Modelling Volatilities of High-Dimensional Time Series with Network Structure and Asymmetry

Yue Pan
Department of Mathematics and Statistics
University of Strathclyde
Glasgow, UK

October 29, 2024

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

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Jiazhu Pan, for his invaluable mentorship and support. His guidance and feedback have been instrumental in shaping the direction and quality of this thesis. His dedication to research has been inspiring throughout my academic and personal growth.

I am also immensely thankful to my parents for their endless support, love, and encouragement. Their belief in my abilities and their constant motivation have been a source of strength and inspiration throughout my time in Glasgow. I could not have reached this milestone without their unwavering support.

Additionally, I would like to extend my sincere thanks to the staff members of the Department of Mathematics and Statistics for their assistance and support. Their help has been invaluable in fostering a collaborative and nurturing academic environment, which has greatly facilitated my research at the University of Strathclyde.

Abstract

This thesis explores innovative methodologies for modelling volatilities of network-based high-dimensional time series that exhibit asymmetry. We start with a law of large numbers and a central limit theorem for triangular arrays of random fields that are non-stationary. We derive key intermediate results to bridge the gap between the proposed limit theorems and their application to the inference of high-dimensional time series under large dimension N and sample size T. These theoretical advancements are exemplified through a maximum likelihood estimation of a network autoregressive model.

Building on this foundation, we propose a threshold network GARCH (TNGARCH) model that incorporates asymmetries in the reaction of conditional variances to positive and negative shocks. Taking integer-valued data into account, we also propose a Poisson TNGARCH (PTNGARCH) model, which has an unknown threshold that can be estimated alongside other parameters. For both models, the stationarity over time is investigated, and the maximum likelihood estimation is proved to be consistent and asymptotically normal for large N and T. The asymptotic properties are tested by simulation studies. For real data analysis, we fit the TNGARCH model to the daily log-returns of stocks from two Chinese stock markets and the PTNGARCH model to the daily counts of car accidents in New York City neighbourhoods. Wald tests are conducted to show the asymmetry in both data sets.

Additionally, we establish unified methodologies for a class of network GARCH models with conditional distributions in the one-parameter exponential family. This theoretical framework is applied to a new negative binomial TNGARCH model. We evaluate its performance against the Poisson TNGARCH model using the same car accident data, employing a probability integral transformation test for comparative analysis.

Keywords: High-dimensional time series, conditional heteroscedasticity, threshold GARCH, integer-valued GARCH, network GARCH, limit theorems, arrays of random fields.

Contents

List of Notations

- $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{N}$ denote the sets of real numbers, integers and non-negative integers, respectively.

- $\|\cdot\|$ denotes a generic norm unless otherwise specified.

- $\mathbb{L}^p(E, \mathcal{E}, m)$ denotes classes of measurable functions:

$$\|f\|_p = \left(\int_E |f(x)|^p dm(x) \right)^{\frac{1}{p}} < \infty.$$

  Specifically, $\|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}$ for a random variable X.

- $\mathbb{L}^\infty(E, \mathcal{E}, m)$ denotes classes of measurable functions:

$$\|f\|_\infty = \inf\{C > 0 : |f| \leq C \textit{ almost everywhere}\} < \infty.$$

- $|A|_c$ denotes the cardinality of a finite set A.

- $(\mathcal{X}, d_x)$ and $(\mathcal{Y}, d_y)$ are two metric spaces equipped with metrics d_x and d_y, then

$$\mathrm{Lip}(f) = \sup_{X_1, X_2 \in \mathcal{X}} \frac{d_y(f(X_1), f(X_2))}{d_x(X_1, X_2)}$$

  denotes the Lipschitz constant of function $f : \mathcal{X} \mapsto \mathcal{Y}$.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Volatility refers to the fluctuation of prices of assets in a financial context. A widely conducted approach to volatility modelling starts from the autoregressive conditional heteroscedasticity (ARCH) model (Engle, 1982), which depicts the conditional variance of the United Kingdom's inflation as linearly dependent upon past realizations. Bollerslev (1986) then proposed a generalized ARCH (GARCH) model to further accommodate the dependence of the conditional variance on its own past. The GARCH model has become one of the most popular models in econometrics, and numerous variations of it have been developed ever since, extending the scope of econometric phenomena that can be explained by GARCH models. In this research, we focus on three directions in which the GARCH model has been extended in the literature:

- GARCH models that depict asymmetry in the response;
- GARCH models for integer-valued data;
- High-dimensional GARCH models for spatio-temporal data.

Aiming to accommodate asymmetry in volatility, GARCH models with threshold structures were proposed: Glosten et al. (1993) fitted their GJR-GARCH to the monthly returns of a stock index and found that the variance responds differently to positive and negative shocks. Zakoïan (1994) and Nelson (1991) also observed asymmetry in the standard deviation and log-transformed variance through their threshold GARCH (TGARCH) and exponential GARCH (EGARCH) models, respectively. In the study of integer-valued data, some authors replaced the conditional Gaussian distribution in the original GARCH model with discrete distributions, including the Poisson distribution (Ferland et al., 2006; Fokianos et al., 2009; Wang et al., 2014), the binomial distribution (Ristić et al., 2016), and the non-negative binomial distribution (Zhu, 2010; Xu et al., 2012). In particular, Wang et al. (2014) found asymmetry in annual earthquake counts through their threshold Poisson autoregressive model.

The aforementioned GARCH variations are limited to univariate cases. Starting from Bollerslev et al. (1988), a series of multivariate GARCH (MGARCH) models have been developed (Bollerslev, 1990; Engle and Kroner, 1995; Tse and Tsui, 2001; Engle, 2002), aiming to simultaneously study the dynamic structure in the conditional covariances between cross-sectional variables. However, the number of parameters of these MGARCH models increases with the dimension, causing significant challenges in statistical inference. Therefore, applications of MGARCH models are often limited to multivariate data of very low dimension, such as two stock indices (Karolyi, 1995) or exchange rates of two currencies (Tse and Tsui, 2001). To circumvent the over-parameterization problem in MGARCH models, some authors use a network to describe cross-sectional relations instead of relying on dynamic conditional covariances. This idea was first applied by Zhu et al. (2017) to their network vector autoregression (NAR), in which the number of parameters is fixed even with increasing dimension. It led to a series of subsequent studies, including Zhou et al. (2020)'s network GARCH (NGARCH) model, Xu et al. (2024)'s dynamic network quantile regression (DNQR), and Armillotta and Fokianos (2024)'s Poisson network autoregressive (PNAR) model. Distinct from traditional multivariate time series models, these network-based models are capable of handling time series with very high dimensions. For example, Zhou et al. (2020) fitted their NGARCH to daily log returns observed simultaneously from hundreds of stocks, and Armillotta and Fokianos (2024)'s PNAR was used to analyse monthly crime numbers from 552 blocks in Chicago.

An open issue related to high-dimensional GARCH models is to establish limit theorems for statistics under increasing dimension. Zhou et al. (2020)'s network GARCH assumed a fixed dimension N. The asymptotic properties of their quasi maximum likelihood estimation (QMLE) hold under increasing temporal sample size, i.e., $T \to \infty$. However, their limit theorems cannot be applied when both $T \to \infty$ and $N \to \infty$. An innovative approach to this problem was proposed by Xu et al. (2024), who regarded samples under large N and large T as random fields, deriving the asymptotic properties using limit theorems for random fields. In this research, we will use the idea of treating high-dimensional time series as random fields and establish methodologies to estimate a series of network GARCH models under large N and large T.

1.2 Research method

With the background being introduced, in this research we make the following contributions to modelling volatilities of high-dimensional time series:

- We propose two new network GARCH models that accommodate asymmetric and potentially integer-valued spatio-temporal data in a large-scale network;
- We propose general methodologies that apply to a wide range of network GARCH models with different conditional distributions and structures;
- We develop limit theorems for non-stationary arrays of random fields, and we apply these results to the proposed models, establishing parameter estimations that are consistent and asymptotically normal when $T \to \infty$ and $N \to \infty$.

As we have mentioned in the background, the first issue in the development of high-dimensional GARCH models is the over-parameterization problem caused by a large N. To address this issue, we adopt the idea of incorporating an observed network into the model. Compared to other parameter-reduction techniques, such as the conditional correlations (Bollerslev, 1990; Tse and Tsui, 2001) and the Factor-GARCH models (Engle et al., 1990; Pan et al., 2010; Li et al., 2016), the network approach is advanced in two aspects: first, it is natural to describe cross-sectional relations by a network;

second, the number of parameters is fixed under increasing N. Further introduction and comments on the network approach and other parameter reduction techniques are made in Section 2.2.

Empirical evidence has shown that bad news and good news have asymmetric effects on predictable volatility (Black, 1976; French et al., 1987). In this research, we use the self-excited threshold structure to capture this asymmetry, as in the GJR-GARCH (2.1.5) by Glosten et al. (1993). This choice is based on the work of Engle and Ng (1993), who fitted different asymmetric GARCH models to daily stock return data, and the GJR-GARCH (Glosten et al., 1993) outperformed the others. We propose a threshold network GARCH model (4.1.1) with a threshold of 0, which is appropriate for analyzing stock returns. However, in the Poisson threshold network GARCH model (5.2.1) that is proposed for non-negative integer-valued data, we follow Wang et al. (2014) and let the threshold value be an unknown integer, which can be estimated simultaneously with other parameters. We also propose a negative binomial threshold network GARCH (6.5.2) as an example of the generalized network GARCH model (6.2.2), following the piece-wise threshold structure of Samia and Chan (2011).

Maximum likelihood estimation is a conventional approach in estimating univariate GARCH models and multivariate GARCH models with fixed N. See Francq and Zakoïan (2004) and Zhou et al. (2020) for example. The asymptotic properties of maximum likelihood estimation are based on limit theorems that do not apply when both $T \to \infty$ and $N \to \infty$. Xu et al. (2024) adopted the limit theorems for random fields in the inference of their dynamic network quantile regression (DNQR) model, establishing a consistent and asymptotically normal instrumental variable estimation that accommodates the large N case. As far as we know, it was the first time limit theorems for random fields were applied in the statistical inference of a high-dimensional model. Xu et al. (2024) used the limit theorems proposed by Jenish and Prucha (2012) under near-epoch dependence (NED), which is a spatial dependence measure of random fields. However, NED cannot be used without an auxiliary random field (see our introduction and comments on NED in Section 2.4.2), as in models (5.2.1) and (6.2.2). In this research, we use the ϵ-weak dependence measure of random fields in the sense of

Doukhan and Louhichi (1999). There is no preceding work in the literature that applies the limit theorems under ϵ-weak dependence to the inference of high-dimensional time series models. We fill the gap in two aspects: first, we extend existing limit theorems for ϵ-weakly dependent random fields (Dedecker et al., 2007; Curato et al., 2022) to accommodate non-stationarity; second, we establish some important results to facilitate the application of proposed limit theorems, e.g., the preservation of ϵ-weak dependence under transformations and infinite shifts.

1.3 Thesis outline

Chapter 2 is a review of preliminary knowledge regarding GARCH models and random fields. First, a series of univariate GARCH models are introduced, including asymmetric GARCH models and integer-valued GARCH models. Then, we introduce the multivariate GARCH models and the over-parameterization problem caused by high dimension. Conditional correlation GARCH and Factor-GARCH are introduced as classic approaches to parameter reduction, followed by the introduction of the network approach. At last, we introduce two dependence measures for random fields, namely NED and ϵ-weak dependence. We comment on the limitations of the NED measure and existing limit theorems under ϵ-weak dependence.

In Chapter 3, we propose a law of large numbers and a central limit theorem for random fields that are weakly dependent with respect to θ- and η-coefficients. Some properties of these coefficients are also derived, including weak dependence under transformations and infinite shifts. These intermediate results fill the gap between the proposed limit theorems and their application to the inference of high-dimensional time series. We establish a maximum likelihood estimation that is consistent and asymptotically normal under large NT. Finally, we apply our results to estimate a network autoregressive model as an example.

In Chapter 4, we propose a threshold network GARCH (TNGARCH) model. The major part of this chapter has been published in Pan and Pan (2024). Compared to Zhou et al. (2020)'s network GARCH, we also consider asymmetry in how conditional variances react to positive and negative shocks. The stationarity of proposed model is checked under fixed N, and the limit theorems of QMLE are investigated under large N and large T. We fit the model to log-returns of four groups of stocks from the Shanghai Stock Exchange and the Shenzhen Stock Exchange. A Wald statistic is proposed to test the existence of the threshold, and a high-dimensional white noise test is carried out to check the model adequacy.

In Chapter 5, we consider an extension of the TNGARCH model to accommodate integer-valued high-dimensional time series, where the conditional distribution is assumed to be Poisson. Unlike the continuous-valued TNGARCH, the threshold value in this Poisson TNGARCH (PTNGARCH) model is unknown. We propose a two-step maximum likelihood estimation (MLE) method to estimate the threshold and other parameters simultaneously. The model is fitted to the daily counts of car accidents in different neighborhoods of New York City.

In Chapter 6, we establish unified methodologies for a class of network GARCH models with conditional distributions in the one-parameter exponential family. The stationarity under fixed N is checked, and we establish a consistent and asymptotically normal maximum likelihood estimation under large N and large T. The results are applied to a new negative binomial TNGARCH model. We fit this model to the same car accident data and compare its performance against the Poisson TNGARCH model through a probability integral transformation test.

Chapter 2

Preliminaries

2.1 Univariate GARCH models

Let $\{y_t : t \in \mathbb{Z}\}$ be a univariate time series. A major purpose of time series models is to forecast the future based on past information. For example, the one-step forecast of y_t based on past information $\mathcal{H}_{t-1}$ is the conditional mean $\mathbb{E}(y_t|\mathcal{H}_{t-1})$. Conventional econometric models assumed constant conditional variance $\mathrm{Var}(y_t|\mathcal{H}_{t-1})$. For example, a first-order autoregressive model is written as:

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{2.1.1}$$

where $\{\varepsilon_t : t \in \mathbb{Z}\}$ is independently and identically distributed (IID) with mean 0 and $\mathrm{Var}(\varepsilon_t) = \sigma^2$. Then the one-step forecast is $\mathbb{E}(y_t|y_{t-1}) = \phi y_{t-1}$ with conditional variance $\mathrm{Var}(y_t|y_{t-1}) = \sigma^2$. However, some econometric forecasters found that randomness associated with forecasts changes widely over time (McNees, 1979), hence the constant conditional variance seems inappropriate. Engle (1982) proposed the ARCH model in order to accommodate time-varying conditional variance in time series forecasting. A first-order ARCH model is written as:

$$\begin{aligned} y_t|\mathcal{H}_{t-1} &\sim N(0, h_t), \\ h_t &= \omega + \alpha y_{t-1}^2, \end{aligned} \tag{2.1.2}$$

where $\omega > 0$ and $\alpha \geq 0$, ensuring the positiveness of h_t. The conditional distribution of y_t is assumed to be normal with mean 0 and variance h_t. In this setting, the conditional variance $\mathrm{Var}(y_t|y_{t-1}) = h_t$ is allowed to change over time.

Bollerslev (1986) proposed a natural generalization of the ARCH model, namely GARCH. A GARCH(1,1) model has the following form:

$$\begin{aligned} y_t|\mathcal{H}_{t-1} &\sim N(0, h_t), \\ h_t &= \omega + \alpha y_{t-1}^2 + \beta h_{t-1}, \end{aligned} \tag{2.1.3}$$

where $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$, ensuring the positiveness of h_t. Unlike the ARCH model, the conditional variance in the GARCH model is related to both y_{t-1}^2 and its own past h_{t-1}. This feature allows the GARCH model to incorporate longer memory in the conditional variance. In fact, it can be regarded as an ARCH(∞) model when $0 < \beta < 1$ (p.309, Bollerslev, 1986). The GARCH model has been widely applied in econometric studies, leading to a series of extensions that accommodate additional features not described by the original GARCH model. Next, we will introduce two types of extended GARCH models: threshold GARCH models, which allow for asymmetry in the conditional variance, and integer-valued GARCH models, which are designed for count data. For readers interested in other types of GARCH variations, we recommend Teräsvirta (2009)'s survey of univariate GARCH-type models.

2.1.1 Asymmetric GARCH models

According to the empirical works by Black (1976) and French et al. (1987) among others, sometimes an unexpected drop in price (bad news) increases predictable volatility more than an unexpected increase in price (good news) of similar magnitude. The original GARCH model (2.1.3) cannot explain this effect, as the impact of y_{t-1} on h_t is not related to the sign of y_{t-1}. Different extensions of the GARCH model have been proposed to accommodate this effect by incorporating different coefficients around the threshold 0. We list two important examples below.

Example 2.1.1. Nelson (1991) proposed an exponential GARCH as follows:

$$\begin{aligned} y_t &= \varepsilon_t \sqrt{h_t}, \\ \log(h_t) &= \omega + \alpha\varepsilon_{t-1} + \gamma\left[|\varepsilon_{t-1}| - \mathbb{E}|\varepsilon_{t-1}|\right] + \beta\log(h_{t-1}), \end{aligned} \tag{2.1.4}$$

where $\{\varepsilon_t : t \in \mathbb{Z}\}$ is IID with mean 0 and variance 1. Nelson (1991) fitted his model to daily returns for a value-weighted market index and obtained a negative estimation of α, indicating that negative shocks generate more volatility than positive shocks.

Example 2.1.2. Glosten et al. (1993) proposed a more natural specification, namely GJR-GARCH, as follows:

$$h_t = \omega + \alpha y_{t-1}^2 + \gamma 1_{\{y_{t-1}>0\}} y_{t-1}^2 + \beta h_{t-1}, \tag{2.1.5}$$

where $1_{\{y_{t-1}>0\}}$ is an indicator that equals 1 if $y_{t-1} > 0$ and 0 otherwise. It is assumed that $\omega > 0$, $\alpha \geq 0$, $\gamma \geq 0$ and $\beta \geq 0$, to ensure the positiveness of h_t. Therefore, the slope of y_{t-1}^2 is $\alpha + \gamma$ on the right side of the threshold 0 and α on the left side. Their empirical results agreed with those of Nelson (1991), as the estimation of γ was negative.

Engle and Ng (1993) conducted an excellent comparison between different specifications of asymmetric ARCH/GARCH models, including the EGARCH and GJR-GARCH, on daily Japanese stock return data. Their empirical results suggested that the GJR-GARCH outperformed the others. Therefore, in Chapter 4, we will adopt a threshold structure similar to the GJR-GARCH, as we will also analyze daily stock return data.

2.1.2 Integer-valued GARCH models

The original GARCH model (2.1.3) assumes a normal conditional distribution of y_t, while in the EGARCH model (2.1.4) and the GJR-GARCH model (2.1.5), the conditional distribution is centered and continuous. These GARCH models are not capable of handling time series of counts, such as the trading volume of houses in the real estate market De Wit et al. (2013), the number of stock transactions Jones et al. (1994), or the daily mortality from COVID-19 Pham (2020).

A natural idea is to consider a discrete conditional distribution of y_t. For example, a conditional Poisson distribution was considered by Heinen (2003), with a GARCH-type autoregressive conditional intensity. A similar specification was also used by the integer GARCH (INGARCH) model of Ferland et al. (2006) and the Poisson autoregression (PAR) of Fokianos et al. (2009). There are other specifications in the literature, according to different features presented by the data. For example, Ristić et al. (2016) considered a conditional binomial distribution for bounded $\mathbb{Z}$-valued data, and Zhu (2010) proposed a negative binomial integer-valued GARCH (NB-INGARCH) model to handle $\mathbb{Z}$-valued data with over-dispersion. We give two examples below.

Example 2.1.3. A GARCH(1,1) model with a conditional Poisson specification has the following form:

$$\begin{aligned} &y_t|\mathcal{H}_{t-1} \sim Poisson(\lambda_t), \\ &\lambda_t = \omega + \alpha y_{t-1} + \beta \lambda_{t-1}, \end{aligned} \tag{2.1.6}$$

where $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$, ensuring the positiveness of λ_t. The conditional Poisson intensity λ_t can be interpreted as the conditional variance or the conditional mean. Accordingly, model (2.1.6) can be named integer-valued GARCH (Ferland et al., 2006) or Poisson autoregression (Fokianos et al., 2009). We tend to use the name INGARCH since the autoregressive structure of the intensity process parallels that of (2.1.3).

Example 2.1.4. An NB-INGARCH(1,1) model from Zhu (2010) is written as follows:

$$\begin{aligned} &y_t|\mathcal{H}_{t-1} \sim NB(K, p_t), \\ &\frac{1-p_t}{p_t} = \lambda_t = \omega + \alpha y_{t-1} + \beta \lambda_{t-1}, \end{aligned} \tag{2.1.7}$$

where $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$, ensuring the positiveness of $\frac{1-p_t}{p_t}$. Here, y_t is the count of failures before K successes, and p_t is the (conditional) probability of success. Although $\lambda_t = \frac{1-p_t}{p_t}$ is not the conditional variance, it is still appropriate to call it a GARCH-type model, since the conditional variance $\frac{K(1-p_t)}{p_t^2}$ changes over time.

Analogous to the continuous-valued case, integer-valued time series can also exhibit asymmetry. For example, Wang et al. (2014) proposed a self-excited threshold Poisson autoregression (SETPAR) model and fitted it to annual counts of earthquakes, with results showing an asymmetric structure around an estimated threshold of 25. With a conditional Poisson distribution of y_t, a SETPAR specification on the conditional intensity λ_t is:

$$\lambda_t = \begin{cases} \omega_1 + \alpha_1 y_{t-1} + \beta_1 \lambda_{t-1} & y_{t-1} \leq r, \\ \omega_2 + \alpha_2 y_{t-1} + \beta_2 \lambda_{t-1} & y_{t-1} > r. \end{cases} \tag{2.1.8}$$

To ensure the positiveness of λ_t, it is supposed that ω_1 and ω_2 are positive, α_1, α_2, β_1 and β_2 are non-negative. Here, the coefficient parameters change when y_{t-1} is on different sides of the threshold r. Unlike the GJR-GARCH, the threshold r is a positive integer to be estimated simultaneously with the coefficient parameters. Wang et al. (2014) suggested a two-step maximum likelihood estimation. Because the threshold is an integer, it is not computationally demanding to search within a range of candidates for the one that maximizes the likelihood.

Davis and Liu (2016) established a unified theory and inference related to a class of GARCH models with conditional distributions in the one-parameter exponential family and the accompanying process of the conditional mean $\mu_t = \mathbb{E}(y_t|\mathcal{H}_{t-1})$ evolving as a function of y_{t-1} and μ_{t-1}. i.e.:

$$\mu_t = g_\theta(y_{t-1}, \mu_{t-1}). \tag{2.1.9}$$

The one-parameter exponential family includes a wide range of continuous and discrete distributions, e.g., the normal distribution, Poisson distribution, and binomial distribution, among others. Therefore, Davis and Liu (2016)'s methodology applies to many classical GARCH models, including the original GARCH (2.1.3), the GJR-GARCH (2.1.5), the Poisson GARCH (2.1.6), and the NB-INGARCH (2.1.7).

2.2 Multivariate GARCH models

Let $\{\mathbf{y}_t : t \in \mathbb{Z}\}$ be an N-dimensional time series where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})'$. The univariate volatility models (2.1.2) and (2.1.3) concern the dynamics of the conditional variance. So it is natural to consider the dynamics of conditional covariance matrix in the multivariate case. Bollerslev et al. (1988) first proposed a multivariate GARCH model as follows:

$$\begin{aligned} \mathbf{y}_t|\mathcal{H}_{t-1} &\sim \mathcal{N}(0, H_t), \\ \mathrm{vech}(H_t) &= C + A\,\mathrm{vech}(\mathbf{y}_{t-1}\mathbf{y}'_{t-1}) + B\,\mathrm{vech}(H_{t-1}). \end{aligned} \tag{2.2.1}$$

The conditional distribution of $\mathbf{y}_{t-1}$ is a multivariate normal distribution $\mathcal{N}(0, H_t)$. H_t is the $N \times N$ conditional covariance matrix, and the $\frac{1}{2}N(N+1)$-dimensional vector $\mathrm{vech}(H_t)$ is the stacked columns of H_t. A total number of $\frac{1}{2}N(N+1) + \frac{1}{2}N^2(N+1)^2$ parameters are contained in A, B, and C.

The MGARCH model of Bollerslev et al. (1988) is very general, as it considers all cross-sectional relations between these conditional variances and covariances. However, the positive definiteness of H_t cannot be ensured by simply restraining A, B and C, and the number of parameters can be excessive even when N is only mildly large. These two problems make the estimation of parameters computationally infeasible in practice. In fact, in the empirical study on the quarterly percentage returns from three assets ($N = 3$), Bollerslev et al. (1988) used a simplified version of (2.2.1) with diagonal matrices A and B. The over-parameterization problem is one of the major obstacles in the development of multivariate GARCH models, and many efforts have been made in the literature to reduce the number of parameters without sacrificing too many cross-sectional relations. For a systematic survey of multivariate GARCH models, we recommend Bauwens et al. (2006). Selected examples of parameter reduced multivariate GARCH models will be introduced in the subsequent sections.

2.2.1 CCC-GARCH and DCC-GARCH

Bollerslev (1990) proposed a constant conditional correlation GARCH (CCC-GARCH) model, which only considers dynamics in the diagonal elements (conditional variances) of the conditional covariance matrix H_t. Specifically, a CCC-GARCH(1,1) model assumes that

$$\begin{aligned} h_{ii,t} &= \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{ii,t-1}, & i &= 1, 2, ..., N; \\ h_{ij,t} &= \rho_{ij}\sqrt{h_{ii,t}h_{jj,t}}, & j &\neq i. \end{aligned} \tag{2.2.2}$$

Each conditional variance $h_{ii,t}$ follows a typical GARCH dynamic, while the conditional covariance $h_{ij,t}$ is derived from $h_{ii,t}$, $h_{jj,t}$, and a constant correlation ρ_{ij}. To ensure the positiveness of $h_{ii,t}$ it suffices that $\omega_i > 0$, $\alpha_i \geq 0$, and $\beta_i \geq 0$ for all $i = 1, 2, \ldots, N$. Additionally, the number of parameters in (2.2.2) is of order $\mathcal{O}(N^2)$, which is significantly reduced compared to model (2.2.1).

Tse and Tsui (2001) proposed a dynamic conditional correlation GARCH (DCC-GARCH) model, extending the CCC-GARCH model by assuming time-varying conditional correlations as follows:

$$\rho_{ij,t} = (1 - \theta_1 - \theta_2)\rho_{ij} + \theta_1\rho_{ij,t-1} + \theta_2\phi_{ij,t-1}, \tag{2.2.3}$$

where $\phi_{ij,t-1}$ is a function of lagged observation of y_{it}, $\theta_1 \geq 0$, $\theta_2 \geq 0$ and $\theta_1 + \theta_2 \leq 1$. The DCC-GARCH model adds more flexibility to the CCC-GARCH while retaining $\mathcal{O}(N^2)$ parameters, since the coefficients θ_1 and θ_2 (2.2.3) are the same for all $i = 1, 2, ..., N$.

Compared to model (2.2.1), the CCC-GARCH imposes a standalone GARCH dynamic with different coefficients on each conditional variance while assuming constant conditional correlations. The DCC-GARCH further imposes standalone GARCH dynamics with the same coefficients on each conditional correlation. However, these models still suffer from the over-parameterization problem with large N despite their significance in parameter reduction.

2.2.2 Factor-GARCH

Another approach to reducing parameters in multivariate GARCH models is using factor models. The idea behind Factor-GARCH is assuming that the problem of modeling the $N \times N$ conditional covariance matrix H_t can be reduced to modeling the $K \times K$ conditional covariance matrix of K common factors, where K is supposed to be much smaller than N. For example, Pan et al. (2010) assumed that

$$\mathbf{y}_t = A\mathbf{x}_t + \mathbf{v}_t, \tag{2.2.4}$$

where $\mathbf{x}_t$ is the K-dimensional vector of factors, $\mathbf{v}_t$ is an IID innovation vector with mean 0 and covariance Σ_v, and A is a $N \times K$ matrix of parameters. It is not hard to obtain that

$$H_t = AH_t^{(f)}A' + \Sigma_v,$$

where $H_t^{(f)} = \text{Var}(\mathbf{x}_t|\mathcal{H}_{t-1})$ is a $K \times K$ matrix. The first factor volatility model was proposed by Engle et al. (1990), followed by a series of extensions including Bollerslev and Engle (1993), Pan et al. (2010), Hu and Tsay (2014), and Li et al. (2016).

The performance of a factor model when fitted to high-dimensional time series largely relies on the estimated number of factors. For example, Li et al. (2016) fitted their factor GARCH model to daily returns of 196 stocks ($N = 196$), and the estimated number of factors was only 1 ($K = 1$), reducing the high-dimensional process to a univariate process. However, such efficiency in dimension reduction cannot be assured and completely depends on the data itself.

2.2.3 Network GARCH

Zhu et al. (2017) proposed an alternative approach to the aforementioned parameter-reduction and dimension-reduction methods by incorporating observed social relationships into a vector autoregression. They regarded $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$ as observations from N nodes in an undirected and weightless network, which should be observed in practice. The cross-sectional relations are represented by edges a_{ij}, where $a_{ij} = 1$ if node i and node j are connected, and $a_{ij} = 0$ otherwise. Moreover, Zhu et al. (2017)'s network AR model has a fixed number of parameters. They fitted their model to weekly $\log(1+x)$-transformed post lengths from $N = 2982$ social media users, where the network was established directly based on the follower-followee relations.

Zhou et al. (2020) saw the merits of network in high-dimensional time series analysis and established a network GARCH(1,1) as follows:

$$h_{it} = \omega + \alpha y_{i,t-1}^2 + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-1}^2 + \beta h_{i,t-1}, \quad i = 1, 2, \cdots, N. \tag{2.2.5}$$

where $w_{ij} = \frac{a_{ij}}{\sum_{k=1}^{N} a_{ik}}$, $\omega > 0$, $\alpha \geq 0$, $\lambda \geq 0$ and $\beta \geq 0$. Compared to conventional multivariate GARCH models, (2.2.5) uses minimum parameterization and depicts the cross-sectional relations in a direct way, i.e. the conditional variance h_{it} of an individual i is associated with past observations on its neighbours.

The parsimony in this idea leads to a series of network time series models, including network quantile autoregression (Zhu et al., 2019; Xu et al., 2024) and Poisson network autoregression (Armillotta and Fokianos, 2024). The network can be established according to empirical needs (Anselin, 1988). For example, Zhu et al. (2019) and Zhou et al. (2020) considered connections between stocks through common shareholders; Xu et al. (2024) saw two companies as connected if their headquarters are in the same city; Armillotta and Fokianos (2024) assumed that crime numbers from two geometrically neighboring locations are related.

2.3 Spatio-temporal GARCH models

Multivariate time series models belong to a much larger family of spatio-temporal models. Spatio-temporal models deal with random variables observed over both time and space. For example, a random variable $y_t(\mathbf{s})$ ($\mathbf{s} \in \mathbb{R}^2$) observed at time t on geographic coordinate $\mathbf{s}$ is spatio-temporal. An observation y_{it} from the network GARCH model (2.2.5) is also spatio-temporal, with the spatial location $\mathbf{s} = i$ ($i \in \mathbb{Z}$).

