(\tau)) \xrightarrow{P} (\theta_0(\tau), r_0(\tau))$.

□

In the proof of the next theorem 4.3.1, we need to use lemma 4.3.3 that shows $\hat{\theta}_r(\tau)$ has a Bahadur representation.

Lemma 4.3.3. *Suppose assumptions A1-A4 and C1-C4,*

$$\sqrt{NT}(\hat{\theta}(\tau) - \theta(\tau)) = \hat{\Omega}_1^{-1}(\tau, r) S_{NT}(\tau, r) + o_p(1),$$

where $S_{NT}(\tau, r) = \frac{1}{\sqrt{NT}} \sum_{it} \{X_{it}(r) \Psi_\tau(y_{it} - F_{it}^{-1}(\tau))\}$ *is the score function and* $\Psi_\tau(u) = \tau - I(u < 0)$.

Proof. Inspired by the proof of Lemma 3 in [14], define $\hat{\Omega}_1(\tau, r) = \frac{1}{NT} \sum_{it} f_{it}(X_{it}^\top(r)\theta_0) X_{it}(r) X_{it}^\top(r)$ and $S_{NT}(\tau, r) = \frac{1}{NT} \sum_{it} \Psi_\tau(y_{it} - X_{it}^\top(r)\theta_0) X_{it}(r)$, where $\Psi_\tau(u) = \tau - I(u < 0)$ is the influence function of the quantile regression model. Due to Lemma 2 that the estimator is a consistent estimate, we define $\alpha = (\beta_0, \gamma, \beta_1, \beta_2, 0)$ and $\beta = (\beta_0, \gamma, \beta_1, \beta_2, \beta_3)$ as subsets of θ that correspond to the regimes $\{X_{it}(r) | y_{i(t-1)} > r\}$ and $\{X_{it}(r) | y_{i(t-1)} \leq r\}$, respectively, such that $\theta_{0\tau} = (\alpha_{0\tau}^\top, \beta_{0\tau}^\top)^\top$ as true parameter vectors.

Also define the following weighted quantities,

$$\hat{\Delta}_{\tau r} = \begin{pmatrix} \sqrt{NT}(\hat{\alpha}_{\tau r} - \alpha_{0\tau}) \\ \sqrt{NT}(\hat{\beta}_{\tau r} - \beta_{0\tau}) \end{pmatrix}, \Delta_\tau = \begin{pmatrix} \sqrt{NT}(\alpha_{\tau r} - \alpha_{0\tau}) \\ \sqrt{NT}(\beta_{\tau r} - \beta_{0\tau}) \end{pmatrix}.$$

Let $y_{it}^* = y_{it} - X_{it}^\top(r)\theta_{0\tau}$ and $y_{itr}^*(\Delta_\tau) = y_{it}^* - X_{it}^\top(r)\Delta_\tau/\sqrt{NT} = y_{it} - X_{it}^\top(r)\theta(\tau)$. Therefore, $\hat{\Delta}_{\tau r} = \min_{\Delta_\tau} \sum_{it} \rho_\tau(y_{itr}^*(\Delta_\tau))$. Let

$$\begin{aligned} V_n(\tau, r, \Delta_\tau) &= \frac{1}{\sqrt{NT}} \sum_{it} \Psi_\tau(y_{itr}^*(\Delta_\tau)) X_{it}(r) \\ &= \frac{1}{\sqrt{NT}} \sum_{it} \Psi_\tau\left(y_{it} - X_{it}^\top(r)\left(\theta_{0\tau} + \frac{\Delta_\tau}{\sqrt{NT}}\right)\right) X_{it}(r), \end{aligned}$$

$$\bar{V}_n(\tau, r, \Delta_\tau) = \frac{1}{\sqrt{NT}} \sum_{it} E[\Psi_\tau\left(y_{it} - X_{it}^\top(r)\left(\theta_{0\tau} + \frac{\Delta_\tau}{\sqrt{NT}}\right)\right) X_{it}(r)].$$

Because $-\Delta_\tau^\top V_n(\tau, r, \lambda\Delta_\tau)$ is an increasing function of $\lambda > 1$ and A3 guarantee the consistency of $\hat{\Omega}_1(\tau, r)$, Lemma 4.3.3 can be proven by applying Lemma 4.3.4 (i.e Lemma A.4 of [30]). This requires the assumptions A1-A4 and C1-C4, lemma 2 and 3 of [40] and lemma A1 and A2 of [14]. □

Lemma 4.3.4. *Let $V_n(\Delta)$ be a vector function that satisfies*

i) $-\Delta^\top V_n(\lambda\Delta) \geq -\Delta^\top V_n(\Delta), \lambda \geq 1$,

ii) $\sup_{\|\Delta\| \leq M} \|V_n(\Delta) + f(F^{-1}(\tau))D\Delta - A_n\| = o_p(1)$, where $\|A_n\| = o_p(1), 0 < M < \infty, f(F^{-1}(\tau)) > 0$, and D is a positive-definite matrix. Suppose that Δ_n is a vector such that $\|V_n(\Delta_n)\| = o_p(1)$. Then, $\|\Delta_n\| = o_p(1)$ and $\Delta_n = \frac{D^{-1}}{f(F^{-1}(\tau))} A_n + o_p(1)$.

Theorem 4.3.1 shows the asymptotic distribution of $\hat{\theta}_r(\tau)$ in the high-dimensional case of the network.

Theorem 4.3.1. *Given assumption A1-A4 and C1-C4,*

$$\sqrt{NT}\{\hat{\theta}_r(\tau) - \theta_r(\tau)\} \xrightarrow{d} B(\tau, r)$$

as min$\{N, T\} \to \infty$, where $B(\tau, r)$ is a bivariate Gaussian process, with mean zero and covariance kernel defined by $K((\tau_i, r_i), (\tau_j, r_j)) = E(B(\tau_i, r_i)B(\tau_i, r_j)) = (\tau_i \wedge \tau_j - \tau_i\tau_j)\Omega_1(\tau_i, r_i)^{-1}\Omega_0(r_i, r_j)\Omega_1(\tau_i, r_j)^{-1}$ with $\tau_i, \tau_j \in B$ and $r_i, r_j \in \mathcal{G}$.

And naturally, we can obtain the following lemma.

Lemma 4.3.5. *Given assumptions C1-C4 and A1-A4, for a fixed pair $(\tau, r) \in \mathcal{G} \times B$*

$$\sqrt{NT}(\hat{\theta}(\tau) - \theta(\tau)) \xrightarrow{d} N(0, \tau(1-\tau)\Sigma_\theta(\tau, r)).$$

Proof. Now we prove Theorem 4.3.1 and Lemma 4.3.5.

Fixed $r \in \mathcal{G}$ for certain $\tau \in B$ given. By Lemma 4.3.3,

$$\sqrt{NT}(\hat{\theta}(\tau) - \theta(\tau)) = \hat{\Omega}_1^{-1} S_{NT}(\tau, r) + o_p(1),$$

where $S_{NT}(\tau, r) = \frac{1}{\sqrt{NT}} \sum_{i,t} X_{it}(r)\Psi_\tau(y_{it} - F_{it}^{-1}(\tau))$.

By the law of iterated expectations $E[X_{it}(r)\Psi_\tau(y_{it} - F_{it}^{-1}(\tau)|\mathcal{F}_t)] = 0$

According to the central limit theorem martingale difference sequences, Slutsky's Theorem and C1-C4 and A1-A4,

$$\frac{1}{\sqrt{NT}}\hat{\Omega}_1^{-1}(\tau, r)\sum_{it} X_{it}(r)\Psi_\tau(y_{it} - F_{it}^{-1}(\tau)) \xrightarrow{d} N(0, \Sigma(\tau, r)),$$

with $\Sigma(\tau, r) = \tau(1-\tau)\Omega_1(\tau, r)^{-1}\Omega_0(r, r)\Omega_1(\tau, r)^{-1}$. This proves Lemma 4.3.5. Next we can extend the results to the corresponding functional process indexed by τ and r, where τ and r belong to B and $\mathcal{G}$ respectively. With Lemma A2 in [14], it is possible that this class of function belongs to the Donsker class. Thus, this process converges in distribution to the Skorohod space $D(B, \mathcal{G})$, equips the uniform norm, to a bivariate Gaussian process with zero mean and covariance kernel

$$K((\tau_i, r_i), (\tau_j, r_j)) = (\tau_i \wedge \tau_j - \tau_i\tau_j)\Omega_1(\tau_i, r_i)^{-1}\Omega_0(r_i, r_j)\Omega_1(\tau_i, r_j)^{-1},$$

for every $i, j = 1, \cdots, n$ with $\tau_i, \tau_j \in B$ and $r_i, r_j \in \mathcal{G}$. Finally, we obtain

$$\sqrt{NT}(\hat{\theta}_r(\tau) - \theta_r(\tau)) = \hat{\Omega}_1^{-1}(\tau, r)S_{NT}(\tau, r) + o_p(1),$$

that converges in distribution to the bivariate Gaussian process $B(\tau, r)$ with mean zero and covariance kernel $K((\tau_i, r_i), (\tau_j, r_j))$. □

By fixing the values of τ and r, the asymptotic distribution of the estimator $\hat{\theta}(\tau)$ can be immediately derived.

4.3.2 Heavy-tailed Distribution and Tail Index Estimator

Heavy tailed phenomena appear in almost all areas of life. Data in many fields such as meteorology, hydrology, environment, telecommunications, insurance, finance, etc. do not meet the normal distribution assumption, but have heavy tails. If the data distribution is not normal and has the power behaviour in tails, we cannot use OLS to obtain the correct parameter estimations. The term robustness of estimation has been recognised as one of the main problem that we need to overcome in statistics. In order to obtain reliable results, the quantile regression can be used. Therefore, before making an estimate, a normality test needs to be performed on the data set and obtain all the information of tail distribution.

4.3.2.1 Normal Test

The Kolmogorov-Smirnov test can be used. Here we skip the introduction for the Kolmogorov-Smirnov test.

4.3.2.2 Hill Estimator

Since the data to be studied does not follow the normal distribution, it is necessary to know the information as much as possible about the tail of the distribution. Therefore, the heavy tail index is an important method for us to obtain the tail information of the distribution. The Hill estimator is used to obtain the tail index. The process is as following [23]'s method: reorder the set X_i in such way that

$$X_i \geq X_j \quad \text{for} \quad i < j$$

i.e. the set X_i is ordered decreasingly. Then [10] shows the Hill estimator $\gamma(m)$ can be obtained by

$$\gamma(m) = \frac{1}{m}\sum_{i=1}^{m} \log \frac{X_i}{X_m}, \tag{4.5}$$

where m is the truncation number of the tail data that we choose.

The Hill estimator depends on the correct choice of m so it is important to choose a reasonable m. The preliminary judgment in experience is to choose 2% -10% of the order statistics. Phillips et al. [37] proposed an optimal choice of m which gives minimum mean squared error and the formula is $m = [\lambda n^{2/3}]$, where [ ] signifies the integer part of its argument and n is the total number of data. [15] show that the parameter λ can be estimated by the following formula:

$$\hat{\lambda} = \left| \frac{\hat{\gamma_2}}{\sqrt{2}(n/m_1)(\hat{\gamma_1} - \hat{\gamma_2})} \right|^{2/3}$$

Here $\hat{\gamma_1}$ and $\hat{\gamma_2}$ are preliminary estimates of γ by using formula (4.5) with data truncations, $m_1 = [n^A]$ and $m_2 = [n^B]$, respectively, where $0 < A < 2/3$ and $2/3 < B < 1$. A and B are advised to be set as 0.6 and 0.9 respectively. Applying these into our Twitter data in Chapter 3, the Hill estimator is about 38 which means the twitter response has a heavy tail.

4.4 Finite Sample Simulation Analysis

4.4.1 Simulation Models

We propose a Monte Carlo experiment to analyze the finite sample properties of the TNQAR model. We generate a set of time series through a given model and estimate the parameters of the corresponding TNQAR model.

Consider the following model:

$$\begin{aligned} y_{it} = 0.3 \quad &+ \quad \sum_{i} Z_{il}\gamma_l + 0.1 n_i^{-1} \sum_{j} a_{ij} y_{i(t-1)} \\ &+ \quad 0.5 y_{i(t-1)} + 0.1 I_{\{y_{i(t-1)}>r_0\}} y_{it} + u_{it}, i = 1, \cdots, N \end{aligned} \tag{4.6}$$

where u_{it} is the quantile equation of the error term, $r_0 = 0$ and $\gamma = (-0.5, 0.3, 0.8, 0, 0)^\top$. By comparing with (4.1), we have $\beta_0(\tau) = 0.3 + F_u^{-1}(\tau), \beta_1(\tau) = 0.1, \beta_2(\tau) = 0.5, \beta_3(\tau) = 0.1$. Note that we fix the dimension of nodal covariates (i.e., Z_i) to be 5.

To generate observations from the TNQAR mechanism (4.1), the following procedures are performed. First, we generate $u_{it}s$ $(1 \leq i \leq N, 1 \leq t \leq T)$ independently from a standard normal distribution $N(0, 1)$ and t-distribution with 5 degrees of freedom in order to discuss the influence of different type of error on the estimation results.

Next, the nodal covariates $Z_i = (Z_{i1}, \cdots, Z_{i5})^\top \in \mathbb{R}^5$ are sampled from a multivariate normal distribution $N(0, \Sigma_z)$, where $\Sigma_z = (\delta_{j_1 j_2})$ and $\delta_{j_1 j_2} = 0.5^{|j_1 - j_2|}$. We set the network size to $N = 100, 500, 1000$ and the sample size to $T = N/10$ and the number of repeated experiments to $R = 1000$. The estimator from the rth replication is recorded as $\hat{\theta}^{(r)}(\tau) = \{\hat{\beta_0}^{(r)}(\tau), \hat{\beta_1}^{(r)}(\tau), \hat{\beta_2}^{(r)}(\tau), \hat{\beta_3}^{(r)}(\tau), \hat{\gamma}^{(r)\top}(\tau)\}$. In order to evaluate the performance of finite sample, we consider the following indicator. The root mean square error (RMSE) of β_j: $RMSE_j(\tau) = \{R^{-1} \sum_r (\hat{\beta}_j^{(r)}(\tau) - \beta_j(\tau))^2\}^{1/2}$, $j = 0, 1, 2, 3$. In addition, inspired by [52], the RMSE for the nodal effect equation γ is determined by $RMSE_\gamma(\tau) = \{(5R)^{-1} \sum_r ||\hat{\gamma}^{(r)}(\tau) - \gamma(\tau)||^2\}^{1/2}$.