A general form of a spatio-temporal model is written as:

$$y_t(\mathbf{s}) = f^{(t,\mathbf{s})} \left(\theta^{(t,\mathbf{s})}, \mathbb{X}_t(\mathbf{s}), \varepsilon_t(\mathbf{s}) \right), \tag{2.3.1}$$

where $f^{(t,\mathbf{s})}$ is a space-time-specific function of the parameters $\theta^{(t,\mathbf{s})}$, a vector of explanatory variables $\mathbb{X}_t(\mathbf{s})$, and an error term $\varepsilon_t(\mathbf{s})$. In practice, it is essential to design a simplified version of model (2.3.1), as the general form typically includes more parameters than observations (Anselin, 1988). For example, Lu et al. (2009) proposed a spatio-temporal model with parameters varying only across spatial locations $\mathbf{s}$. The network GARCH model (2.2.5) is a more simplified version of (2.3.1), with parameters unchanged over both time and space. The Spatial Bilateral BEKK GARCH model proposed by Billio et al. (2023) incorporates space-varying parameters embedded within time-varying network structures.

The spatio-temporal GARCH model of Hølleland and Karlsen (2020) is a direct extension of the original GARCH model (2.1.3). It is written as:

$$
\begin{aligned}
y_t(\mathbf{s}) &= \varepsilon_t(\mathbf{s})\sqrt{h_t(\mathbf{s})}, \\
h_t(\mathbf{s}) &= \omega + \sum_{i=1}^{p} \sum_{\mathbf{u}\in\Delta_{1i}} \alpha_i(\mathbf{u}) y_{t-i}^2(\mathbf{s}-\mathbf{u}) + \sum_{i=1}^{q} \sum_{\mathbf{u}\in\Delta_{2i}} \beta_i(\mathbf{u}) h_{t-i}(\mathbf{s}-\mathbf{u}),
\end{aligned} \tag{2.3.2}
$$

where $\mathbf{s} \in \mathbb{Z}^d$ is a d-dimensional spatial location, $\omega > 0$, $\Delta_{1i} = \left\{\mathbf{u} \in \mathbb{Z}^d : \alpha_i(\mathbf{u}) \geq 0\right\}$ and $\Delta_{2i} = \left\{\mathbf{u} \in \mathbb{Z}^d : \beta_i(\mathbf{u}) \geq 0\right\}$. The original GARCH model (2.1.3) only considers the temporal heterogeneity of the conditional variance process, while model (2.3.2) further incorporates spatial effects.

Autoregressive spatio-temporal models have been extensively studied (Fan et al., 2003; Lu et al., 2009, 2024), but the study of spatio-temporal GARCH models is still in its early stages. Since high-dimensional time series data is a special case of spatio-temporal data where the location $\mathbf{s}$ is a single index $i \in \mathbb{Z}$, the network-based GARCH models proposed in this research are spatio-temporal GARCH models with highly specific structures. Similar to existing network-based models (Zhu et al., 2017; Zhou et al., 2020; Xu et al., 2024; Armillotta and Fokianos, 2024), we employ an explicit network structure to represent the spatial heterogeneity of conditional variances. Furthermore, it is the connection between high-dimensional time series data and spatio-temporal data that inspires us to consider the limit theorems for random fields in the estimation of high-dimensional time series models.

2.4 Limit theorems for random fields

In this research, we are interested in statistical inference for asymmetric network GARCH models considering the case when $N \to \infty$. N-dimensional time series model with $N \to \infty$ are also called spatio-temporal models by some authors. The empirical study of Zhou et al. (2020) showed that the network GARCH model can handle data with a large N; however, their quasi-maximum likelihood estimation was established with a fixed N. Their proof of asymptotic normality cannot be applied when $N \to \infty$ since it relied on the strict stationarity of the N-dimensional random vector $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$. Zhu et al. (2017) employed ordinary least squares estimation for their network AR model, and Armillotta and Fokianos (2024) adopted a quasi-maximum likelihood estimation for their Poisson network AR model, both allowing $N \to \infty$. However, their methods cannot be applied to GARCH-type models, and no parameter estimation for high-dimensional GARCH models in the existing literature remains valid when $N \to \infty$, as far as we know.

2.4.1 Limit theorems in the inference of high-dimensional models

Limit theorems are essential in statistical inference for time series models. Particularly, the classical law of large numbers (LLN) and the central limit theorem (CLT) are fundamental in establishing consistent and asymptotically normal parameter estimation.

Example 2.4.1. Let $\{X_t : t \in \mathbb{Z}\}$ be a time series of IID random variables with mean μ and variance σ^2, the classic law of large numbers and the central limit theorem are written as:

$$\begin{aligned} &(LLN) \quad \bar{X}_T \overset{a.s.}{\to} \mu; \\ &(CLT) \quad \sqrt{T}(\bar{X}_T - \mu) \overset{d}{\to} N(0, \sigma^2) \end{aligned}$$

as $T \to \infty$, where $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$ is the sample mean.

The IID assumption required by the classical limit theorems in Example 2.4.1 is too stringent in time series analysis. Limit theorems corresponding to different types of sequential dependence have been developed and applied in the estimation of time series models. For example, Zhu et al. (2017) and Zhou et al. (2020) used the central limit theorem for martingale difference sequences (Proposition 7.8, Hamilton, 2020). However, if the limit theorems for dependent time series are used in the inference for high-dimensional models, severe assumptions on limiting behaviors as $N \to \infty$ are often required. See, for examples, Assumption (C3) in Zhu et al. (2017), Assumption (C2) in Zhu et al. (2019), and Assumption (B3) in Armillotta and Fokianos (2024).

Remark. $\{X_t : t \in \mathbb{Z}\}$ is a martingale difference sequence if $\mathbb{E}(X_t | X_{t-1}, X_{t-2}, ...) = 0$ and $\mathbb{E}\|X_t\| < \infty$. This concept defines a measure of dependence of X_t on its past. However, if X_t is an N-dimensional vector, being a martingale difference sequence tells no information about the cross-sectional dependence.

In contrast, Xu et al. (2024) employed an instrumental variable estimation for their dynamic network quantile regression (DNQR) model that accommodates $N \to \infty$, and it was proved to be consistent and asymptotically normal by using the limit theorems for random fields. The dependence measure they use is the near-epoch dependence (NED), which was extended by Jenish and Prucha (2012) to random fields.

Remark. A random field is a collection of random variables $X_{\mathbf{i}} : \mathbf{i} \in D$ where $X_{\mathbf{i}}$ is indexed by $\mathbf{i}$ in a lattice D. For example, letting $\mathbf{i} = t \in \mathbb{Z}$, a time series $\{X_t : t \in \mathbb{Z}\}$ is a random field on the lattice $D = \mathbb{Z}$; letting $\mathbf{i} = (i, t) \in \mathbb{Z}^2$, a spatio-temporal process $\{y_{it} : (i, t) \in \mathbb{Z}^2\}$ is a random field on the lattice $D = \mathbb{Z}^2$.

Comparing to the conventional measures of serial dependence such as martingale difference, using a dependence measure for random fields seems to be more appropriate when investigating the spatio-temporal dependence.

2.4.2 Near-epoch dependent random fields

The limit theorems of Jenish and Prucha (2012) were established for random fields that are NED on a mixing random field. According to their definition, $\{X_{\mathbf{i}} : \mathbf{i} \in D\}$ is said to be $\mathbb{L}^p$-NED ($p \geq 1$) on $\{\varepsilon_{\mathbf{i}} : \mathbf{i} \in D\}$ if $\sup_{\mathbf{i} \in D} \|X_{\mathbf{i}}\|_p < \infty$, and

$$\|X_{\mathbf{i}} - E(X_{\mathbf{i}} | \mathcal{F}_{\mathbf{i}}(s)\|_p \leq d_{\mathbf{i}} \psi(s),$$

where $\mathcal{F}_{\mathbf{i}}(s) := \sigma\{\varepsilon_{\mathbf{j}} : \|\mathbf{j}-\mathbf{i}\| \leq s\}$, $\{\psi(s) : s \geq 1\}$ are non-negative constants such that $\lim_{s\to\infty}\psi(s) = 0$, and $\{d_{\mathbf{i}} : \mathbf{i} \in D\}$ are finite positive constants. Moreover, it is called uniformly NED if d_i's are uniformly bounded. It is called NED of size$-\mu$ if $\psi(s) = \mathcal{O}(s^{-\mu})$ for some $\mu > 0$.

NED measures the approximability of a random field by another random field that is mixing in the sense of Jenish and Prucha (2009). Therefore, NED becomes an invalid concept if such an auxiliary random field can not be appropriately identified. In the DNQR model of Xu et al. (2024), there exists a auxiliary random field consisting of IID uniformly distributed random variables $\{U_{it}\}$. On the other hand, in GARCH models with conditional Poisson distribution (Fokianos et al., 2009) or negative binomial distribution (Zhu, 2010), there is no appropriate auxiliary random field.

2.4.3 ϵ-weakly dependent random fields

Let $U \subseteq D$ and $V \subseteq D$ be two sub-lattices, with $\rho(U,V) = \min\{\|\mathbf{i}-\mathbf{j}\| : \mathbf{i} \in U, \mathbf{j} \in V\}$ defining their distance. $\mathfrak{X}_U$ denotes a collection of random variables $\{X_{\mathbf{i}} : \mathbf{i} \in U\}$. Doukhan and Louhichi (1999) introduced a dependence measure

$$\epsilon_{u,v}(s) = \sup\left\{\frac{|\operatorname{Cov}(f(\mathfrak{X}_U), g(\mathfrak{X}_V))|}{\Psi(f,g)} : f \in \mathcal{F}, g \in \mathcal{G}, |U|_c = u, |V|_c = v, \rho(U,V) \geq s\right\},$$

where $\mathcal{F}$ and $\mathcal{G}$ are two classes of real-valued functions and Ψ is a positive bounded real-valued function of $f \in \mathcal{F}$ and $g \in \mathcal{G}$. The ϵ-coefficient $\epsilon_{u,v}(s)$ is specified by Ψ and regularity conditions on $\mathcal{F}$ and $\mathcal{G}$, and it measures the dependence between two groups of random variables (with cardinality u and v respectively) that are s-apart spatially. In what follows, we give two examples of dependence coefficients.

Example 2.4.2. (Dedecker and Doukhan, 2003) ϵ-coefficients become θ-coefficients if $\mathcal{F}$ is a class of bounded functions, $\mathcal{G}$ is a class of Lipschitz continuous functions, and $\Psi(f,g) = v\|f\|_\infty \operatorname{Lip}(g)$.

Example 2.4.3. (Doukhan and Louhichi, 1999) ϵ-coefficients become η-coefficients if $\mathcal{F}_u$ and $\mathcal{G}_v$ are classes of bounded and Lipschitz continuous functions, and $\Psi(f,g) = v\|f\|_\infty \operatorname{Lip}(g) + u\|g\|_\infty \operatorname{Lip}(f)$.

Other types of ϵ-coefficients can be looked up in Dedecker et al. (2007). The random field $\{X_{\mathbf{i}} : \mathbf{i} \in D\}$ is said to be $(\mathcal{F}, \mathcal{G}, \Psi)$-dependent (or ϵ-dependent for short) if $\lim_{s\to\infty} \epsilon(s) = 0$. $(\mathcal{F}, \mathcal{G}, \Psi)$-dependence is named θ- or η-dependence, according to different specifications of dependence coefficients in above examples.

Comparing to the concept of NED, ϵ-dependence measures the spatial dependence in a more direct way, as it does not require an auxiliary random field. However, there is no attempt in the literature on applying limit theorems under ϵ-dependence to the inference of high-dimensional time series models. In fact, existing limit theorems for random fields under ϵ-weak dependence is not as flexible as that under NED (Jenish and Prucha, 2012). For example, the most recent result is a central limit theorem proposed by Curato et al. (2022) for θ-lex weakly dependent random fields, where their θ-lex dependence considers the covariance between $f(\mathfrak{X}_U)$ and $g(\mathfrak{X}_V)$, with $v = 1$ and all indices in U are lexicographically smaller than the one in V (see Definition 2.1 in Curato et al. (2022) for details).

Example 2.4.4. $\{D_n : n \geq 1\}$ is a series of finite subsets of $\mathbb{Z}^m$ such that

$$\lim_{n\to\infty} |D_n|_c = \infty, \qquad \lim_{n\to\infty} \frac{|\partial D_n|_c}{|D_n|_c} = 0,$$

where $\partial D_n = \{\mathbf{i} \in D_n : \exists \mathbf{j} \notin D_n, \|\mathbf{i} - \mathbf{j}\| = 1\}$. $\{X_{\mathbf{i}} : \mathbf{i} \in \mathbb{Z}^m\}$ is a strictly stationary centered real-valued random field such that $\mathbb{E}|X_{\mathbf{i}}|^{2+\delta} < \infty$ for some $\delta > 0$, and the dependence coefficients $\theta(s) = \mathcal{O}(s^{-\mu})$ for some $\mu > m(1 + 1/\delta)$. Then as $n \to \infty$,

$$\frac{1}{\sqrt{|D_n|_c}} \sum_{\mathbf{i}\in D_n} X_{\mathbf{i}} \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = \sum_{\mathbf{j}\in\mathbb{Z}^m} \mathbb{E}(X_0 X_{\mathbf{j}}|\mathcal{I})$ with $\mathcal{I}$ being the σ-algebra of shift invariant sets as defined in Dedecker (1998).

For the aforementioned high-dimensional models (Zhu et al., 2017; Zhou et al., 2020; Xu et al., 2024; Armillotta and Fokianos, 2024), stationarity over t can be proved under mild conditions. However, the CLT above requires strict stationarity over both i and t, which is too stringent for high-dimensional time series. Indeed, the type II strict stationarity proposed by Zhu et al. (2017) can be proved for some high-dimensional network models (Zhu et al., 2017; Armillotta and Fokianos, 2024), but it is not a widely accepted definition of stationarity in a spatio-temporal sense.

Chapter 3

Limit Theorems of Weakly Dependent Random Fields

3.1 Introduction

High-dimensional models have drawn much attention recently in science, social science, econometrics, among other fields. High-dimensional models, or spatio-temporal models, have their merits in describing dependence over both time and space. However, since the samples from high-dimensional models form a two-dimensional panel that grows in two directions, the limit theorems used in the inference of univariate or fixed-dimensional time series are no longer valid. Note that a spatio-temporal model can be regarded as a random process running on a two-dimensional lattice, i.e., a random field. Therefore, we seek limit theorems for random fields that could potentially provide useful tools in the inference of high-dimensional models.

Limit theorems for random fields have been extensively studied in the literature. Jenish and Prucha (2009) proposed limit theorems for arrays of random fields under α- and ϕ-mixing. Compared to previous limit theorems for mixing random fields (Bolthausen, 1982; Guyon, 1995; Dedecker, 1998), their limit theorems are more general in the sense that they accommodate arrays of random fields that are non-stationary and have asymptotically unbounded moments. However, the mixing property may fail to hold for integer-valued time series (see (3.6.2) in Dedecker et al. (2007)). Even a simple AR(1) model with Bernoulli-distributed innovation is not mixing (Gorodetskii, 1978; Andrews, 1984). To solve this flaw of the mixing property, Doukhan and Louhichi (1999) introduced a new concept of weak dependence, which can be extended to random fields. See Chapter 2 in Dedecker et al. (2007) for details.

However, existing limit theorems under weakly dependence either require stationarity (Dedecker et al., 2007; El Machkouri et al., 2013; Curato et al., 2022) or are only established for single-indexed sequences (Neumann, 2013; Merlevède et al., 2019). For example, El Machkouri et al. (2013) proposed a CLT for random fields that are Bernoulli shifts of IID innovations, which could be regarded as a special case of the models that our CLT can handle. Neumann (2013) and Merlevède et al. (2019) proposed a CLT and a functional CLT, respectively, for non-stationary triangular arrays of random variables. Since they are both limited to random sequences along a single time index, their results are not applicable to high-dimensional time series that we will discuss later, and, moreover, none of them is robust against asymptotically unbounded moments. In Theorem 3.1 and Theorem 3.2, we will propose a law of large numbers and a central limit theorem for weakly dependent triangular arrays of random fields, which are not necessarily stationary and have potentially asymptotically unbounded moments.

Another flaw of existing literature is the lack of applications of limit theorems for random fields to the inference of high-dimensional time series. The property of weak dependence that we use in this chapter has been proved to be preserved under transformations with certain conditions. See, for example, Proposition 2.4 in Curato et al. (2022). In Proposition 3.2, we will also show that weak dependence can be preserved under infinite shifts. Facilitated by these properties of weak dependence, we are able to apply our limit theorems to establish a maximum likelihood estimator for high-dimensional time series, with consistency and asymptotic normality being proved in Proposition 3.4 and Proposition 3.5. With these new results, we have built a sufficient theoretical basis for making asymptotic inference in a wide range of high-dimensional time series models that can be treated as weakly dependent random fields under reasonably general conditions. As an example of the application of our general theory, we obtained the asymptotic normality of the estimation for network autoregressive (NAR) models in Proposition 3.6 without assuming second-type stationarity and under less restrictive conditions on networks (ref. Zhu et al. (2017)).

The rest of this chapter is organized as follows. In Section 3.2, we will introduce the concept of weak dependence for arrays of random fields and investigate its heredity under transformation and infinite shift. Our LLN and CLT for arrays of random fields will be presented in Section 3.3. In Section 3.4, we will provide the conditions for high-dimensional time series to be weakly dependent and establish the asymptotic properties of the MLE. The proofs of all our results are included in Section A.1.

3.2 Weakly dependent random fields

Random fields are random processes running on multi-dimensional lattices. Considering a metric space $(\mathbb{T}, \rho)$, one could easily define on $\mathbb{T}$ an infinitely countable lattice $D \subset \mathbb{T}$, which satisfies the following assumption throughout this chapter:

Assumption 3.2.1. *Defined on the metric space* $(\mathbb{T}, \rho)$*, the lattice* $D \subset \mathbb{T}$ *is infinitely countable, and there exists a minimum distance* $\rho_0 = \inf_{i,j \in D} \rho(i, j)$*, and without loss of generality we assume* $\rho_0 \geq 1$*.*

This minimum distance assumption is required to ensure the growth of sample size with the expansion of sample region on D. A simple example that satisfies Assumption 3.2.1 is $\mathbb{T}$ being a d-dimensional Euclidean space $\mathbb{R}^d$ and $D = \mathbb{Z}^d$ being an infinite lattice with minimum distance of 1. Let $\{D_n : D_n \subset D, n \geq 1\}$ be a series of sub-lattices of D with finite sizes. In the rest of the section, we consider a random field $\{X_{i,n} : i \in D_n, n \geq 1\}$, with $X_{i,n}$ takes its value in a Banach space $(\mathcal{X}, \|\cdot\|)$.

In their Definition 2.2, Dedecker et al. (2007) defined the $(\mathcal{F}, \mathcal{G}, \Psi)$-coefficients that measure the dependence between two separated groups of random variables on $\mathbb{Z}$. Dedecker et al. (2007) also remarked that their definition could be easily extended to general metric sets of indices. Curato et al. (2022) defined their θ-lex dependence coefficients for random fields on $\mathbb{R}^m$. In Section 3.2.1 we will adopt the weak dependence by Dedecker et al. (2007), and extend it for triangular arrays of random fields with indices on any lattice D that satisfies Assumption 3.2.1.

3.2.1 Weak dependence

Let $\mathcal{F}_u$ and $\mathcal{G}_v$ denote two classes of functions from $\mathcal{X}^u$ to $\mathbb{R}$ and $\mathcal{X}^v$ to $\mathbb{R}$ respectively. If Ψ is some function mapping from $\mathcal{F}_u \times \mathcal{G}_v$ to $\mathbb{R}_+$, and $\mathfrak{X}$ and $\mathfrak{Y}$ are two arbitrary random variables in $\mathcal{X}^u$ and $\mathcal{X}^v$, then we can define the measurement of dependence between $\mathfrak{X}$ and $\mathfrak{Y}$ by

$$\epsilon(\mathfrak{X}, \mathfrak{Y}) = \sup \left\{ \frac{|\operatorname{Cov}(f(\mathfrak{X}), g(\mathfrak{Y}))|}{\Psi(f, g)} : f \in \mathcal{F}_u, g \in \mathcal{G}_v \right\}$$

Given any $U_n \subset D_n$ with cardinality $|U_n|_c = u$ and $V_n \subset D_n$ with cardinality $|V_n|_c = v$, let $\mathfrak{X}_{U_n} = (X_{i,n})_{i \in U_n}$, $\mathfrak{X}_{V_n} = (X_{i,n})_{i \in V_n}$. Then the dependence coefficient of random fields $\{X_{i,n} : i \in D_n, n \geq 1\}$ is defined by

$$\epsilon_{n,u,v}(s) = \sup \left\{ \epsilon(\mathfrak{X}_{U_n}, \mathfrak{X}_{V_n}) : |U_n|_c = u, |V_n|_c = v, \rho(U_n, V_n) \geq s \right\}, \tag{3.2.1}$$

where $\rho(U_n, V_n) := \min_{i \in U_n, j \in V_n} \rho(i,j)$ measure the distance between U_n and V_n.

Remark. For any functions $f \in \mathcal{F}_u$ and $g \in \mathcal{F}_v$ such that $\Psi(f,g) < \infty$, inequality below follows directly from (3.2.1):

$$|\operatorname{Cov}(f(\mathfrak{X}_{U_n}), g(\mathfrak{X}_{V_n}))| \leq C \epsilon_{n,u,v}(\rho(U_n, V_n)) \tag{3.2.2}$$

for some constant $0 < C < \infty$ that is related to f, g and Ψ.

Remark. Similar to Definition 1 in Jenish and Prucha (2009), we introduce the following notations:

$$\bar{\epsilon}_{u,v}(s) = \sup_n \epsilon_{n,u,v}(s), \qquad \bar{\epsilon}(s) = \sup_{u,v} \bar{\epsilon}_{u,v}(s).$$

Now we are ready to give a formal definition of weak dependence as follows:

Definition 3.1. *The random fields $\{X_{i,n} : i \in D_n, n \geq 1\}$ in Banach space $(\mathcal{X}, \|\cdot\|)$ are said to be ϵ-weakly dependent if* $\lim_{s \to \infty} \bar{\epsilon}(s) = 0$.

Using specifications (2.2.3) and (2.2.7) of Ψ in Dedecker et al. (2007), two variations of weak dependence coefficients can be defined according to different regularity conditions on $\mathcal{F}_u$ and $\mathcal{G}_v$:

- If $\mathcal{F}_u$ is a class of bounded functions and $\mathcal{G}_v$ is a class of Lipschitz continuous functions, then we can define the θ-dependence coefficient as

$$\theta_{n,u,v}(s) = \sup\left\{\epsilon(\mathfrak{X}_{U_n}, \mathfrak{X}_{V_n}) : |U|_c = u, |V|_c = v, \rho(U_n, V_n) \geq s\right\}, \tag{3.2.3}$$

  by letting $\Psi(f,g) = v\|f\|_\infty \operatorname{Lip}(g)$.

- If $\mathcal{F}_u$ and $\mathcal{G}_v$ are both classes of bounded Lipschitz continuous functions, we can also define the η-dependence coefficient as

$$\eta_{n,u,v}(s) = \sup\left\{\epsilon(\mathfrak{X}_{U_n}, \mathfrak{X}_{V_n}) : |U|_c = u, |V|_c = v, \rho(U_n, V_n) \geq s\right\}, \tag{3.2.4}$$

  by letting $\Psi(f,g) = u\|g\|_\infty \operatorname{Lip}(f) + v\|f\|_\infty \operatorname{Lip}(g)$.

In this chapter we focus on θ and η coefficients. For readers who may be interested in other variations, we refer to Dedecker et al. (2007) section 2.2. From now on, we will use ϵ to denote a generic dependence coefficient despite of cases when specific notations are necessary.

In some particular cases, we can compare the $\epsilon-$weak dependence to the mixing (Jenish and Prucha, 2009) and the near-epoch dependence (NED) (Jenish and Prucha, 2012), which are two widely used concepts of dependence of random fields in the literature. In the case of random processes (i.e. the dimension of D is 1), an AR(1) process with non-smooth innovation is η-weakly dependent but it is not mixing (Andrews, 1984), and Proposition 1 in Doukhan et al. (2012) shows that η-weak dependence implies α-mixing for integer-valued processes. Moreover, by Example 3.4.2 and Proposition 3.3 in Section 3.4, we show that uniform $\mathbb{L}^1$-NED on IID random variables implies η-weak dependence, in the case when $D = \mathbb{Z}^2$.

3.2.2 Heredity of weak dependence

Before we establish the limit theorems in the next section, it is important to investigate the heredity of weak dependence, which is essential for us to apply the limit theorems in the inference of high-dimensional models. See for example, the instrumental variable quantile regression estimation for the dynamic network quantile regression (DNQR) by Xu et al. (2024) and the quasi maximum likelihood estimation for the threshold network GARCH (TNGARCH) by (Pan and Pan, 2024). In this section, we will show that the θ and η weak dependence can be preserved under locally Lipschitz transformations and infinite shifts.

Proposition 3.1 below is a natural extension of Proposition 2.1 and Proposition 2.2 in Dedecker et al. (2007) to arrays of random fields. It shows that weak dependence is inherited under transformations satisfying condition (3.2.5). A simple example is any Lipschitz-continuous function when $a = 1$.

Proposition 3.1. *Let $\{X_{i,n} : i \in D_n, n \geq 1\}$ be a $\mathbb{R}^{d_x}$-valued random field with $\sup_n \sup_{i\in D_n} \|X_{i,n}\|_p < \infty$ for some $p > 1$, and $H : \mathbb{R}^{d_x} \mapsto \mathbb{R}$ is a function such that*

$$|H(x) - H(y)| \leq c\|x - y\|(\|x\|^{a-1} + \|y\|^{a-1}) \tag{3.2.5}$$

for some $c \in (0, +\infty)$, $a \in [1, p)$, and any $x, y \in \mathbb{R}^{d_x}$. Suppose that $\{Y_{i,n} : i \in D_n, n \geq 1\}$ are transformed from $\{X_{i,n} : i \in D_n, n \geq 1\}$ by letting $Y_{i,n} = H(X_{i,n})$. If $\{X_{i,n} : i \in D_n, n \geq 1\}$ are weakly dependent with coefficients $\bar{\theta}_x(s)$ or $\bar{\eta}_x(s)$, then $\{Y_{i,n} : i \in D_n, n \geq 1\}$ are also weakly dependent with $\bar{\theta}_y(s) \leq C\bar{\theta}_x(s)^{\frac{p-a}{p-1}}$ or $\bar{\eta}_y(s) \leq C\bar{\eta}_x(s)^{\frac{p-a}{p-1}}$ for some constant $C > 0$.

For the heredity of weak dependence under shifts, we consider $D \subset \mathbb{Z}^d$, equipped with distance measure $\rho(i, j)$ for any $i, j \in D$. Let $\{\varepsilon_i : i \in D\}$ be a $\mathbb{R}$-valued random field on D. Let $H_{i,n} : \mathbb{R}^D \mapsto \mathbb{R}$ be a measurable function, and random fields $\{X_{i,n} : i \in D_n, n \geq 1\}$ are defined by $X_{i,n} := H_{i,n}((\varepsilon_j)_{j\in D})$. For each $h \in \mathbb{N}$, and for any $(x_j)_{j\in D}$

and $(y_j)_{j\in D}$ such that $x_j \neq y_j$ if and only if $\rho(i,j)=h$, $H_{i,n}$ satisfies that

$$\begin{aligned}&|H_{i,n}((x_j)_{j\in D}) - H_{i,n}((y_j)_{j\in D})|\\ \le& B_{i,n}(h)(\max_{\rho(i,j)\neq h}|x_j|^l \vee 1)\sum_{\rho(i,j)=h}|x_j-y_j|\end{aligned} \tag{3.2.6}$$

almost surely, where $l \geq 0$ and $\{B_{i,n}(h) : i \in D_n, n \geq 1\}$ are positive constants satisfying that

$$C_B := \sup_{n\geq 1}\sup_{i\in D_n}\sum_{h=0}^{\infty} B_{i,n}(h)h^{d-1} < \infty. \tag{3.2.7}$$

In Proposition 3.2 below, we investigate the preservation of weak dependence from the $\{\varepsilon_i : i \in D\}$ to $\{X_{i,n} : i \in D_n, n \geq 1\}$.

Proposition 3.2. *Let $\{X_{i,n} \in \mathbb{R} : i \in D_n, n \geq 1\}$ be an array of Bernoulli shifts of $\{\varepsilon_i \in \mathbb{R} : i \in D\}$, such that $X_{i,n} = H_{i,n}((\varepsilon_j)_{j\in D})$ and $H_{i,n} : \mathbb{R}^D \mapsto \mathbb{R}$ satisfies conditions (3.2.6) and (3.2.7). Assume that $\sup_{i\in D}\mathbb{E}|\varepsilon_i|^p < \infty$ with $p > l+1$. If the random field $\{\varepsilon_i \in \mathbb{R} : i \in D\}$ is weakly dependent with coefficients $\bar{\theta}_\varepsilon(s)$ or $\bar{\eta}_\varepsilon(s)$, then $\{X_{i,n} \in \mathbb{R} : i \in D_n, n \geq 1\}$ are also weakly dependent with coefficients*

$$\bar{\theta}(r) = C\inf_{0<s\leq[r/2]}\left\{C(s)\vee\left[s^d\bar{\theta}_\varepsilon(r-2s)^{\frac{p-1-l}{p-1}}\right]\right\}, \tag{3.2.8}$$

or

$$\bar{\eta}(r) = C\inf_{0<s\leq[r/2]}\left\{C(s)\vee\left[s^d\bar{\eta}_\varepsilon(r-2s)^{\frac{p-1-l}{p-1}}\right]\right\}, \tag{3.2.9}$$

where $C(s) = \sup_{n\geq 1}\sup_{i\in D_n}\sum_{h\geq s}B_{i,n}(h)h^{d-1}$ and $C>0$ is a constant.

It is not easy to find the exact infinum in (3.2.8) and (3.2.9). However, the dependence coefficients of the outputs have upper bounds in explicit forms, if the dependence coefficients and $B_{i,n}$ decay in a regular manner.

Example 3.2.1. Let the dependence coefficients of the input $\bar{\eta}_\varepsilon(r) = \mathcal{O}(r^{-\mu})$ for some $\mu > \frac{p-1}{p-1-l}d$, and $B_{i,n}(h) = \mathcal{O}(h^{-b})$ for some $b \geq \frac{p-1-l}{p-1}\mu$. Then the dependence coefficients of the output are bounded by:

$$\bar{\eta}(r) \leq Cr^{d-\frac{p-1-l}{p-1}\mu}. \tag{3.2.10}$$

Example 3.2.2. Assume that $d = 2$, let the dependence coefficients of the input $\bar{\eta}_\varepsilon(r) = \mathcal{O}(r^{-\mu})$ for some $\mu > 0$, and $B_{i,n}(h) = \mathcal{O}(e^{-bh})$ for some $b \geq \frac{p-1-l}{p-1}\mu$. Then the dependence coefficients of the output are bounded by:

$$\bar{\eta}(r) \leq C(\log r)^2 r^{-\frac{p-1-l}{p-1}\mu}. \tag{3.2.11}$$

With θ coefficients we have the same results. The proofs of (3.2.10) and (3.2.11) are given in section A.1.1.

3.3 Limit theorems for weakly dependent random fields

In this section, we investigate the asymptotic behaviour of a weakly dependent random field $\{X_{i,n} : i \in D_n, n \geq 1\}$ on $D \subset \mathbb{R}^d$ $(d \geq 1)$, which satisfies Assumption 3.2.1. $(D_n)_{n\in\mathbb{Z}}$ is a series of sample regions on D with finite cardinality, i.e. $|D_n|_c < \infty$, and $\lim_{n\to\infty} |D_n|_c = \infty$ represents the expansion of sample region as $n \to \infty$.

In section 3.3.1 we propose a law of large numbers (in $\mathbb{L}^1$) for weakly dependent random fields in general. In section 3.3.2 we proposed a central limit theorem for θ and η weakly dependent random fields. Recently Curato et al. (2022) proposed a CLT for θ-lex weakly dependent random fields that are strictly stationary; The CLT of Neumann (2013) requires bounded moments instead of stationarity, but it only applies to random sequences. Our limit theorems, however, are more general compared to theirs in the following aspects:

- Stationarity is not required;
- Our limit theorems accommodate arrays of random fields with asymptotically unbounded moments;
- The lattice D is not required to be evenly spaced like $\mathbb{Z}^d$.

Our proofs of LLN and CLT are based on Jenish and Prucha (2009), where they derived limit theorems for α and ϕ mixing random fields with asymptotically unbounded moments.

3.3.1 Law of large numbers

Assumption 3.3.1 below helps our LLN to accommodate random fields with asymptotically unbounded l-th moments, by setting $c_{i,n} = \mathbb{E}|X_{i,n}|^l \vee 1$. For random fields with uniform bounded moments, we can simply set $c_{i,n} = 1$. Assumption 3.3.2 puts restriction on the decaying rate of dependence coefficient.

Assumption 3.3.1. *There exist positive constants* $\{c_{i,n} : i \in D_n, n \geq 1\}$ *such that*

$$\sup_n \sup_{i \in D_n} \mathbb{E} \left| \frac{X_{i,n}}{c_{i,n}} \right|^l < \infty \tag{3.3.1}$$

for some $l > 1$.

Remark. By Hölder's inequality and Markov's inequality, (3.3.1) implies the $\mathbb{L}^p$ uniform integrability for any $0 < p < l$. i.e.

$$\lim_{k \to \infty} \sup_n \sup_{i \in D_n} \mathbb{E} \left[\left| \frac{X_{i,n}}{c_{i,n}} \right|^p 1 \left(\left| \frac{X_{i,n}}{c_{i,n}} \right| \geq k \right) \right] = 0. \tag{3.3.2}$$

See page 216 in Billingsley (2008) for the definition.

Assumption 3.3.2. *The dependence coefficient of* $\{X_{i,n} : i \in D_n, n \geq 1\}$ *satisfies* $\bar{\epsilon}_{1,1}(s) = \mathcal{O}(s^{-\alpha})$ *with* $\alpha > d$.

Now we are ready to present our LLN as follows.

Theorem 3.1. *Let* $\{X_{i,n} \in \mathbb{R} : i \in D_n, n \geq 1\}$ *be a random field on* $D \subset \mathbb{R}^d (d \geq 1)$, *where* $(D_n)_{n \geq 1}$ *is a sequence of finite sub-lattices of* D *with* $\lim_{n \to \infty} |D_n|_c = \infty$. *If Assumption 3.2.1, Assumption 3.3.1 and Assumption 3.3.2 are satisfied, then as* $n \to \infty$,

$$\frac{1}{M_n |D_n|_c} \sum_{i \in D_n} (X_{i,n} - \mathbb{E} X_{i,n}) \to 0$$

in $\mathbb{L}^1$, *where* $M_n = \sup_{i \in D_n} c_{i,n}$.

3.3.2 Central limit theorem

Let $S_n = \sum_{i \in D_n} X_{i,n}$ and $\sigma_n^2 = \mathrm{Var}(S_n)$. We need the following assumptions to state our CLT. In Assumptions 3.3.4 and 3.3.5, notice that the condition on θ-coefficients is slightly weaker than that on η-coefficients, since θ-weak dependence is actually more stringent than η-weak dependence according to (3.2.3) and (3.2.4). Assumption 3.3.6 is a standard condition in the limit theory literature, as maintained in Bolthausen (1982), Jenish and Prucha (2009), and Jenish and Prucha (2012). It is required to prove that σ_n^2 is asymptotically proportional to $|D_n|_c$ as $n \to \infty$, which ensures that no single summand dominates the sum.

Assumption 3.3.3. *There exist positive constants $\{c_{i,n} : i \in D_n, n \geq 1\}$ such that*

$$\sup_n \sup_{i \in D_n} \mathbb{E} \left| \frac{X_{i,n}}{c_{i,n}} \right|^m < \infty \tag{3.3.3}$$

for some $m > 2$.

Assumption 3.3.4. *With the same $m > 2$ in Assumption 3.3.3, the θ-coefficient of $\{X_{i,n} : i \in D_n, n \geq 1\}$ satisfies:*

(a). For all $u + v \leq 4$, $\bar{\theta}_{u,v}(s) = \mathcal{O}(s^{-\alpha})$ with $\alpha > \frac{m-1}{m-2} d$;

(b). $\bar{\theta}_{\infty,1}(s) := \sup_u \bar{\theta}_{u,1}(s) = \mathcal{O}(s^{-\beta})$ with $\beta > d$.

Assumption 3.3.5. *With the same $m > 2$ in Assumption 3.3.3, the η-coefficient of $\{X_{i,n} : i \in D_n, n \geq 1\}$ satisfies:*

(a). For all $u + v \leq 4$, $\bar{\eta}_{u,v}(s) = \mathcal{O}(s^{-\alpha})$ with $\alpha > \frac{m-1}{m-2} d$;

(b). $\bar{\eta}_{\infty,1}(s) := \sup_u \bar{\eta}_{u,1}(s) = \mathcal{O}(s^{-\beta})$ with $\beta > 2d$.

Assumption 3.3.6. *Let $M_n = \sup_{i \in D_n} c_{i,n}$, assume that*

$$\liminf_{n \to \infty} (|D_n|_c)^{-1} M_n^{-2} \sigma_n^2 > 0.$$

Our CLT is given as a theorem below.

Theorem 3.2. *Let $\{X_{i,n} \in \mathbb{R} : i \in D_n, n \geq 1\}$ be a zero-mean random field on $D \subset \mathbb{R}^d (d \geq 1)$, where $(D_n)_{n \in \mathbb{N}_+}$ is a sequence of finite sub-lattices of D with $\lim_{n\to\infty} |D_n|_c = \infty$. If Assumption 3.2.1, Assumption 3.3.3 and Assumption 3.3.6 hold true, and the dependence coefficient of $\{X_{i,n} \in \mathbb{R} : i \in D_n, n \geq 1\}$ satisfies either Assumption 3.3.4 or Assumption 3.3.5, then as $n \to \infty$,*

$$\sigma_n^{-1} S_n \xrightarrow{d} N(0, 1).$$

Theorem 3.2 only applies to scalar-valued random fields, limiting its application in the inference of high-dimensional time series, whereas vector-valued statistics like the maximum likelihood estimator (3.4.8) are very common. Facilitated by the transformation-invariance of ϵ-weak dependence in Proposition 3.1, Theorem 3.2 can be easily extended to arrays of vector-valued random fields using a standard Cramér-Wold device.

Corollary 3.2.1. *Let $\{X_{i,n} \in \mathbb{R}^k : i \in D_n, n \geq 1\}$ be an array of vector-valued zero-mean random fields. By regarding $|\cdot|$ in Assumption 3.3.3 as Euclidean norm, and replacing Assumption 3.3.6 by*

$$\liminf_{n\to\infty} (|D_n|_c)^{-1} M_n^{-2} \lambda_{min}(\Sigma_n) > 0$$

where $\lambda_{min}(\Sigma_n)$ is the smallest eigenvalue of $\Sigma_n := \mathrm{Var}(S_n)$, then as $n \to \infty$:

$$\Sigma_n^{-1/2} S_n \xrightarrow{d} N(0, I_k).$$

We now compare Theorem 3.2 with existing CLTs for weakly dependent random variables in the literature. El Machkouri et al. (2013) developed a CLT for a class of stationary random fields that are Bernoulli shifts of IID innovations. Their results are derived through a coupling technique, based on p-stability, which is a dependence measure different from ours. Our CLT can also deal with Bernoulli shifts as a special case through a similar coupling technique; see Example 3.4.1. For non-stationary triangular arrays of random sequences, Neumann (2013) proposed a CLT under weak dependence, while Merlevède et al. (2019) developed a functional CLT for martingale-like sequences.

Their CLTs are limited to the case when $d = 1$ and require Lindeberg conditions (see (2.2) in Neumann (2013) and (3.1) in Merlevède et al. (2019)) that could be violated by asymptotically unbounded moments, which are allowed by Assumption 3.3.3 in our CLT.

Indeed, Jenish and Prucha (2009) proposed a CLT for triangular arrays of random fields without stationarity and bounded moments. However, the α- or ϕ-mixing it requires does not hold for many integer-valued models or models with discrete-valued innovations, as we have pointed out in Section 3.1. For instance, the integer-valued bilinear model is not mixing, but it is proved to be θ-weakly dependent Doukhan et al. (2006). Jenish and Prucha (2012) later extended Jenish and Prucha (2009)'s CLT for random fields under $\mathbb{L}^2$-NED, which is still stronger than what our CLT requires. In Example 3.4.2, we will show that $\mathbb{L}^1$-NED on an IID random field is sufficient for η-weak dependence.

3.4 Applications to high-dimensional time series

In order to apply our results, we treat a high-dimensional time series as a random field with spatial (i.e. cross sectional) index i and time index t. Following Xu et al. (2024) and Pan and Pan (2024), we set

$$D = \{(i,t) : i \in \mathbb{Z}, t \in \mathbb{Z}\}$$

as an infinitely countable lattice on $\mathbb{R}^2$, equipped with distance measure $\rho((i,t),(u,v)) := \max\{|i-u|, |t-v|\}$, and D satisfies Assumption 3.2.1. Note that i here represents individual i, while the $i \in D$ we used in previous sections is a location on lattice D. We will continue to use the same notation under these two scenarios since we don't think it would cause any confusion. $\{X_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$ are NT samples generated by a high-dimensional time series. With the specification of D above, we furthermore specify a series of sample regions $\{D_{NT} : NT \geq 1\}$ where

$$D_{NT} = \{(i,t) : 1 \leq i \leq N, 1 \leq t \leq T\},$$

of which the cardinality $|D_{NT}|_c = NT$ expands as $NT \to \infty$. In this setting, we transform the samples of high-dimensional time series into an array of random fields $\{X_{it} : (i,t) \in D_{NT}, NT \geq 1\}$. The same setting of sample regions was used by Xu et al. (2024) on a dynamic network quantile regression model, which is an example of high-dimensional time series.

After we have built the tools of limit theorems on weakly dependent random fields, in Section 3.4.1 we will propose general conditions when a high-dimensional time series model is η-weakly dependent. Then in Section 3.4.2 the proposed limit theorems will be applied to prove the consistency and asymptotic normality of MLE under certain restrictions on the likelihood function.

3.4.1 Examples of η-weakly dependent high-dimensional time series

We consider a series of samples $\{X_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ from a high-dimensional time series, with innovations $\{\xi_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ that satisfy:

Assumption 3.4.1. *The innovations ξ_{it}'s are independently and identically distributed (IID) across i and t. They are also independent from X_{it} for any i and t.*

Let $\mathcal{F}_{it}(s) = \sigma\{\xi_{it} : (j,\tau) \in D_{NT}, \rho((i,t),(j,\tau)) \leq s\}$ for $s > 0$, then we can define $\{X_{it}^{(s)} : (i,t) \in D_{NT}, NT \geq 1\}$ with $X_{it}^{(s)}$ being $\mathcal{F}_{it}(s)$-measurable. Based on this definition, $X_{it}^{(s)}$ is independent from $X_{j\tau}^{(s)}$ if $\rho((i,t),(j,\tau)) > 2s$. In the assumption below we assume that X_{it} can be approximated by $X_{it}^{(s)}$:

Assumption 3.4.2. $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \mathbb{E}\left|X_{it} - X_{it}^{(s)}\right| \leq C\delta(s)$ *for some constant* $C > 0$, *where* $\delta(s) \geq 0$ *and* $\lim_{s\to\infty} \delta(s) = 0$.

Note that we do not need X_{it} to be a Bernoulli shift of ξ_{it} as it is required in El Machkouri et al. (2013). The Bernoulli shift assumption is strict in practice since this type of model is assumed to be some specific transformation of IID innovations, whereas our results accommodate random fields with more complicated structures, as long as they can be approximated by neighbouring innovations in a way as in Assumption 3.4.2. Nevertheless, we take Bernoulli shifts as a special instance in Example 3.4.1 below.

Example 3.4.1. If X_{it} is a Bernoulli shift in the form $X_{it} = H((\xi_{i-l,t-l})_{l>0})$, according to (11) in Doukhan and Truquet (2007), we can define

$$X_{it}^{(s)} = H((\xi_{i-l,t-l}^{(s)})_{l>0})$$

with

$$\xi_{i-l,t-l}^{(s)} = \begin{cases} \xi_{i-l,t-l} & if \quad l \leq s, \\ 0 & if \quad l > s. \end{cases}$$

In Example 3.4.2 we give a way to construct the $\mathcal{F}_{it}(s)$-measurable approximation in general, adopting the definition of near-epoch-dependence, see Definition 1 in Jenish and Prucha (2012). The CLT of Jenish and Prucha (2012) is based on $\mathbb{L}^2$-NED, which is stringent in practice since it will either degenerate to $\mathbb{L}^1$-NED (see Theorem 17.9 in Davidson (1994)) or require bounded high-order moments (see Lemma A.2 in Xu and Lee (2015)) after multiplication. Example 3.4.2 indicates that, comparing to the CLT of Jenish and Prucha (2012), ours also accommodates $\mathbb{L}^1$-NED random fields.

Example 3.4.2. If we define

$$X_{it}^{(s)} = \mathbb{E}[X_{it}|\mathcal{F}_{it}(s)],$$

then Assumption 3.4.2 is equivalent to the uniform $\mathbb{L}^1$-NED on IID innovations, with coefficient $\delta(s)$.

Proposition 3.3. *Under Assumptions 3.4.1 and 3.4.2,* $\{X_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ *is η-weakly dependent with $\bar{\eta}(s) \leq C\delta(s/2)$.*

Remark. The limit theorems have requirements on the decaying rate of $\bar{\eta}(s)$ as $s \to \infty$, see Assumptions 3.3.2 and 3.3.5. Proposition 3.3 allows us to check the rate of $\delta(s)$ alternatively. With careful specification of $X_{it}^{(s)}$, $\delta(s)$ could be derived in explicit form, making it easier to check the decaying rate in practice.

3.4.2 Maximum likelihood estimation (MLE)

In this section, we will investigate the asymptotic properties (i.e. consistency and asymptotic normality) of MLE for parameters in a high-dimensional time series model with increasing sample size, i.e., $|D_{NT}|_c = NT \to \infty$.

Assume that the model of interest is characterized by an array of parameters θ in a specific parameter space $\Theta \subset \mathbb{R}^k$, such that the true parameter $\theta_0 \in \Theta$. Based on samples $\{X_{it} \in \mathbb{R} : (i,t) \in D_{NT}\}$, we could construct log likelihood functions in the form

$$L_{NT}(\theta) := \frac{1}{NT} \sum_{(i,t)\in D_{NT}} l_{it}(\theta), \tag{3.4.1}$$

where $l_{it}(\theta) = \log f_{it}(x;\theta)$, and $f_{it}(x;\theta)$ denotes the density (or probability mass) function of X_{it} with parameter θ. Note that the parameter θ is not necessarily a vector of real numbers for a model with finite number of real parameters. It could be an element of an abstract metric space.

To discuss the estimation of parameter θ, We need the following assumptions regarding the parameter space Θ and the likelihood function:

Assumption 3.4.3. *The parameter space and likelihood function of the model satisfy*

(a). $\Theta \subset \mathbb{R}^k$ is compact;

(b). The functions $l_{it}(\theta)$ are continuous on Θ, and are measurable for each $\theta \in \Theta$;

(c). The true parameter θ_0 lies in the interior of Θ. And for any $\delta > 0$,

$$\sup_{NT\geq 1} \sup_{\substack{\theta\in\Theta \\ \|\theta-\theta_0\|\geq\delta}} \{\mathbb{E}[L_{NT}(\theta)] - \mathbb{E}[L_{NT}(\theta_0)]\} < 0.$$

Usually in practice the exact likelihood function cannot be calculated, and the estimate of θ_0 could only be obtained through an approximation of (3.4.1). i.e.

$$\hat{\theta}_{NT} := \operatorname*{argmax}_{\theta\in\Theta} \tilde{L}_{NT}(\theta), \tag{3.4.2}$$

where

$$\tilde{L}_{NT}(\theta) := \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \tilde{l}_{it}(\theta). \tag{3.4.3}$$

To consider the consistency of the MLE (3.4.2), we need Assumption 3.4.4 below regarding the convergence of the approximated likelihood function (3.4.3) to the exact likelihood function (3.4.1).

Assumption 3.4.4. *For any* $\theta \in \Theta$, $|L_{NT}(\theta) - \tilde{L}_{NT}(\theta)| \xrightarrow{p} 0$ *as* $NT \to \infty$.

And we also need Assumption 3.4.5 below to apply the LLN to $\{l_{it}(\theta)\}$.

Assumption 3.4.5. *(a). The functions* $l_{it}(\theta)$ *are uniformly* $\mathbb{L}^p$*-bounded for some* $p > 1$, *i.e.*

$$\sup_{NT \geq 1} \sup_{(i,t) \in D_{NT}} \sup_{\theta \in \Theta} \|l_{it}(\theta)\|_p < \infty.$$

(b). For any $\theta \in \Theta$, *the array of functions* $\{l_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ *are weakly dependent with coefficients* $\bar{\theta}(s) = \mathcal{O}(s^{-\alpha})$ *for some* $\alpha > 2$.

The following proposition gives the consistency of MLE for a high-dimensional time series model with expanding sample sizes or/and expanding dimensions.

Proposition 3.4. *If Assumptions 3.4.3, 3.4.4 and 3.4.5 are satisfied, then the MLE (3.4.2) is consistent, i.e.*

$$\hat{\theta}_{NT} \xrightarrow{p} \theta_0 \quad as \quad NT \to \infty.$$

As for the asymptotic normality of $\hat{\theta}_{NT}$, we need additional assumptions on $\tilde{L}_{NT}(\theta)$ and $L_{NT}(\theta)$ as in Assumption 3.4.6 below. Besides, Assumptions 3.4.7(a) and 3.4.7(b) are required for the LLN of $\left\{\frac{\partial^2}{\partial\theta\partial\theta'} l_{it}(\theta_0)\right\}$, as Assumptions 3.4.7(c), 3.4.7(d) and 3.4.7(e) for the CLT of $\left\{\frac{\partial l_{it}(\theta_0)}{\partial\theta}\right\}$.

Assumption 3.4.6. *As* $NT \to \infty$:

(a). $\sqrt{NT}\left\|\frac{\partial\tilde{L}(\theta_0)}{\partial\theta} - \frac{\partial L(\theta_0)}{\partial\theta}\right\| \xrightarrow{p} 0$;

(b). $\sup_{\|\theta-\theta_0\|<\xi}\left\|\frac{\partial^2\tilde{L}(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 L(\theta_0)}{\partial\theta\theta'}\right\| = O_p(\xi)$.

Remark. In the inference of a specific model, the convergences in Assumptions 3.4.4 and 3.4.6 may require extra restrictions on the diverging pattern of N and T. For example, for the TNGARCH model in Chapter 4, it is required that $T \to \infty$, $N \to \infty$ and $N = o(T)$. However, these extra restrictions do not cause any issue in applying our limit theorems, which only require that $NT \to \infty$.

Assumption 3.4.7. *(a).* $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}l_{it}(\theta_0)\right\|_p < \infty$ *for some* $p > 1$*;*

(b). $\left\{\frac{\partial^2}{\partial\theta\partial\theta'}l_{it}(\theta_0) : (i,t)\in D_{NT}, NT\geq 1\right\}$ *are weakly dependent with coefficients* $\bar{\theta}(s) = \mathcal{O}(s^{-\alpha})$ *for some* $\alpha > 2$*;*

(c). $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{\partial l_{it}(\theta_0)}{\partial\theta}\right\|_{p'} < \infty$ *for some* $p' > 2$*;*

(d). $\left\{\frac{\partial l_{it}(\theta_0)}{\partial\theta} : (i,t)\in D_{NT}, NT\geq 1\right\}$ *are weakly dependent with coefficients* $\bar{\theta}(s) = \mathcal{O}(s^{-\alpha'})$ *for some* $\alpha' > 2\vee\frac{2p'-2}{p'-2}$*;*

(e). $\inf_{NT\geq 1}\lambda_{min}(B_{NT}) > 0$ *and* $\inf_{NT\geq 1}\lambda_{min}(B_{NT}^{-1/2}A_{NT}) > 0$*, where* $A_{NT} = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta\partial\theta'}L_{NT}(\theta_0)\right]$*,* $B_{NT} = \mathrm{Var}\left[\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\right]$ *and* $\lambda_{min}(\cdot)$ *denotes the smallest eigenvalue.*

The asymptotic normality of the MLE can be stated as follows.

Proposition 3.5. *If Assumptions 3.4.3 to 3.4.7 are satisfied, then* (3.4.2) *is asymptotically normal, i.e.*

$$\sqrt{NT}(B_{NT}^{-1/2}A_{NT})(\hat{\theta}_{NT}-\theta_0)\xrightarrow{d} N(0, I_k) \quad as \quad NT\to\infty.$$

3.4.3 Application to a network autoregressive model

In this section we will apply our methodology in previous sections to the estimation of a specific high-dimensional time series model. Zhu et al. (2017) proposed a network autoregressive (NAR) model. They established an ordinary-least-squares (OLS) estimation that was proved to be consistent and asymptotically normal when $\min\{N,T\}\to\infty$, and when $N\to\infty$ as T is fixed. The model is defined as

$$y_{it} = \beta_0 + \beta_1\sum_{j=1}^{N} w_{ij}y_{j,t-1} + \beta_2 y_{i,t-1} + Z_i'\gamma + \varepsilon_{it}, \tag{3.4.4}$$

where the $\mathbb{R}$-valued random variable y_{it} is observed both spatially over $i = 1,2,...,N$ and temporally over $t = 1,2,...,T$. Z_i is a $\mathbb{R}^m$-valued covariates vector, which is t-invariant and observable for each individual i. The innovations ε_{it}'s are IID with mean zero and variance σ^2. $\{Z_i\}$ and $\{\varepsilon_{it}\}$ are mutually independent. One feature of the NAR model is using network structure to describe the spatial dependence. Such network is represented by a directed graph, with each edge $a_{ij} = 1$ if node i connects to j and $a_{ij} = 0$ otherwise. In (3.4.4), the effect of each neighboured node j is weighted by $w_{ij} = \frac{a_{ij}}{\sum_{k=i}^{N} a_{ik}}$.

However, the weak dependence of $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ cannot be established under specification (3.4.4). For example, considering the case when $\beta_1 = \beta_2 = 0$, (3.4.4) becomes

$$y_{it} = \beta_0 + Z_i'\gamma + \varepsilon_{it}.$$

For any $s \geq 1$, $|\text{Cov}(y_{i,t+s}, y_{it})| = |\text{Cov}(Z_i'\gamma, Z_i'\gamma)| > 0$, which does not decay to 0 as $s \to \infty$ since $|\text{Cov}(Z_i'\gamma, Z_i'\gamma)|$ is s-invariant. Therefore, in order to apply our limit theorems, in this section we will investigate the weak dependence of the NAR model conditioning on $\mathcal{Z} = (Z_1', Z_2', ..., Z_N')'$. The asymptotic properties of proposed parameter estimates will also be discussed conditioning on $\mathcal{Z}$.

Denote $\theta = (\beta', \gamma')'$ the parameter vector, where $\beta = (\beta_0, \beta_1, \beta_2)'$ and $\gamma = (\gamma_1, \gamma_2, ..., \gamma_m)'$. Let

$$\mathbf{x}_{it} = \begin{pmatrix} 1 \\ \sum_{j=1}^{N} w_{ij} y_{jt} \\ y_{it} \\ Z_i \end{pmatrix} \in \mathbb{R}^{m+3},$$

then the quasi log likelihood function conditioning on $\mathcal{Z}$ (under Gaussian density) is written

$$\begin{cases} L_{NT}(\theta) = \frac{1}{NT} \sum_{(i,t) \in D_{NT}} l_{it}(\theta), \\ l_{it}(\theta) = -(y_{it} - \mathbf{x}_{i,t-1}'\theta)^2. \end{cases} \tag{3.4.5}$$

Given observations of y_{it} and Z_i at $t = 0, 1, ..., T$ and $i = 1, 2, ..., N$, the quasi maximum likelihood estimation (QMLE) could be directly evaluated as follows:

$$\hat{\theta}_{NT} = \left(\sum_{(i,t) \in D_{NT}} \mathbf{x}_{i,t-1}\mathbf{x}_{i,t-1}' \right)^{-1} \left(\sum_{(i,t) \in D_{NT}} \mathbf{x}_{i,t-1} y_{i,t} \right). \tag{3.4.6}$$

It has the same form with OLS of Zhu et al. (2017). Notice that

$$\begin{aligned} \frac{\partial L_{NT}(\theta_0)}{\partial \theta} &= \tfrac{2}{NT} \sum_{(i,t)\in D_{NT}} \varepsilon_{it} \mathbf{x}_{i,t-1}, \\ \frac{\partial^2 L_{NT}(\theta_0)}{\partial \theta \partial \theta'} &= -\tfrac{2}{NT} \sum_{(i,t)\in D_{NT}} \mathbf{x}_{i,t-1} \mathbf{x}'_{i,t-1}. \end{aligned} \tag{3.4.7}$$

Then (3.4.6) could be rewritten as

$$\hat{\theta}_{NT} = \theta_0 - \left(\frac{\partial^2 L_{NT}(\theta_0)}{\partial \theta \partial \theta'} \right)^{-1} \frac{\partial L_{NT}(\theta_0)}{\partial \theta}. \tag{3.4.8}$$

Based on general method in Section 3.4.2, the following assumptions are required to investigate the asymptotic properties of the QMLE in this particular case of the NAR model.

Assumption 3.4.8. *(a). The innovations ε_{it}'s are IID with mean zero and variance σ^2, they are also independent from Z_i and $\mathbf{x}_{i,t-1}$;*

(b). $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \mathbb{E} |\varepsilon_{it}|^p < \infty$ *and* $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \mathbb{E} \|Z_{it}\|^p < \infty$ *for some $p > 2$;*

(c). $|\beta_1| + |\beta_2| < 1$.

Assumption 3.4.9. *Let $G = \beta_1 W + \beta_2 I_N$.*

(a). The elements of G^k satisfy

$$|G^k(i,j)| \leq C_1 \rho_1^k |j-i|^{-\alpha-2}$$

for some constants $C_1 > 0$, $0 < \rho_1 < 1$, $\alpha > 4 \vee \frac{2p-2}{p-2}$ and $p > 2$;

(b). The diagonal elements of $(GG')^k$ satisfy

$$\max_i \{ |(GG')^k(i,i)| \} \leq C_2 \rho_2^k$$

for some constants $C_2 > 0$ and $0 < \rho_2 < 1$.

Assumption 3.4.10. $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \|y_{it}\|_p < \infty$.

Assumption 3.4.11. $\Sigma_{NT} = \frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}_{\mathcal{Z}}\left[\mathbf{x}_{i,t-1}\mathbf{x}'_{i,t-1}\right]$ *satisfies that*

$$\inf_{NT\geq 1} \lambda_{min}(\Sigma_{NT}) > 0.$$

In Assumption 3.4.8, the conditions on the innovations and nodal covariates are the same as condition (C1) in Zhu et al. (2017), except that the finite fourth-order moments of ε_{it} and Z_i are not required in our method. Assumption 3.4.9(a) puts restrictions on the connectivity between nodes. For example, Assumption 3.4.9(a) indicates that the effect of node j on node i through the k-step connection between them weakens with $|i-j|$ and also the length of the connection k. Assumption 3.4.9, together with the bound condition in Assumption 3.4.10 are crucial in verifying weak dependence in the proof of Proposition 3.6.

Proposition 3.6. *If Assumptions 3.4.8 to 3.4.11 are satisfied, then (3.4.6) is consistent and follows asymptotically a normal distribution conditioning on $\mathcal{Z}$:*

$$\sqrt{NT}(\Sigma_{NT})^{1/2}(\hat{\theta}_{NT} - \theta_0) \xrightarrow{d} N(0, \sigma^2 I_{m+3})$$

when $NT \to \infty$.

Since ε_{it}'s are IID, σ^2 could be consistently estimated by

$$\hat{\sigma}^2 = \frac{1}{NT}\sum_{(i,t)\in D_{NT}} \left(y_{it} - \mathbf{x}'_{i,t-1}\hat{\theta}_{NT}\right)^2. \tag{3.4.9}$$

As it will be verified in the proof of Proposition 3.6, following convergence

$$\frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbf{x}_{i,t-1}\mathbf{x}'_{i,t-1} \xrightarrow{p} \Sigma_{NT}$$

allows us to estimate Σ_{NT} by

$$\hat{\Sigma}_{NT} = \frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbf{x}_{i,t-1}\mathbf{x}'_{i,t-1}, \tag{3.4.10}$$

where $\mathbf{x}_{i,t-1}\mathbf{x}'_{i,t-1}$ is a $(m+3)\times(m+3)$ matrix:

$$\begin{pmatrix} 1 & \sum_{j=1}^{N} w_{ij}y_{j,t-1} & y_{i,t-1} & Z_i' \\ \sum_{j=1}^{N} w_{ij}y_{j,t-1} & \big(\sum_{j=1}^{N} w_{ij}y_{j,t-1}\big)^2 & \big(\sum_{j=1}^{N} w_{ij}y_{j,t-1}\big)y_{i,t-1} & \big(\sum_{j=1}^{N} w_{ij}y_{j,t-1}\big)Z_i' \\ y_{i,t-1} & \big(\sum_{j=1}^{N} w_{ij}y_{j,t-1}\big)y_{i,t-1} & y_{i,t-1}^2 & y_{i,t-1}Z_i' \\ Z_i & \big(\sum_{j=1}^{N} w_{ij}y_{j,t-1}\big)Z_i & y_{i,t-1}Z_i & Z_iZ_i' \end{pmatrix}.$$

The covariance matrices in Theorem 3 and Proposition 2 of Zhu et al. (2017) are estimated by statistics in the same form as (3.4.10). Nevertheless, our results are derived on a totally different theoretical basis comparing to their OLS estimation which is actually the MLE because they assumed normality of innovation terms.

Chapter 4

Threshold Network GARCH

4.1 Introduction

The network GARCH model (2.2.5) was proposed by Zhou et al. (2020), who fitted their model to daily log returns of stocks on Chinese stock markets. Their analysis shows that the predictable volatility of one stock is positively related to the log returns of all its neighboring stocks on the network, which can be established according to common shareholders. An explicit assumption in model (2.2.5) is that the prediction of stock volatility is not affected by whether today's stock price is rising or falling. However, there exists empirical evidence against this assumption (Black (1976) and French et al. (1987), among others). The EGARCH model (2.1.4) by Nelson (1991) and the GJR-GARCH model (2.1.5) by Glosten et al. (1993) were established to identify the asymmetry in how predictable volatility adapts to positive and negative news. In their empirical studies on stock return data, they found that negative returns generate more volatility than positive ones. A comparison between different asymmetric GARCH models was conducted by Engle and Ng (1993), and the results suggested using the GJR-GARCH model when analyzing stock return data. In this chapter, we propose a threshold network GARCH model (TNGARCH) that incorporates a self-excited threshold similar to that in the GJR-GARCH model (2.1.5).