We simulate a time series with a sample size of 100+T from model (4.6), $\mathbb{Y}_0$ obtained by (2.4) in [50]. We only save the last T values to remove the effect of the initial value. We adopt two different kinds of adjacency matrices, see example 1, 2 in [50] and [52] for details.

4.4.2 Simulation Results

Now we start analyzing the results of the finite sample simulation. Detailed results are shown in 4.1. When β_0, τ increase from 0.05 to 0.95, the RMSE start with becoming smaller and then becoming larger, and reach the minimum at τ=0.5. For example, in the case of a normal distribution with $N = 100$, the RMSE drops from 2.13 to 0.59 and then rises to 2.39.

For a fixed τ, when N and T increase, the corresponding RMSE is decreased. For example, at $\tau = 0.5$ in Example 1 for the normal distribution, the RMSE of β_0 decreases from 0.59 to 0.14 as N changes from 100 to 1000. This shows that the estimator will become more accurate when the sample size becomes larger. In addition, the RMSE of the t-distribution is generally larger than the RMSE of the normal distribution. It is worth noting that the RMSE of γ does not change much when N becomes large, both in the normal distribution and in the t-distribution. Theoretically $\beta_1, \beta_2, \beta_3$, and γ do not change with τ, so $\beta_1, \beta_2, \beta_3$, and γ in $\tau = 0.05$ and those in other τ situation are the same, then only the case of $\tau = 0.05$ is listed here to represent the results of $\beta_1, \beta_2, \beta_3$, and γ in all τ. In addition, the incentive fee is shown to have a negative correlation with hedge fund returns at $\tau = 0.05$ and $\tau = 0.25$, but this phenomenon is not appear at upper tail and median.

Table 4.1: Simulation results for Example 1 with 1000 replications

	β_0	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
N=100	Z	2.13 (88.00)	0.96 (91.30)	0.59 (93.00)	1.20 (89.70)	2.39 (84.80)
	T	2.40 (87.20)	1.00 (91.00)	0.57 (94.50)	1.24 (91.50)	2.66 (87.40)
N=500	Z	2.22 (100.00)	0.90 (100.00)	0.20 (100.00)	0.99 (100.00)	2.30 (100.00)
	T	2.47 (100.00)	0.95 (100.00)	0.24 (100.00)	1.07 (100.00)	2.55 (100.00)
N=1000	Z	2.24 (100.00)	0.92 (100.00)	0.14 (100.00)	0.95 (100.00)	2.28 (100.00)
	T	2.58 (100.00)	0.98 (100.00)	0.11 (100.00)	1.03 (100.00)	2.59 (100.00)
	$\tau = 0.05$	β_1	β_2	β_3	γ	
Z	N=100	0.41 (95.00)	0.12 (97.00)	0.33 (96.80)	0.45	
	N=500	0.16 (100.00)	0.05 (100.00)	0.14 (100.00)	0.44	
	N=1000	0.11 (100.00)	0.04 (100.00)	0.11 (100.00)	0.44	
T	N=100	0.39 (94.80)	0.12 (95.70)	0.25 (96.50)	0.46	
	N=500	0.20 (100.00)	0.04 (100.00)	0.10 (100.00)	0.44	
	N=1000	0.09 (100.00)	0.03 (100.00)	0.08 (100.00)	0.44	

4.5 Empirical Application

Next, we apply our methodology to study the return rates of global hedge funds that have the common strategy. The data set includes 915 hedge funds and the sample interval is from January 2007 to December 2009. The number of samples in this time series is 36. The corresponding response is the monthly rate of return. The network density of these hedge funds is 0.11%.

Figure 4.1 shows the time series of average returns of global hedge funds. It can be seen that hedge funds experienced a large loss from July to October in 2008 due to the global financial crisis. In order to construct the network structure, the strategies used by hedge funds are collected. For ith and jth funds, $a_{ij} = 1$ if they use the same hedging strategy, otherwise $a_{ij} = 0$. In addition, two variables that do not change over time are considered. They are management fees and incentive fees. These two variables are normalized into the [0,1] interval.

Descriptive statistics show that during the entire sample period, the mean of the return rate is 0.49%, the skewness and kurtosis are 1.27 and 61.47, and the p-value of Jarque-Bera normality test is less than 2.2×10^{-16}. It can be seen that the monthly return rate of the hedge fund does not follow the normal distribution and has a high kurtosis and positive skew. In addition, the Hill estimator is 3.3, which means that the distribution of hedge fund returns is heavy-tailed. Because it does not satisfy the error term following the normal distribution, it is one of the reasons that we consider using the quantile regression method.

We applied the LM test in (3.48) to the hedge fund return, which showed significant for the hedge fund return. The LM value is 16.18115, which less than 95% quantile of the chi-square distribution with a degree of freedom of one (3.841459). The LM test results prove that the hedge fund return is nonlinear, so it is reasonable to use the proposed TNQAR model for modelling the hedge fund return.

We use threshold value r=0 according to the most general idea and $\tau = 0.05, 0.25, 0.5, 0.75$ and 0.95 to fit the return of the global hedge fund, where the estimates of these parameters are shown in the Table 4.2. It is worth noting that the β_3 (i.e. threshold effect) is significant in α=0.10 (the significant level), in the median and lower tail case (i.e. $\tau = 0.5, 0.25$ and 0.05), but not very significant in the upper tail (i.e. $\tau = 0.75$ and 0.95), which indicates that the return of global hedge funds, in low conditional quantile level, present the threshold effect.

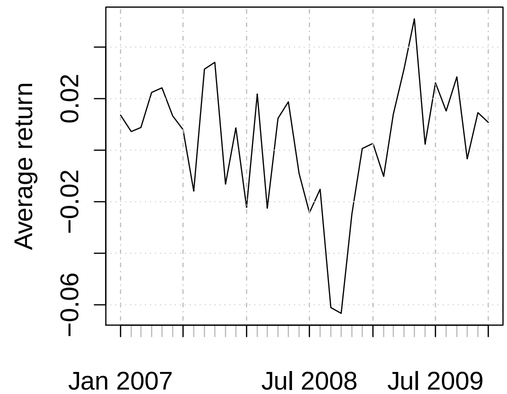


Figure 4.1: Average return on global hedge funds

In addition, the management fee is significantly (<0.01) correlated at the lower and median tail case (i.e. τ=0.05, 0.25 and 0.5), but not significant (p value 0.28 and 0.56) at the upper tail case (i.e. τ=0.75 and 0.95), and the incentive fee is significantly (<0.01) correlated at the upper tail and not significantly at the lower tails. This shows that the return of global hedge fund is more relevant to the management fee in the conditional quantile level at τ=0.05, 0.25 and 0.5, and is more closely related to the incentive fee in the conditional quantile at τ=0.75 and 0.95.

Table 4.2: Empirical analysis

	$\tau = 0.05$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.95$	
	Est.	p-value	Est.	p-value	Est.	p-value	Est.	p-value	Est.	p-value
β_0	-0.05 (0.00)	<0.01	-0.01 (0.00)	<0.01	0 (0.00)	0.99	0.01 (0.00)	0.99	0.04 (0.00)	0.99
β_1	0.66 (0.01)	0.99	0.22 (0.01)	0.99	0.12 (0.00)	0.99	0.12 (0.01)	0.99	0.18 (0.06)	0.99
β_2	0.61 (0.08)	0.99	-0.25 (0.06)	<0.01	0.14 (0.01)	0.99	0.01 (0.02)	0.75	-1.32 (0.08)	<0.01
β_3	-0.5 (0.07)	<0.01	0.44 (0.06)	0.99	0.07 (0.01)	0.99	0.18 (0.02)	0.99	1.49 (0.07)	0.99
Management Fee	0.00 (0.01)	0.44	0 (0.00)	0.14	0 (0.00)	0.93	0.01 (0.00)	0.99	0.01 (0.01)	0.73
Incentive Fee	-0.08 (0.01)	<0.01	-0.01 (0.00)	<0.01	0 (0.00)	0.99	0.02 (0.00)	0.99	0.09 (0.00)	0.99

4.6 Conclusion

In this chapter, we introduce the Threshold Network Quantile Autoregressive (TNQAR) model. We give a sufficient condition for a geometric ergodicity. We discussed the parameter estimation method of this model and provided the asymptotic properties of this model. In addition, we introduced a tail index estimator, i.e. the Hill estimator, to obtain the tail information of a distribution. In addition, we also carry out some simulation experiments to analyze the finite sample properties of the TNQAR model. In the empirical application, we applied our methodology to modelling the returns of global hedge funds.

Chapter 5

Markov-Switching Network Autoregression Model

5.1 Introduction

In this chapter, we are interested in the network vector Autoregression model with random coefficients. Similar random coefficient problems are discussed in Chapter 3, where coefficients in the NAR model are affected by a threshold value. The difference between the two models is that the coefficients, in the class of threshold models, are affected by endogenous variables, while the coefficients, in the class of Markov-switching models, are affected by exogenous variables. What they have in common is that they both describe the dynamic characteristics of regime switching and can explain sudden changes.

The Markov-Switching Network Autoregression Model (MS-NAR) belongs to the category of the Markov-switching model. The Markov-switching model proposed by [16] is a relatively popular type of nonlinear time series model in econometrics. Much work has done by [16], [17], [12], [19], [18]. [27] discussed the related literature overview of Markov-switching.

[31] applied Markov-switching to the VAR model and established the MS-VAR model, which makes the estimated parameters in the traditional VAR model change with the change of the regime. [49] discusses the stationarity conditions of the MS-VAR model. However, as the dimension of the VAR model increases, the number of parameters that need to be estimated becomes extremely large, causing

the estimation method to fail.

In this chapter, we propose a new high-dimensional Markov-switching model. In this model, the autoregressive coefficients switch between the two states. The number of parameters do not increase as the number of dimensions increases. We study the probability property (ergodicity and stationary) of the MS-NAR model first, and then discuss its maximum likelihood estimation. To illustrate the nature of our model and estimation methods, we perform a series of simulations and fit a simple MS-NAR model into a real data set.

The rest of this chapter is organized as follows. In the second section, we introduce our model and provide the stationary condition of MS-NAR model. The third section introduces the estimation method of the model. The fourth section reports on our simulation study.

5.2 Markov-Switching Network Autoregression Model

5.2.1 Definition of the Model

We consider a first-order Network Vector autoregressive process in which the autoregressive parameters change with state:

$$y_{it} = \beta_0 + Z_i^\top \gamma + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} y_{j(t-1)} + \beta_2(s_t) y_{i(t-1)} + \varepsilon_{it}, \tag{5.1}$$

where

- $n_i = \sum_{j \neq i} a_{ij}$ is the out-degree for the ith node. [47]
- ε_{it} is the error term independent of Z_is, which follows normal distribution with $E(\varepsilon_{it}) = 0$ and $\text{var}(\varepsilon_{it}) = \sigma^2$.
- $Z_i = (Z_{i1}, \cdots, Z_{ip})^\top$ is a random vector with p-dimensional node feature. $\gamma = (\gamma_1, \cdots, \gamma_p)^\top$ is the corresponding coefficient of p-dimension. $\beta_0 + Z_i^\top \gamma$ constitutes the intercept of the ith node.

- The regime process s_t taking values in $\{1, \cdots, K\}$ is characterized as an unobservable, irreducible, aperiodic K-state Markov chain with s_t independent of ε_τ for all t and τ.

- The state transition is driven by a stationary first-order K-state Markov chain $\{s_t\}$, and its transition probability matrix is $P = (p_{ij})_{K\times K}$, where the probability of transitioning from state i to state j is

$$p_{ij} = P(s_t = j|s_{t-1} = i), \quad i, j = 1, \cdots, K.$$

- In addition, the initial distribution of the first-order Markov chain s_t is

$$P(s_0 = j) = \pi_j, \quad j = 1, \cdots, K.$$

  Denote $\boldsymbol{\pi} = [\pi_1, \cdots, \pi_K]^\top$.

Remark 5.1. The state s_t is independent of the node, that is, the state of all nodes is uniform.

We can write the model 5.1 in vector form as:

$$\mathbb{Y}_t = \mathcal{B}_0 + \mathbb{G}(s_t)\mathbb{Y}_{t-1} + u_t, \tag{5.2}$$

where

- $\mathbb{Y}_t = (y_{1t}, y_{2t}, \cdots, y_{Nt})^\top$,
- $\mathcal{B}_0 = \beta_0\mathbf{1} + \mathbb{Z}\gamma$, where $\mathbf{1} = (1, \cdots, 1)^\top$ and $\mathbb{Z} = (Z_1, \cdots, Z_N)^\top$,
- $\mathbb{G}(s_t) = \beta_1 W + \beta_2(s_t)I_N$, where $W = \{n_i^{-1}a_{ij}\}_{N\times N}$,
- $u_t = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})^\top$ and u_t satisfies

$$E(u_t) = 0, \quad E(u_t u_t^\top) = \sigma^2 I_N.$$

5.2.2 Stationarity

Next we will show the stationarity of the MS-NAR model. Note that when N is fixed, the model 5.2 is a special case of the MS-VAR model, so the stationarity condition of the MS-VAR model also applies to model 5.2.

Assumption 5.2.1. $max_{i=1,\cdots,K}\{\sum_{j=1}^{K} p_{ij}(|\beta_1| + |\beta_2(j)|)^2\} < 1$

Theorem 5.2.1. *Under assumption 5.2.1, model 5.2 has a stationary solution, and the solution can be expressed as follows:*

$$\mathbb{Y}_t = \sum_{j=0}^{\infty} \Pi_j(s_t) V_{t-j}, \tag{5.3}$$

where $V_t = \mathcal{B}_0 + u_t$ *and*

$$\Pi_j(s_t) = \begin{cases} I_N & \text{if } j = 0 \\ \mathbb{G}(s_t)\mathbb{G}(s_{t-1})\cdots\mathbb{G}(s_{t-j+1}) & \text{if } j \geq 1 \end{cases} \tag{5.4}$$

Regarding Theorem 5.2.1, we have the following remarks.

It can be seen that the stationarity depends on β_1, $\beta_2(s_t)$ and P. It is not necessary to satisfy the stationary condition of [50], that is $|\beta_1| + |\beta_2| < 1$, in both states, as long as the mean magnitude of the coefficient matrix is less than 1. Even if $|\beta_1| + |\beta_2| > 1$ in a certain state, as long as the corresponding transition probability is small enough, the overall stationary of model 5.2 will not be affected.