Recalling from model (2.2.5), the network consists of N nodes and is denoted by an adjacency matrix $A = (a_{ij})_{N\times N}$, where $a_{ij} = 1$ if nodes i and j are connected, and $a_{ij} = 0$ otherwise. Besides, $a_{ii} = 0$ as self-connection is not allowed. For each individual $i = 1, 2, ..., N$, the predicted volatility at time t is related to the returns of all its neighbours through

$$\sum_{j=1}^{N} w_{ij} y_{j,t-1}^2$$

in the sense of Cliff and Ord (1972), where $w_{ij} = \frac{a_{ij}}{\sum_{k=1}^{N} a_{ik}}$ is the (i, j)-th component of the row-normalized adjacency matrix W.

A TNGARCH (1,1) model is written as follows:

$$\begin{aligned} y_{it} &= \varepsilon_{it}\sqrt{h_{it}}, \\ h_{it} &= \omega + \left(\alpha^{(1)} 1_{\{y_{i,t-1}\geq 0\}} + \alpha^{(2)} 1_{\{y_{i,t-1}<0\}}\right) y_{i,t-1}^2 + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-1}^2 + \beta h_{i,t-1}, \quad (4.1.1) \\ i &= 1, 2, \cdots, N, \end{aligned}$$

where $1_{\{\cdot\}}$ is the indicator function. To assure the positiveness of h_{it}, it is assumed that $\omega > 0$ while $\alpha^{(1)}, \alpha^{(2)}, \lambda, \beta \geq 0$. $\{\varepsilon_{it}\}$ is a white noise process satisfying the following assumption:

Assumption 4.1.1. $\{\varepsilon_{it} : i = 1, 2, ..., N; t \in \mathbb{Z}\}$ *are IID across* i *and* t*, sharing an non-degenerate distribution with mean 0 and variance 1.*

If $\alpha^{(1)} \neq \alpha^{(2)}$, then the effect of $y_{i,t-1}^2$ on the predicted volatility h_{it} changes depending on whether $y_{i,t-1} \geq 0$ or $y_{i,t-1} < 0$. Otherwise, (4.1.1) degenerates to (2.2.5).

Stationarity conditions of this model will be derived in Section 4.2 with fixed N. The asymptotic properties of QMLE will be investigated in Section 4.3, in the case when $T \to \infty$ and $N \to \infty$. Then we will propose a Wald statistic in Section 4.4.1 to test the existence of threshold effect. In Section 4.5, our methodology is tested upon simulated data that are generated based on four different kinds of network structure. We observed an asymmetry that is different from existing literature, in how much the volatility responds to good news and bad news at individual level by applying our model to high-dimensional time series of log returns in Section 4.6.

4.2 Stationarity with fixed N

To derive the conditions under which model (4.1.1) is strictly stationary, we rewrite the conditional variance process in vector form

$$\mathbf{h}_t = \omega \mathbf{1}_N + B_{t-1}\mathbf{h}_{t-1} \tag{4.2.1}$$

with notations as follows:

$$\begin{aligned}
&\mathbf{h}_t = (h_{1t}, h_{2t}, ..., h_{Nt})' \in \mathbb{R}^N, \\
&\mathbf{1}_N = (1, 1, ..., 1)' \in \mathbb{R}^N, \\
&B_{t-1} = \alpha^{(1)} R_{t-1} E_{t-1} + \alpha^{(2)}(I_N - R_{t-1})E_{t-1} + \lambda W E_{t-1} + \beta I_N, \\
&R_{t-1} = \text{diag}\left\{1_{\{y_{1,t-1}\geq 0\}}, 1_{\{y_{2,t-1}\geq 0\}}, ..., 1_{\{y_{N,t-1}\geq 0\}}\right\}, \\
&E_{t-1} = \text{diag}\left\{\varepsilon^2_{1,t-1}, \varepsilon^2_{2,t-1}, ..., \varepsilon^2_{N,t-1}\right\}.
\end{aligned}$$

Since $y_{it} = \varepsilon_{it}\sqrt{h_{it}}$, $y_{it} \geq 0$ is equivalent to $\varepsilon_{it} \geq 0$. Hence

$$R_{t-1} = \text{diag}\left\{1_{\{\varepsilon_{1,t-1}\geq 0\}}, 1_{\{\varepsilon_{2,t-1}\geq 0\}}, ..., 1_{\{\varepsilon_{N,t-1}\geq 0\}}\right\}.$$

In this case, the random matrices $\{B_t\}$ are i.i.d. and model (4.2.1) is a generalized autoregressive equation by Definition 1.4 in Bougerol and Picard (1992). It is easy to verify that $\mathbb{E}(\log^+ \|B_0\|_*) < \infty$. Therefore, the top Lyapunov exponent associated to $\{B_t\}$ is well-defined as follows:

$$\gamma := \inf\left\{\mathbb{E}\left(\frac{1}{t+1}\log\|B_t B_{t-1}...B_0\|_*\right),\ t \in \mathbb{N}\right\}, \tag{4.2.2}$$

where $\|\cdot\|_*$ is an operator norm of $N \times N$ matrices, corresponding to any norm on $\mathbb{R}^N$ through

$$\|M\|_* = \sup\left\{\|M\mathbf{x}\| / \|\mathbf{x}\|;\ \mathbf{x} \in \mathbb{R}^N,\ \mathbf{x} \neq 0\right\}.$$

According to Theorem 3.2 in Bougerol and Picard (1992), the series

$$\mathbf{h}_t = \omega \mathbf{1}_N + \omega \sum_{k=1}^{\infty} B_{t-1} ... B_{t-k} \mathbf{1}_N \tag{4.2.3}$$

is the unique strictly stationary and ergodic solution of model (4.2.1) if and only if the Lyapunov exponent $\gamma < 0$. Under this condition, process $\{\mathbf{y}_t\}$ is also strictly stationary and ergodic where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})' \in \mathbb{R}^N$ since we could easily construct a continuous function $\Lambda : \mathbb{R}^N \to \mathbb{R}^N$ according to (4.1.1) such that $\mathbf{y}_t = \Lambda(\mathbf{h}_t)$. Besides, since $y_{it} = \varepsilon_{it}\sqrt{h_{it}}$, the almost sure convergence of (4.2.3) guarantees that $\mathbb{E}(h_{it}) < \infty$ for any i. Thus, $\mathbb{E}\|\mathbf{y}_t\|^2 = \sum_{i=1}^{N} \mathbb{E}(h_{it}) < \infty$ with $\|\cdot\|$ being an Euclidean norm.

By the subadditive ergodic theorem in Kingman (1973),

$$\gamma = \lim_{t\to\infty} \frac{1}{t+1} \log \|B_t B_{t-1} ... B_0\|_*$$

almost surely. In this case, γ could be approximated through computer simulation technique given a specific distribution of ε_{it}. For the purpose of reducing computation complexity, we derive a sufficient condition that is simple and much easier to verify.

Theorem 4.1. *Under Assumption 4.1.1, model* (4.2.1) *has a unique strictly stationary and ergodic solution in the form* (4.2.3) *if*

$$\max\left\{\alpha^{(1)}, \alpha^{(2)}\right\} + \beta + \lambda < 1. \tag{4.2.4}$$

4.3 Parameter estimation with $T \to \infty$ and $N \to \infty$

Following the settings in Section 3.4, let $D := \{(i,t) : i \in \mathbb{Z}, t \in \mathbb{Z}\}$ be a lattice on space $\mathbb{R}^2$, and $\rho((i,t),(j,\tau)) := \max\{|i-j|, |t-\tau|\}$ measures the distance between any two locations $(i,t),(j,\tau) \in D$. Assume we have observations $\{y_{it}, 1 \le i \le N, 1 \le t \le T\}$ from model (4.1.1) with respect to true parameters $\theta_0 := (\omega_0, \alpha_0^{(1)}, \alpha_0^{(2)}, \lambda_0, \beta_0)' \in \mathbb{R}^5$. Then these observations could be regarded as triangular array of random fields $\{y_{it} : (i,t) \in D_{NT}, NT \ge 1\}$ with $\{D_{NT}, NT \ge 1\}$ being a series of finite rectangular lattices $D_{NT} := \{(i,t) : 1 \le i \le N, 1 \le t \le T\}$.

Based on the infinite past of observations, the quasi log-likelihood function (ignoring constants) is

$$\begin{aligned} L_{NT}(\theta) &= \tfrac{1}{NT}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}(\theta), \\ l_{it}(\theta) &= \log\sigma_{it}^2(\theta) + \tfrac{y_{it}^2}{\sigma_{it}^2(\theta)}, \end{aligned} \tag{4.3.1}$$

where σ_{it}^2 is generated from model (4.1.1) as

$$\sigma_{it}^2 = \omega + \left\{\alpha^{(1)}1_{\{y_{i,t-1}\geq 0\}} + \alpha^{(2)}1_{\{y_{i,t-1}<0\}}\right\} y_{i,t-1}^2 + \lambda d_i^{-1}\sum_{j=1}^{N} a_{ij}y_{j,t-1}^2 + \beta\sigma_{i,t-1}^2,$$

and $\theta := (\omega, \alpha^{(1)}, \alpha^{(2)}, \lambda, \beta)' \in \mathbb{R}^5$ is the parameter vector. Since the evaluation of the exact value of (4.3.1) is infeasible in practice, it is convenient to approximate (4.3.1) with

$$\begin{aligned} \tilde{L}_{NT}(\theta) &= \tfrac{1}{NT}\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{l}_{it}(\theta), \\ \tilde{l}_{it}(\theta) &= \log\tilde{\sigma}_{it}^2(\theta) + \tfrac{y_{it}^2}{\tilde{\sigma}_{it}^2(\theta)}, \end{aligned} \tag{4.3.2}$$

where $\tilde{\sigma}_{it}^2$ is also generated from model (4.1.1) but with initial value $\tilde{\sigma}_{i0}^2 = 0$. And the QMLE of $\theta \in \Theta$ is given by

$$\hat{\theta}_{NT} := \underset{\theta\in\Theta}{\operatorname{argmin}}\, \tilde{L}_{NT}(\theta),$$

where "argmin" is the argument of the minimum.

Remark. A negative-valued constant is ignored in the log-likelihood function (4.3.1). Therefore the QMLE $\hat{\theta}_{NT}$ is the argument that minimizes $\tilde{L}_{NT}(\theta)$.

Firstly we investigate the weak dependence of $\{\sigma_{it}^2(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ and $\{y_{it}^2 : (i,t) \in D_{NT}, NT \geq 1\}$ with assumptions below, utilizing the connection between uniform NED (Jenish and Prucha, 2012) and η-weak dependence (see Example 3.4.2). Assumption 4.3.2 is also required by Zhou et al. (2020) to prove the asymptotic properties in the case when N is fixed. Assumption 4.3.3 puts restriction on the sparsity of the network. Similar restrictions on the network structure could also be seen in Assumption 3 by Xu and Lee (2015) and Assumption 3.2 by Xu et al. (2024).

Assumption 4.3.1. $\kappa_4 := \mathbb{E}\varepsilon_{it}^4 < \infty$ *such that* $\kappa_4\left(\max\{\alpha^{(1)}, \alpha^{(2)}\} + \beta + \lambda\right)^2 < 1.$

Assumption 4.3.2. *Θ is a compact subset of $\mathbb{R}^5$ such that all $\theta \in \Theta$ satisfy that $\omega > 0, \alpha^{(1)} \geq 0, \alpha^{(2)} \geq 0, \lambda \geq 0, \beta \geq 0$, (4.2.4) and Assumption 4.3.1, and the true parameter $\theta_0 \in \Theta$ is an interior point of Θ.*

Assumption 4.3.3. *The row-normalized adjacency matrix W satisfies one of following conditions:*

(a). $w_{ij} = \mathcal{O}(|i-j|^{-\frac{\mu+2}{2}})$ for some $\mu > 0$;

(b). $w_{ij} \neq 0$ if $|i-j| \leq K$ for some constant $K \geq 1$, and $w_{ij} = 0$ otherwise.

Remark. In Assumption 4.3.3, w_{ij} that measures the power of the connection between two arbitrary nodes i and j is restricted by $|i-j|$, which does not represent the distance between node i and node j. Assumption 4.3.3 is simply a technical restriction on the structure of the matrix W, similar to the Assumption 3.2 in Xu et al. (2024). Assumption 5.3.3 in Chapter 5 and Assumption (NB4) in Chapter 6 are also purely technical. Of course, as we have mentioned in Section 2.3, for spatio-temporal models, i and j often represent spatial locations rather than just two indices. For example, Xu and Lee (2015) has similar assumption as ours, except that w_{ij} is restricted by the Euclidean distance between i and j, where i and j are vector-valued spatial locations.

Recalling from Section 3.4.1, we could define a σ-algebra

$$\mathcal{F}_{it}(s) := \sigma\left\{\varepsilon_{it} : (j,\tau) \in D_{NT}, \rho((i,t),(j,\tau)) \leq s\right\}$$

for all $(i,t) \in D_{NT}$, $NT \geq 1$ and $s > 0$.

Lemma 4.3.1. *If (4.2.4), Assumptions 4.1.1, 4.3.1, 4.3.2 and 4.3.3(a) are satisfied, then for all $\theta \in \Theta$ we have*

$$\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \left\|\sigma_{it}^2(\theta) - \mathbb{E}(\sigma_{it}^2(\theta)|\mathcal{F}_{it}(s))\right\|_2 \leq Cs^{-\mu}$$

for some constant $C > 0$. If Assumption 4.3.3(b) holds instead of 4.3.3(a),

$$\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \left\|\sigma_{it}^2(\theta) - \mathbb{E}(\sigma_{it}^2(\theta)|\mathcal{F}_{it}(s))\right\|_2 \leq C\rho^s$$

for some constant $0 < \rho < 1$.

Remark. By Assumptions 4.1.1 and 4.3.1, the same results in Lemma 4.3.1 could be directly derived for $\{y_{it}^2 : (i,t) \in D_{NT}, NT \geq 1\}$ since $y_{it}^2 = \varepsilon_{it}^2 \sigma_{it}^2(\theta_0)$.

Remark. Since we have shown that $\{y_{it}^2 : (i,t) \in D_{NT}, NT \geq 1\}$ and $\{\sigma_{it}^2(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ are uniformly $\mathbb{L}^2$-NED, by Proposition 3.3 they are also η-weakly dependent with η-coefficients asymptotically equivalent to the NED coefficients.

It is essential for us to obtain the weak dependence of $\{l_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$, as well as their first and second order derivatives at θ_0, so that we could utilize the limit theorems (Theorem 3.1 and Theorem 3.2) to prove the consistency and asymptotic normality of $\hat{\theta}_{NT}$. Therefore we need the assumptions below aside from those required by Lemma 4.3.1. Particularly, Assumption 4.3.5 is a constraint on the decaying rate of dependence coefficients, which is required by Assumption 3.3.5. Apparently, in the second case in Lemma 4.3.1, we do not need Assumption 4.3.5.

Assumption 4.3.4. $\mathbb{E}|\varepsilon_{it}|^{2r} < \infty$ *and* $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \mathbb{E}|\sigma_{it}(\theta)|^{2r} < \infty$ *for some* $r > 2$.

Remark. Since $y_{it}^2 = \varepsilon_{it}^2 \sigma_{it}^2(\theta_0)$ and ε_{it} is independent from $\sigma_{it}(\theta_0)$, we also have

$$\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \mathbb{E}|y_{it}|^{2r} < \infty$$

based on the assumption above.

Assumption 4.3.5. *The* μ *in Assumption 4.3.3 satisfies* $\mu > 4 \vee \frac{2(r-1)}{r-2}$.

Assumption 4.3.6. $\inf_{NT\geq 1} \lambda_{min}(\Sigma_{NT}) > 0$ *where*

$$\Sigma_{NT} := \frac{\kappa_4 - 1}{NT} \sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\frac{1}{\sigma_{it}^4(\theta_0)} \frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0) \frac{\partial}{\partial\theta'}\sigma_{it}^2(\theta_0)\right].$$

Theorem 4.2. *Under Assumptions required by Lemma 4.3.1, the quasi-maximum likelihood estimator* $\hat{\theta}_{NT}$ *is consistent, i.e.*

$$\hat{\theta}_{NT} \xrightarrow{p} \theta_0$$

as $T \to \infty$ and $N \to \infty$. If $N = o(T)$, Assumptions 4.3.4, 4.3.5 and 4.3.6 also hold, then

$$\sqrt{NT}\Sigma_{NT}^{1/2}(\hat{\theta}_{NT} - \theta_0) \xrightarrow{d} N(0, (\kappa_4 - 1)^2 I_5).$$

As we will show in the proof of Proposition 4.1, κ_4 and Σ_{NT} above could be approximated by

$$\hat{\kappa}_4 := \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{y_{it}^4}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})} \tag{4.3.3}$$

and

$$\widehat{\Sigma}_{NT} := \frac{\hat{\kappa}_4 - 1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{1}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta'}\right] \tag{4.3.4}$$

respectively. The latter could be calculated recursively as

$$\frac{\partial}{\partial\theta}\tilde{\sigma}_{it}^2(\hat{\theta}_{NT}) = \tilde{\mathbf{u}}_{i,t-1} + \hat{\beta}\frac{\partial}{\partial\theta}\tilde{\sigma}_{i,t-1}^2(\hat{\theta}_{NT})$$

where

$$\tilde{\mathbf{u}}_{i,t-1} = \begin{pmatrix} 1 \\ y_{i,t-1}^2 1_{\{\hat{\varepsilon}_{i,t-1}\geq 0\}} \\ y_{i,t-1}^2 1_{\{\hat{\varepsilon}_{i,t-1}< 0\}} \\ \sum_{j=1}^{N} w_{i,j} y_{j,t-1}^2 \\ \tilde{\sigma}_{i,t-1}^2(\hat{\theta}_{NT}) \end{pmatrix}.$$

4.4 Tests on threshold effect and residuals

4.4.1 A Wald test for the threshold effect

Given a null hypothesis

$$H_0 : \Gamma\theta_0 = \eta \tag{4.4.1}$$

where Γ is an $s \times 5$ matrix with rank s and η is an s-dimensional vector, we could define a Wald test statistic as follows:

$$W_{NT} := (\Gamma\hat{\theta}_{NT} - \eta)'\left\{\frac{\Gamma}{NT}(\hat{\kappa}_4 - 1)^2\widehat{\Sigma}_{NT}^{-1}\Gamma'\right\}^{-1}(\Gamma\hat{\theta}_{NT} - \eta), \tag{4.4.2}$$

where $\hat{\kappa}_4$ and $\widehat{\Sigma}_{NT}$ are defined in (4.3.3) and (4.3.4).

By the asymptotic normality of $\hat{\theta}_{NT}$, W_{NT} could also be proved to follow a canonical asymptotic distribution as in the following theorem.

Proposition 4.1. *Under the same assumptions required by Theorem 4.2, as $T \to \infty$, $N \to \infty$ and $N = o(T)$, the Wald test statistic defined in (4.4.2) asymptotically follows a χ^2 distribution with degree of freedom s, i.e.*

$$W_{NT} \xrightarrow{d} \chi^2_s.$$

4.4.2 A white noise test on the residuals

There has been a large literature investigating high-dimensional time series models, including Xu and Lee (2015), Zhu et al. (2017) and Xu et al. (2024) among others, but none of them has used diagnostic tools to check the model adequacy. In this section, we will introduce a high-dimensional white noise test developed by Li et al. (2019) that can be applied to the diagnostic of high-dimensional models including ours.

Assume we have residuals $\{\mathbf{r}_t : 1 \leq t \leq T\}$, where $\mathbf{r}_t := (r_{1t}, ..., r_{Nt})'$. We want to test whether $\{\mathbf{r}_t : 1 \leq t \leq T\}$ are high-dimensional white noises, i.e. there exists a matrix P such that

$$H_0 : \mathbf{r}_t = P\mathbf{z}_t, \tag{4.4.3}$$

where $\mathbf{z}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$. The test statistic is the sum of squared singular values of first q lagged sample autocovariance matrices:

$$G_q := \sum_{\tau=1}^{q} tr(\hat{S}_\tau \hat{S}'_\tau), \tag{4.4.4}$$

where $\hat{S}_\tau = \frac{1}{T}\sum_{t=1}^{T} \mathbf{r}_t \mathbf{r}'_{t-\tau}$ with $\mathbf{r}_t = \mathbf{r}_{t+T}$ when $t \leq 0$.

If P is unknown, the sample covariance matrix of $\mathbf{r}_t$ is $\hat{S}_0 = \frac{1}{T}\sum_{t=1}^{T} \mathbf{r}_t \mathbf{r}'_t$. According to (2.8) in Li et al. (2019), we reject (4.4.3) if

$$\frac{G_q - \frac{N^2 q}{T}\hat{s}_1^2}{\sqrt{\frac{2N^2 q}{T^2}\left(\hat{s}_2 - \frac{N}{T}\hat{s}_1^2\right)^2}} > Z_\alpha$$

where $\hat{s}_1 = \frac{1}{N}tr(\hat{S}_0)$, $\hat{s}_2 = \frac{1}{N}tr(\hat{S}_0^2)$ and Z_α is the upper-α quantile of standard normal distribution.

Note that $\{\mathbf{r}_t : 1 \leq t \leq T\}$ being white-noise means that the residuals are uncorrelated over t. However, it does not indicate that the residuals are uncorrelated over both i and t. The latter indicates a stronger adequacy of high-dimensional model. We could assume that $P = I_N$ in the null hypothesis, and by (2.5) in Li et al. (2019), we reject $H_0 : \mathbf{r}_t = \mathbf{z}_t$ if

$$\frac{G_q - \frac{N^2 q}{T}}{\sqrt{\frac{2N^2 q}{T^2} + \frac{4N^3 q^2(\kappa_4 - 3)}{T^3} + \frac{8N^3 q^2}{T^3}}} > Z_\alpha.$$

4.5 Simulation study

4.5.1 Network simulation

The symmetric matrix A in model 4.2.1 represents an undirected network structure, the pattern of which varies over different application scenarios. In this simulation study, we tend to use four different mechanisms of simulating corresponding network. The network structure in Example 4.5.1 adapts to Assumption 4.3.3(b), which is required by geometric NED as we have shown in Lemma 4.3.1. Simulation mechanisms introduced in Examples 4.5.2 – 4.5.4 are for testing the robustness of our estimation, against network structures that may violate Assumption 4.3.3.

Example 4.5.1. **(D-neighbourhood)** For each node $i \in \{1, 2, ..., N\}$, it is connected to node j only if j is inside i's D-neighbourhood. That is, in the adjacency matrix, $a_{ij} = 1$ if $0 < |i - j| \leq D$ and $a_{ij} = 0$ otherwise. Figure 4.1(a) is a visualization of such a network with $N = 100$ and $D = 10$.

Example 4.5.2. **(Random)** For each node $i \in \{1, 2, ..., N\}$, we generate D_i from uniform distribution $U(0, 5)$, and then draw $[D_i]$ samples randomly from $\{1, 2, ..., N\}$ to form a set S_i ($[x]$ denotes the integer part of x). $A = (a_{ij})$ could be generated by letting $a_{ij} = 1$ if $j \in S_i$ and $a_{ij} = 0$ otherwise. In a network simulated with such mechanism, as it is indicated in Figure 4.1(b), there is no significantly influential node (i.e. node with extremely large in-degree).

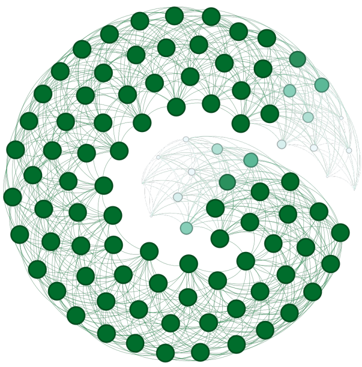
Example 4.5.3. **(Power-law)** According to Clauset et al. (2009), for each node i in such a network, D_i is generated the same way as in Example 4.5.2. Instead of uniformly selecting $[D_i]$ samples from $\{1, 2, ..., N\}$, these samples are collected w.r.t. probability $p_i = s_i / \sum_{i=1}^{N} s_i$ where s_i is generated from a discrete power-law distribution $\mathbb{P}\{s_i = x\} \propto x^{-a}$ with scaling parameter $a = 2.5$. As shown in Figure 4.1(c), a few nodes have much larger in-degrees while most of them have less than 2. Compared to Example 4.5.2, network structure with power-law distribution exhibits larger gaps between the influences of different nodes. This type of network is suitable for modeling social media such as Twitter and Instagram, where celebrities have huge influence while the ordinary majority has little.

Example 4.5.4. **(K-blocks)** As it was proposed in Nowicki and Snijders (2001), in a network with stochastic block structure, all nodes are divided into blocks and nodes from the same block are more likely to be connected compared to those from different blocks. To simulate such structure, these N nodes are randomly divided into K groups by assigning labels $\{1, 2, ..., K\}$ to every node with equal probability. For any two nodes i and j from the same group, let $\mathbb{P}(a_{ij} = 1) = 0.5$ while for those two from different groups, $\mathbb{P}(a_{ij} = 1) = 0.001/N$. Hence, it is very unlikely for nodes to be connected across groups. Our simulated network successfully mimics this characteristic as Figure 4.1(d) shows clear boundaries between groups. Block network also has its advantage from a practical perspective. For instance, the price of one stock is highly relevant to those in the same industry sector.

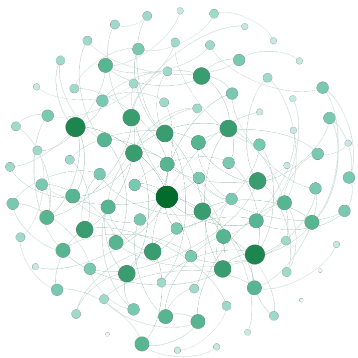
In the next section, the simulation study is carried out on datasets that are generated according to the process (4.1.1) in conjunction with three types of adjacency matrices in Examples 4.5.1 – 4.5.4.

4.5.2 Simulation results

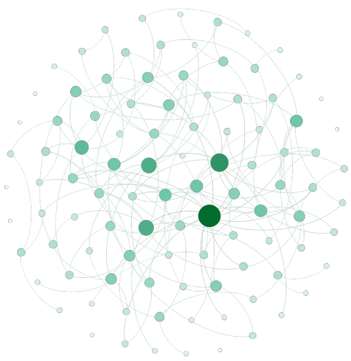
Setting the true parameters θ_0 as $(0.1, 0.1, 0.2, 0.2, 0.2)'$, we generate data according to process (4.1.1) with different sample sizes T and number of dimensions N. In our setting, T increases from 50 to 4000, while N also increases at relatively slower rates of $\mathcal{O}(\sqrt{T})$ and $\mathcal{O}(T/\log(T))$ respectively, as it is showed in the following table:



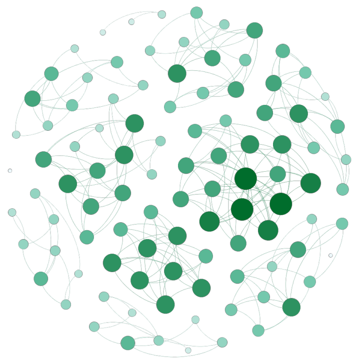
(a). Example 4.5.1 ($D = 10$)



(b). Example 4.5.2



(c). Example 4.5.3



(d). Example 4.5.4 ($K = 10$)

Figure 4.1: Visualized network structures with $N = 100$

T	200	500	1000	2000
$N \approx \sqrt{T}$	14	22	31	44
$N \approx T/\log(T)$	37	80	144	263

For each combination of (T, N), $M = 1000$ datasets will be simulated independently, according to (4.1.1). Based on the m-th $(m = 1, 2, ..., M)$ dataset, the estimation of θ_0 will be carried out and the estimation result is denoted as $\hat{\theta}_m = (\hat{\theta}_{km})' = (\hat{\omega}_m, \hat{\alpha}_m^{(1)}, \hat{\alpha}_m^{(2)}, \hat{\lambda}_m, \hat{\beta}_m)'$. For $k = \{1, 2, 3, 4, 5\}$, the following two measurements are used to evaluate the performance of simulation results:

1. **root-mean-square error**: $RMSE_k = \sqrt{M^{-1} \sum_{m=1}^{M} (\hat{\theta}_{km} - \theta_{k0})^2}$,
2. **coverage probability**: $CP_k = M^{-1} \sum_{m=1}^{M} 1_{\{\theta_{k0} \in CI_{km}\}}$.

CI_{km} is the 95% confidence interval defined as

$$CI_{km} = \left(\hat{\theta}_{km} - z_{0.975} \widehat{SE}_{km}, \hat{\theta}_{km} + z_{0.975} \widehat{SE}_{km} \right),$$

where the estimated standard error $\widehat{SE}_{km}$ could be calculated as the square root of k-th diagonal element of $(NT)^{-1}(\hat{\kappa}_4 - 1)\widehat{\Sigma}_{NT}^{-1}$ and $z_{0.975}$ is the 0.975th quantile of a standard normal distribution. In order to eliminate the effect of starting points, a different initial guess of θ is used for each m.

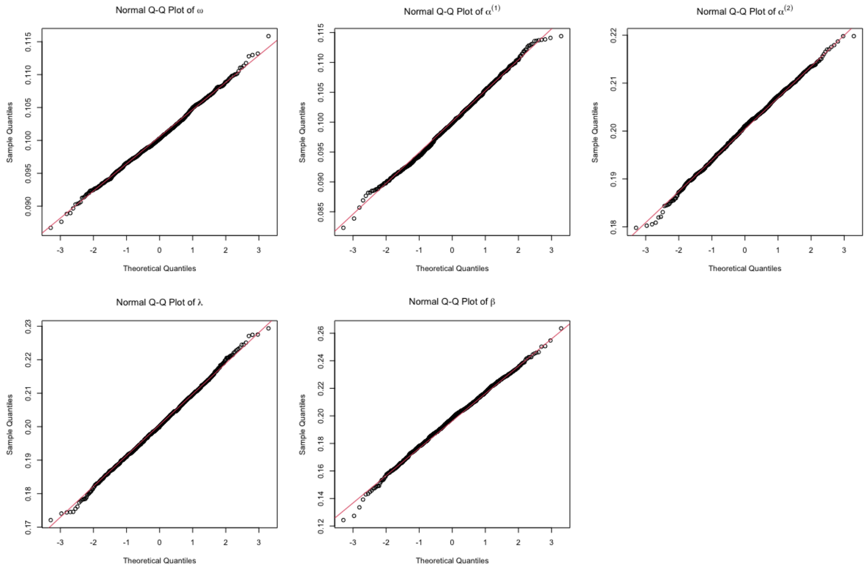
The results of root-mean-square errors with coverage probabilities in the parentheses are reported in Table 4.1 and Table 4.2 respectively, under different network structures and sample sizes. the consistency of the estimator is obvious since RMSE drops towards zero when T and N increases. Additionally, $\widehat{SE}$ provides reliable estimates of true standard errors since the coverage probabilities are close the theoretical value of 95%. Moreover, in Figures 4.2 to 4.5 we draw the normal Q-Q plots for the estimation results when $T = 2000, N = 44$ and $T = 2000, N = 263$ under different network structures. These Q-Q plots provide additional evidence for the asymptotic normality of $\hat{\theta}_{NT}$ in Proposition 4.2. In conclusion, the asymptotic properties of our estimator in Theorem 4.2 are well supported by our simulation results, even for network structures in Examples 4.5.2 – 4.5.4 that may violate Assumption 4.3.3.

	T	N	ω	$\alpha^{(1)}$	$\alpha^{(2)}$	λ	β
Example 4.5.1	200	14	0.0197 (0.92)	0.0301 (0.94)	0.0378 (0.95)	0.0371 (0.95)	0.0964 (0.91)
	500	22	0.0100 (0.94)	0.0152 (0.95)	0.0192 (0.95)	0.0218 (0.95)	0.0492 (0.93)
	1000	31	0.0061 (0.95)	0.0086 (0.96)	0.0114 (0.96)	0.0143 (0.96)	0.0299 (0.96)
	2000	44	0.0042 (0.92)	0.0053 (0.96)	0.0066 (0.96)	0.0094 (0.95)	0.0201 (0.93)
Example 4.5.2	200	14	0.0173 (0.93)	0.0295 (0.94)	0.0386 (0.94)	0.0344 (0.94)	0.0854 (0.93)
	500	22	0.0086 (0.95)	0.0149 (0.95)	0.0191 (0.95)	0.0177 (0.95)	0.0414 (0.96)
	1000	31	0.0047 (0.95)	0.0088 (0.95)	0.0110 (0.95)	0.0107 (0.94)	0.0263 (0.95)
	2000	44	0.0028 (0.96)	0.0051 (0.96)	0.0067 (0.95)	0.0058 (0.95)	0.0144 (0.96)
Example 4.5.3	200	14	0.0169 (0.92)	0.0299 (0.93)	0.0388 (0.94)	0.0330 (0.94)	0.0852 (0.91)
	500	22	0.0077 (0.95)	0.0153 (0.95)	0.0190 (0.95)	0.0166 (0.95)	0.0413 (0.94)
	1000	31	0.0047 (0.94)	0.0092 (0.94)	0.0117 (0.93)	0.0099 (0.96)	0.0252 (0.95)
	2000	44	0.0027 (0.94)	0.0052 (0.96)	0.0065 (0.95)	0.0057 (0.96)	0.0152 (0.94)
Example 4.5.4	200	14	0.0226 (0.91)	0.0308 (0.94)	0.0383 (0.94)	0.0486 (0.93)	0.1063 (0.90)
	500	22	0.0095 (0.94)	0.0152 (0.94)	0.0194 (0.95)	0.0196 (0.95)	0.0478 (0.94)
	1000	31	0.0057 (0.94)	0.0092 (0.94)	0.0113 (0.96)	0.0118 (0.94)	0.0286 (0.94)
	2000	44	0.0034 (0.95)	0.0053 (0.96)	0.0066 (0.96)	0.0078 (0.95)	0.0169 (0.95)

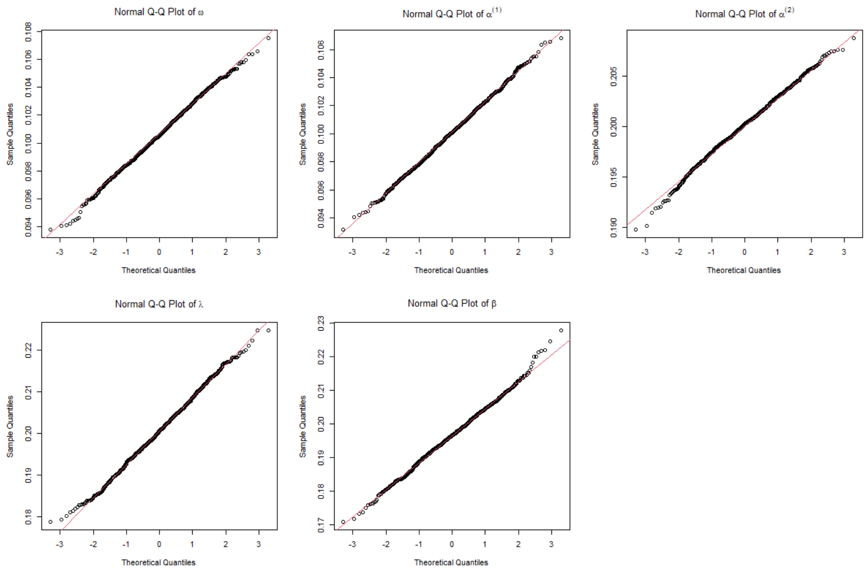
Table 4.1: Simulation results with different network structures ($N \approx \sqrt{T}$).

	T	N	ω	$\alpha^{(1)}$	$\alpha^{(2)}$	λ	β
Example 4.5.1	200	37	0.0131 (0.94)	0.0188 (0.94)	0.0236 (0.95)	0.0301 (0.95)	0.0624 (0.93)
	500	80	0.0065 (0.94)	0.0077 (0.96)	0.0100 (0.95)	0.0174 (0.94)	0.0303 (0.93)
	1000	144	0.0041 (0.93)	0.0042 (0.96)	0.0054 (0.94)	0.0119 (0.95)	0.0173 (0.92)
	2000	263	0.0023 (0.95)	0.0022 (0.95)	0.0028 (0.94)	0.0079 (0.95)	0.0088 (0.93)
Example 4.5.2	200	37	0.0102 (0.94)	0.0183 (0.94)	0.0237 (0.93)	0.0201 (0.95)	0.0521 (0.93)
	500	80	0.0044 (0.93)	0.0077 (0.95)	0.0102 (0.95)	0.0086 (0.95)	0.0229 (0.94)
	1000	144	0.0024 (0.93)	0.0042 (0.94)	0.0054 (0.94)	0.0048 (0.95)	0.0127 (0.93)
	2000	263	0.0013 (0.94)	0.0021 (0.96)	0.0028 (0.94)	0.0025 (0.95)	0.0066 (0.94)
Example 4.5.3	200	37	0.0103 (0.94)	0.0188 (0.94)	0.0225 (0.95)	0.0199 (0.95)	0.0505 (0.94)
	500	80	0.0042 (0.94)	0.0076 (0.96)	0.0096 (0.96)	0.0088 (0.95)	0.0225 (0.95)
	1000	144	0.0022 (0.94)	0.0041 (0.95)	0.0051 (0.95)	0.0048 (0.95)	0.0119 (0.94)
	2000	263	0.0011 (0.95)	0.0022 (0.94)	0.0027 (0.96)	0.0024 (0.95)	0.0061 (0.95)
Example 4.5.4	200	37	0.0129 (0.93)	0.0190 (0.95)	0.0231 (0.95)	0.0282 (0.94)	0.0615 (0.92)
	500	80	0.0053 (0.93)	0.0080 (0.95)	0.0095 (0.97)	0.0107 (0.97)	0.0267 (0.93)
	1000	144	0.0027 (0.94)	0.0042 (0.95)	0.0053 (0.95)	0.0055 (0.96)	0.0134 (0.94)
	2000	263	0.0014 (0.93)	0.0022 (0.95)	0.0027 (0.95)	0.0029 (0.94)	0.0072 (0.93)

Table 4.2: Simulation results with different network structures ($N \approx T/\log(T)$).

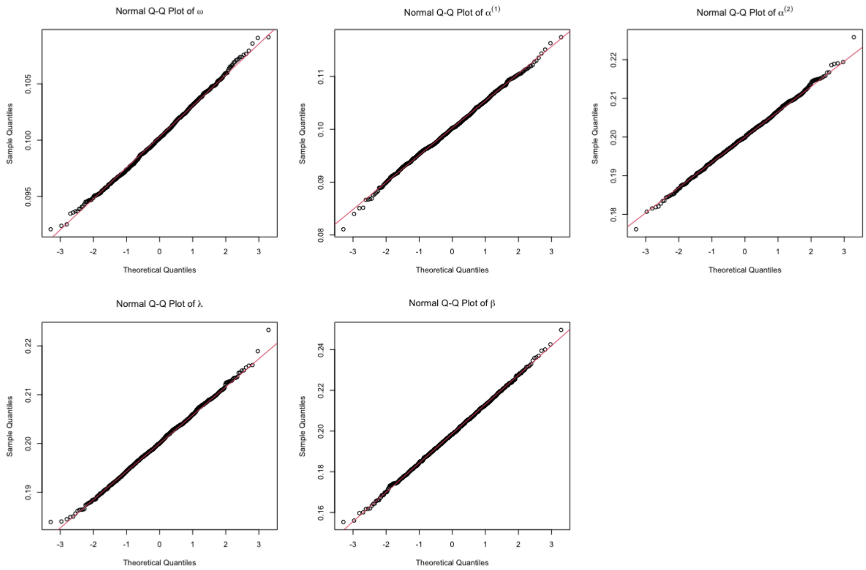


(a). $T = 2000, N = 44$

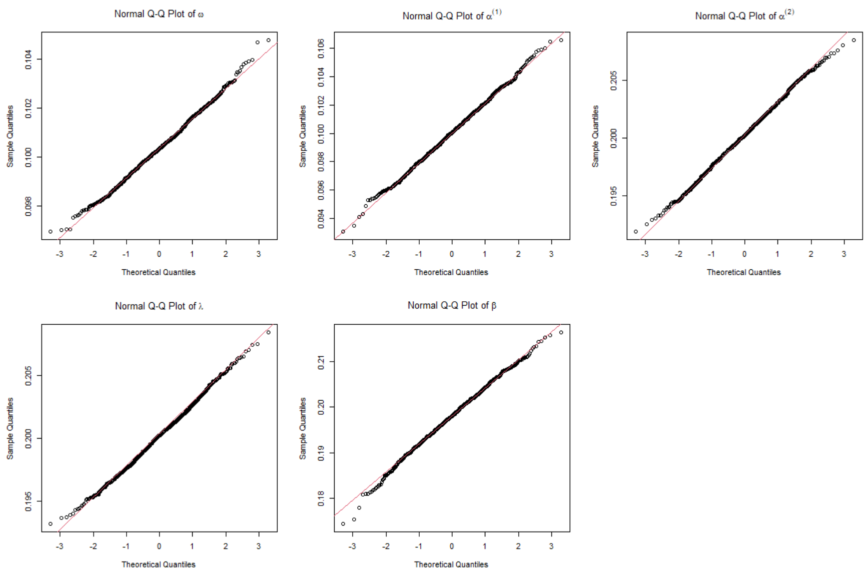


(b). $T = 2000, N = 263$

Figure 4.2: Q-Q plots of estimates for Example 4.5.1.

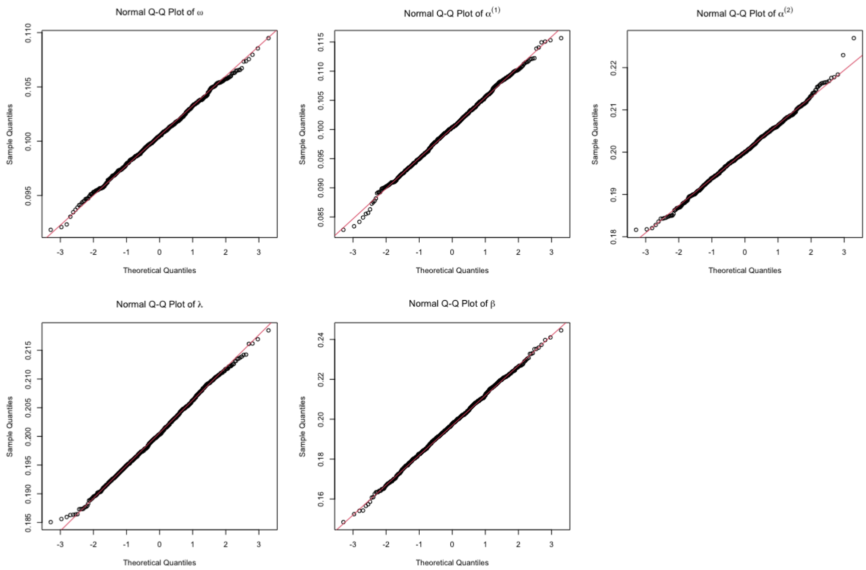


(a). $T = 2000, N = 44$

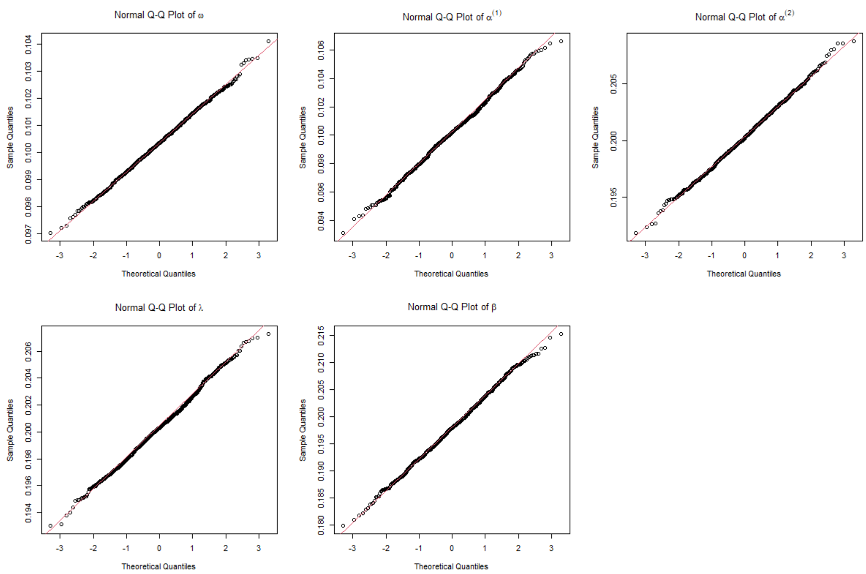


(b). $T = 2000, N = 263$

Figure 4.3: Q-Q plots of estimates for Example 4.5.2.

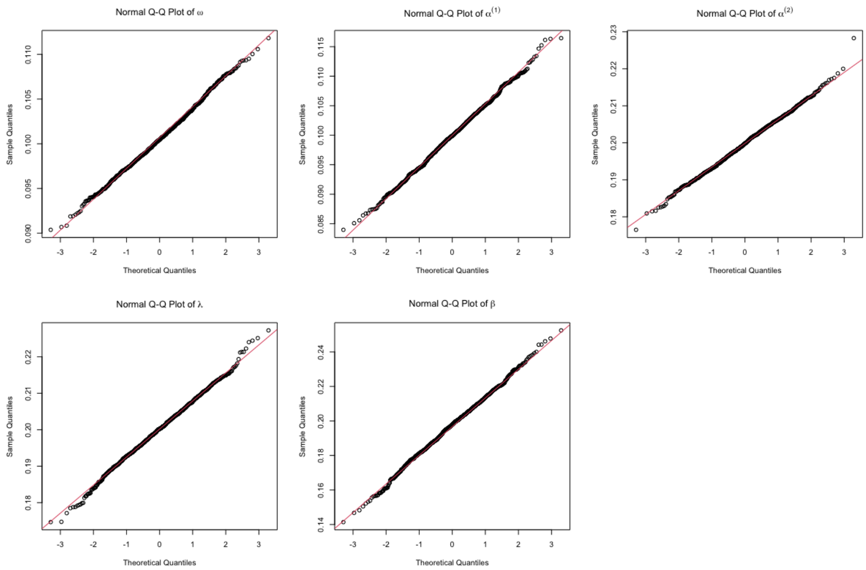


(a). $T = 2000, N = 44$

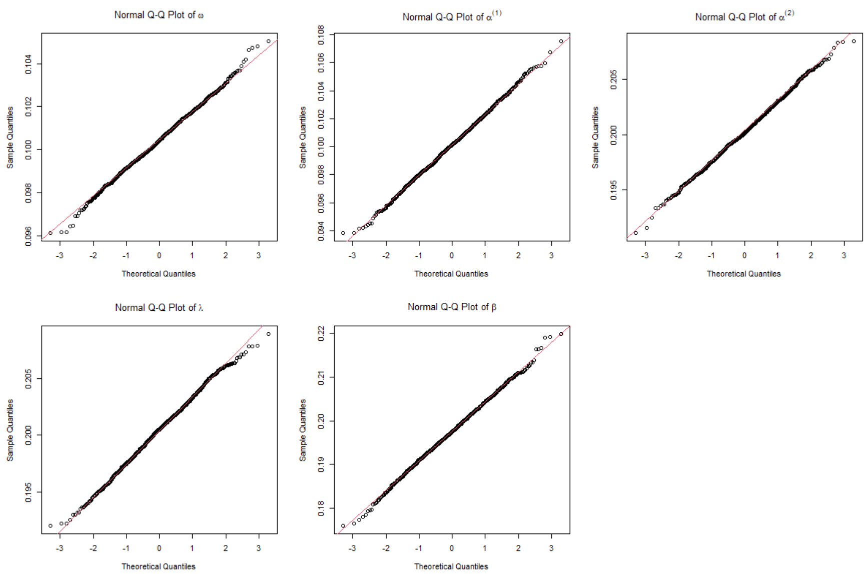


(b). $T = 2000, N = 263$

Figure 4.4: Q-Q plots of estimates for Example 4.5.3.



(a). $T = 2000, N = 44$



(b). $T = 2000, N = 263$

Figure 4.5: Q-Q plots of estimates for Example 4.5.4.

4.6 Empirical data analysis

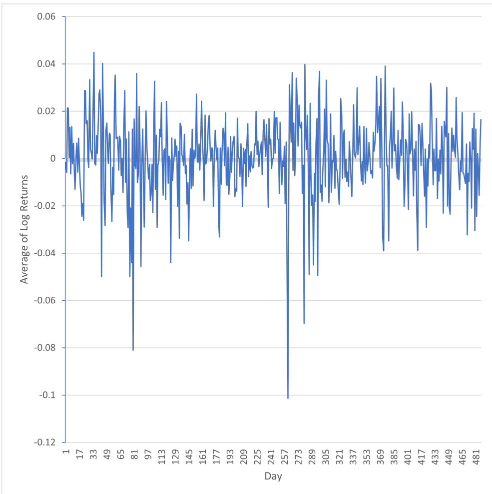
In addition to simulation studies, we want to test our model using real data from Chinese Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE). The dataset consists of daily log returns of 286 stocks, which are observed in two consecutive years of 2019 and 2020 ($T = 487$ except for closing days). These stocks come from four industry sectors as follows:

- 75 stocks from automotive industry sector;
- 73 stocks from financial industry sector;
- 68 stocks from information industry sector;
- 70 stocks from pharmaceutical industry sector.

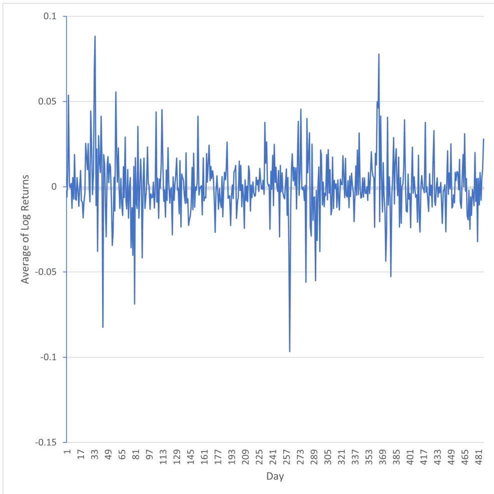
And our model is tested within each sector, in which the number of stocks is approximately $T/\log(T) \approx 79$. Hence the estimates and inferences could be trusted according to the simulation study.

As an initial impression of data from each category the time plots of daily average log returns are presented in Figure 4.6. We also have the shareholder information of each stock, based on which two stocks are considered as connected when they share at least one common shareholder among their top ten shareholders. By this principle, four adjacency matrices are constructed and visualized as Figure 4.7 for four different industry sectors. Although it is quite intuitive to tell from Figure 4.7 the sparsity of these four networks, we tend to use the network density (ND) as a quantified measurement, which is defined by the ratio of the number of existing edges to the number of potential connections: $ND := 100\% \times \frac{\sum_{i=1}^{N} d_i}{N(N-1)}$.

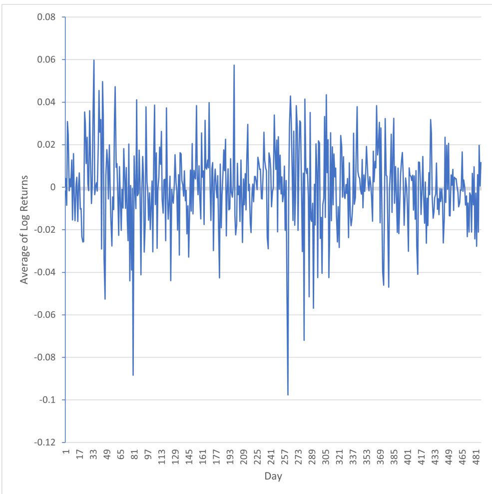
The results of parameter estimation is summarized in Table 4.3. Positive estimates of λ indicate positive correlation between the return of a stock and the returns of its neighbours, however it is worth noticing that the estimated network effect λ for automotive industry sector is much smaller than those from other sectors. As indicated in Figure 4.7(a), this could be caused by the sparsity of the network structure as the data from automotive industry has the lowest network density compared to others.



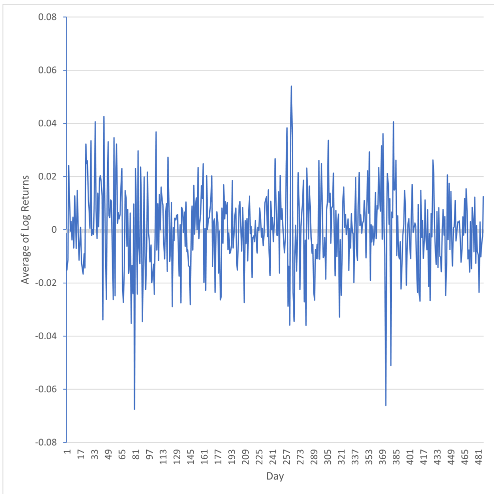
(a). Automotive industry



(b). Financial industry

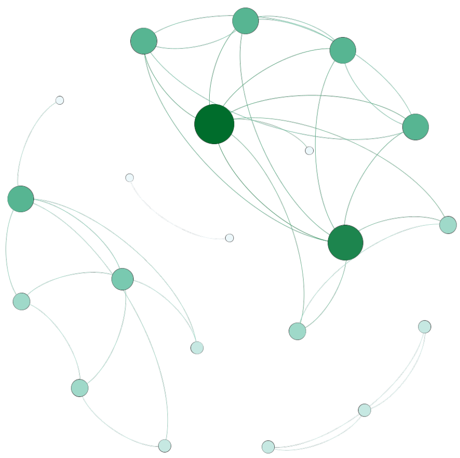


(c). Information industry

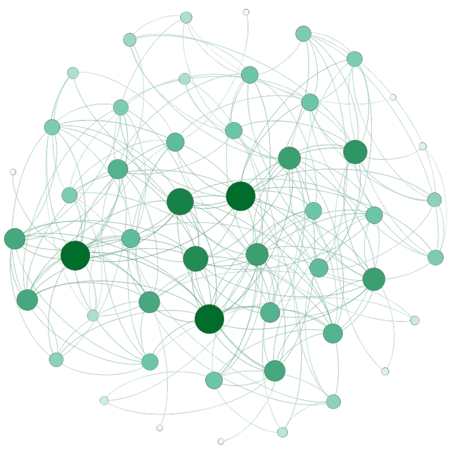


(d). Pharmaceutical industry

Figure 4.6: Average log returns of stocks from different industry sectors

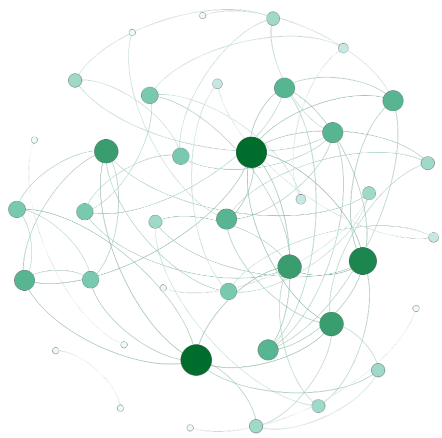


(a). Automotive industry (ND = 1.26%)



(b). Financial industry (ND = 8.11%)

(c). Information industry (ND = 1.58%)



(d). Pharmaceutical industry (ND = 2.82%)

Figure 4.7: Visualization of networks for stocks from different industry sectors

Comparing with other parameters, the estimates of β are much larger for all four categories. Strong memory of volatility has been observed in many econometric studies on daily data, and such persistence would be stronger with data sampled at higher frequency according to Nelson (1991).

Automotive Industry			**Financial Industry**		
Parameter	Estimation	SE	Parameter	Estimation	SE
ω	0.000099	5.83e-07	ω	0.000043	3.12e-06
$\alpha^{(1)}$	0.199408	1.08e-02	$\alpha^{(1)}$	0.247765	1.41e-02
$\alpha^{(2)}$	0.136423	1.01e-02	$\alpha^{(2)}$	0.202237	1.47e-02
λ	0.004591	4.71e-03	λ	0.010469	5.35e-03
β	0.727756	1.17e-02	β	0.737272	1.09e-02
Information Industry			**Pharmaceutical Industry**		
Parameter	Estimation	SE	Parameter	Estimation	SE
ω	0.000105	6.39e-06	ω	0.000063	4.15e-06
$\alpha^{(1)}$	0.172737	9.34e-03	$\alpha^{(1)}$	0.180950	1.05e-02
$\alpha^{(2)}$	0.122312	8.86e-03	$\alpha^{(2)}$	0.131722	1.06e-02
λ	0.009475	4.03e-03	λ	0.012929	4.06e-03
β	0.745699	1.11e-02	β	0.753305	1.11e-02

Table 4.3: Estimation results based on daily log-returns (2019&2020) of stocks from four industries.

We now conduct a Wald test on the existence of threshold effect based on the estimated parameters. By letting $\Gamma := (0, 1, -1, 0, 0)$ and $\eta := 0$ in (4.4.1), we can make a null hypothesis as follows:

$$H_0 : \alpha_0^{(1)} = \alpha_0^{(2)}.$$

As it is indicated in Table 4.4, we could reject the null hypothesis with strong confidence and conclude that there exists extremely significant threshold effect within each industry sector.

Automotive Industry	**Financial Industry**	**Information Industry**	**Pharmaceutical Industry**
1.09e-10	2.16e-07	3.8e-06	3.17e-06

Table 4.4: p-values of Wald test on $H_0 : \alpha_0^{(1)} = \alpha_0^{(2)}$

Using the diagnostic tool introduced in Section 4.4.2, we could check the model adequacy by inspecting the correlations between residual vectors $\mathbf{r}_t = \left[\frac{y_{1t}}{\tilde{\sigma}_{1t}(\hat{\theta}_{NT})}, ..., \frac{y_{Nt}}{\tilde{\sigma}_{Nt}(\hat{\theta}_{NT})}\right]'$.

We will test null hypothesis $H_0 : \mathbf{r}_t = P\mathbf{z}_t$ with P being unknown and $P = I_N$ respectively, the results are summarized in Table 4.5. In all sectors, we can not reject the hypothesis that the residual vectors are high-dimensional white noises with $\mathbb{E}\mathbf{r}_t = 0$ and $Var(\mathbf{r}_t) = PP'$ over t. However, the stronger hypothesis $H_0 : \mathbf{r}_t = \mathbf{z}_t$ is rejected, as there exist correlations between residuals $\left\{\frac{y_{it}}{\tilde{\sigma}_{it}(\hat{\theta}_{NT})}\right\}$ with different i. We might be able to eliminate such deficiency in the adequacy of our model by heterogeneous parameterization with coefficients as ω_i. $\alpha_i^{(1)}$, $\alpha_i^{(2)}$, λ_i and β_i, or by considering a dynamic network structure. However, the purpose of the introduction of network structure is to reduce the number of parameters of high-dimensional time series. Besides, deriving limit theorems for models with heterogeneous parameters or dynamic network could be theoretically challenging.

	Automotive Industry	**Financial Industry**	**Information Industry**	**Pharmaceutical Industry**
P is unknown	Not rejected	Not rejected	Not rejected	Not rejected
$P = I_N$	Rejected	Rejected	Rejected	Rejected

Table 4.5: Results of high-dimensional white noise test on $H_0 : \mathbf{r}_t = P\mathbf{z}_t$ with $q = 3$ and $\alpha = 0.01$.

On the other hand, our results on asymmetric effect of positive and negative news are quite different compared to what was derived from univariate data in the literature. For instance, in a study by Engle and Ng (1993) on the daily returns of Japanese stock index TOPIX, it was found that negative news would have larger impact on future volatility. Such a phenomenon is reasonable in the stock market since investors would lose confidence to a certain asset when it performs badly, hence they would adjust their portfolio and add more uncertainty to the future. However, it is not necessarily the case if we take into consideration the whole picture instead of looking at one individual and ignoring possible impact of its neighbours in the same system. In our estimation results, $\alpha^{(1)}$ are uniformly larger than $\alpha^{(2)}$, indicating a larger impact of good news on volatility. A more precise conclusion would be that the volatility of one individual is more sensitive to its own good news, which actually does not contradict the conclusion of Engle and Ng (1993), since in the univariate case, how much proportion of the "bad news" effect is actually contributed by bad performance in systematic perspective remains unknown. Our results show that good news has larger "local influence" as it is indicated by $\alpha^{(1)}$, while there is a possibility that bad news, despite of having less "local influence", spreads faster and has larger "global influence" on the neighbours through network connection. Such potential leads to a future extension of our model that the threshold effect could be further applied on the coefficient λ, allowing good news and bad news to have asymmetric network effect.

Chapter 5

Poisson Threshold Network GARCH

5.1 Introduction

Integer-valued time series can be observed in a wide range of scientific fields, such as the yearly trading volume of houses on real estate market De Wit et al. (2013), number of transactions of stocks Jones et al. (1994), or the daily mortality from COVID-19 Pham (2020). A first idea to model integer-valued time series is using a simple first-order autoregressive model (2.1.1). However in model (2.1.1) y_t is not necessarily an integer given integer-valued y_{t-1} and ε_t, due to the multiplication structure αy_{t-1}. Circumventing such problem by replacing the ordinary multiplication αy_{t-1} by the (binomial) thinning operation $\alpha \circ y_{t-1}$ where $\alpha \circ y | y \sim Bin(y, \alpha)$, McKenzie (1985) and Al-Osh and Alzaid (1987) proposed an integer-valued counterpart of the AR model (INAR), which was ground-breaking and led to various extensions of thinning-based linear models including integer-valued moving average model (INMA) (Al-Osh and Alzaid, 1988) and INARMA model McKenzie (1988) among others. An alternative approach to the multiplication problem, is to consider the regression of the conditional mean $\lambda_t := \mathbb{E}(y_t | \mathcal{H}_{t-1})$ where H_{t-1} is the σ-algebra generated by historical information up to $t - 1$. Based on this idea, integer-valued GARCH-type models (INGARCH) were proposed by (Heinen, 2003; Ferland et al., 2006; Fokianos et al., 2009) with conditional Poisson distribution of y_t, e.g., the Poisson autoregression (2.1.6). In this chapter we will construct a model based on the Poisson INGARCH model. Other variations of INGARCH models with different specifications of conditional distribution include negative binomial INGARCH (Zhu, 2010; Xu et al., 2012) and generalized Poisson INGARCH (Zhu, 2012) among others.