Because we used the sufficient condition of the stationarity in [49], it is conceivable that stationary conditions can continue to be weakened.

Proof. Set $\Phi = \begin{bmatrix} \mathbb{G}(1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbb{G}(K) \end{bmatrix}_{KN\times KN}$, $\hat{\Phi} = \begin{bmatrix} \|\mathbb{G}(1)\| & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|\mathbb{G}(K)\| \end{bmatrix}_{K\times K}$, where $\|\cdot\|$ denotes the 1-norm of a vector (the sum of absolute values of all elements). Notice that $\|\beta_1 W + \beta_2(s_t)I\| \leq |\beta_1|\|W\| + |\beta_2(s_t)| = |\beta_1| + |\beta_2(s_t)|$ for $s_t = \{1, \cdots, K\}$ since $\|W\|$ and $\|I\|$ are both equal to 1.

According to the sufficient condition for the stationarity $\|\hat{\Phi}^2 P\| < 1$ in [49], $\|\hat{\Phi}^2 P\| = max_{i=1,\cdots,K}\{\sum_{j=1}^{K} p_{ij}\|\beta_1 W + \beta_2(j)I\|^2\} \leq max_{i=1,\cdots,K}\{\sum_{j=1}^{K} p_{ij}(|\beta_1| + |\beta_2(j)|)^2\} < 1$ meets the stationarity condition of $\{\mathbb{Y}_t\}$.

□

5.3 Maximum Likelihood Estimation of MS-NAR Model

For the MS-NAR model, $\mathbb{Y}_t$ is the time series we are preparing to study. We will refer to (5.1) as the Markov-Switching Network Vector Autoregresion (MS-NAR) model. Next, in the estimation method, we consider the case of $K = 2$. The model 5.1 can be represented as a matrix form:

$$\mathbb{Y}_t = \mathbb{X}_{t-1}\tilde{\theta}(s_t) + u_t, \tag{5.5}$$

where

- $\mathbb{X}_t = (\mathbf{1}, W\mathbb{Y}_t, \mathbb{Y}_t, \mathbb{Z})^\top$
- $\tilde{\theta}(s_t) = (\beta_0, \beta_1, \beta_2(s_t), \gamma^\top)^\top$

Assume that $f(\mathbb{Y}_t|s_t, \mathfrak{Y}_{t-1}, \theta)$ is the conditional probability density function of the random vector $\mathbb{Y}_t$, where $\mathfrak{Y}_{t-1} = \{\mathbb{Y}_{t-1}, \mathbb{Y}_{t-2}, \cdots\}$ and $\theta = (\tilde{\theta}(s_t)^\top, p_{11}, p_{22}, \pi_1, \pi_2, \sigma^2)^\top$, the parameter to be estimated. In the following, we will abbreviate the above density function $f(\mathbb{Y}_t|s_t, \mathfrak{Y}_{t-1}, \theta)$ into $f(\mathbb{Y}_t|s_t, \mathfrak{Y}_{t-1})$.

There are currently three methods for estimating the parameters of the MS-NAR model, namely Maximum Likelihood Estimation(MLE) in [16], EM algorithm in [17], and Gibbs sampler in [1].

In general, the EM algorithm is difficult to implement when there are AR items in the model and the Gibbs sampling algorithm requires a lot of computation. Therefore, we use the maximum likelihood (MLE) algorithm to estimate the parameters of the model. Therefore, we first need to obtain the likelihood function. Under the above assumption of the model, the corresponding conditional probability density function can be obtained:

$$f(\mathbb{Y}_t|s_t, \mathfrak{Y}_{t-1}) = (2\pi)^{-N/2}(\Sigma)^{-1/2}\text{exp}\left[-\frac{1}{2}(\mathbb{Y}_t - \mathbb{X}_{t-1}\tilde{\theta}(s_t))^\top\Sigma^{-1}(\mathbb{Y}_t - \mathbb{X}_{t-1}\tilde{\theta}(s_t))\right], \tag{5.6}$$

where $\Sigma = diag(\sigma^2, \cdots, \sigma^2)$.

Since s_t is an unobservable random variable, we cannot use $f(\mathbb{Y}_t|s_t, \mathfrak{Y}_{t-1})$ to construct a likelihood function when taking maximum likelihood estimation, so we need to get the density function $f(\mathbb{Y}_t|\mathfrak{Y}_{t-1})$.

The joint density function of $\{\mathbb{Y}_t, s_t\}$ under the condition of the past information set $\mathfrak{Y}_{t-1}$:

$$f(\mathbb{Y}_t, s_t|\mathfrak{Y}_{t-1}) = f(\mathbb{Y}_t|s_t, \mathfrak{Y}_{t-1}) \times P(s_t|\mathfrak{Y}_{t-1}), \tag{5.7}$$

so

$$f(\mathbb{Y}_t|\mathfrak{Y}_{t-1}) = \sum_{s_t=1}^{2} f(\mathbb{Y}_t, s_t|\mathfrak{Y}_{t-1}). \tag{5.8}$$

For i= 1,2, we call $P(s_t = i|\mathfrak{Y}_{t-1})$ the prediction probabilities of s_t and $P(s_t = i|\mathfrak{Y}_t)$ the filtering probabilities of s_t, so we have the following two equations:

$$P(s_t = i|\mathfrak{Y}_{t-1}) = p_{0i}P(s_{t-1} = 0|\mathfrak{Y}_{t-1}) + p_{1i}P(s_{t-1} = 1|\mathfrak{Y}_{t-1}) \tag{5.9}$$

and

$$P(s_t = i|\mathfrak{Y}_t) = \frac{P(s_t = i|\mathfrak{Y}_{t-1})f(\mathbb{Y}_t|s_t = i, \mathfrak{Y}_{t-1})}{f(\mathbb{Y}_t|\mathfrak{Y}_{t-1})} \tag{5.10}$$

by the Bayes theorem.

With the initial values $P(s_0 = i) = \pi_i$, we can derive the following log-likelihood function by iterating the equations (5.6)-(5.10):

$$L_T(\theta) = \sum_{t=1}^{T} \ln f(\mathbb{Y}_t|\mathfrak{Y}_{t-1}) \tag{5.11}$$

which is a complex function of θ. Some numerical search algorithms can be used to calculate this MLE $\hat{\theta}_T$. There are many programs to calculate this MLE, such as the GAUESS program or the R program, both of which apply the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. We use the second program to estimate the parameters of the model. The filtering and prediction probabilities are easily calculated by putting $\hat{\theta}_T$ into the above formula.

In order to calculate the smooth probability $P(s_t = i|\mathfrak{Y}_T)$, which is based on all the information in the sample, we use the algorithm of [26]. Notice that

$$\begin{aligned} P(s_t &= i|s_{t+1} = j, \mathfrak{Y}_T) && (5.12)\\ &= P(s_t = i|s_{t+1} = j, \mathfrak{Y}_t) && (5.13)\\ &= \frac{p_{ij}P(s_t = j|\mathfrak{Y}_t)}{P(s_{t+1} = j|\mathfrak{Y}_t)}. && (5.14) \end{aligned}$$

For i=1,2, the smooth probability can be expressed as

$$
\begin{aligned}
P(s_t &= i|\mathfrak{Y}_T) \\
&= P(s_{t+1}=0|\mathfrak{Y}_T)P(s_t=i|s_{t+1}=0,\mathfrak{Y}_T) \\
&\quad +P(s_{t+1}=1|\mathfrak{Y}_T)P(s_t=i|s_{t+1}=1,\mathfrak{Y}_T) \\
&= P(s_t=i|\mathfrak{Y}_t) \\
&\quad \times\left(\frac{p_{i0}P(s_{t+1}=0|\mathfrak{Y}_T)}{P(s_{t+1}=0|\mathfrak{Y}_t)}+\frac{p_{i1}P(s_{t+1}=1|\mathfrak{Y}_T)}{P(s_{t+1}=1|\mathfrak{Y}_t)}\right)
\end{aligned} \tag{5.15}
$$

In the above iteration on (5.9)-(5.15), the parameter vector θ was a fixed, known vector. Once the iteration has been completed for $t = 1, 2, \cdots, T$ for a given fixed θ, the value of the log likelihood implied by that value of θ is then known from (5.11). The value of θ that maximizes the log likelihood can be found using the numerical methods.

If the transition probabilities are restricted only by the conditions that $p_{ij} \leq 0$ and $(p_{i1} + p_{i2} + \cdots + p_{iN}) = 1$ for all i and j, then it is shown in [17] that the maximum likelihood estimates for the transition probabilities satisfy

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} P\{s_t = j, s_{t-1} = i|\mathfrak{Y}_T; \hat{\theta}\}}{\sum_{t=2}^{T} P\{s_{t-1} = i|\mathfrak{Y}_T; \hat{\theta}\}},$$

where $\hat{\theta}$ denotes the maximum likelihood estimates.

Remark 5.2. $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N}(\mathbb{Y}_t - \mathbb{X}_{t-1}\theta(j))^\top(\mathbb{Y}_t - \mathbb{X}_{t-1}\theta(j))P(s_t = j|\mathfrak{Y}_T; \hat{\theta})$.

5.4 Simulation

In this section, the Monte Carlo method is used to study the finite sample properties of the MLE method for the model (5.1). We are going to consider three examples in [50]. The main difference is the generation mechanism of the adjacency matrix. Besides, they are all similar. Data generation code and simulation code can be learned from GitHub webpage (https://github.com/mojianxiaocai/MS-simulation/tree/testing). The MLE is estimated by optimizing the log likelihood function and use the "optim" function in the R program, which applies the BFGS algorithm.

In particular, for each example, the random error ε_{it} is generated from the standard normal distribution N(0,1). The covariate $Z_i = (Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5}) \in \mathbb{R}^5$

is generated from a multivariate normal distribution with mean zero and covariance $\sigma_Z = (\delta_{j_1,j_2})$, which $\delta_{j_1,j_2} = 0.5^{|j_1-j_2|}$. Also, for each example, the parameters to be evaluated are set to $(\beta_0, \beta_1, \beta_2(1), \beta_2(2), p_{11}, p_{22}, \sigma_1, \sigma_2)^\top = (0.3, 0.2, 0.3, 0.7, 0.7, 0.4, 1, 1)^\top$ and γ is fixed to be $(-0.5, 0.3, 0.8, 0, 0)^\top$. In order to generate $\mathbb{Y}_t$, the initial value $\mathbb{Y}_0$ is randomly generated according to Proposition 1 in [50]. When $\mathbb{Y}_0$ is given, $\mathbb{Y}_t$ can be generated by model (5.1).

5.4.1 Example 1

We used the first example in [50] as our adjacency matrix - Dyad Independence Model. Fixed sample size T=20,30,100 and network size N=100,500,1000, we use the true parameters of the model as the initial values. The random repetition of the experiment R=1000 times with relative convergence tolerance being $1e^{-4}$. Table 5.2 reports the maximum likelihood estimates (MLEs) for the model 5.1 and their mean square error (MSE) (in parentheses) as well as average absolute error (AAE) for different N and T. The estimates of the parameters and the mean square error are averaged from the results of all replicate experiments. It can be seen from Table 5.2 that with the same sample size N, the mean square error (MSE) is significantly reduced as the number of samples T increases from 20 to 100. In addition, for the same sample size T, the MSE decreases as the network size N increases from 100 to 500. Table 5.3-5.5 show the distribution of maximum likelihood estimators for model 5.1 for different sample size T and network size N. Figure 5.1 provides the boxplot of the AAE of MLEs for model 5.1 for different sample sizes and network sizes. As the picture shows, AAE decreases as the sample sizes T increases from 20 to 100 and AAE did not increase or decrease significantly with the increase in network size. This shows that our estimation method will be more accurate as the sample size T increases, and as the network size N increases, it will not cause the estimation method to fail.

5.4.2 Example 2

We next consider the second example of [50], which is the stochastic block model. In order to generate this block network structure, we follow [34] to put each node with the block label from 1 to K, where $K \in \{5, 10, 20\}$. Let us set $P(a_{ij} = 1) =$

$0.3N^{-0.3}$ when i and j are in the same block, otherwise $P(a_{ij} = 1) = 0.3N^{-1}$. This means that nodes within the same block have a higher probability of being connected than between blocks. We use $(\beta_0, \beta_1, \beta_2(1), \beta_2(2), p_{11}, p_{22}, \sigma_1, \sigma_2, \gamma^\top)^\top = (0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 2, 2, 0.5, 0.5, 0.5, 0.5, 0.5)^\top$ as our initial value. The experiment was randomly repeated $R = 1000$ times with convergence relative tolerance 1e-4. Fixed T=30 and $N = 100, 300, 500$.

Table 5.6 gives the results of MLE and the mean square error (in parentheses) and the average absolute error of these estimators for different K (network densities) and different N (network size). As can be seen from Table 5.6, if T is fixed, AAE decreases as N increases. Therefore, when we meet a limited sample size, the averaged performance of the estimators can be improved by increasing network size N.

In addition, when N is fixed, there is no significant change in AAE for different network densities K, indicating that the network densities has no significant impact on the estimation results. The statement can also be verified by Figure 5.2.

Figure 5.2 shows the boxplot of AAE for Maximum likelehood estimator of model (2.1) for different network densities and network sizes for example 2. It can be seen from Figure 5.2 that as the network size N increases, the boxplot moves down as a whole.

5.4.3 Example 3

In this case, we consider the third example of [50], which is the Power-Law Distribution model, detailed in [9]. In order to generate the adjacency matrix A, first generate its in-degree $d_i = \sum_j a_{ji}$ for each node according to the discrete power-law distribution, that is, $P(d_i = k) = ck^{-\alpha}$. c is a normalizing constant. We set the exponent parameter $\alpha \in \{1.2, 2, 3\}$. Next, we randomly select d_i nodes as the followers of the ith node. We use $(\beta_0, \beta_1, \beta_2(1), \beta_2(2), p_{11}, p_{22}, \sigma_1, \sigma_2, \gamma^\top)^\top = (0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 2, 2, 0.5, 0.5, 0.5, 0.5, 0.5)^\top$ as our initial value. The random repetition of the experiment R=1000 times with relative convergence tolerance being $1e^{-4}$. Finally, fixed T=30 and N=100,300,500.