The application of preceding integer-valued models are all limited to one-dimensional time series, and the development of multi-dimensional integer-valued GARCH-type models is still at its early stage. e.g. the bivariate INGARCH models (Lee et al., 2018; Cui and Zhu, 2018; Cui et al., 2020) and other multivariate INGARCH models (Fokianos et al., 2020; Lee et al., 2023) on low-dimensional time series of counts. As for high-dimensional integer-valued time series, there exist several counterparts of the network GARCH model proposed by Zhou et al. (2020), such as the Poisson network autoregressive model (PNAR) by Armillotta and Fokianos (2024) and the grouped PNAR model by Tao et al. (2024). The PNAR of Armillotta and Fokianos (2024) allows for integer-valued time series with increasing network dimension. However, their model adopted an ARCH-type structure without considering the autoregressive term on the conditional mean, and moreover, there is no threshold structure in their model to capture asymmetric characteristics of volatilities. The grouped PNAR Tao et al. (2024) has a GARCH structure indeed, but its network dimension is fixed and not applicable to ultra high dimensional data. In this chapter we propose a Poisson threshold network GARCH model (PTNGARCH) that are distinguished in following aspects:

- A threshold structure is designed in our PTNGARCH so that it is capable of capturing asymmetric properties of high-dimensional volatilities for discrete data. The threshold effect can also be tested under such a framework.
- Our PTNGARCH includes an autoregressive term on the conditional mean so that it provides a parsimonious description of dynamic volatilities of high-dimensional count time series.
- Asymptotic theory, when both sample size and network dimension are large, of

maximum likelihood estimation for our model is established by the limit theorems for weakly dependent random fields in Chapter 3.

5.2 Stationarity under fixed N

Recalling the TNGARCH model (4.1.1), we consider an non-directed and weightless network with N nodes, represented by adjacency matrix A with its entry $a_{ij} = 1$ if there is a connection between node i and j, and $a_{ij} = 0$ otherwise. Correspondingly we have the row-normalized adjacency matrix W with its entry $w_{ij} = \frac{a_{ij}}{\sum_{j=1}^{N} a_{ij}}$. Distinguished from the model (4.1.1), PTNGARCH deals with $\mathbb{N}$-valued data. Let y_{it} be a non-negative integer-valued observation on node i at time t, and $\mathcal{H}_{t-1}$ denotes the σ-algebra consisting of all available information up to $t-1$. In our Poisson threshold network GARCH model, we suppose that y_{it} follows a conditional (on $\mathcal{H}_{t-1}$) Poisson distribution with (i,t)-varying mean λ_{it}. That is, a PTNGARCH(1,1) model has following form:

$$
\begin{aligned}
& y_{it}|\mathcal{H}_{t-1} \sim Poisson(\lambda_{it}), \\
& \lambda_{it} = \omega + \left(\alpha^{(1)} 1_{\{y_{i,t-1} \geq r\}} + \alpha^{(2)} 1_{\{y_{i,t-1} < r\}}\right) y_{i,t-1} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \beta \lambda_{i,t-1}, \quad (5.2.1) \\
& i = 1, 2, \cdots, N.
\end{aligned}
$$

The threshold parameter r is an positive integer, and $1_{\{\cdot\}}$ denotes an indicator function. To assure the positiveness of conditional variance, we need to assume that $\omega > 0$, $\alpha^{(1)} \geq 0$, $\alpha^{(2)} \geq 0$, $\xi \geq 0$ and $\beta \geq 0$.

Remark. Notice that in (5.2.1) we model the dynamics of conditional mean λ_{it}, which is the reason why the name “Poisson autoregressive” is sometimes used in the literature (Fokianos et al., 2009; Wang et al., 2014). Some authors still use the name “GARCH” since the mean is equal to the variance under Poisson distribution, and the dynamics of conditional mean are GARCH-like. We tend to keep the name “GARCH” to align with the TNGARCH model in Chapter 4.

Let $\{M_{it} : i = 1, 2, ..., N, t \in \mathbb{Z}\}$ be independent Poisson processes with unit intensities. Depending on λ_{it}, y_{it} can be interpreted as a Poisson distributed random variable $M_{it}(\lambda_{it})$, which is the number of occurrences during the time interval $(0, \lambda_{it}]$. i.e. $\mathbb{P}(y_{it} = n|\lambda_{it} = \lambda) = \frac{\lambda^n}{n!}e^{-\lambda}$. We could rewrite (5.2.1) in vectorized form as follows:

$$\begin{aligned} \mathbb{Y}_t &= \mathbb{M}_t(\Lambda_t), \\ \Lambda_t &= \omega \mathbf{1}_N + A(\mathbb{Y}_{t-1})\mathbb{Y}_{t-1} + \beta\Lambda_{t-1}, \end{aligned} \tag{5.2.2}$$

where $\mathbb{M}_t := (M_{1t}(\lambda_{1t}), M_{2t}(\lambda_{2t}), ..., M_{Nt}(\lambda_{Nt}))' \in \mathbb{N}^N$, and

$$\begin{aligned} \Lambda_t &= (\lambda_{1t}, \lambda_{2t}, ..., \lambda_{Nt})' \in \mathbb{R}^N, \\ \mathbf{1}_N &= (1, 1, ..., 1)' \in \mathbb{R}^N, \\ A(\mathbb{Y}_{t-1}) &= \alpha^{(1)}S(\mathbb{Y}_{t-1}) + \alpha^{(2)}(I_N - S(\mathbb{Y}_{t-1})) + \xi W, \\ S(\mathbb{Y}_{t-1}) &= \text{diag}\left\{1_{\{y_{1,t-1}\geq r\}}, 1_{\{y_{2,t-1}\geq r\}}, ..., 1_{\{y_{N,t-1}\geq r\}}\right\}. \end{aligned}$$

Assumption 5.2.1. *The coefficients $\alpha^{(1)}$ and $\alpha^{(2)}$ satisfy:*

(a). $\alpha^{(1)} \leq \alpha^{(2)}$*;*

(b). $\alpha^{(2)} \leq \left(1 + \frac{1}{r-1}\right)\alpha^{(1)}$ *when* $r > 1$*.*

Now we are ready to give a sufficient condition for model (5.2.2) to have a strictly stationary solution.

Theorem 5.1. *If Assumption 5.2.1 is satisfied and*

$$\max\{\alpha^{(1)}, \alpha^{(2)}\} + \xi + \beta < 1,$$

then there exists a strictly stationary process $\{\mathbb{Y}_t : t \in \mathbb{Z}\}$ that satisfies (5.2.2) and has finite first order moment.

5.3 Parameter estimation with $T \to \infty$ and $N \to \infty$

Assume that the model of interest is characterized by an array of parameters $\nu = (\theta', r)'$ with $\theta = (\omega, \alpha^{(1)}, \alpha^{(2)}, \xi, \beta)'$ and the parameter space $\Theta \times \mathbb{Z}_+$. The samples

$\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ are generated by model (5.2.1) with respect to true parameters $\nu_0 = (\omega_0, \alpha_0^{(1)}, \alpha_0^{(2)}, \xi_0, \beta_0, r_0)'$.

Based on the infinite past of observations, the log-likelihood function (ignoring constants) is

$$\begin{cases} L_{NT}(\nu) = \frac{1}{NT}\sum_{(i,t)\in D_{NT}} l_{it}(\nu), \\ l_{it}(\nu) = y_{it}\log\lambda_{it}(\nu) - \lambda_{it}(\nu) \end{cases} \tag{5.3.1}$$

where $\lambda_{it}(\nu)$ is generated from model (5.2.1) as

$$\begin{aligned} \lambda_{it}(\nu) =& \omega + \alpha^{(1)} 1_{\{y_{i,t-1}\geq r\}} y_{i,t-1} + \alpha^{(2)} 1_{\{y_{i,t-1}<r\}} y_{i,t-1} \\ &+ \xi \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \beta\lambda_{i,t-1}(\nu). \end{aligned} \tag{5.3.2}$$

In practice, (5.3.1) cannot be evaluated without knowing the true values of λ_{i0} for $i = 1,2,...,N$. Therefore, we approximate (5.3.1) by (5.3.3) below, using specified initial values $\lambda_{i0} = \tilde{\lambda}_{i0}, i = 1,2,...,N$:

$$\begin{cases} \tilde{L}_{NT}(\nu) = \frac{1}{NT}\sum_{(i,t)\in D_{NT}} \tilde{l}_{it}(\nu), \\ \tilde{l}_{it}(\nu) = y_{it}\log\tilde{\lambda}_{it}(\nu) - \tilde{\lambda}_{it}(\nu). \end{cases} \tag{5.3.3}$$

And the maximum likelihood estimates (MLE) are evaluated by

$$\hat{\nu}_{NT} = \underset{\nu\in\Theta\times\mathbb{Z}_+}{\operatorname{argmax}}\, \tilde{L}_{NT}(\nu). \tag{5.3.4}$$

However, the solution that maximizes the target function $\tilde{L}_{NT}(\nu)$ cannot be directly obtained by solving $\frac{\partial \tilde{L}_{NT}(\nu)}{\partial \nu} = 0$, since $r \in \mathbb{Z}_+$ is discrete, therefore the partial derivative of $\tilde{L}_{NT}(\nu)$ w.r.t. r is invalid. According to Wang et al. (2014), such an optimization problem with integer-valued parameter r could be broken up into two steps as follows:

1. Find

$$\hat{\theta}_{NT}^{(r)} = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \tilde{L}_{NT}(\theta, r)$$

for each r in a predetermined range $[\underline{r}, \bar{r}] \subset \mathbb{Z}_+$.

2. Find

$$\hat{r}_{NT} = \underset{r \in [\underline{r}, \bar{r}]}{\text{argmax}} \, \tilde{L}_{NT}(\hat{\theta}_{NT}^{(r)}, r).$$

Then $\hat{\nu}_{NT} = \left(\hat{\theta}_{NT}^{(\hat{r}_{NT})'}, \hat{r}_{NT}\right)'$ would be the optimizer of $\tilde{L}_{NT}(\nu)$.

Assumption 5.3.1 is a regularity condition on the parameter space. Assumptions 5.3.2 and 5.3.3 are necessary for obtaining η-weak dependence of $\{l_{it}(\nu) : (i,t) \in D_{NT}, NT \geq 1\}$. Then the consistency of MLE in Theorem 5.2 could be proved based on the LLN of η-weakly dependent arrays of random fields in Theorem 3.1.

Assumption 5.3.1. *The parameter space $\Theta \times \mathbb{Z}_+$ satisfies:*

(a). Θ is compact and θ_0 is an interior point of Θ;

(b). For any $\theta \in \Theta$, the conditions in Theorem 5.1 are satisfied.

Assumption 5.3.2. *(a). $\sup_{NT \geq 1} \sup_{(i,t) \in D_{NT}} \|y_{it}\|_{2p} < \infty$ for some $p > 1$;*

(b). The array of random fields $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is η-weakly dependent with coefficients $\bar{\eta}_y(r) := \mathcal{O}(r^{-\mu_y})$ for some $\mu_y > 2\frac{2p-1}{p-1}$.

Assumption 5.3.3. *For any $i = 1, 2, ..., N$ and $j = 1, 2, ..., N$, there exist constants $C > 0$ and $b > \mu_y$ such that $w_{ij} \leq C|j - i|^{-b}$. That is, the power of connection between two nodes i and j decays as $|i - j|$ grows.*

Theorem 5.2. *If Assumptions 5.3.1, 5.3.2 and 5.3.3 are satisfied, then the MLE defined by (5.3.4) is consistent:*

$$\hat{\nu}_{NT} \xrightarrow{p} \nu_0$$

as $T \to \infty$ and $N \to \infty$.

Since $\hat{r}_{NT}$ is an integer-valued consistent estimate of r_0, $\hat{r}_{NT}$ will eventually be equal to r_0 when the sample size NT becomes sufficiently large. Therefore, $\hat{\nu}_{NT} = \left(\hat{\theta}_{NT}^{(\hat{r}_{NT})'}, \hat{r}_{NT}\right)'$ is asymptotically equal to $\left(\hat{\theta}_{NT}^{(r_0)'}, r_0\right)'$. In this way, the problem of investigating the asymptotic distribution of $\hat{\nu}_{NT}$ degenerates to investigating the asymptotic distribution of $\hat{\theta}_{NT}^{(r_0)}$.

Theorem 5.3. *Assume that all conditions in Theorem 5.2 are satisfied, with* $\mu_y > \frac{6p-3}{p-1} \vee \frac{(4p-3)(2p-1)}{2(p-1)^2}$ *in Assumption 5.3.2(b) instead. If the smallest eigenvalue* $\lambda_{min}(\Sigma_{NT})$ *of*

$$\Sigma_{NT} := \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\frac{1}{\lambda_{it}(\nu_0)} \frac{\partial \lambda_{it}(\nu_0)}{\partial \theta} \frac{\partial \lambda_{it}(\nu_0)}{\partial \theta'}\right]$$

satisfies that

$$\inf_{NT\geq 1} \lambda_{min}(\Sigma_{NT}) > 0, \tag{5.3.5}$$

then $\hat{\theta}_{NT}^{(r_0)}$ *is asymptotically normal, i.e.*

$$\sqrt{NT}\Sigma_{NT}^{1/2}(\hat{\theta}_{NT}^{(r_0)} - \theta_0) \xrightarrow{d} N(0, I_5)$$

as $T \to \infty$*,* $N \to \infty$ *and* $N = o(T)$*.*

Remark. In the proof of Proposition 5.1 we will show that, Σ_{NT} could be consistently estimated by

$$\widehat{\Sigma}_{NT} = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\nu}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\nu}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\nu}_{NT})}{\partial \theta'}\right]$$

in practice.

Based on Theorem 5.2 and Theorem 5.3, for sufficiently large sample region such that $\hat{r}_{NT} = r_0$, we are able to design a Wald test with null hypothesis

$$H_0 : \Gamma\theta_0 = \eta, \tag{5.3.6}$$

where Γ is an $s \times 5$ matrix with rank s and η is an s-dimensional vector. For example, to test the existence of a threshold effect, simply let $\Gamma := (0, 1, -1, 0, 0)$ and $\eta := 0$, and the null hypothesis (5.3.6) becomes

$$H_0 : \alpha_0^{(1)} = \alpha_0^{(2)}.$$

Corresponding to the asymptotic normality of $\hat{\theta}_{NT}^{(r_0)}$ in Theorem 5.3, we define a

Wald test statistic as follows:

$$W_{NT} := (\Gamma\hat{\theta}_{NT}^{(r_0)} - \eta)' \left\{ \frac{\Gamma}{NT} \widehat{\Sigma}_{NT}^{-1} \Gamma' \right\}^{-1} (\Gamma\hat{\theta}_{NT}^{(r_0)} - \eta), \tag{5.3.7}$$

where

$$\widehat{\Sigma}_{NT} = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\nu}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\nu}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\nu}_{NT})}{\partial \theta'} \right].$$

And in Proposition 5.1 below, W_{NT} is proved to have an asymptotic χ^2-distribution with s degrees of freedom.

Proposition 5.1. *Under the same assumptions required by Theorem 5.3, as $T \to \infty$, $N \to \infty$ and $N = o(T)$, the Wald test statistic defined in (5.3.7) asymptotically follows a χ^2 distribution with degree of freedom s, i.e.*

$$W_{NT} \xrightarrow{d} \chi_s^2.$$

5.4 Simulation study and empirical data analysis

5.4.1 Simulation study

Set the true parameters $\nu_0 = (0.5, 0.7, 0.6, 0.1, 0.1, 5)'$ of the data generating process (5.2.1). For the sample region $D_{NT} = \{(i,t) : i = 1, 2, ..., N; t = 1, 2, ..., T\}$, let T increase from 200 to 2000, while N also increases at relatively slower rates of $\mathcal{O}(\sqrt{T})$ and $\mathcal{O}(T/\log(T))$ respectively, as shown in the following table:

T	200	500	1000	2000
$N \approx \sqrt{T}$	14	22	31	44
$N \approx T/\log(T)$	37	80	144	263

For each network size N, the adjacency matrix A is simulated according to four different mechanisms in Example 4.5.1 to Example 4.5.4 in Section 4.5.

Remark. Particularly, in the empirical analysis we will study the dataset of car collisions across different neighbourhoods that are distributed on five boroughs of New York City. These boroughs are separated by rivers (except for Brooklyn and Queens), and neighbourhoods within the same borough are more likely to share a borderline while cross-borough connections are very rare. Therefore the network constructed with New York City neighbourhoods follows the block structure in Example 4.5.4 with $N = 20$ and $K = 5$.

Based on a simulated network, the data is generated according to (5.2.1), and the true parameters are estimated by the MLE (5.3.4). To monitor the finite performance of MLE, data generation and parameter estimation are repeated for $M = 1000$ times, for each combination of sample size (N, T). The m-th replication produces the estimates $\hat{\theta}_m = (\hat{\omega}_m, \hat{\alpha}_m^{(1)}, \hat{\alpha}_m^{(2)}, \hat{\xi}_m, \hat{\beta}_m)'$ and $\hat{r}_m$. Root-mean-square errors (RMSE) and coverage probabilities (CP) with different sample sizes and network simulation mechanisms, are reported in Tables 5.1 and 5.2; We also report the mean estimates of the threshold r_0 at the last columns of both tables.

	T	N	ω	$\alpha^{(1)}$	$\alpha^{(2)}$	ξ	β	$\bar{r}$
Example 4.5.1	200	14	0.0696 (0.94)	0.0203 (0.94)	0.0278 (0.93)	0.0170 (0.95)	0.0256 (0.93)	5.028
	500	22	0.0367 (0.96)	0.0100 (0.95)	0.0138 (0.95)	0.0101 (0.93)	0.0127 (0.95)	5
	1000	31	0.0238 (0.95)	0.0058 (0.95)	0.0081 (0.95)	0.0062 (0.97)	0.0074 (0.95)	5
	2000	44	0.0153 (0.95)	0.0035 (0.95)	0.0047 (0.95)	0.0041 (0.96)	0.0045 (0.95)	5
Example 4.5.2	200	14	0.0454 (0.95)	0.0200 (0.95)	0.0264 (0.94)	0.0119 (0.96)	0.0245 (0.94)	5.045
	500	22	0.0284 (0.95)	0.0101 (0.95)	0.0134 (0.95)	0.0072 (0.94)	0.0126 (0.95)	5.002
	1000	31	0.0162 (0.97)	0.0059 (0.96)	0.0077 (0.97)	0.0044 (0.94)	0.0074 (0.95)	5
	2000	44	0.0112 (0.96)	0.0034 (0.96)	0.0047 (0.95)	0.0029 (0.94)	0.0043 (0.96)	5
Example 4.5.3	200	14	0.0511 (0.96)	0.0200 (0.95)	0.0272 (0.94)	0.0131 (0.95)	0.0246 (0.95)	5.034
	500	22	0.0349 (0.95)	0.0102 (0.95)	0.0135 (0.96)	0.0084 (0.95)	0.0127 (0.96)	5.001
	1000	31	0.0146 (0.95)	0.0060 (0.95)	0.0079 (0.95)	0.0038 (0.95)	0.0077 (0.94)	5
	2000	44	0.0104 (0.95)	0.0035 (0.95)	0.0048 (0.94)	0.0025 (0.95)	0.0043 (0.96)	5
Example 4.5.4	200	14	0.0882 (0.95)	0.0205 (0.95)	0.0273 (0.95)	0.0227 (0.94)	0.0256 (0.93)	5.013
	500	22	0.0379 (0.94)	0.0102 (0.95)	0.0136 (0.95)	0.0096 (0.95)	0.0124 (0.95)	5
	1000	31	0.0218 (0.95)	0.0060 (0.95)	0.0078 (0.95)	0.0055 (0.95)	0.0073 (0.96)	5
	2000	44	0.0118 (0.94)	0.0035 (0.96)	0.0047 (0.95)	0.0029 (0.95)	0.0043 (0.96)	5

Table 5.1: Simulation results with different network structures ($N \approx \sqrt{T}$).

From Tables 5.1 and 5.2 we can tell, that the RMSEs of $\hat{\theta}_{NT}$ decrease asymptotically toward zero, and the mean of $\hat{r}_{NT}$ is equal to $r_0 = 5$ for sufficiently large sample size. These results support the consistency of MLE (5.3.4) in Theorem 5.2. The reported CPs are close to the value 0.95, showing that $\widehat{SE}$ provides a reliable estimation of the true standard error of $\hat{\theta}_{NT}$. Moreover, in Figures 5.1 to 5.4 we draw the normal Q-Q plots for the estimation results when $T = 2000, N = 44$ and $T = 2000, N = 263$ respectively, under different network structures. These Q-Q plots provide additional

	T	N	ω	$\alpha^{(1)}$	$\alpha^{(2)}$	ξ	β	$\bar{r}$
Example 4.5.1	200	37	0.0537 (0.95)	0.0124 (0.95)	0.0164 (0.95)	0.0143 (0.94)	0.0158 (0.94)	5.002
	500	80	0.0287 (0.96)	0.0054 (0.94)	0.0071 (0.95)	0.0078 (0.95)	0.0066 (0.95)	5
	1000	144	0.0201 (0.95)	0.0029 (0.94)	0.0040 (0.93)	0.0055 (0.95)	0.0036 (0.94)	5
	2000	263	0.0136 (0.95)	0.0015 (0.94)	0.0019 (0.95)	0.0038 (0.95)	0.0019 (0.93)	5
Example 4.5.2	200	37	0.0347 (0.95)	0.0121 (0.95)	0.0170 (0.95)	0.0089 (0.95)	0.0161 (0.93)	5.008
	500	80	0.0140 (0.95)	0.0053 (0.95)	0.0070 (0.95)	0.0035 (0.95)	0.0066 (0.95)	5
	1000	144	0.0073 (0.95)	0.0029 (0.93)	0.0036 (0.95)	0.0020 (0.94)	0.0036 (0.93)	5
	2000	263	0.0041 (0.95)	0.0014 (0.95)	0.0020 (0.94)	0.0011 (0.95)	0.0018 (0.96)	5
Example 4.5.3	200	37	0.0385 (0.95)	0.0124 (0.94)	0.0168 (0.95)	0.0092 (0.95)	0.0152 (0.95)	5.003
	500	80	0.0144 (0.95)	0.0054 (0.95)	0.0071 (0.94)	0.0036 (0.95)	0.0067 (0.95)	5
	1000	144	0.0073 (0.94)	0.0029 (0.94)	0.0035 (0.96)	0.0019 (0.94)	0.0035 (0.95)	5
	2000	263	0.0037 (0.95)	0.0015 (0.95)	0.0019 (0.96)	0.0009 (0.95)	0.0018 (0.95)	5
Example 4.5.4	200	37	0.0498 (0.95)	0.0120 (0.95)	0.0165 (0.94)	0.0129 (0.94)	0.0148 (0.96)	5.011
	500	80	0.0176 (0.94)	0.0055 (0.94)	0.0071 (0.94)	0.0045 (0.94)	0.0069 (0.94)	5
	1000	144	0.0083 (0.97)	0.0028 (0.95)	0.0036 (0.96)	0.0022 (0.96)	0.0034 (0.95)	5
	2000	263	0.0048 (0.95)	0.0015 (0.95)	0.0019 (0.95)	0.0012 (0.96)	0.0019 (0.95)	5

Table 5.2: Simulation results with different network structures ($N \approx T/\log(T)$).

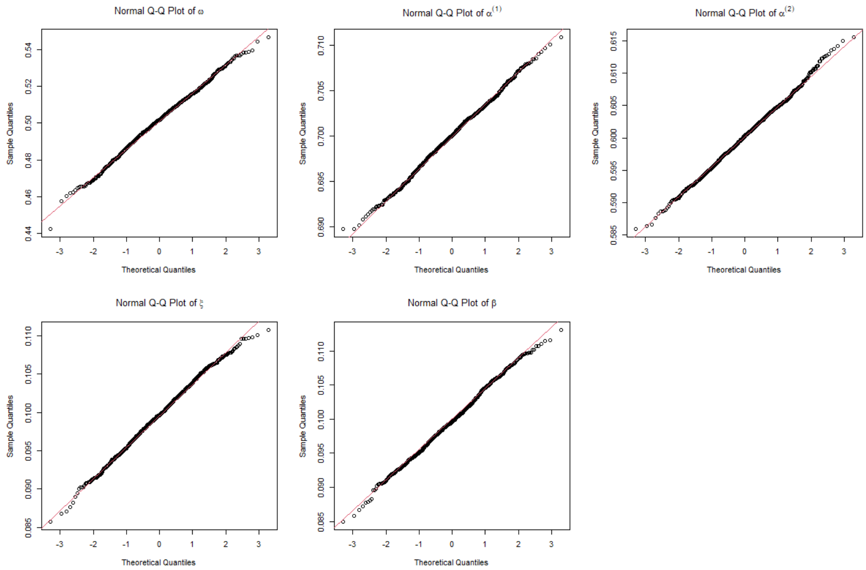
evidence for the asymptotic normality of $\hat{\theta}_{NT}$ in Theorem 5.3.

5.4.2 Analysis of daily numbers of car accidents in New York City

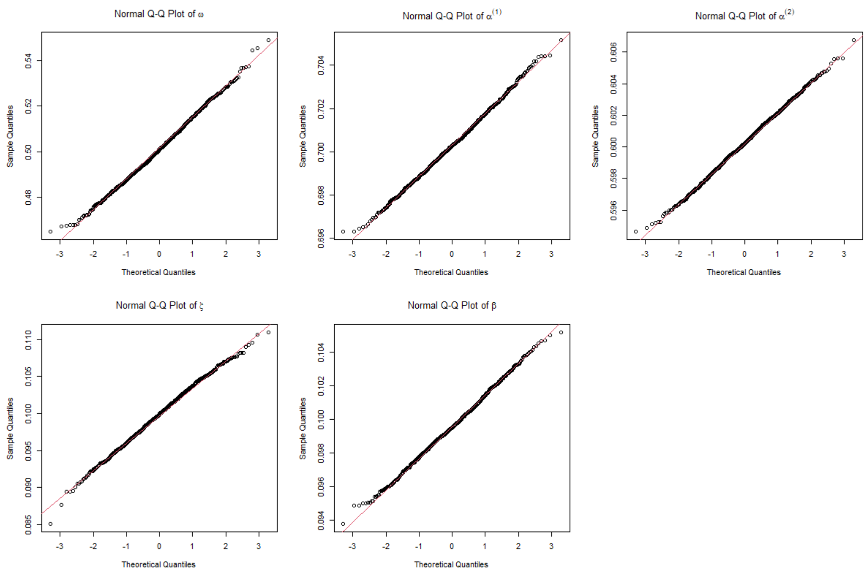
New York City Police Department (NYPD) publishes and regularly updates the detailed data of motor vehicle collisions that have occurred city-wide. These data are openly accessible on the NYPD website [1] and contain sufficient information for us to apply our model. We collect all records from February 16th 2021 to June 30th 2022, each record includes the date when an accident happened, and the zip code of where it happened. We classified all records into 41 neighbourhoods according to the correspondence between zip codes and the geometric locations they represent. Re-grouping the data by neighbourhoods and the date of occurrence, we obtain a high-dimensional time series with $N = 41$ and $T = 500$.

Two neighbourhoods are regarded as connected nodes if they share a borderline. Base on the geometric information, we are able to construct a reasonable network with 41 nodes, which is visualized in Figure 5.5. In Figure 5.6 we plot histograms of daily numbers of car accidents in 9 randomly selected neighbourhoods. The shapes of the histograms of sampled data show potential Poisson distribution. Moreover, in Figure 5.7 we could easily observe volatility clustering in the daily numbers of car accident in

[1] https://www1.nyc.gov/site/nypd/stats/traffic-data/traffic-data-collision.page

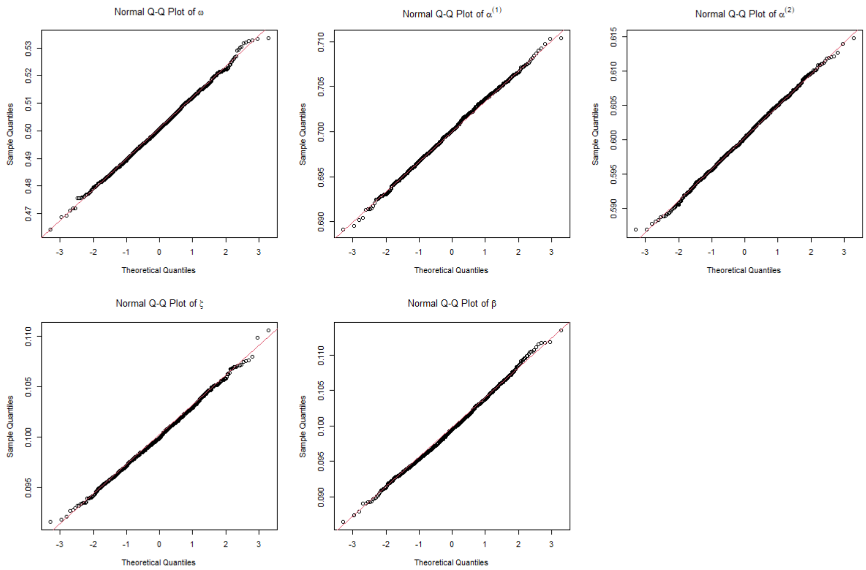


(a). $T = 2000, N = 44$

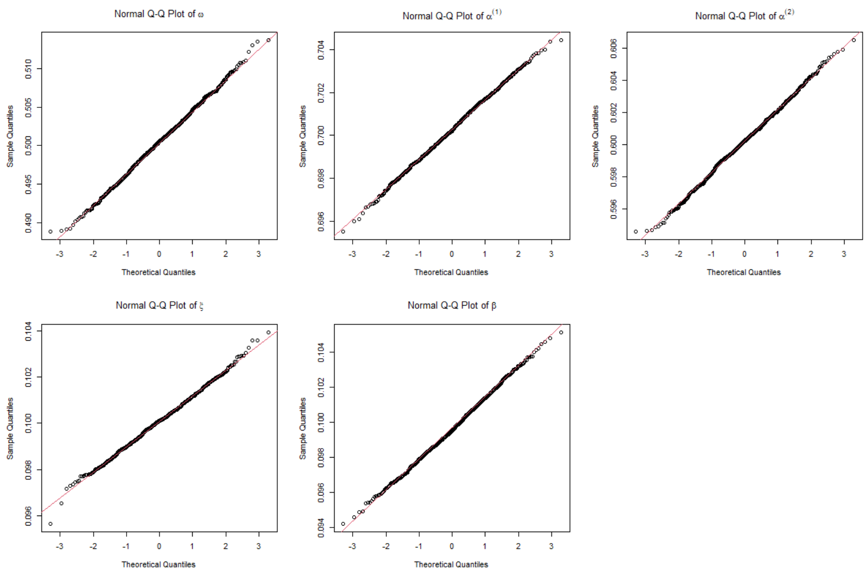


(b). $T = 2000, N = 263$

Figure 5.1: Q-Q plots of estimates for Example 4.5.1.