Table 5.7 shows the MLE with the mean square error in parentheses and AAE of MLE for for different α (network densities) and different N (network size). This table shows MSE of the estimators and AAE drop as network size N and α

increase.

Figure 5.3 shows the boxplot of AAE for Maximum likelihood estimator of model (2.1) for different network densities and network sizes for example 3. The result is basically similar to example 2, and increasing the sample size N will improve the performance of the MLE. The difference is that the increase in α will also improve the performance of the MLE.

5.5 Real Data Example

In this section, we applied the MS-NAR model to illustrate our methodology for Twitter data, which has been applied into TNAR model in Chapter 3. What interests us is whether the activities of the users are affected by nonlinearity. In other words, we want to know whether there is a potential two-state Markov chain effect, in addition to the normal linear NAR model effect.

Specifically, a total of N=9908 active followers of the Strathclyde official Twitter account are recorded for a continuous T=8 weeks. The network structure is defined to be the followee-follower relationship. The resulting network density is around 4.0%. The histogram of responses is plotted in Figure 3.1. It can be observed that the distribution of in-degrees is much skewer than the out-degrees, which indicates the existence of influential network users.

The response y_{it} is defined as the $log(1+x)$-transformed tweets length for the ith user at the tth week. The lag 1 term $y_{i(t-1)}$ is used to account for the autoregressive effect of t_{it}. In order to characterize the user bahaviors, two covariates Z_{it} are taken into account. They are the out-degrees and in-degrees. If the out-degree is high, it means that the user is very concerned about others. If the in-degree is high, the user is enthusiastic to follow others. The user features can also be characterized by other labels, such as the gender or age.

Table 5.1 shows the detailed estimation results of MS-NAR model defined by 5.5. The initial values of the parameters in the Twitter data set are the parameters estimated by the TNAR model. The $\hat{\beta}_2(2)$ is around -0.22, which means there exists a nonlinear effect. The two states are diametrically opposed. There is a positive momentum effect for user behaviors in state 1 , while the momentum effect is negative in state 2. For the nodal covariates, it can be observed that the

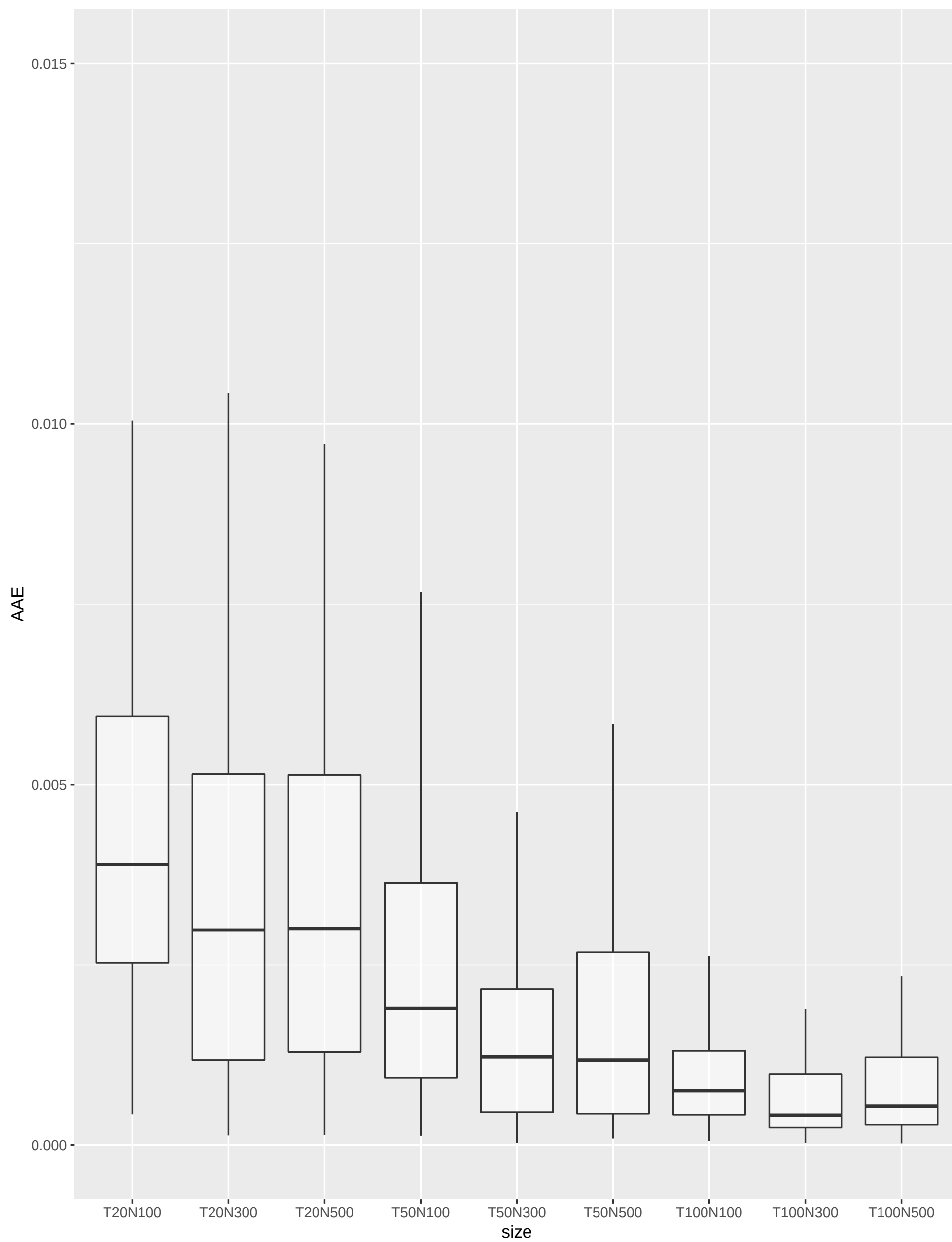


Figure 5.1: Boxplot of AAE for Maximum likelihood estimator of model 5.1 for different sample sizes and network sizes.

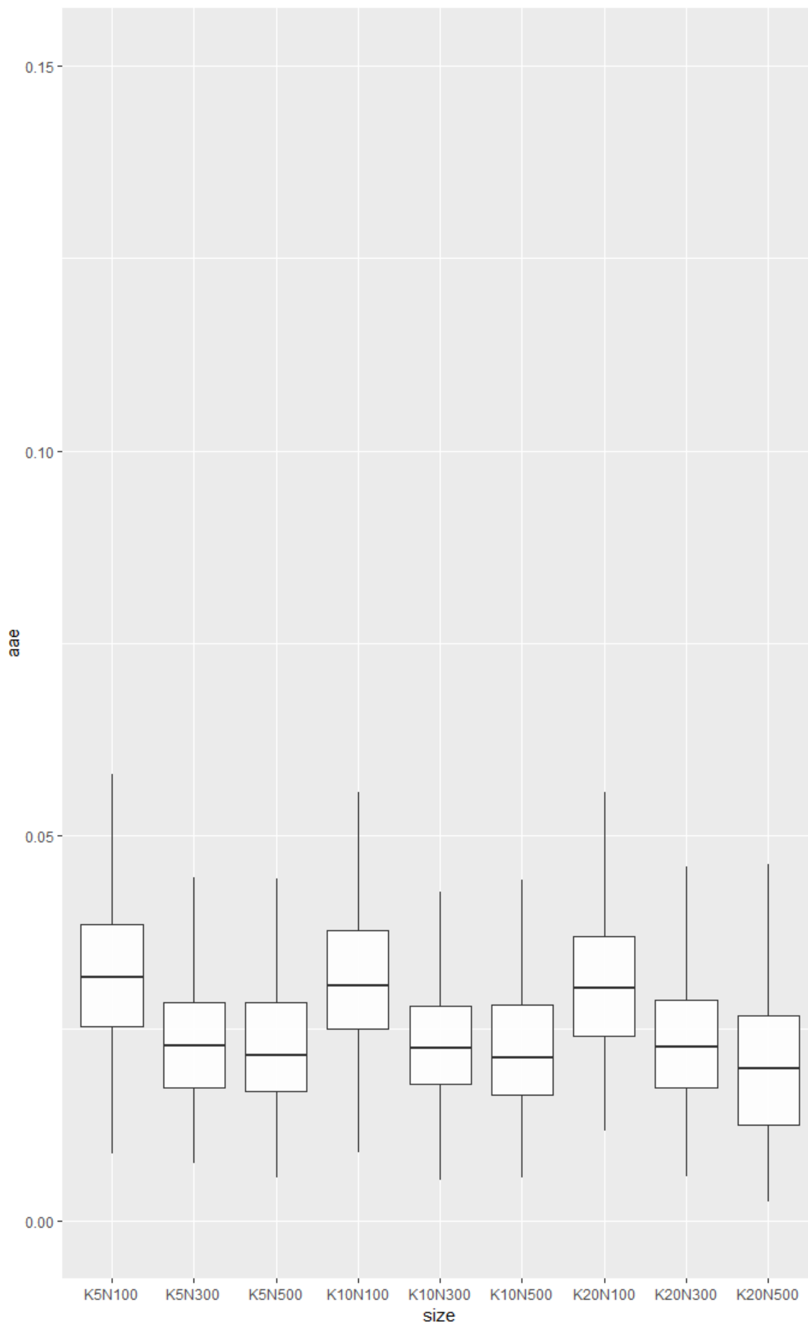


Figure 5.2: Boxplot of AAE for Maximum likelihood estimator of model 5.1 for different network densities and network sizes.

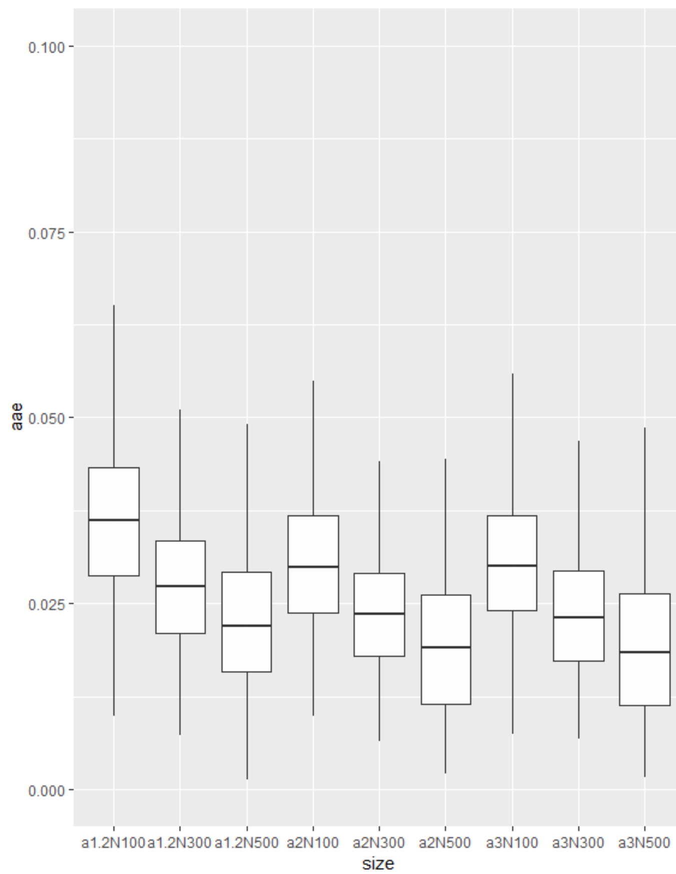


Figure 5.3: Boxplot of AAE for Maximum likelihood estimator of model 5.1 for different network densities and network sizes.

in-degree and out-degree are both positively related to the user's activeness level.

In addition, the transition probabilities appear to be different for two states. The process tends to stay longer in state 2 than in state 1. The expected duration in state 1 is $\frac{1}{1-p_{11}} = 1.98$ while in the other one is $\frac{1}{1-p_{22}} = 7.35$.

Table 5.1: The estimated parameters for the Twitter Data Set

Regression coefficient	Estimate
$\hat{\beta}_0$	3.66
$\hat{\beta}_1$	9.22×10^{-2}
$\hat{\beta}_2(1)$	9.81×10^{-2}
$\hat{\beta}_2(2)$	-0.22
$\hat{\gamma}_1$	1.16×10^{-3}
$\hat{\gamma}_2$	1.76×10^{-3}
$\hat{p_{11}}$	0.496
$\hat{p_{22}}$	0.864

5.6 Conclusion

This chapter considers a new type of nonlinear network autoregression model. The NAR model is extended to another nonlinear form, that is, Markov switching form. This is a natural generalization of a traditional NAR model [50].

The stationary conditions of the MS-NAR model is provided and the maximum likelihood estimation is also investigated. Many simulations were carried out to estimate the model parameters with different sample sizes and different initial values. An application of the MS-NAR model to Twitter data is presented.

Table 5.2: Simulation results for Example 1 with 1000 replications. Maximum Likelihood estimates with their mean square errors ($\times 10^2$) in parentheses for different N and T.

		$T = 100$			$T = 50$			$T = 20$		
	Value	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$
β_0	0.3	0.30(0.04)	0.30(0.01)	0.30(0.00)	0.30(0.15)	0.29(0.03)	0.30(0.02)	0.31(0.27)	0.30(0.04)	0.30(0.04)
β_1	0.2	0.19(0.03)	0.20(0.01)	0.19(0.00)	0.20(0.14)	0.21(0.01)	0.20(0.01)	0.19(0.03)	0.20(0.07)	0.20(0.04)
$\beta_2(1)$	0.3	0.31(0.01)	0.30(0.00)	0.30(0.00)	0.30(0.02)	0.31(0.00)	0.30(0.00)	0.30(0.02)	0.20(0.01)	0.30(0.01)
$\beta_2(2)$	0.7	0.71(0.01)	0.70(0.01)	0.70(0.00)	0.70(0.03)	0.7(0.01)	0.70(0.01)	0.70(0.04)	0.70(0.02)	0.70(0.01)
p_{11}	0.7	0.70(0.45)	0.75(0.31)	0.70(0.00)	0.69(0.97)	0.7(0.99)	0.65(0.28)	0.67(2.28)	0.67(2.02)	0.66(1.96)
p_{22}	0.4	0.41(0.16)	0.46(0.33)	0.30(0.04)	0.39(1.76)	0.38(1.50)	0.40(0.82)	0.42(1.33)	0.41(2.15)	0.41(2.04)
δ_1	1	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.05)	1.00(0.02)	1.00(0.01)
δ_2	1	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.04)	0.99(0.04)	1.00(0.01)	1.00(0.02)	1.00(0.03)	1.00(0.02)
γ_1	-0.5	-0.48(0.03)	-0.50(0.00)	-0.50(0.00)	-0.50(0.05)	-0.49(0.02)	-0.5(0.01)	-0.50(0.23)	-0.50(0.03)	-0.50(0.02)
γ_2	0.3	0.29(0.02)	0.31(0.01)	0.30(0.00)	0.30(0.06)	0.30(0.01)	0.30(0.01)	0.31(0.29)	0.30(0.02)	0.30(0.02)
γ_3	0.8	0.80(0.01)	0.80(0.00)	0.79(0.00)	0.80(0.09)	0.8(0.01)	0.81(0.02)	0.80(0.04)	0.80(0.04)	0.80(0.03)
γ_4	0	0.00(0.01)	0.00(0.00)	0.00(0.00)	0.00(0.04)	0.00(0.00)	0.00(0.01)	0.00(0.04)	0.00(0.03)	0.00(0.01)
γ_5	0	-0.01(0.01)	0.00(0.01)	0.00(0.00)	0.00(0.04)	0.00(0.00)	0.00(0.01)	0.00(0.06)	0.00(0.02)	0.00(0.01)
AAE		0.0010	0.008	0.004	0.0028	0.0018	0.0009	0.0192	0.0039	0.0020

Table 5.3: Distribution of the maximum likelihood estimates of model 5.1 for different N and T. Results are based on 1000 simulations

Network Size	Parameter	value	mean	Lower	25%	Median	75%	Upper
	β_0	0.3	0.308	0.221	0.273	0.306	0.345	0.402
	β_1	0.2	0.194	0.107	0.159	0.194	0.230	0.279
	$\beta_2(1)$	0.3	0.296	0.259	0.286	0.297	0.308	0.329
	$\beta_2(2)$	0.7	0.696	0.653	0.680	0.696	0.712	0.739
	p_{11}	0.7	0.674	0.403	0.590	0.702	0.789	0.876
$T = 20; N = 100$	p_{22}	0.4	0.419	0.188	0.306	0.404	0.544	0.702
	δ_1	1	0.998	0.965	0.985	0.998	1.010	1.034
	δ_2	1	0.996	0.951	0.979	0.996	1.012	1.042
	γ_1	-0.5	-0.503	-0.555	-0.524	-0.505	-0.481	-0.445
	γ_2	0.3	0.305	0.241	0.281	0.305	0.333	0.362
	γ_3	0.8	0.802	0.739	0.776	0.799	0.830	0.874
	γ_4	0	-0.001	-0.052	-0.021	0.001	0.020	0.051
	γ_5	0	0.000	-0.042	-0.021	0.001	0.021	0.043
	β_0	0.3	0.303	0.273	0.291	0.303	0.313	0.335
	β_1	0.2	0.197	0.150	0.180	0.198	0.214	0.240
	$\beta_2(1)$	0.3	0.299	0.281	0.290	0.298	0.308	0.318
	$\beta_2(2)$	0.7	0.699	0.677	0.691	0.696	0.707	0.723
	p_{11}	0.7	0.666	0.403	0.567	0.688	0.785	0.870
$T = 20; N = 300$	p_{22}	0.4	0.405	0.158	0.314	0.380	0.476	0.680
	δ_1	1	0.997	0.973	0.988	0.997	1.007	1.018
	δ_2	1	1.003	0.974	0.991	1.003	1.016	1.003
	γ_1	-0.5	-0.498	-0.527	-0.506	-0.499	-0.487	-0.468
	γ_2	0.3	0.299	0.272	0.289	0.300	0.311	0.323
	γ_3	0.8	0.801	0.765	0.787	0.799	0.814	0.839
	γ_4	0	0.000	-0.030	-0.014	0.000	0.013	0.030
	γ_5	0	-0.003	-0.026	-0.013	-0.002	0.006	0.019
	β_0	0.3	0.304	0.267	0.292	0.304	0.321	0.337
	β_1	0.2	0.198	0.155	0.186	0.199	0.214	0.230
	$\beta_2(1)$	0.3	0.298	0.279	0.291	0.299	0.305	0.315
	$\beta_2(2)$	0.7	0.698	0.676	0.690	0.700	0.705	0.721
	p_{11}	0.7	0.661	0.350	0.600	0.685	0.782	0.876
$T = 20; N = 500$	p_{22}	0.4	0.411	0.188	0.311	0.407	0.537	0.659
	δ_1	1	0.998	0.980	0.994	0.997	1.004	1.011
	δ_2	1	0.999	0.979	0.991	0.998	1.008	1.023
	γ_1	-0.5	-0.502	-0.523	-0.513	-0.503	-0.493	-0.475
	γ_2	0.3	0.301	0.276	0.294	0.300	0.312	0.323
	γ_3	0.8	0.801	0.771	0.787	0.804	0.814	0.931
	γ_4	0	0.002	-0.019	-0.006	0.001	0.009	0.021
	γ_5	0	-0.001	-0.017	-0.008	0.000	0.006	0.015

Table 5.4: Distribution of the maximum likelihood estimates of model 5.1 for different N and T. Results are based on 1000 simulations

Network Size	Parameter	value	mean	Lower	25%	Median	75%	Upper
	β_0	0.3	0.304	0.240	0.280	0.302	0.327	0.366
	β_1	0.2	0.197	0.136	0.175	0.197	0.219	0.258
	$\beta_2(1)$	0.3	0.299	0.274	0.290	0.299	0.308	0.323
	$\beta_2(2)$	0.7	0.698	0.668	0.687	0.699	0.709	0.725
	p_{11}	0.7	0.687	0.522	0.631	0.695	0.751	0.830
$T = 50; N = 100$	p_{22}	0.4	0.394	0.163	0.302	0.391	0.489	0.627
	δ_1	1	0.998	0.975	0.989	0.998	1.006	1.002
	δ_2	1	0.998	0.965	0.985	0.999	1.011	1.032
	γ_1	-0.5	-0.501	-0.542	-0.515	-0.501	-0.485	-0.462
	γ_2	0.3	0.302	0.260	0.286	0.301	0.319	0.344
	γ_3	0.8	0.801	0.753	0.781	0.800	0.820	0.854
	γ_4	0	0.000	-0.037	-0.013	0.001	0.014	0.033
	γ_5	0	0.000	-0.032	-0.013	0.001	0.013	0.032
	β_0	0.3	0.299	0.276	0.288	0.302	0.309	0.320
	β_1	0.2	0.201	0.177	0.191	0.200	0.211	0.223
	$\beta_2(1)$	0.3	0.300	0.288	0.294	0.300	0.307	0.313
	$\beta_2(2)$	0.7	0.700	0.685	0.693	0.700	0.708	0.825
	p_{11}	0.7	0.696	0.546	0.642	0.702	0.742	0.825
$T = 50; N = 300$	p_{22}	0.4	0.383	0.130	0.294	0.399	0.461	0.552
	δ_1	1	1.00	0.989	0.995	1.000	1.005	1.011
	δ_2	1	0.998	0.981	0.992	1.000	1.005	1.016
	γ_1	-0.5	-0.499	-0.519	-0.505	-0.498	-0.491	-0.480
	γ_2	0.3	0.301	0.283	0.293	0.300	0.308	0.320
	γ_3	0.8	0.798	0.773	0.790	0.798	0.808	0.822
	γ_4	0	0.001	-0.014	-0.006	0.001	0.007	0.016
	γ_5	0	0.000	-0.014	-0.007	-0.001	0.005	0.017
	β_0	0.3	0.301	0.282	0.293	0.301	0.310	0.321
	β_1	0.2	0.199	0.177	0.191	0.200	0.207	0.220
	$\beta_2(1)$	0.3	0.299	0.291	0.295	0.300	0.303	0.308
	$\beta_2(2)$	0.7	0.699	0.688	0.695	0.699	0.704	0.711
	p_{11}	0.7	0.701	0.530	0.653	0.714	0.752	0.812
$T = 50; N = 500$	p_{22}	0.4	0.396	0.209	0.302	0.406	0.458	0.607
	δ_1	1	1.000	0.990	0.996	1.001	1.004	1.008
	δ_2	1	1.001	0.987	0.995	1.002	1.007	1.017
	γ_1	-0.5	-0.501	-0.514	-0.507	-0.502	-0.495	-0.484
	γ_2	0.3	0.299	0.285	0.294	0.300	0.305	0.314
	γ_3	0.8	0.802	0.784	0.795	0.801	0.810	0.821
	γ_4	0	0.001	-0.016	-0.005	0.001	0.007	0.016
	γ_5	0	0.000	-0.013	-0.005	0.000	0.006	0.011

Table 5.5: Distribution of the maximum likelihood estimates of model 5.1 for different N and T. Results are based on 1000 simulations.

Network Size	Parameter	value	mean	Lower	25%	Median	75%	Upper
	β_0	0.3	0.302	0.265	0.287	0.303	0.316	0.339
	β_1	0.2	0.198	0.163	0.183	0.198	0.213	0.232
	$\beta_2(1)$	0.3	0.299	0.286	0.293	0.299	0.305	0.314
	$\beta_2(2)$	0.7	0.700	0.684	0.693	0.700	0.706	0.717
	p_{11}	0.7	0.695	0.600	0.660	0.696	0.735	0.786
$T = 100; N = 100$	p_{22}	0.4	0.392	0.247	0.330	0.396	0.452	0.535
	δ_1	1	1.000	0.984	0.994	1.000	1.006	1.015
	δ_2	1	0.999	0.980	0.991	0.999	1.008	1.019
	γ_1	-0.5	-0.500	-0.522	-0.509	-0.500	-0.491	-0.477
	γ_2	0.3	0.300	0.276	0.290	0.300	0.310	0.325
	γ_3	0.8	0.801	0.772	0.788	0.701	0.714	0.832
	γ_4	0	0.000	-0.022	-0.009	0.000	0.009	0.020
	γ_5	0	0.000	-0.021	-0.009	0.000	0.008	0.020
	β_0	0.3	0.300	0.286	0.295	0.300	0.305	0.313
	β_1	0.2	0.200	0.179	0.194	0.201	0.206	0.218
	$\beta_2(1)$	0.3	0.300	0.291	0.296	0.300	0.303	0.309
	$\beta_2(2)$	0.7	0.699	0.689	0.695	0.699	0.703	0.710
	p_{11}	0.7	0.691	0.590	0.660	0.697	0.734	0.767
$T = 100; N = 300$	p_{22}	0.4	0.385	0.247	0.341	0.396	0.433	0.496
	δ_1	1	1.000	0.993	0.997	0.999	1.002	1.006
	δ_2	1	0.999	0.988	0.995	0.999	1.004	1.010
	γ_1	-0.5	0.501	-0.513	-0.507	-0.501	-0.495	-0.488
	γ_2	0.3	0.300	0.287	0.295	0.300	0.307	0.314
	γ_3	0.8	0.800	0.783	0.794	0.800	0.807	0.817
	γ_4	0	0.000	-0.010	-0.006	-0.001	0.005	0.013
	γ_5	0	0.001	-0.010	-0.003	0.001	0.005	0.012
	β_0	0.3	0.3010	0.2813	0.2926	0.3024	0.3091	0.3161
	β_1	0.2	0.1982	0.1811	0.1892	0.1975	0.2063	0.2204
	$\beta_2(1)$	0.3	0.3000	0.2924	0.2968	0.3003	0.3024	0.3070
	$\beta_2(2)$	0.7	0.6999	0.6929	0.6968	0.6996	0.7027	0.2094
	p_{11}	0.7	0.7043	0.6165	0.6689	0.7030	0.7490	0.7815
$T = 100; N = 500$	p_{22}	0.4	0.3781	0.1903	0.3201	0.3959	0.4374	0.5112
	δ_1	1	0.9995	0.9936	0.9969	0.9991	1.0026	1.0053
	δ_2	1	1.0001	0.9882	0.9962	1.0004	1.0045	1.0114
	γ_1	-0.5	-0.4993	-0.5085	-0.5040	-0.4986	-0.4948	-0.4882
	γ_2	0.3	0.2999	0.2882	0.2951	0.2996	0.3046	0.3130
	γ_3	0.8	0.7995	0.7881	0.7939	0.7990	0.8043	0.8140
	γ_4	0	-0.0007	-0.0079	-0.0046	-0.0012	0.0021	0.0096
	γ_5	0	-0.0001	-0.0082	-0.0039	0.0003	0.0039	0.0071

Table 5.6: Simulation results for Example 2 with 1000 replications. Maximum Likelihood estimates with their mean square errors ($\times 10^2$) in parentheses for different N and K. T is fixed at 30.