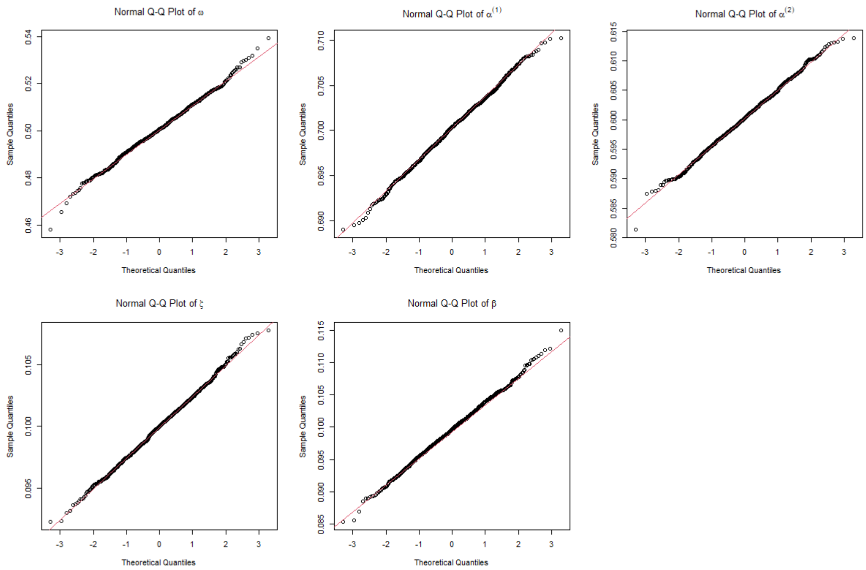


(a). $T = 2000, N = 44$

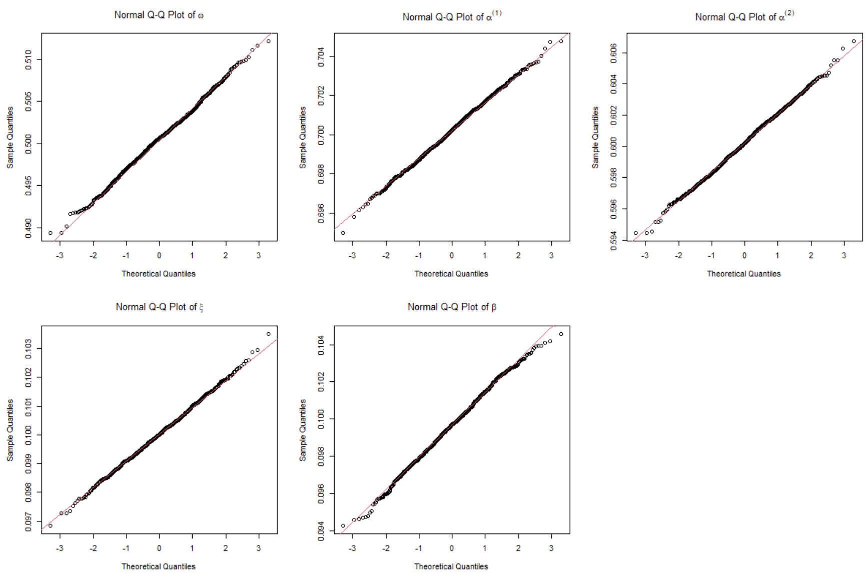


(b). $T = 2000, N = 263$

Figure 5.2: Q-Q plots of estimates for Example 4.5.2.

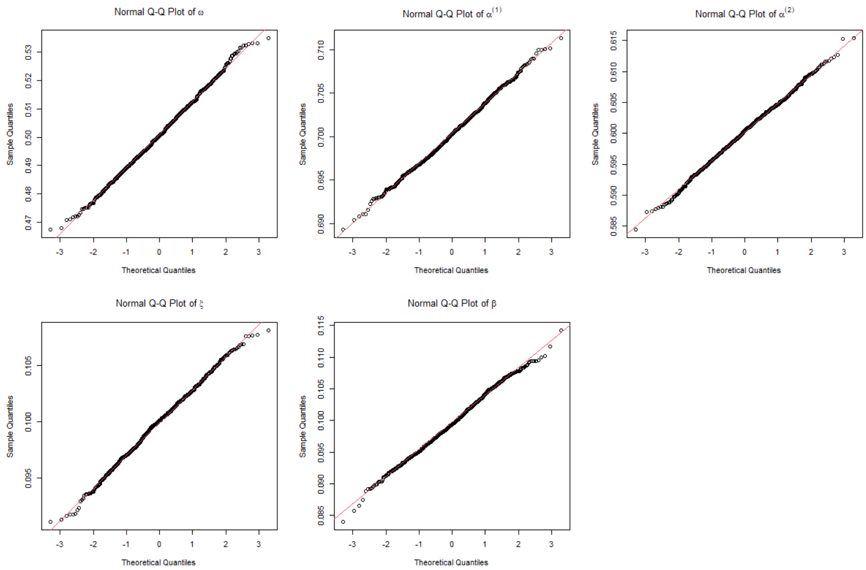


(a). $T = 2000, N = 44$

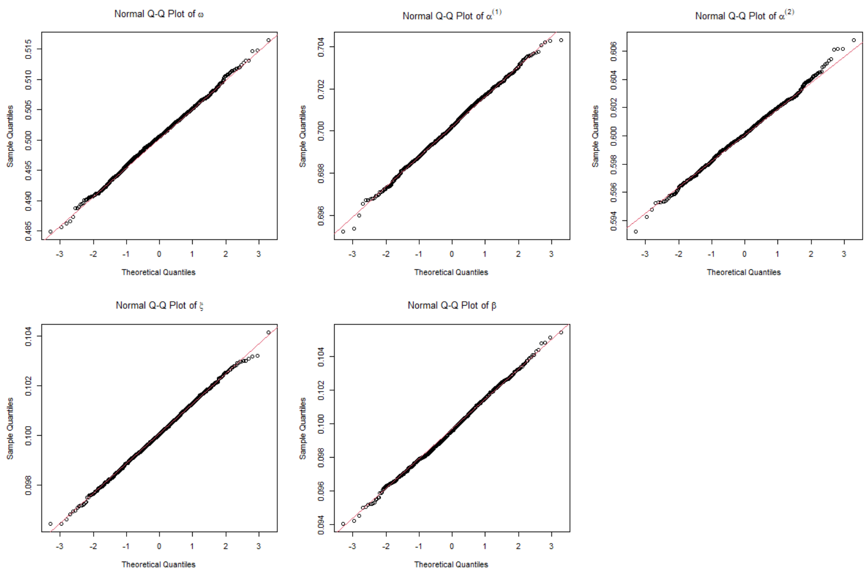


(b). $T = 2000, N = 263$

Figure 5.3: Q-Q plots of estimates for Example 4.5.3.



(a). $T = 2000, N = 44$



(b). $T = 2000, N = 263$

Figure 5.4: Q-Q plots of estimates for Example 4.5.4.

four selected neighbourhoods of NYC, indicating potential autoregressive structure in the conditional heteroscedasticity of the data.

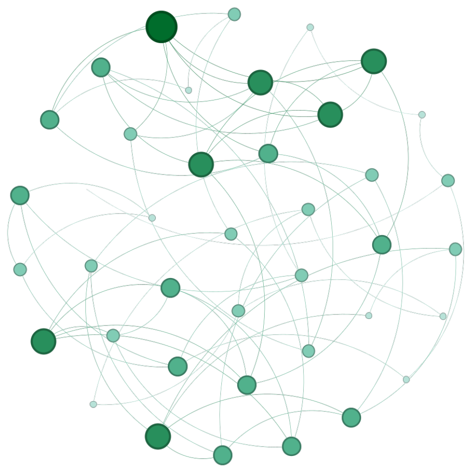


Figure 5.5: Network of 41 neighbourhoods in New York City

The estimation results are reported in Table 5.3. Firstly, it is worthy of note that $\alpha^{(1)}$ is slightly smaller than $\alpha^{(2)}$, which means that the conditional variance of the number of car accidents in these neighbourhoods are less affected by the number on the previous day if it is above the threshold $r = 10$. Secondly, the volatility in the number of car accidents in one area is also affected by its geometrically neighboured areas. Besides, the estimated value of β is significantly larger than other coefficients, indicating a strong persistence in volatility that leads to volatility clustering. Moreover, we utilize the Wald test to further investigate the existence of threshold effect. Let $\Gamma := (0, 1, -1, 0, 0)$ and $\eta := 0$ in (5.3.6), then the null hypothesis becomes

$$H_0 : \alpha_0^{(1)} = \alpha_0^{(2)}.$$

The Wald statistic (5.3.7) $W_{NT} = 18.94$, which suggests the rejection of H_0 at significant level below 0.01 according to Proposition 5.1.

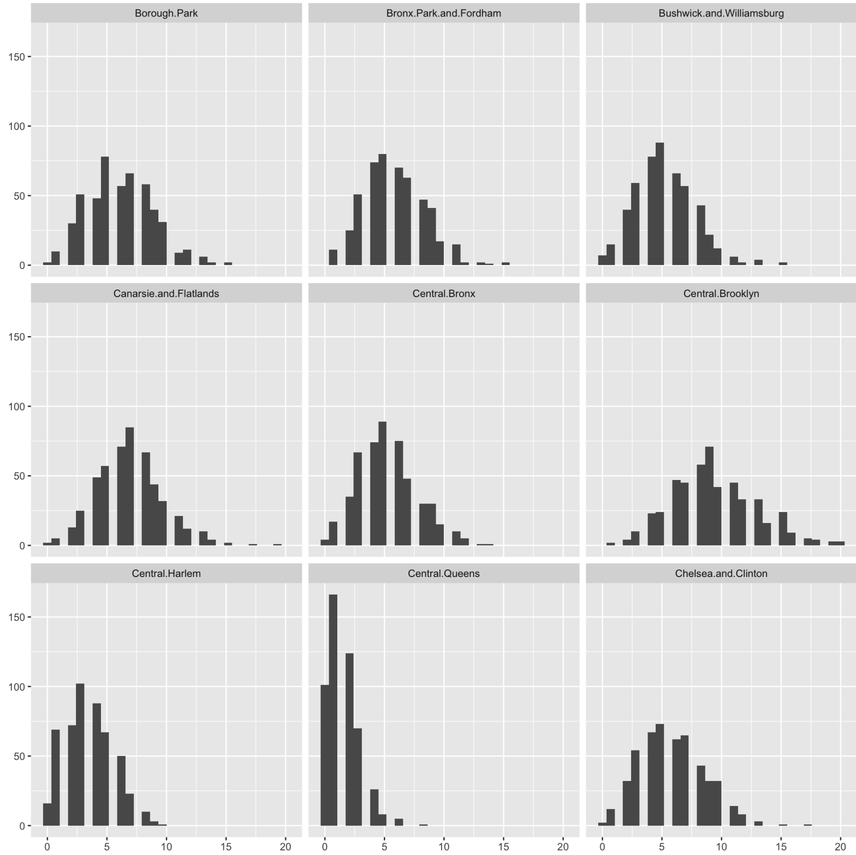


Figure 5.6: Distributions of daily occurrences of car accidents in selected neighbourhoods.

	ω	$\alpha^{(1)}$	$\alpha^{(2)}$	ξ	β	r
Estimation	0.018693	0.126472	0.135026	0.002727	0.862244	10
SE	4.12e-03	4.40e-03	4.68e-03	1.09e-03	4.73e-03	\

Table 5.3: Estimation results based on daily number of car accidents in 41 neighbourhoods of NYC.

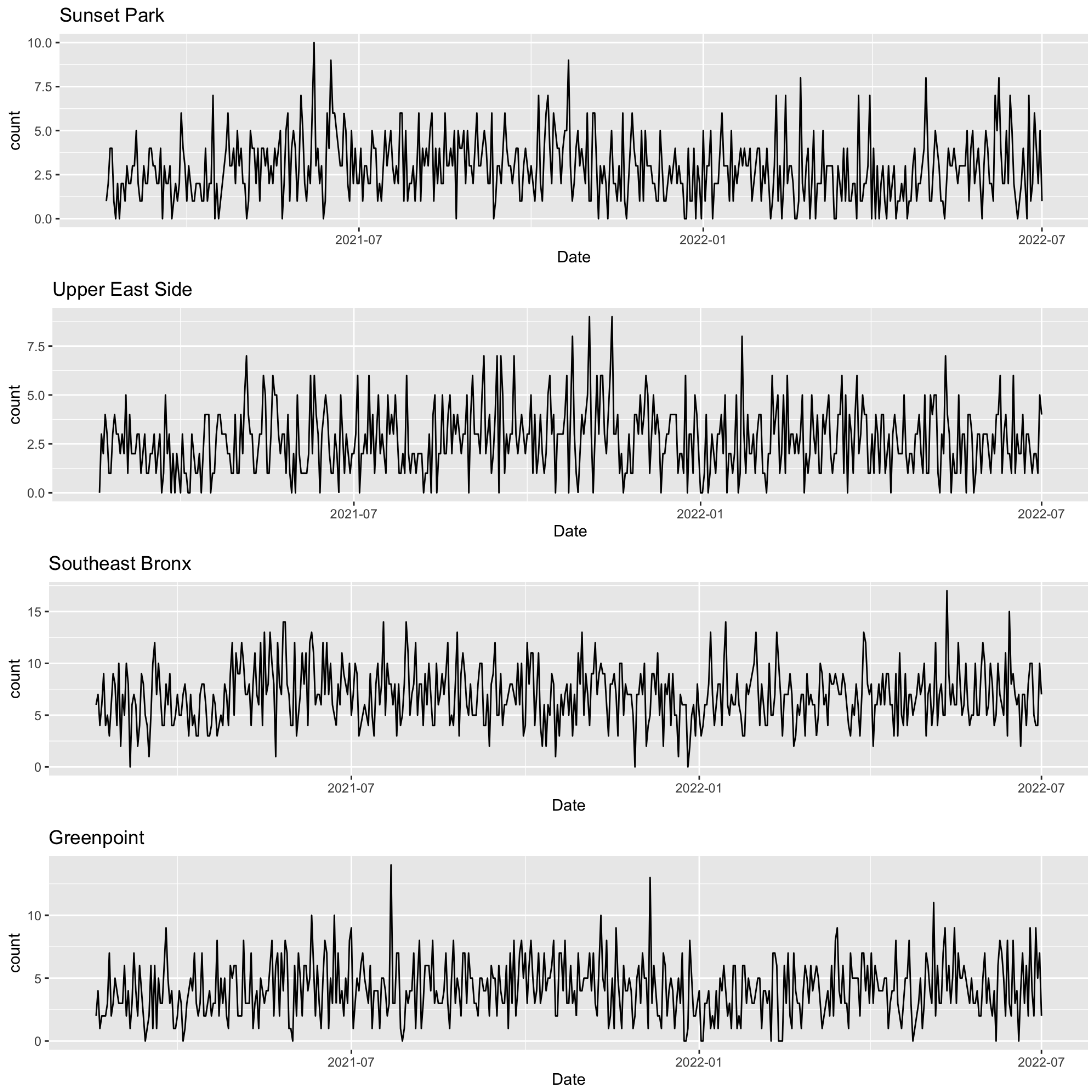


Figure 5.7: Daily occurrences of car accidents in 4 neighbourhoods.

Chapter 6

Network GARCH Models in the One-Parameter Exponential Family

6.1 Introduction

Corresponding to different application scenarios, GARCH-type models in existing literature are designed with different conditional distributions and iterative structures that drive the hidden conditional mean (or variance) processes. Davis and Liu (2016) established general theory and inference for a class of univariate GARCH models that have conditional distributions belonging to the one-parameter exponential family and the conditional means defined through (linear or non-linear) iterated random functions of their lagged values and past observations. In this chapter, we will consider high-dimensional GARCH models with conditional distributions in the one-parameter exponential family. As we have mentioned, one major challenge faced by multi-dimensional GARCH models is that the number of parameters increases as the spatial dimension expands, causing problems in establishing a feasible estimation method. As we have done in preceding chapters, to circumvent this problem, we suppose that the simultaneously observed individuals are connected through a network structure, and each conditional mean (or variance) is driven by a weighted average lag-1 values from its neighbours on the network (see (6.2.2)). This idea of utilizing a large scale network is popular in the literature of high-dimensional time series, starting from Zhu et al. (2017)'s network AR model, followed by other AR-type models (Xu et al., 2024) and GARCH-type models (Tao et al., 2024; Armillotta and Fokianos, 2024; Pan and Pan, 2024).

This chapter is organized as follows: In Section 6.3 the stationarity shall be discussed under a fixed-dimension setting, utilizing the method of geometric moment contraction established by Wu and Shao (2004) for Markov chains driven by iterated random functions. Then in Section 6.4 we will establish maximum likelihood estimation that is consistent and asymptotically normal under increasing size of temporal and spatial dimensions, facilitated by the limit theorems of weakly dependent random fields in Chapter 3. As far as we know, among all the studies on high-dimensional GARCH-type models, only Pan and Pan (2024) and Armillotta and Fokianos (2024) consider the case of increasing size of spatial dimension. However, the threshold network GARCH model of Pan and Pan (2024) is limited to continuous data and the Poisson network autoregression of Armillotta and Fokianos (2024) has a simple ARCH-type conditional intensity process. Our methodology accommodates both continuous and integer-valued data, and it is feasible under non-linear structures in the conditional mean process. In Section 6.5, we will test our methodology on a new negative binomial threshold network GARCH model, with simulation studies and real data analysis carried out as well.

6.2 Network GARCH in one-parameter exponential family

Adopting the settings in preceding chapters, we consider an non-directed and weightless network with N nodes, represented by adjacency matrix A and corresponding row-normalized adjacency matrix W. For any node i in this network, let y_{it} be an non-negative observation at time t, and $\mathcal{H}_{t-1}$ denote the σ-algebra consisting of all available information up to $t-1$. In this chapter, we assume that the conditional distribution of $y_{it}|\mathcal{H}_{t-1}$ belongs to the one-parameter exponential family (OPE) with parameter η_{it}.

For any random variable y adapts to this family of distribution, the probability density function of y follows the form

$$f(y|\eta) = h(y) \exp\{\eta y - \mathcal{A}(\eta)\}, \qquad y \geq 0. \tag{6.2.1}$$

η is called the natural parameter. The function $h(\cdot)$ is non-negative and independent from η. The first order derivative of function $\mathcal{A}(\cdot)$ exists and $\mathcal{B}(\cdot) := \mathcal{A}'(\cdot)$; Both functions in (6.2.1) are known. Based on the density function (6.2.1) we have the conditional mean $\mathbb{E}(y|\eta) = \mathcal{B}(\eta)$ and the conditional variance $\mathrm{Var}(y|\eta) = \mathcal{B}'(\eta)$.

Remark. Since we assume that $y \geq 0$ in (6.2.1) throughout this chapter, then $\mathcal{A}$ is a strictly increasing function since $\mathcal{A}'(\eta) = \mathcal{B}(\eta) = \mathbb{E}(y|\eta) > 0$. Moreover, to ensure that $\mathrm{Var}(y|\eta) = \mathcal{B}'(\eta) > 0$, $\mathcal{B}$ is also assumed to be strictly increasing.

An NGARCH-OPE(1,1) model has following form:

$$\begin{aligned} y_{it}|\mathcal{H}_{t-1} &\sim OPE(\eta_{it}), \\ \mu_{it} &= g_\theta\left(y_{i,t-1}, \sum_{j=1}^{N} w_{ij} y_{j,t-1}, \mu_{i,t-1}\right) \end{aligned} \tag{6.2.2}$$

for $i = 1, 2, ..., N$. Similar to the settings of TNGARCH and PTNGARCH in previous chapters, we assume that the conditional distributions $y_{it}|\mathcal{H}_{t-1}$ are independent for $i = 1, 2, ..., N$ and $t \in \mathbb{Z}$. Characterized by θ-parameterized function $g_\theta(\cdot)$, the conditional mean process $\mu_{it} := \mathcal{B}(\eta_{it})$ has a GARCH-type structure with an extra network term. The parameter θ takes value that ensures the positiveness of the conditional mean process μ_{it}. With carefully specified OPE and g_θ, model (6.2.2) covers a large class of $\mathbb{R}$-valued and integer-valued network GARCH, as well as univariate GARCH-type models ($N = 1$). We will give three examples below.

Example 6.2.1. To model unbounded non-negative integer-valued data, we can set OPE as a Poisson distribution with parameter λ, (6.2.1) becomes

$$f(y|\eta) = \frac{1}{y!} \exp\{\eta y - e^\eta\}, \tag{6.2.3}$$

with $\eta = \log(\lambda)$, $\mathcal{A}(\eta) = e^\eta$ and $\mathcal{B}(\eta) = e^\eta$. The grouped network Poisson AR (Tao et al., 2024) and Poisson network AR (Armillotta and Fokianos, 2024) are special cases of (6.2.2).

Example 6.2.2. We can also model bounded non-negative integer-valued data, by setting OPE as a binomial distribution with number of trials n and probability of success p, then (6.2.1) becomes

$$f(y|\eta) = \binom{n}{y} \frac{e^{\eta y}}{(1+e^\eta)^n}, \tag{6.2.4}$$

with $\eta = \log \frac{p}{1-p}$, $\mathcal{A}(\eta) = n \log(1+e^\eta)$ and $\mathcal{B}(\eta) = n\left(1 - \frac{1}{1+e^\eta}\right)$. Hence (6.2.2) also covers the binomial ARCH of Ristić et al. (2016).

Example 6.2.3. For non-negative integer-valued data with over-dispersion, we could set OPE as a negative binomial distribution with number of success K and probability of success p, then (6.2.1) becomes

$$f(y|\eta) = \binom{y+K-1}{y} (1-e^\eta)^K e^{\eta y}, \tag{6.2.5}$$

with $\eta = \log(1-p)$, $\mathcal{A}(\eta) = -K\log(1-e^\eta)$ and $\mathcal{B}(\eta) = \frac{Ke^\eta}{1-e^\eta}$. The negative binomial GARCH of Zhu (2010) is a univariate special case of (6.2.2).

6.3 Stationarity under fixed N

Let F_μ be the cumulative distribution function of OPE in (6.2.2) with $\mu = \mathcal{B}(\eta)$ and its inverse $F_\mu^{-1}(u) := \inf\{q \geq 0 : F_\mu(q) \geq u\}$ for any $u \in (0,1)$. With the N-dimensional vectors $\mathbf{x} := (x_1, x_2, ..., x_N)$ and $\mathbf{u} := (u_1, u_2, ..., u_N)'$ where $u_i \in (0,1)$ for $i = 1, 2, ..., N$, we could define an N-dimensional function

$$\mathbb{G}(\mathbf{x}, \mathbf{u}) := \begin{pmatrix} g_{\theta_0}\left(F_{x_1}^{-1}(u_1), \sum_{j=1}^N w_{1j} F_{x_j}^{-1}(u_j), x_1\right) \\ g_{\theta_0}\left(F_{x_2}^{-1}(u_2), \sum_{j=1}^N w_{2j} F_{x_j}^{-1}(u_j), x_2\right) \\ \cdots \\ g_{\theta_0}\left(F_{x_N}^{-1}(u_N), \sum_{j=1}^N w_{Nj} F_{x_j}^{-1}(u_j), x_N\right) \end{pmatrix}. \tag{6.3.1}$$

For all i and t, let U_{it}'s be independent and identically distributed (IID) random variables that follow uniform distribution on $[0, 1]$. With

$$\mathbb{X}_t := (\mu_{1t}, \mu_{2t}, ..., \mu_{Nt})',$$
$$\mathbb{Y}_t := (y_{1t}, y_{2t}, ..., y_{Nt})',$$
$$\mathbb{U}_t := (U_{1t}, U_{2t}, ..., U_{Nt})',$$

we could define an N-dimensional Markov chain $\{\mathbb{X}_t\}$ based on model (6.2.2) as follows:

$$\mathbb{X}_t = \mathbb{G}(\mathbb{X}_{t-1}, \mathbb{U}_t). \tag{6.3.2}$$

The Markov chain (6.3.2) could be regarded as an iterated random function (IRF) system, where the random function $\mathbb{G}_{\mathbb{U}_t}(\cdot) := \mathbb{G}(\cdot, \mathbb{U}_t)$ is defined on a complete and separable metric space $(\mathcal{X}, |\cdot|_\infty)$ with $\mathcal{X} := \mathbb{R}_+^N$, and the $\mathbb{U}_t$'s are IID random vectors that take values in another measurable space $[0, 1]^N$. Therefore the stationarity of (6.3.2) could be investigated using the methods of Wu and Shao (2004), who established convergence of IRF to its stationary distribution in the sense of geometric moment contraction (GMC).

For any starting point $\mathbb{X}_0 = \mathbf{x} \in \mathcal{X}$ of (6.3.2), we can define an process $\{\mathbb{X}_t(\mathbf{x}) : t \geq 0\}$ as

$$\mathbb{X}_t(\mathbf{x}) := \mathbb{G}_{\mathbb{U}_t} \circ \mathbb{G}_{\mathbb{U}_{t-1}} \circ ... \circ \mathbb{G}_{\mathbb{U}_1}(\mathbf{x}).$$

If the stationary distribution of (6.3.2) exists and is denoted by π, then (6.3.2) could also be represented by $\mathbb{X}_t(\mathbf{x})$ if $\mathbf{x} \sim \pi$. Let $\mathbf{x}' \in \mathcal{X}$ be another π-distributed starting point that is independent from $\mathbf{x}$. According to (2) in Wu and Shao (2004), the process (6.3.2) is said to be geometric moment contracting if there exist constants $\alpha > 0$, $C = C(\alpha) > 0$ and $\rho = \rho(\alpha) \in (0, 1)$ such that

$$\mathbb{E} \left|\mathbb{X}_t(\mathbf{x}) - \mathbb{X}_t(\mathbf{x}')\right|_\infty^\alpha \leq C\rho^t \tag{6.3.3}$$

for all $t \in \mathbb{N}$. Similar to $\mathbb{X}_t(\mathbf{x})$, we define a backward iteration process

$$\mathbb{Z}_t(\mathbf{x}) := \mathbb{G}_{\mathbb{U}_1} \circ \mathbb{G}_{\mathbb{U}_2} \circ ... \circ \mathbb{G}_{\mathbb{U}_t}(\mathbf{x}),$$

which has the same distribution with $\mathbb{X}_t(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$ since $\mathbb{U}_1, \mathbb{U}_2, ...\mathbb{U}_t$ are IID. Therefore, if there exists a random vector $\mathbb{Z}_\infty$ such that $\mathbb{Z}_t(\mathbf{x}) \stackrel{a.s.}{\to} \mathbb{Z}_\infty$ for all $\mathbf{x} \in \mathcal{X}$, then $\mathbb{X}_t(\mathbf{x}) \stackrel{d}{\to} \mathbb{Z}_\infty$.

Assumption 6.3.1. *Let $\mathbb{S}_0$ be the range of $\left(y_{it}, \sum_{j=1}^N w_{ij} y_{jt}, \mu_{it}\right)$ for all $(i,t) \in D_{NT}, NT \geq 1$. Then for any (a,b,c) and (a',b',c') in $\mathbb{S}_0$,*

$$|g_\theta(a,b,c) - g_\theta(a',b',c')| \leq \rho_1|a-a'| + \rho_2|b-b'| + \rho_3|c-c'|, \tag{6.3.4}$$

where the constants ρ_1, ρ_2, ρ_3 are non-negative and $\rho_1 + \rho_2 + \rho_3 < 1$. Moreover, $g_\theta(0,0,0) < \infty$.

Theorem 6.1. *With Assumption 6.3.1, the following statements hold for the process (6.3.2):*

(a). There exists a random vector $\mathbb{Z}_\infty$ such that, for all $\mathbf{x} \in \mathcal{X}$, $\mathbb{Z}_t(\mathbf{x}) \stackrel{a.s.}{\to} \mathbb{Z}_\infty$. $\mathbb{Z}_\infty$ does not depend on $\mathbf{x}$ and follows distribution π, which is the stationary distribution of (6.3.2).

(b). The Markov chain (6.3.2) is geometric moment contracting with unique stationary distribution π, and $\mathbb{E}_\pi \|\mathbb{X}_t\| < \infty$.

6.4 Maximum likelihood estimation

Based on a series of samples $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ from (6.2.2), we will investigate the consistency and asymptotic normality of MLE as the size of the sample region $NT \to \infty$. Assume that the model of interest is characterized by an array of k parameters θ in a parameter space Θ that is a compact subset of $\mathbb{R}^k$, and the true parameter $\theta_0 \in \Theta$. Based on samples $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$, we could construct a log likelihood function in the form

$$L_{NT}(\theta) = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} l_{it}(\theta) = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\eta_{it}(\theta) y_{it} - \mathcal{A}(\eta_{it}(\theta))\right], \tag{6.4.1}$$

where $\eta_{it}(\theta) = \mathcal{B}^{-1}(\mu_{it}(\theta))$, and $\mu_{it}(\theta)$ is obtained through iteration

$$\mu_{it}(\theta) = g_\theta \left(y_{i,t-1}, \sum_{j=1}^{N} w_{ij} y_{j,t-1}, \mu_{i,t-1}(\theta) \right).$$

The exact value of (6.4.1) cannot be calculated solely depending on the samples, since the starting values $\mu_{i0}(\theta)$ for $i = 1, 2, ..., N$ are not observable. In practice, the the estimate of θ_0 is often obtained through an approximation of (6.4.1), i.e.

$$\hat{\theta}_{NT} = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \tilde{L}_{NT}(\theta), \tag{6.4.2}$$

where the approximated likelihood is

$$\tilde{L}_{NT}(\theta) = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \tilde{l}_{it}(\theta) = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\tilde{\eta}_{it}(\theta) y_{it} - \mathcal{A}(\tilde{\eta}_{it}(\theta))\right], \tag{6.4.3}$$

with $\tilde{\eta}_{it}(\theta) = \mathcal{B}^{-1}(\tilde{\mu}_{it}(\theta))$, and $\tilde{\mu}_{it}(\theta)$ being obtained through iteration

$$\tilde{\mu}_{it}(\theta) = g_\theta \left(y_{i,t-1}, \sum_{j=1}^{N} w_{ij} y_{j,t-1}, \tilde{\mu}_{i,t-1}(\theta) \right),$$

with prior setting of initial values $\tilde{\mu}_{i0}$ for $i = 1, 2, ..., N$. We need Assumption 6.4.1 below regarding the convergence of the approximated likelihood (6.4.3) to the exact likelihood (6.4.1).

Assumption 6.4.1. *For any $\theta \in \Theta$, $|L_{NT}(\theta) - \tilde{L}_{NT}(\theta)| \xrightarrow{p} 0$ as $NT \to \infty$.*

To establish the consistency of our MLE, we firstly assume that the random fields of observations $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ have uniform bounded moments, and are weakly dependent.

Assumption 6.4.2. $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}} \mathbb{E}\,|y_{it}|^{2p} < \infty$ *for some* $p > 1$*, and the array of random fields* $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ *is* η*-weakly dependent with* $\bar{\eta}_y(r) = \mathcal{O}(r^{-\mu})$ *with* $\mu > \frac{3(2p-1)}{p-1} \vee \frac{(4p-3)(2p-1)}{2(p-1)^2}$.

The element (i,j) of W, w_{ij} measures the power of connection between any two nodes i and j in a network. In the following assumption we assume that w_{ij} decays as the distance $|i-j|$ grows.

Assumption 6.4.3. $w_{ij} \leq C|i-j|^{-\alpha}$ *for some constants* $C > 0$ *and* $\alpha \geq \frac{2(p-1)}{2p-1}\mu + 2$.

Comparing to the contracting assumption on g_θ in Assumption 6.3.1, in this section g_θ is only required to be partially contracting with respect to its third argument.

Assumption 6.4.4. *Let* $\mathbb{S}_0$ *be the range of* $\left(y_{it}, \sum_{j=1}^N w_{ij}y_{jt}, \mu_{it}\right)$ *for all* $(i,t) \in D_{NT}$*,* $NT \geq 1$*. Then for any* (a,b,c) *and* (a',b',c') *in* $\mathbb{S}_0$*,*

$$|g_\theta(a,b,c) - g_\theta(a',b',c')| \leq C_1|a-a'| + C_2|b-b'| + \rho|c-c'| \tag{6.4.4}$$

for some constants $C_1 > 0$*,* $C_2 > 0$ *and* $0 < \rho < 1$.