		$K=5$			$K=10$			$K=20$		
	Value	$N=100$	$N=300$	$N=500$	$N=100$	$N=300$	$N=500$	$N=100$	$N=300$	$N=500$
β_0	0.3	0.3017(0.02)	0.2999(0.01)	0.2999(0.01)	0.3019(0.04)	0.3008(0.01)	0.2998(0.01)	0.3018(0.04)	0.3005(0.01)	0.3010(0.01)
β_1	0.2	0.2015(0.01)	0.2004(0.01)	0.2003(0.01)	0.2004(0.01)	0.2002(0.00)	0.2001(0.00)	0.1994(0.01)	0.2002(0.00)	0.2002(0.00)
$\beta_2(1)$	0.3	0.3077(0.19)	0.5066(7.16)	0.2993(0.01)	0.3045(0.29)	0.4929(6.61)	0.2996(0.01)	0.2997(0.08)	0.4955(7.01)	0.2991(0.00)
$\beta_2(2)$	0.7	0.6835(0.21)	0.4908(7.26)	0.6995(0.01)	0.6893(0.31)	0.5050(6.66)	0.6995(0.01)	0.6965(0.09)	0.5014(7.09)	0.6990(0.00)
p_{11}	0.7	0.6387(0.71)	0.5363(5.27)	0.6803(1.11)	0.6637(2.25)	0.5398(4.81)	0.6803(1.18)	0.6786(0.98)	0.5423(5.41)	0.6838(0.57)
p_{22}	0.4	0.4774(1.02)	0.5373(4.30)	0.3856(1.91)	0.4077(1.71)	0.5340(4.29)	0.3856(2.02)	0.3993(1.56)	0.5312(5.07)	0.3954(1.02)
δ_1	1	0.9988(0.01)	1.0008(0.02)	1.0000(0.001)	0.9981(0.02)	0.9949(0.81)	1.0000(0.01)	0.9978(0.02)	0.9998(0.01)	0.9986(0.00)
δ_2	1	0.9983(0.02)	0.9992(0.02)	0.9994(0.001)	0.9984(0.04)	0.9998(0.01)	0.9994(0.01)	0.9987(0.04)	(0.9996(0.01))	0.9986(0.01)
γ_1	-0.5	-0.5024(0.03)	-0.5015(0.02)	-0.4996(0.01)	-0.5038(0.06)	-0.5014(0.02)	-0.5007(0.01)	-0.5027(0.05)	-0.5016(0.02)	-0.5001(0.01)
γ_2	0.3	0.3018(0.03)	0.3014(0.02)	0.2999(0.01)	0.3028(0.06)	-0.3012(0.02)	0.3006(0.01)	0.3008(0.06)	0.3017(0.02)	0.3005(0.01)
γ_3	0.8	0.8041(0.04)	0.8009(0.03)	0.8005(0.02)	0.8049(0.08)	0.8009(0.04)	0.8003(0.02)	0.8030(0.08)	0.8021(0.03)	0.8005(0.01)
γ_4	0	0.0039(0.03)	0.0006(0.02)	0.0008(0.01)	0.0003(0.05)	0.0000(0.02)	0.0005(0.01)	0.0009(0.04)	-0.0009(0.02)	-0.0006(0.01)
γ_5	0	-0.0005(0.02)	-0.0004(0.01)	-0.0003(0.01)	-0.0007(0.06)	-0.0009(0.01)	-0.0002(0.01)	-0.0017(0.03)	0.0007(0.01)	0.0008(0.01)
AAE		0.075	0.069	0.023	0.078	0.068	0.023	0.077	0.068	0.020

Table 5.7: Simulation results for Example 3 with 1000 replications. Maximum Likelihood estimates with their mean square errors ($\times 10^2$) in parentheses for different N and α. T is fixed at 30.

		α=1.2			α=2.0			α=3.0		
	Value	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$
β_0	0.3	0.3123(0.33)	0.3048(0.15)	0.3021(0.11)	0.3014(0.05)	0.3016(0.03)	0.3000(0.03)	0.3013(0.06)	0.3010(0.02)	0.2999(0.01)
β_1	0.2	0.1894(0.33)	0.1936(0.26)	0.1984(0.08)	0.2001(0.02)	0.1997(0.01)	0.1999(0.01)	0.2004(0.02)	0.2002(0.00)	0.2001(0.00)
$\beta_2(1)$	0.3	0.3194(0.65)	0.2999(0.02)	0.2994(0.01)	0.3007(0.07)	0.2989(0.01)	0.2996(0.01)	0.2979(0.03)	0.2990(0.01)	0.2994(0.01)
$\beta_2(2)$	0.7	0.6782(0.66)	0.6968(0.07)	0.6996(0.01)	0.6923(0.18)	0.6986(0.03)	0.6998(0.01)	0.6962(0.06)	0.6990(0.03)	0.6997(0.01)
p_{11}	0.7	0.6711(1.16)	0.6807(1.03)	0.6875(1.04)	0.6828(1.15)	0.6862(0.89)	0.6897(0.91)	0.6840(1.15)	0.6921(0.97)	0.6866(1.05)
p_{22}	0.4	0.4064(2.12)	0.3930(1.95)	0.3882(1.68)	0.3860(1.89)	0.3945(1.56)	0.3932(1.59)	0.3858(1.97)	0.3937(1.88)	0.3972(1.72)
δ_1	1	0.9986(0.02)	1.0007(0.02)	0.9985(0.01)	0.9998(0.05)	0.9996(0.01)	0.9982(0.01)	0.9984(0.05)	0.9991(0.01)	0.9984(0.01)
δ_2	1	0.9972(0.04)	1.0002(0.02)	0.9975(0.01)	0.9982(0.05)	0.9999(0.01)	0.9982(0.01)	0.9991(0.04)	0.9999(0.01)	0.9988(0.01)
γ_1	-0.5	-0.5034(0.04)	-0.5014(0.02)	-0.4997(0.01)	-0.5021(0.06)	-0.5010(0.02)	-0.5008(0.01)	-0.5027(0.07)	-0.5006(0.02)	-0.5001(0.01)
γ_2	0.3	0.3024(0.05)	0.3008(0.02)	0.3000(0.01)	0.3007(0.05)	0.3011(0.02)	0.3005(0.01)	0.3006(0.08)	0.2999(0.02)	0.3003(0.01)
γ_3	0.8	0.8017(0.08)	0.8015(0.04)	0.8013(0.02)	0.8044(0.12)	0.8013(0.03)	0.8001(0.02)	0.8042(0.12)	0.8019(0.03)	0.8011(0.02)
γ_4	0	-0.0007(0.04)	0.0002(0.02)	-0.0006(0.001)	-0.0008(0.07)	-0.0007(0.01)	0.0003(0.01)	-0.0013(0.06)	-0.0003(0.02)	0.0000(0.01)
γ_5	0	-0.0005(0.04)	-0.0001(0.02)	0.0001(0.01)	0.0000(0.05)	0.0005(0.01)	0.0001(0.01)	0.0020(0.06)	0.0000(0.02)	-0.0001(0.01)
AAE		0.041	0.028	0.023	0.031	0.024	0.020	0.031	0.024	0.019

Chapter 6

Twitter Network Data Collection

In this chapter, we provide details of how to extract and process the raw data from Twitter website. This is divided into the following steps:

1. Background of data collection
2. Get permission to extract data from Twitter website
3. Data processing

6.1 Background

The aim of our data collection is to obtain the target user's weekly number of tweets and the relationship network between the target users through the API interface provided by Twitter, and finally store the data in an excel file to be able to read into R and used for analysis. Before we start to extract data, we need permission of Twitter by registering an account on the Twitter API. The twitter API is located at `https://developer.twitter.com/en/docs/twitter-for-websites/log-in-with-twitter/login-in-with-twitter` as shown in Figure 6.1. There are no legal issues in the data collection process, since the API is provided by Twitter.

6.2 Get the OAuth Information

First, open the Twitter Developer Platform and log in by clicking the "log in" button in the upper right corner (if you don't have a Twitter account, you will

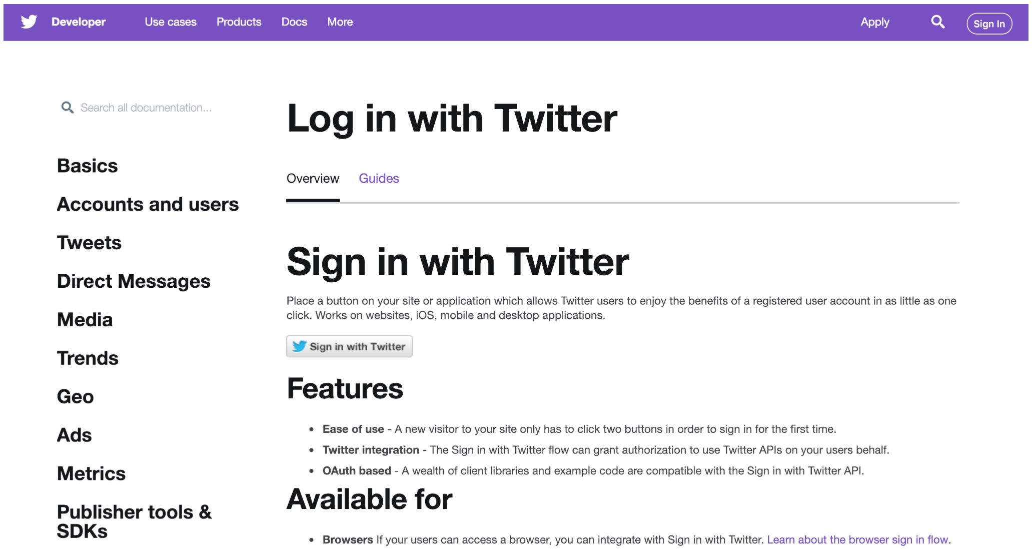


Figure 6.1: Twitter API

have to register one). After logging in, you can click "Apps" in the upper right corner to enter the "API" console and then click the "Create an app" button, as Figure 6.2 shown.

Fill in the app name and other related APP information, click "Create", you can see your own APP in the API console. Under the "Keys and tokens" column, it contains the OAuth information that we need next, i.e. API key, API secret key, Access token and Access token secret, as shown in Figure 6.3. It is worth noting that each account has limited amount of data that can be downloaded each day. Due to the amount data we need, more than one account was created for this data extraction.

6.3 Data Processing

The implementation of this project is divided into the following three steps.

1) Obtain the user-id of the followers of Strathclyde Official Twitter and save it to txt file (details in section 6.3.1).

2) Obtain the number of tweets per week of Strathclyde Official Twitter followers and save it to Excel file.

3) Obtain the relationship matrix between the followers of Strathclyde Official

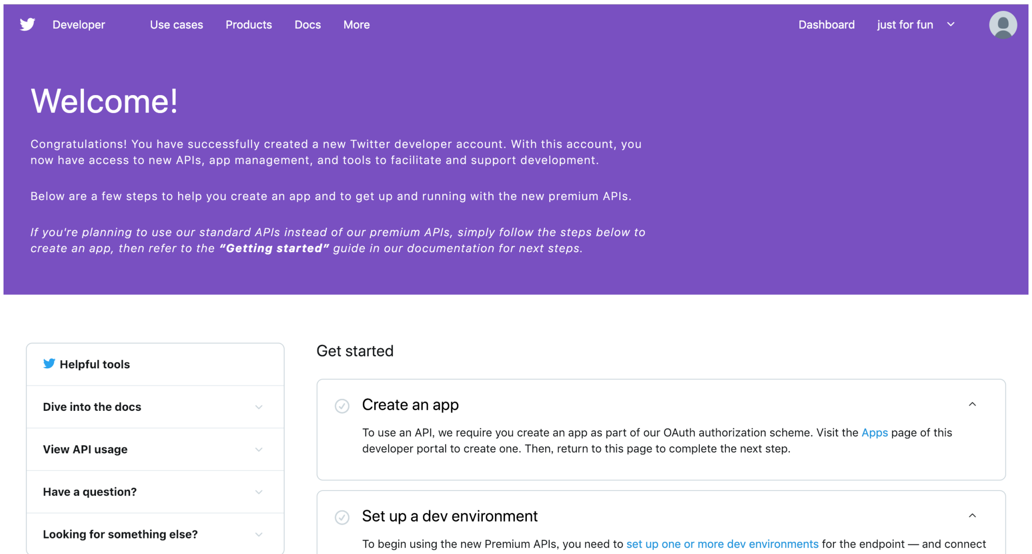


Figure 6.2: API Console

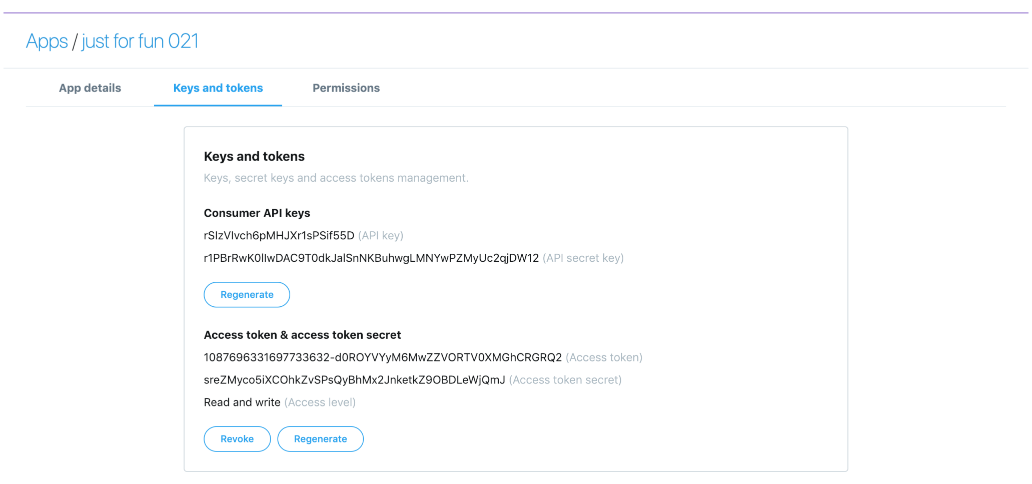
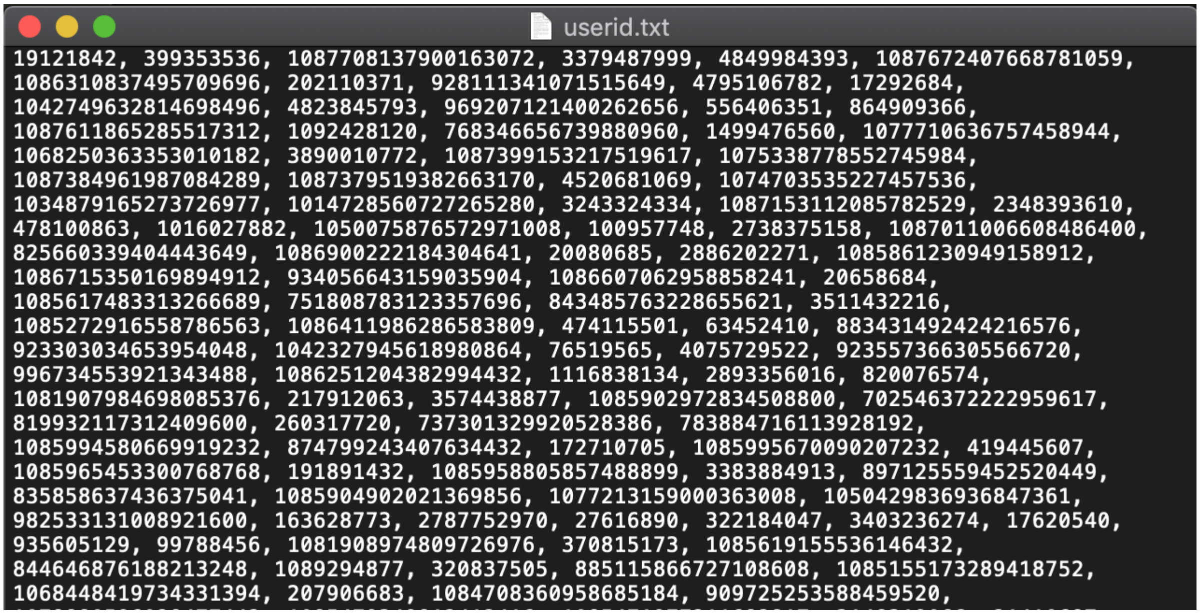


Figure 6.3: OAuth Information



```
userid.txt
19121842, 399353536, 1087708137900163072, 3379487999, 4849984393, 1087672407668781059,
1086310837495709696, 202110371, 928111341071515649, 4795106782, 17292684,
1042749632814698496, 4823845793, 969207121400262656, 556406351, 864909366,
1087611865285517312, 1092428120, 768346656739880960, 1499476560, 1077710636757458944,
1068250363353010182, 3890010772, 1087399153217519617, 1075338778552745984,
1087384961987084289, 1087379519382663170, 4520681069, 1074703535227457536,
1034879165273726977, 1014728560727265280, 3243324334, 1087153112085782529, 2348393610,
478100863, 1016027882, 1050075876572971008, 100957748, 2738375158, 1087011006608486400,
825660339404443649, 1086900222184304641, 20080685, 2886202271, 1085861230949158912,
1086715350169894912, 934056643159035904, 1086607062958858241, 20658684,
1085617483313266689, 751808783123357696, 843485763228655621, 3511432216,
1085272916558786563, 1086411986286583809, 474115501, 63452410, 883431492424216576,
923303034653954048, 1042327945618980864, 76519565, 4075729522, 923557366305566720,
996734553921343488, 1086251204382994432, 1116838134, 2893356016, 820076574,
1081907984698085376, 217912063, 3574438877, 1085902972834508800, 702546372222959617,
819932117312409600, 260317720, 737301329920528386, 783884716113928192,
1085994580669919232, 874799243407634432, 172710705, 1085995670090207232, 419445607,
1085965453300768768, 191891432, 1085958805857488899, 3383884913, 897125559452520449,
835858637436375041, 1085904902021369856, 1077213159000363008, 1050429836936847361,
982533131008921600, 163628773, 2787752970, 27616890, 322184047, 3403236274, 17620540,
935605129, 99788456, 1081908974809726976, 370815173, 1085619155536146432,
844646876188213248, 1089294877, 320837505, 885115866727108608, 1085155173289418752,
1068448419734331394, 207906683, 1084708360958685184, 909725253588459520,
```

Figure 6.4: User ID Output Result

Twitter. If the i-th account follows the j-th account, then the element in the i-th row and j-th column of the relationship matrix is defined as 1, otherwise 0.

6.3.1 Obtain the User-ID of the Followers

Our object is all the followers of the official Twitter account of Strathclyde University. These users formed a social network. With the user-ID of these users, the program can automatically extract all the tweet data of these users, and then calculate the number of weekly tweets we need. The code and output results are shown in Appendix C.1 and Figure 6.4, respectively.

6.3.2 Obtain the Number of Tweets Per Week by Strathclyde Offical Twitter Followers

Next, we extract the number of weekly tweets of each follower with user id and username attached. The code is shown in Appendix C.2.

In the above code, we remove the inactive accounts, that is, the account that have the number of followers less than 80, the number of followees less than 40, and the total number of tweets less than 200. The details of the resulting data

userid	username	week 1	week 2	week 3	week 4	week 5	week 6	week 7	week 8
219287702	GaiaWind133	1401	439	738	0	0	0	0	0
1255519842	Rtudaortiz	646	106	212	624	441	709	214	190
436902627	morganspice20	3003	1468	2249	1185	1632	1246	864	1532
3352192444	strochsfcladies	3856	380	623	834	283	1193	379	1233
1009099808	chiaramannarjno	1777	604	241	553	811	1231	255	727
787433122233192448	neilston_madras	4526	4007	1419	946	1090	1103	1316	525

Figure 6.5: Number of Weekly Tweets Results

table are shown in Figure 6.5.

6.3.3 Obtain the Relationship Matrix Between the Followers

After obtaining the number of weekly tweets for all followers, we develop a relationships matrix between all the followers in this section. Since we removed some inactive accounts in 6.3.2, the list of user ids here is different from that obtained in 6.3.1. We need to re-acquire it from the previous file obtained in 6.3.2. The code is shown in Appendix C.3.

Next, we develop the relationship matrix of all followers and the code is shown in Appendix C.4. The resulting matrix is a sparse matrix since the network density (ND) is around 4%.

6.4 Summary

This chapter describes how to use the API to get data and save it to a file. If you want to get additional data through the API, you can try to get it using the methods in this chapter. For detailed principles of OAuth authentication, please see Appendix B.

Bibliography

[1] James H. Albert and Siddhartha Chib. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics*, 11(1):1–15, 1993.

[2] Andrew Ang and Geert Bekaert. Regime switches in interest rates. *Journal of Business and Economic Statistics*, 20(2):163–182, 2002.

[3] Albert László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[4] E. J Bedrick and C. L Tsai. Model selection for multivariate regression in small samples. *Biometrics*, 50(1):226–231, 1994.

[5] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley, 2015.

[6] Trevor Stanley Breusch and Adrian Rodney Pagan. The Lagrange Multiplier test and to model applications specification in econometrics. *The Review of Economic Studies*, 47(1):239–253, 1980.

[7] Kung-Sik Chan. Testing for threshold autoregression. *The Annals of Statistics*, 18(4):1886–1894, 1990.

[8] Victor Chernozhukov and Christian Hansen. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398, 2008.

[9] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[10] Jon Danielsson, Lerby Murat Ergun, Laurens de Haan, and Casper G de Vries. Tail index estimation: Quantile driven threshold selection. *Available at SSRN 2717478*, 2016.

[11] Walter Enders and Stan Hurn. Asymmetric price adjustment and the Phillips curve. *Journal of Macroeconomics*, 24(3):395–412, 2002.

[12] Charles Engel and James D. Hamilton. Long swings in the exchange rate: Are they in the data and do markets know it? *The American Economic Review*, 8(4):689–713, 1990.

[13] Jianqing Fan and Qiwei Yao. *Nonlinear time series: Nonparametric and parametric methods*. Springer, New York, 2003.

[14] Antonio F. Galvao, Gabriel Montes-Rojas, and Jose Olmo. Threshold quantile autoregressive models. *Journal of Time Series Analysis*, 32(3):253–267, 2011.

[15] Peter Hall and A. H. Welsh. Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1):331–341, 1985.

[16] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357, 1989.

[17] James D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70, 1990.

[18] James Douglas Hamilton. *Time series analysis*. Princeton University Press, Princeton, New Jersey, 1994.

[19] B. E. Hansen. The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 7(1 S):S61–S82, 1992.

[20] Bruce E Hansen. Inference when a nuisance parametre is not identified under the null hypothesis. *Econometrica*, 64(2):413–430, 1996.

[21] Bruce E Hansen. Threshold autoregression in economics. *Statistics and Its Interface*, 4:123–127, 2011.

[22] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 50(4):1029–1054, 1982.

[23] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

[24] Paul W. Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs: Rejoinder. *Journal of the American Statistical Association*, 76(373):62, 1981.

[25] J.R.M. Hosking. Lagrange-multiplier tests of multivariate time-series models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(2):219–230, 1981.

[26] C J Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60:1–22, 1994.

[27] Chang-Jin Kim and Charles R. Nelson. *State-space models with regime switching.* MIT Press, Cambridge, Massachusetts, 1999.

[28] Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[29] Roger Koenker and Zhijie Xiao. Quantile autoregression. *Journal of the American Statistical Association*, 101(475):980–990, 2006.

[30] Roger Koenker and Quanshui Zhao. Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12(5):793–813, 1996.

[31] Hans-Martin Krolzig. *Markov-switching vector autoregressions : Modelling, statistical inference, and application to business cycle analysis*, volume 454. 1997.

[32] P. Newbold and C. W. J. Granger. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 137(2):131–165, 1974.

[33] Whitney K Newey and Ken D West. A simple positive-definite heteroskedasticity and autocorrelation-consistent covariance matrix. *Econometrica*, 55:703–708, 1987.

[34] Krzysztof Nowicki and Tom A.B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

[35] Esa Nummelin. *General irreducible Markov chains and non-negative operators*, volume 83. Cambridge University Press, 2004.

[36] Jiazhu Pan and Qiwei Yao. Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379, 2008.

[37] Peter C.B. Phillips, James W. McFarland, and Patrick C. McMahon. Robust tests of forward exchange market efficiency with empirical evidence from the 1920s. *Journal of Applied Econometrics*, 11(1):1–22, 1996.

[38] C. Radhakrishna Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44(1):50–57, 1948.

[39] Samuel David Silvey. The Lagrangian multiplier test. *The Annals of Mathematical Statistics*, 30(2):389–407, 1959.

[40] Liangjun Su and Zhijie Xiao. Testing for parameter stability in quantile regression models. *Statistics and Probability Letters*, 78(16):2768–2775, 2008.

[41] Howell Tong. On a threshold model in pattern recognition and signal processing,(ed) c. *Chen. Sijhoff and Noonhoff, Amsterdam*, 1978.

[42] Howell Tong. *Threshold models in non-linear time series analysis*. New York: Springer, 1983.

[43] Howell Tong and Keng S Lim. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(3):245–292, 1980.

[44] Ruey S. Tsay. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405):231–240, 1989.

[45] Richard L. Tweedie. Sufficient conditions for ergodicity and recurence of Markov chains on a general state space. *Stochastic Processes and their Applications*, 3(4):385–403, 1975.

[46] Yuchung J. Wang and George Y. Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19, 1987.

[47] Stanley Wasserman and Katherine Faust. *Social network analysis : methods and applications*, volume 24. 1994.

[48] C. S. Wong and W. K. Li. Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika*, 84(2):407–418, 1997.

[49] Minxian Yang. Some properties of vector autoregressive processes with Markov-switching coefficients. *Econometric Theory*, 16(1):23–43, 2000.

[50] Xuening Zhu, Rui Pan, Guodong Li, Yuewen Liu, and Hansheng Wang. Network vector autoregression. *Annals of Statistics*, 45(3):1096–1123, 2017.

[51] Xuening Zhu, Weining Wang, Hansheng Wang, and Wolfgang K. Härdle. Network Quantile Autoregression. 2019.

[52] Xuening Zhu, Weining Wang, Hansheng Wang, and Wolfgang Karl Härdle. Network quantile autoregression. *Journal of Econometrics*, 212(1):345–358, 2019.

Appendix A

Proof

A.1 The Spectral Radius of W

We denote the eigenvalues of W by $\lambda_1, \lambda_2, \ldots, \lambda_N$. Its spectral radius $\rho(W)$ is defined as

$$\rho(W) = \mathbf{max}\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_N|\}.$$

Theorem A.1.1.

$$\rho(W) = 1.$$

Proof. It is easy to know $W\mathbf{1} = \mathbf{1}$, that is, 1 is one of the eigenvalues of W, which means max$|\lambda_i| \geq$1. On the other hand, we want to prove $\mathbf{max}|\lambda_i| \leq$1. Suppose there is a λ_M with $|\lambda_M| > 1$. Then we know $\lambda_M I - W$ is a strictly diagonally dominant matrix since $|\lambda_M| > \sum_{j \neq i} |w_{ij}| = 1$ for all i, where w_{ij} denotes the element in the ith row and jth column. According to Levy-Desplanques theorem, a strictly diagonally dominant matrix is non-singular, so $|\lambda_M I - W| \neq 0$, which means λ_M is not the eigenvalue of W. $\square$

The above proof of Theorem A.1.1 uses the following theorem.

Theorem A.1.2 (Levy-Desplanques theorem)**.** *A strictly diagonally dominant matrix is non-singular.*

Proof. Let det(A)=0, then a non-zero vector $\mathbf{x}$ exists such that $A\mathbf{x} = \mathbf{0}$; let M be the index such that $|\mathbf{x}_M| = \max(|x_1|, |x_2|, \ldots, |x_n|)$, so that $|x_j| \leq |x_M|$, for every

j. We have

$$a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MM}x_M + \cdots + a_{Mn}x_n = 0,$$

which implies

$$|a_{MM}||x_M| = |a_{MM}x_M| = |\sum_{j\neq M} a_{Mj}x_j| \leq \sum_{j\neq M} |a_{Mj}||x_j| \leq |x_M| \sum_{j\neq M} |a_{Mj}|,$$

that is

$$|a_{MM}| \leq \sum_{j\neq M} |a_{Mj}|$$

in contrast with strictly diagonally dominance definition.

□

A.2 Proof of LM Value

Proof. To test this hypothesis and obtain the LM value in 3.48, we need to calculate the maximum value of the function:

$$f(\theta, \lambda_1) = L(\theta) + \lambda_1 h_1(\theta). \tag{A.1}$$

Differentiate the function (A.1) and let it be zero, $\frac{\partial f}{\partial \theta} = \frac{\partial \mathbf{L}(\theta)}{\partial \theta} + \lambda_1 \frac{\partial h_1 \theta}{\partial \theta} = 0$ and $\frac{\partial f}{\partial \lambda_1} = h_1(\theta) = 0$.

Let

$$\mathbf{D}(\theta) = \begin{pmatrix} \mathbf{D_1}(\theta) \\ \mathbf{D_2}(\theta) \end{pmatrix}, \tag{A.2}$$

where $\mathbf{D_1}(\theta) = \frac{\partial \mathbf{L}(\theta)}{\partial \beta_3}$ and $\mathbf{D_2}(\theta) = \left(\frac{\partial \mathbf{L}(\theta)}{\partial \beta_0}, \frac{\partial \mathbf{L}(\theta)}{\partial \beta_1}, \frac{\partial \mathbf{L}(\theta)}{\partial \beta_2}, \frac{\partial \mathbf{L}(\theta)}{\partial \gamma_1}, \cdots, \frac{\partial \mathbf{L}(\theta)}{\partial \gamma_p}\right)^\top$.

We know

$$\begin{aligned} \varepsilon_t &= \mathbb{Y}_t - \mathcal{B}_0 - G\mathbb{Y}_{t-1} \\ &= \mathbb{Y}_t - (\beta_0 + \gamma^\top Z) - (\beta_1 W + \beta_2 I + \beta_3 J_{t-1})\mathbb{Y}_{t-1} \\ &\sim N(\mathbf{0}, \Sigma_0), \end{aligned}$$

where

$$\Sigma_0 = \begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{pmatrix},$$

so

$$f(\varepsilon_t) = \prod_{t=1}^{T} (2\pi)^{-n/2} |\Sigma_0|^{-1/2} exp\{-\frac{1}{2}\varepsilon_t' \Sigma_0^{-1} \varepsilon_t\}$$

and

$$\begin{aligned}
\mathbf{L}(\theta) &= -\frac{nT}{2}\mathsf{log}(2\pi) - \frac{T}{2}\mathsf{log}|\Sigma_0| - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t' \Sigma_0^{-1} \varepsilon_t \\
&= -\frac{nT}{2}\mathsf{log}(2\pi) - \frac{T}{2}\mathsf{log}|\Sigma_0| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \varepsilon_t' \varepsilon_t \\
&= -\frac{nT}{2}\mathsf{log}(2\pi) - \frac{nT}{2}\mathsf{log}(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} \varepsilon_{ti}^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial \mathbf{L}(\theta)}{\partial \beta_3} &= \frac{\partial(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t'\varepsilon_t)}{\partial \beta_3} = -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\frac{\partial(\varepsilon_t'\varepsilon_t)}{\partial \beta_3} \\
&= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial \varepsilon_{ti}}{\partial \beta_3}\varepsilon_{ti} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}\varepsilon_{ti} I_{\{Y_{(t-1)i<r}\}}.
\end{aligned}$$

Since

$$\begin{aligned}
\frac{\partial \varepsilon_{ti}}{\partial \beta_3} &= \frac{\partial(Y_{it} - \beta_{0i} - M_i - \beta_3 Y_{(t-1)i} I_{\{Y_{(t-1)i<r}\}})}{\partial \beta_3} \\
&= -Y_{(t-1)i} I_{\{Y_{(t-1)i<r}\}},
\end{aligned}$$

where M_i is the ith row of the matrix $\gamma^\top Z + (\beta_1 W + \beta_2 I)Y_{t-1}$, we can also get $\frac{\partial \varepsilon_{ti}}{\partial \beta_2} = -Y_{(t-1)i}, \frac{\partial \varepsilon_{ti}}{\partial \beta_1} = -\sum_{j=1}^{N} W_{ij} Y_{(t-1)j}, \frac{\partial \varepsilon_{ti}}{\partial \beta_0} = -1, \frac{\partial \varepsilon_{ti}}{\partial \gamma_P} = -Z_{Pi}, P = 1, \cdots, p$.

Review

$$\mathbf{D_2}(\theta) = \left(\frac{\partial \mathbf{L}(\theta)}{\partial \beta_0}, \frac{\partial \mathbf{L}(\theta)}{\partial \beta_1}, \frac{\partial \mathbf{L}(\theta)}{\partial \beta_2}, \frac{\partial \mathbf{L}(\theta)}{\partial \gamma_1}, \cdots, \frac{\partial \mathbf{L}(\theta)}{\partial \gamma_p}\right)^\top,$$

where

$$\begin{aligned}
\frac{\partial \mathbf{L}(\theta)}{\partial \beta_0} &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{ti}, \\
\frac{\partial \mathbf{L}(\theta)}{\partial \beta_1} &= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial \varepsilon_{ti}}{\partial \beta_1}\varepsilon_{ti} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\sum_{j=1}^{N} W_{ij} Y_{(t-1)j}\right)\varepsilon_{ti}, \\
\frac{\partial \mathbf{L}(\theta)}{\partial \beta_2} &= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial \varepsilon_{ti}}{\partial \beta_2}\varepsilon_{ti} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}\varepsilon_{ti},
\end{aligned}$$

$$\frac{\partial \mathbf{L}(\theta)}{\partial \gamma_P} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Z_{Pi}\varepsilon_{ti}, \text{where } P = 1, \cdots, p.$$

In addition,

$$\mathbf{I}(\theta) = \begin{pmatrix} \mathbf{I_{11}}(\theta) & \mathbf{I_{12}}(\theta) \\ \mathbf{I_{21}}(\theta) & \mathbf{I_{22}}(\theta) \end{pmatrix}, \tag{A.3}$$

where

$$\begin{aligned} \mathbf{I_{11}}(\theta) &= E\left(-\frac{\partial^2 L(\theta)}{\partial \beta_3^2}\right) = \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}^2 I_{\{Y_{(t-1)i<r}\}}, \\ \mathbf{I_{12}}(\theta) &= E\begin{pmatrix} -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \gamma_1} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \gamma_p} \end{pmatrix} = \{\mathbf{I_{21}}(\theta)\}^\top. \end{aligned}$$

Next, we calculate the elements of $\mathbf{I_{12}}(\theta)$ separately,

$$\begin{aligned} E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_0}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i} I_{\{Y_{(t-1)i<r}\}}, \\ E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_1}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} \left(\sum_{j=1}^{N} W_{ij} Y_{(t-1)j}\right) Y_{(t-1)i} I_{\{Y_{(t-1)i<r}\}}, \\ E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \beta_2}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}^2 I_{\{Y_{(t-1)i<r}\}}, \\ E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_3 \partial \gamma_P}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}^2 I_{\{Y_{(t-1)i<r}\}} Z_{Pi}, P = 1, \cdots, p. \end{aligned}$$

Finally, we calculate the various elements of $\mathbf{I_{22}}(\theta)$,

$$\mathbf{I_{22}}(\theta) = E\begin{pmatrix} -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0^2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \beta_2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \gamma_1} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \gamma_p} \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1^2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \beta_2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \gamma_1} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_1 \partial \gamma_p} \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2^2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \gamma_1} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_2 \partial \gamma_p} \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \beta_2} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1^2} & \cdots & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_1 \partial \gamma_p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_p \partial \beta_0} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_p \partial \beta_1} & -\frac{\partial^2 \mathbf{L}(\theta)}{\partial \gamma_p \partial \beta_2} & -\frac{\partial^2 L(\theta)}{\partial \gamma_p \partial \gamma_1} & \cdots & -\frac{\partial^2 L(\theta)}{\partial \gamma_p^2}, \end{pmatrix}$$

where

$$E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0^2}\right) = \frac{NT}{\sigma^2}, E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial \beta_0 \partial \beta_1}\right) = \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} Y_{(t-1)j},$$

$$
\begin{aligned}
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\beta_0\partial\beta_2}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}, \\
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\beta_0\partial\gamma_P}\right) &= \frac{T}{\sigma^2}\sum_{i=1}^{N} Z_{Pi}, \text{where } P = 1, \cdots, p, \\
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\beta_1^2}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\sum_{j=1}^{N} W_{ij}Y_{(t-1)j}\right)^2, \\
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\beta_1\partial\beta_2}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\sum_{j=1}^{N} W_{ij}Y_{(t-1)j}\right)Y_{(t-1)i}, \\
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\beta_1\partial\gamma_P}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\sum_{j=1}^{N} W_{ij}Y_{(t-1)j}\right)Z_{Pi}, \text{where } P = 1, \cdots, p, \\
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\beta_2^2}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}^2, \\
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\beta_2\partial\gamma_P}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N} Y_{(t-1)i}Z_{Pi}, \\
E\left(-\frac{\partial^2 \mathbf{L}(\theta)}{\partial\gamma_P\partial\gamma_Q}\right) &= \frac{T}{\sigma^2}\sum_{i=1}^{N} Z_{Pi}Z_{Qi}, \text{where } P, Q = 1, \cdots, p.
\end{aligned}
$$

$\square$

Appendix B

The Principle of OAuth Authentication

In order to crawl these data, we need to log in with the username and password of Twitter, go inside the website, and grab as much data as possible by focusing on a large number of users. Of course, due to the privacy settings of some users, we can't capture all the user data, but those users whose properties are public can basically meet the needs of our analysis. Twitter provides an API interface for us to use, which provides us with great convenience in crawling data. To ensure security, the website authenticates users based on OAuth authentication.

In order to get the data of the Twitter website, we first need to log in to the website and then obtain the permission to complete the crawling of the information resources. Twitter website uses OAuth authentication to control the permissions of third-party applications, so we will briefly introduce the knowledge of OAuth authentication.

The purpose of the OAuth protocol is to provide a simple, secure, and open standard for the authorization of third-party applications. Through this protocol, a third-party application can obtain the authorization of the website for the user resource without knowing the username and password. Because OAuth authentication is simple and secure, more and more Internet service providers provide API access interfaces and use OAuth authentication to authorize users.

The process of OAuth authentication is shown in Figure B.1 and the steps of this process are described as follows:

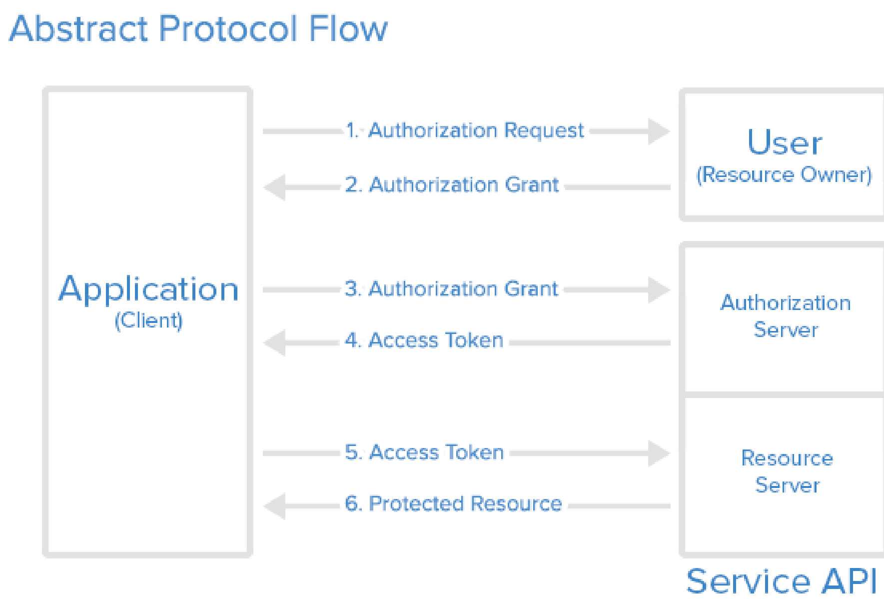


Figure B.1: The Process of OAuth Authentication

1. The client asks the user to give authorization;

2. User agrees to grant authorization;

3. Request a token from the authentication server based on the authorization obtained in the previous step;

4. The authentication server authenticates the authorization and issues the token after confirmation;

5. The client uses the token to request resources from the resource server;

6. The resource server uses the token to confirm the correctness of the token to the authentication server, and provides resources after confirmation.

Twitter's OAuth authentication is consistent with the basic authentication process. We only need to get the Access Token and attach an Access Token in each HTTP request that needs to authenticate from the Twitter API.

Appendix C

Useful Code

C.1 Code for User ID of the Followers

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
import time
import sys
import csv

#http://www.tweepy.org/
import tweepy

#Get your Twitter API credentials and enter them here
consumer_key = "**"
consumer_secret = "**"
access_key = "**"
access_secret = "**"

#http://tweepy.readthedocs.org/en/v3.1.0/getting_started.html#api
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_key, access_secret)
api = tweepy.API(auth)
```

```
ids = []
for page in tweepy.Cursor(api.followers_ids,
screen_name="UniStrathclyde").pages():
ids.extend(page)
```

C.2 Code for Number of Weekly Tweets

```
#method to get a user's last 100 tweets
def get_tweets(userid):
user=api.get_user(userid)
username = user.screen_name
print "writing to {0}_tweets.csv".format(userid)
if (user.followers_count>80 and user.friends_count>40 and
user.statuses_count>200):
number_of_tweets = 200
tweets = api.user_timeline(id = username,count = number_of_tweets)
tweets_for_time=[tweet.created_at for tweet in tweets]
tweets_length=[len(tweet.text) for tweet in tweets]
length=[0,0,0,0,0,0,0,0]
days=datetime.date(2017,2,14)-
tweets_for_time[len(tweets_for_time)-1].date()
if days.days>56:
for i in range(len(tweets_for_time)):
days=datetime.date(2017,2,14)-tweets_for_time[i].date()
if days.days<7:
length[0]=length[0]+tweets_length[i]
elif 7<=days.days<14:
length[1]=length[1]+tweets_length[i]
elif 14<=days.days<21:
length[2]=length[2]+tweets_length[i]
elif 21<=days.days<28:
length[3]=length[3]+tweets_length[i]
elif 28<=days.days<35:
length[4]=length[4]+tweets_length[i]
```

```
elif 35<=days.days<42:
length[5]=length[5]+tweets_length[i]
elif 42<=days.days<49:
length[6]=length[6]+tweets_length[i]
elif 49<=days.days<56:
length[7]=length[7]+tweets_length[i]

with open('xuhui1.csv', 'ab') as f:
writer = csv.writer(f)
data = [userid,username,length[0],length[1],length[2],length[3],
length[4],length[5],length[6],length[7]]
writer.writerow(data)
else:
print "user {0} have more than 200 tweets in 8 weeks".format(userid)




for i in range(len(ids)):
get_tweets(ids[i])
```

C.3 Code for Follower User ID

```
with open('**.csv', 'r') as f:
data = f.readlines()
nameid=[]
for line in data:
line=line.replace('\n','').split(',')
nameid.append(line[0])

del nameid[0]

int_nameid=[int(i) for i in nameid]
```

C.4 Code for Relationship Matrix

```
for i in range(numberapi):
auth = tweepy.OAuthHandler(consumer_key[i], consumer_secret[i])
auth.set_access_token(access_key[i], access_secret[i])
api[i] = tweepy.API(auth)

d=(len(int_nameid))
c=range(numberapi)*int(len(ids)/numberapi)

def get_follows(userid,api):
b=[int(userid)]
user=api.get_user(userid)
username = user.screen_name
follower_id = api.followers_ids(id = username)
a=np.zeros(d)
set1 = set(int_nameid)
set2 = set(follower_id)
set3 = set1 & set2
followee=list(set3)
print (set3)
if len(set3)==0:
with open( 'unfollow.csv', 'ab') as f:
writer = csv.writer(f)
writer.writerow(b)
print "user {0} unfollow".format(userid)
return
else:
for i in range(len(set3)):
n=int_nameid.index(followee[i])
print n
a[n]=1
b.extend(a)
```

```
with open('matrix.csv', 'ab') as f:
writer = csv.writer(f)
writer.writerow(b)
print "user {0} finished".format(userid)

for j in range(len(ids)):
get_follows(ids[j],api[c[int(j/13)]])
time.sleep(20)
```