Facilitated by the assumptions above, in Lemma 6.4.1 below we show that the unobserved random fields $\{\mu_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ are also weakly dependent for any $\theta \in \Theta$.

Lemma 6.4.1. *If Assumptions 6.4.2, 6.4.3 and 6.4.4 are satisfied, then the* η*-coefficients of* $\{\mu_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ *satisfy*

$$\bar{\eta}^{(0)}_\mu(r) \leq Cr^{2-\mu}$$

for some constant $C > 0$*. And* $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\sup_{\theta\in\Theta} \mathbb{E}\,|\mu_{it}(\theta)|^{2p} < \infty$.

In the proof of consistency, we need to verify the convergence $L_{NT}(\theta) - \mathbb{E}L_{NT}(\theta) \xrightarrow{p} 0$, which requires the weak dependence of likelihood functions $\{l_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$. With the Lipschitz continuity assumption below, we obtain the weak dependence of $l_{it}(\theta)$'s in Lemma 6.4.2 thereafter.

Assumption 6.4.5. *Let $\mathbb{S}_\mu$ be the range of μ_{it} for all $(i,t) \in D_{NT}, NT \geq 1$, the functions $\mathcal{B}^{-1}$ and $\mathcal{A} \circ \mathcal{B}^{-1}$ are Lipschitz continuous on $\mathbb{S}_\mu$.*

Lemma 6.4.2. *Beside of all the conditions of Lemma 6.4.1, if Assumption 6.4.5 is also satisfied, then the η-coefficients of $\{l_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ satisfy*

$$\bar{\eta}_l^{(0)}(r) \leq Cr^{2-\frac{2(p-1)}{2p-1}\mu}$$

for some constant $C > 0$. And $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \sup_{\theta\in\Theta} \mathbb{E}\,|l_{it}(\theta)|^p < \infty$.

Assumption 6.4.6 below is required for the true parameters to be uniquely identifiable, i.e. θ_0 is the unique maximizer of $\mathbb{E}|L_{NT}(\theta)|$.

Assumption 6.4.6. *For any $\delta > 0$,*

$$\sup_{NT\geq 1} \sup_{\substack{\theta\in\Theta \\ \|\theta-\theta_0\|\geq\delta}} \{\mathbb{E}[L_{NT}(\theta)] - \mathbb{E}[L_{NT}(\theta_0)]\} < 0.$$

Lemma 6.4.2 allows us to use the LLN for weakly dependent random fields (Theorem 3.1 in Chapter 3), together with Assumption 6.4.1 and the identifiability Assumption 6.4.6 we can prove the consistency of $\hat{\theta}_{NT}$ as follows:

Theorem 6.2. *If Assumptions 6.4.1 to 6.4.6 are satisfied, then the MLE defined by (6.4.2) is consistent, that is*

$$\hat{\theta}_{NT} \xrightarrow{p} \theta_0$$

as $NT \to \infty$.

As for the asymptotic normality of $\hat{\theta}_{NT}$, we need additional assumptions on the approximation $\tilde{L}_{NT}(\theta)$ of $L_{NT}(\theta)$ as in Assumption 6.4.7 below.

Assumption 6.4.7. *As $NT \to \infty$:*

(a). $\sqrt{NT}\left\|\frac{\partial\tilde{L}(\theta_0)}{\partial\theta} - \frac{\partial L(\theta_0)}{\partial\theta}\right\| \xrightarrow{p} 0$;

(b). $\sup_{\|\theta-\theta_0\|<\xi}\left\|\frac{\partial^2\tilde{L}(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'}\right\| = O_p(\xi)$.

There are two essential parts in the proof of asymptotic normality of $\hat{\theta}_{NT}$: Firstly to establish the limit distribution of the score function

$$\sqrt{NT}V_{NT}^{-1/2}\frac{\partial L_{NT}(\theta_0)}{\partial\theta} \xrightarrow{d} \mathcal{N}(0, I_k), \tag{6.4.5}$$

where $V_{NT} = \text{Var}\left[\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\right]$. We also need to verify the convergence of the Hessian matrix $\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}$ to its expectation, i.e.

$$\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'} + H_{NT} \xrightarrow{p} 0, \tag{6.4.6}$$

where $H_{NT} = -\mathbb{E}\left[\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right]$.

Based on (6.2.2), we write

$$\frac{\partial\mu_{it}(\theta_0)}{\partial\theta} = g_{\theta_0}^{(1)}\left(y_{i,t-1}, \sum_{j=1}^{N} w_{ij}y_{j,t-1}, \mu_{i,t-1}(\theta_0), \frac{\partial\mu_{i,t-1}(\theta_0)}{\partial\theta}\right)$$

for some function $g_{\theta_0}^{(1)}$. Let $\mathbb{S}_1$ be the range of $\left(y_{it}, \sum_{j=1}^{N} w_{ij}y_{jt}, \mu_{it}(\theta_0), \frac{\partial\mu_{it}(\theta_0)}{\partial\theta}\right)$ for all $(i,t) \in D_{NT}, NT \geq 1$. In the following assumption, we assume that $g_{\theta_0}^{(1)}$ is Lipschitz continuous on $\mathbb{S}_1$ and partially contracting with respect to the fourth argument.

Assumption 6.4.8. *For any (a,b,c,d) and (a',b',c',d') in $\mathbb{S}_1$,*

$$\left\|g_{\theta_0}^{(1)}(a,b,c,d) - g_{\theta_0}^{(1)}(a',b',c',d')\right\| \leq C_1|a-a'|+C_2|b-b'|+C_3|c-c'|+\rho|d-d'| \tag{6.4.7}$$

for some constants $C_1 > 0$, $C_2 > 0$, $C_3 > 0$ and $0 < \rho < 1$.

Lemma 6.4.3. *Beside of all the conditions of Lemma 6.4.1, if Assumption 6.4.8 is also satisfied, then the η-coefficients of $\{\frac{\partial\mu_{it}(\theta_0)}{\partial\theta} : (i,t) \in D_{NT}, NT \geq 1\}$ satisfy*

$$\bar{\eta}_{\mu}^{(1)}(r) \leq Cr^{2-\mu}$$

for some constant $C > 0$. And $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\mathbb{E}\left|\frac{\partial\mu_{it}(\theta_0)}{\partial\theta}\right|^{2p} < \infty$.

Assumption 6.4.9. *Let $\mathbb{S}_\mu$ be the range of μ_{it} for all $(i,t) \in D_{NT}, NT \geq 1$, the functions $(\mathcal{B}^{-1})'$ and $(\mathcal{A}\circ\mathcal{B}^{-1})'$ are Lipschitz continuous on $\mathbb{S}_\mu$.*

Assumption 6.4.10. *Following bounds exist almost surely:*

$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{\partial\mu_{it}(\theta_0)}{\partial\theta}\right\|<\infty;$$
$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left|(\mathcal{B}^{-1})'(\mu_{it}(\theta_0))\right|<\infty;$$
$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left|(\mathcal{A}\circ\mathcal{B}^{-1})'(\mu_{it}(\theta_0))\right|<\infty.$$

Lemma 6.4.4. *Beside of all the conditions of Lemma 6.4.3, if Assumptions 6.4.9 and 6.4.10 are also satisfied, then the η-coefficients of $\left\{\frac{\partial l_{it}(\theta_0)}{\partial\theta}:(i,t)\in D_{NT}, NT\geq 1\right\}$ satisfy*

$$\bar{\eta}_l^{(1)}(r)\leq Cr^{2-\frac{2(p-1)}{2p-1}\mu}$$

for some constant $C>0$. And $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\mathbb{E}\left|\frac{\partial l_{it}(\theta_0)}{\partial\theta}\right|^{2p}<\infty$.

As for the second order derivative of $\mu_{it}(\theta_0)$ we have

$$\frac{\partial^2\mu_{it}(\theta_0)}{\partial\theta\partial\theta'}=g_{\theta_0}^{(2)}\left(y_{i,t-1},\sum_{j=1}^{N}w_{ij}y_{j,t-1},\mu_{i,t-1}(\theta_0),\frac{\partial\mu_{i,t-1}(\theta_0)}{\partial\theta},\frac{\partial^2\mu_{i,t-1}(\theta_0)}{\partial\theta\partial\theta'}\right)$$

for some function $g_{\theta_0}^{(2)}$. Analogous to Assumption 6.4.8, we assume that $g_{\theta_0}^{(2)}$ is partially contracting with respect to the fifth argument on $\mathbb{S}_2$, which denotes the range of inputs of $g_{\theta_0}^{(2)}$.

Assumption 6.4.11. *For any (a,b,c,d,e) and (a',b',c',d',e') in $\mathbb{S}_2$,*

$$\left\|g_{\theta_0}^{(2)}(a,b,c,d,e)-g_{\theta_0}^{(2)}(a',b',c',d',e')\right\|\leq C_1|a-a'|+C_2|b-b'|+C_3|c-c'|+C_4|d-d'|+\rho|e-e'|$$

for some constants $C_1>0$, $C_2>0$, $C_3>0$, $C_4>0$ and $0<\rho<1$.

Lemma 6.4.5. *Beside of all the conditions of Lemma 6.4.4, if Assumption 6.4.11 is also satisfied, then the η-coefficients of $\{\frac{\partial^2\mu_{it}(\theta_0)}{\partial\theta\partial\theta'}:(i,t)\in D_{NT}, NT\geq 1\}$ satisfy*

$$\bar{\eta}_\mu^{(2)}(r)\leq Cr^{2-\mu}$$

for some constant $C > 0$*. Moreover,* $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\mathbb{E}\left|\frac{\partial^2\mu_{it}(\theta_0)}{\partial\theta\partial\theta'}\right|^{2p} < \infty$*.*

Assumption 6.4.12. *Let* $\mathbb{S}_\mu$ *be the range of* μ_{it} *for all* $(i,t)\in D_{NT}, NT\geq 1$*, the functions* $(\mathcal{B}^{-1})'$ *and* $(\mathcal{A}\circ\mathcal{B}^{-1})'$ *are Lipschitz continuous on* $\mathbb{S}_\mu$*.*

Assumption 6.4.13. *Following bounds exist almost surely:*

$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{\partial^2\mu_{it}(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty;$$
$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left|(\mathcal{B}^{-1})''(\mu_{it}(\theta_0))\right| < \infty;$$
$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left|(\mathcal{A}\circ\mathcal{B}^{-1})''(\mu_{it}(\theta_0))\right| < \infty.$$

Lemma 6.4.6. *Beside of all the conditions of Lemma 6.4.5, if Assumptions 6.4.12 and 6.4.13 are also satisfied, then the* η*-coefficients of* $\left\{\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta\partial\theta'} : (i,t)\in D_{NT}, NT\geq 1\right\}$ *satisfy*

$$\bar{\eta}_l^{(2)}(r) \leq Cr^{2-\frac{2(p-1)}{2p-1}\mu}$$

for some constant $C > 0$*. Moreover,* $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\mathbb{E}\left|\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta\partial\theta'}\right|^{2p} < \infty$*.*

Assumption 6.4.14. θ_0 *is an interior point of the parameter space* Θ*.*

Assumption 6.4.15. $\inf_{NT\geq 1}\lambda_{min}(V_{NT}) > 0$ *and* $\inf_{NT\geq 1}\lambda_{min}(V_{NT}^{-1/2}H_{NT}) > 0$*, where* $V_{NT} = \mathrm{Var}\left[\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\right]$*,* $H_{NT} = -\mathbb{E}\left[\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right]$ *and* $\lambda_{min}(\cdot)$ *denotes the smallest eigenvalue.*

With Lemma 6.4.4 and Assumption 6.4.15, we can prove (6.4.5) according to the Corollary 3.2.1 in Chapter 3.1. (6.4.6) could also be verified with the result of Lemma 6.4.6 by Theorem 3.1 in Chapter 3.1. Combining (6.4.5), (6.4.6) and Assumption 6.4.7, we obtain the asymptotic normality of $\hat{\theta}_{NT}$ as follows:

Theorem 6.3. *If Assumptions 6.4.1 to 6.4.15 are satisfied, then the MLE defined by (6.4.2) is asymptotically normal, that is*

$$\sqrt{NT}(V_{NT}^{-1/2}H_{NT})(\hat{\theta}_{NT}-\theta_0) \xrightarrow{d} \mathcal{N}(0, I_k)$$

as $NT\to\infty$*.*

6.5 A negative binomial threshold network GARCH

Suppose the conditional distribution of y_{it} in (6.2.2) is the negative binomial distribution:

$$y_{it}|\mathcal{H}_{t-1} \sim \mathcal{NB}(K, p_{it}), \tag{6.5.1}$$

which belongs to the one-parameter exponential family with $\eta_{it} = \log(1 - p_{it})$, $\mathcal{A}(x) = -K\log(1 - e^x)$ and $\mathcal{B}(x) = \frac{Ke^x}{1-e^x}$. Notice that the conditional mean $\mathcal{B}(\eta_{it}) = \frac{K(1-p_{it})}{p_{it}}$ is smaller than the conditional variance $\frac{K(1-p_{it})}{p_{it}^2}$, therefore the negative binomial distribution is an appropriate alternative to the Poisson distribution under over-dispersed data. Let $\mu_{it} = \frac{K(1-p_{it})}{p_{it}}$, following the idea of Samia and Chan (2011) and Davis and Liu (2016) of linking the conditional mean process to a piece-wise linear stochastic function, we define g_θ as follows:

$$\mu_{it} = \omega + \alpha^{(1)} y_{i,t-1} + \alpha^{(2)}(y_{i,t-1} - r)^+ + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \beta\mu_{i,t-1}, \tag{6.5.2}$$

where x^+ denotes the positive part of x, and $r > 0$ is an integer-valued threshold parameter. To ensure that $\mu_{it} > 0$, we suppose that $\omega > 0$, $\alpha^{(1)} \geq 0$, $\alpha^{(1)} + \alpha^{(2)} \geq 0$, $\lambda \geq 0$, $\beta \geq 0$.

(6.5.1) and (6.5.2) together define a negative binomial threshold network GARCH model (NBTNGARCH), which is an extension of the negative binomial integer-valued GARCH by Zhu (2010) to high-dimensional network data with threshold effects. Notice that (6.5.2) can be viewed as two linear regression on different regimes separated by the threshold r:

$$\begin{cases} \mu_{it} = \omega + \alpha^{(1)} y_{i,t-1} + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \beta\mu_{i,t-1} & y_{i,t-1} < r; \\ \mu_{it} = \left(\omega - \alpha^{(2)} r\right) + \left(\alpha^{(1)} + \alpha^{(2)}\right) y_{i,t-1} + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \beta\mu_{i,t-1} & y_{i,t-1} \geq r. \end{cases}$$

For each $i = 1, 2, ..., N$, apart from the network structure, (6.5.2) is a special case of the self-excited threshold autoregression by Wang et al. (2014) in that the coefficient β is regime-invariant.

6.5.1 Stationarity and estimation of NBTNGARCH

Proposition 6.1. *Based on the model defined by* (6.5.1) *and* (6.5.2)*, if*

$$\max\{\alpha^{(1)}, \alpha^{(1)} + \alpha^{(2)}\} + \lambda + \beta < 1, \tag{6.5.3}$$

then the N-dimensional process $\mathbb{X}_t = (\mu_{1t}, \mu_{2t}, ..., \mu_{Nt})'$ is geometric moment contracting with a unique stationary distribution and finite first order moment.

Let $(\theta', r)'$ be the array of parameters to be estimated where $\theta = (\omega, \alpha^{(1)}, \alpha^{(2)}, \lambda, \beta)'$. Moreover, $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ are the samples of size NT, generated by (6.5.1) and (6.5.2) with true parameters $\omega_0, \alpha_0^{(1)}, \alpha_0^{(2)}, \lambda_0, \beta_0$ and r_0.

Based on the infinite past of observations, the log-likelihood function (ignoring constants) is

$$\begin{cases} L_{NT}(\theta, r) = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} l_{it}(\theta, r), \\ l_{it}(\theta, r) = y_{it} \log \mu_{it}(\theta, r) - (y_{it} + K) \log\left(\mu_{it}(\theta, r) + K\right), \end{cases} \tag{6.5.4}$$

where $\mu_{it}(\theta, r)$ is generated according to process (6.5.2). Based on finite samples $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$, we obtain the approximated log-likelihood function:

$$\begin{cases} \tilde{L}_{NT}(\theta, r) = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \tilde{l}_{it}(\theta, r), \\ \tilde{l}_{it}(\theta, r) = y_{it} \log \tilde{\mu}_{it}(\theta, r) - (y_{it} + K) \log\left(\tilde{\mu}_{it}(\theta, r) + K\right), \end{cases} \tag{6.5.5}$$

where $\tilde{\mu}_{it}(\theta, r)$ is generated by (6.5.2) with initial values $\mu_{i0} = \tilde{\mu}_{i0}, i = 1, 2, ..., N$. Therefore the MLE is the maximizer of $\tilde{L}_{NT}(\theta, r)$, which can be obtained through a two-step algorithm as follows according to Wang et al. (2014):

1. For each r in a predetermined range $[\underline{r}, \bar{r}] \subset \mathbb{Z}_+$, find $\hat{\theta}_{NT}^{(r)} = \text{argmax}_{\theta\in\Theta} \tilde{L}_{NT}(\theta, r)$ where $\Theta \subset \mathbb{R}^5$ is the parameter space of coefficients;
2. Find $\hat{r}_{NT} = \text{argmax}_{r\in[\underline{r},\bar{r}]} \tilde{L}_{NT}(\hat{\theta}_{NT}^{(r)}, r)$.

Then $\left(\hat{\theta}_{NT}^{(\hat{r}_{NT})'}, \hat{r}_{NT}\right)'$ is the optimizer of $\tilde{L}_{NT}(\theta, r)$.

We need the following assumptions:

(NB1). θ_0 is an interior point of the parameter space Θ, which is a compact subset of $\mathbb{R}^5$ such that $\omega > 0$, $\alpha^{(1)} \geq 0$, $\alpha^{(1)} + \alpha^{(2)} \geq 0$, $\lambda \geq 0$, $\beta \geq 0$, and $\max\{\alpha^{(1)}, \alpha^{(1)} + \alpha^{(2)}\} + \lambda + \beta < 1$ for all $\theta \in \Theta$;

(NB2). $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \mathbb{E}\,|y_{it}|^{2p} < \infty$ for some $p > 1$;

(NB3). The array of random fields $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is η-weakly dependent with $\bar{\eta}_y(r) = \mathcal{O}(r^{-\mu})$ with $\mu > \frac{3(2p-1)}{p-1} \vee \frac{(4p-3)(2p-1)}{2(p-1)^2}$;

(NB4). $w_{ij} \leq C(|i-j|^{-\gamma})$ for some $\gamma \geq \frac{2(p-1)}{2p-1}\mu + 2$;

(NB5). $\inf_{NT\geq 1} \lambda_{min}(\Sigma_{NT}) > 0$, where

$$\Sigma_{NT} = \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\frac{K}{\mu_{it}^2(\theta_0, r_0) + K\mu_{it}(\theta_0, r_0)} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta'}\right].$$

Proposition 6.2. *If (NB1) to (NB4) are satisfied, the MLE* $\left(\hat{\theta}_{NT}^{(\hat{r}_{NT})'}, \hat{r}_{NT}\right)'$ *is consistent as* $T \to \infty$ *and* $N \to \infty$*, i.e.*

$$\left(\hat{\theta}_{NT}^{(\hat{r}_{NT})'}, \hat{r}_{NT}\right)' \xrightarrow{p} (\theta_0, r_0)'.$$

If (NB5) is also satisfied and $N = o(T)$*, then* $\hat{\theta}_{NT}^{(r_0)}$ *is asymptotically distributed as follows:*

$$\sqrt{NT}\Sigma_{NT}^{1/2}(\hat{\theta}_{NT}^{(r_0)} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I_5).$$

Remark. With $\hat{r}_{NT}$ being an integer-valued consistent estimate of r_0, $\hat{r}_{NT}$ will eventually be equal to r_0 as $T \to \infty$ and $N \to \infty$, hence $\hat{\theta}_{NT}^{(\hat{r}_{NT})}$ and $\hat{\theta}_{NT}^{(r_0)}$ are asymptotically equal.

6.5.2 Simulation study

Set the true parameters $\nu_0 = (0.5, 0.6, 0.1, 0.1, 0.1, 5)'$ and the number of successes $K = 100$ in (6.5.1). Let T increases from 200 to 2000, while N also increases at relatively slower rates of $\mathcal{O}(T/\log(T))$. For each network size N, the adjacency matrix A is simulated according to four different mechanisms in Example 4.5.1 to Example

4.5.4. Based on a simulated network, the data is generated according to (6.5.2). To monitor the finite performance of MLE, data generation and parameter estimation are repeated for $M = 1000$ times, for each combination of sample size (N, T). Root-mean-square errors (RMSE) and coverage probabilities (CP) with different sample sizes and network simulation mechanisms are reported in Table 6.1; We also report the mean estimates of the threshold r_0 at the last column of the table.

	T	N	ω	$\alpha^{(1)}$	$\alpha^{(2)}$	λ	β	$\bar{r}$
Example 4.5.1	200	37	0.0500 (0.95)	0.0153 (0.92)	0.0491 (0.90)	0.0171 (0.95)	0.0169 (0.94)	4.94
	500	80	0.0284 (0.94)	0.0062 (0.93)	0.0221 (0.87)	0.0101 (0.94)	0.0073 (0.95)	5.01
	1000	144	0.0190 (0.95)	0.0032 (0.94)	0.0100 (0.92)	0.0070 (0.95)	0.0040 (0.94)	5.01
	2000	263	0.0132 (0.95)	0.0017 (0.93)	0.0046 (0.95)	0.0050 (0.95)	0.0020 (0.95)	5
Example 4.5.2	200	37	0.0294 (0.95)	0.0151 (0.93)	0.0476 (0.91)	0.0096 (0.94)	0.0179 (0.93)	4.95
	500	80	0.0136 (0.94)	0.0065 (0.92)	0.0216 (0.89)	0.0043 (0.95)	0.0075 (0.95)	5.03
	1000	144	0.0078 (0.95)	0.0033 (0.93)	0.0099 (0.91)	0.0024 (0.95)	0.0041 (0.94)	5.01
	2000	263	0.0040 (0.94)	0.0016 (0.94)	0.0046 (0.94)	0.0012 (0.94)	0.0020 (0.94)	5
Example 4.5.3	200	37	0.0362 (0.94)	0.0151 (0.93)	0.0435 (0.92)	0.0107 (0.94)	0.0172 (0.94)	4.87
	500	80	0.0148 (0.93)	0.0064 (0.92)	0.0209 (0.87)	0.0044 (0.95)	0.0072 (0.95)	5.01
	1000	144	0.0077 (0.95)	0.0032 (0.94)	0.0102 (0.90)	0.0025 (0.94)	0.0039 (0.95)	5
	2000	263	0.0040 (0.92)	0.0016 (0.94)	0.0047 (0.94)	0.0012 (0.96)	0.0021 (0.94)	5
Example 4.5.4	200	37	0.0465 (0.95)	0.0155 (0.92)	0.0459 (0.92)	0.0161 (0.94)	0.0175 (0.94)	4.89
	500	80	0.0175 (0.95)	0.0067 (0.91)	0.0202 (0.89)	0.0056 (0.95)	0.0076 (0.94)	4.99
	1000	144	0.0093 (0.94)	0.0032 (0.93)	0.0105 (0.89)	0.0030 (0.95)	0.0039 (0.95)	5.01
	2000	263	0.0047 (0.95)	0.0016 (0.94)	0.0046 (0.93)	0.0015 (0.95)	0.0021 (0.94)	5

Table 6.1: Simulation results with different network structures ($N \approx T/\log(T)$).

From Table 6.1 we can tell, that the RMSEs of $\hat{\theta}_{NT}$ decrease toward zero, and the mean of $\hat{r}_{NT}$ is equal to $r_0 = 5$ for sufficiently large sample size. These results support the consistency of MLE in Proposition 6.2. The reported CPs are close to the value 0.95, showing that $\widehat{SE}$ provides a reliable estimation of the true standard error of $\hat{\theta}_{NT}$. Moreover, in Figures 6.1 to 6.4 we draw the normal Q-Q plots for the estimation results when $T = 2000, N = 263$ under different network structures. These Q-Q plots provide additional evidence for the asymptotic normality of $\hat{\theta}_{NT}$ in Proposition 6.2.

6.5.3 Revisiting the data of car accidents in New York City

We have fitted our Poisson threshold network GARCH model to the daily number of car accidents in New York City in Section 5.4.2. In this section we will fit the non-negative binomial threshold network GARCH to the same dataset, then we will use the diagnostic tool of non-randomized probability integral transform (PIT) to determine whether or

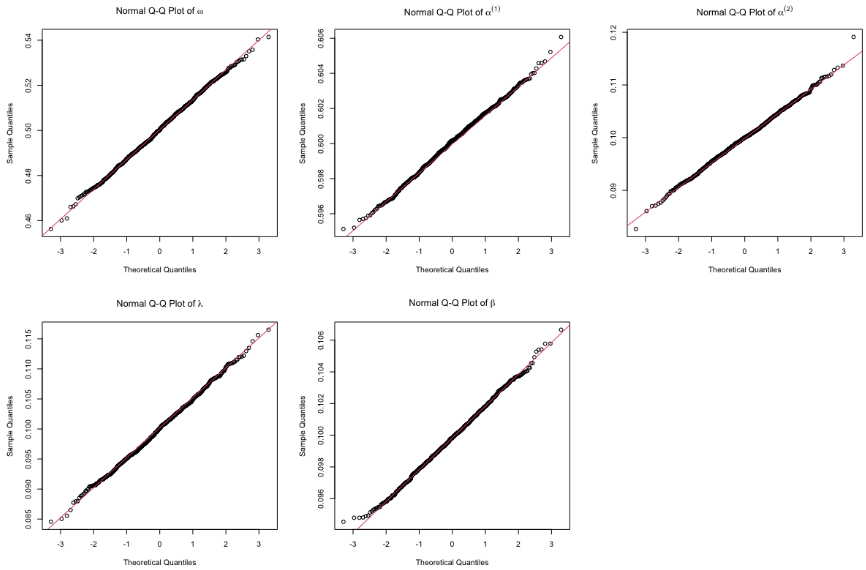


Figure 6.1: Q-Q plots of estimates for Example 4.5.1.

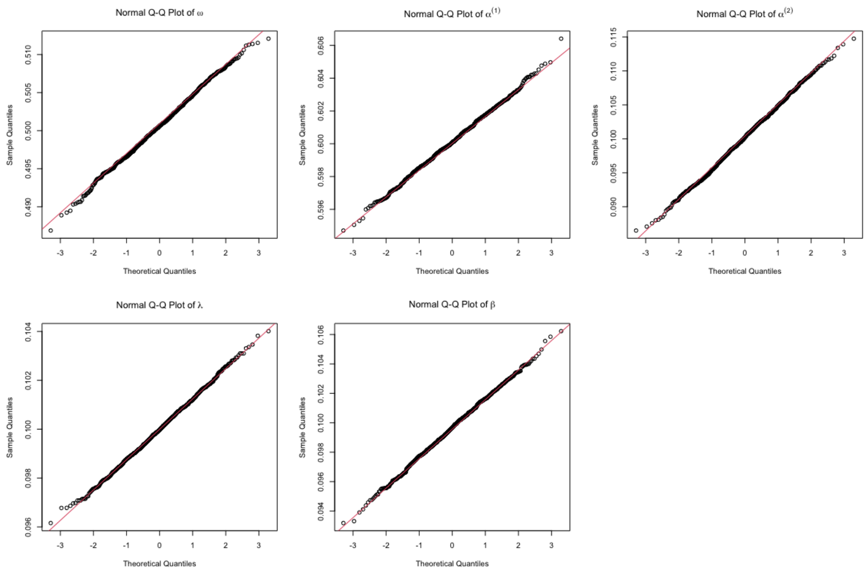


Figure 6.2: Q-Q plots of estimates for Example 4.5.2.

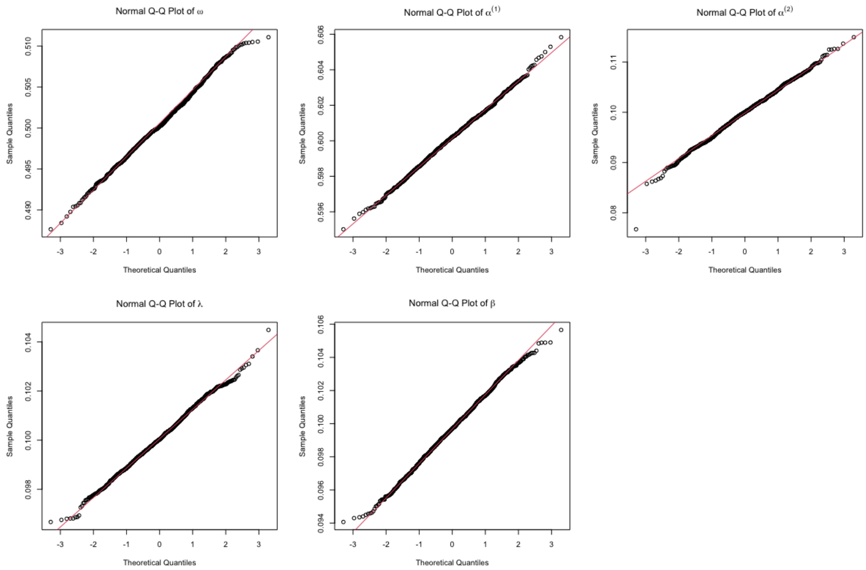


Figure 6.3: Q-Q plots of estimates for Example 4.5.3.

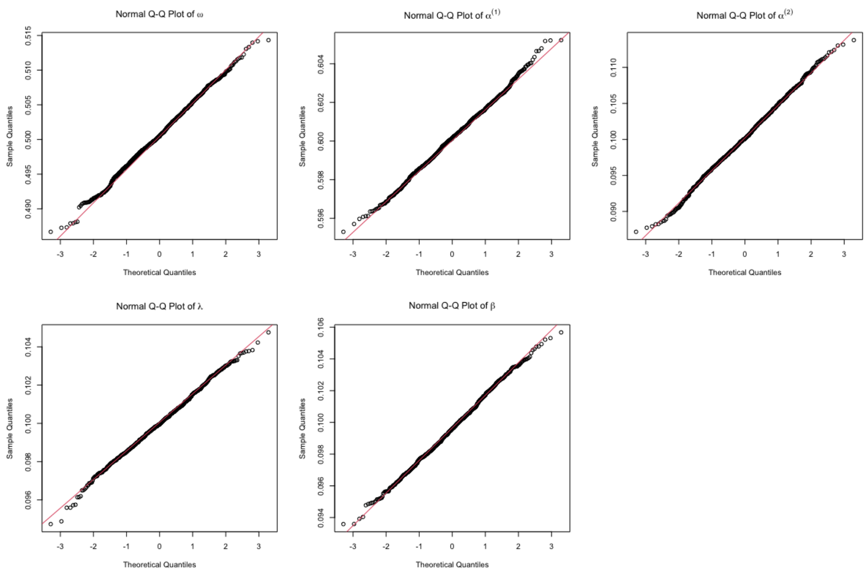


Figure 6.4: Q-Q plots of estimates for Example 4.5.4.

not the non-negative binomial distribution is a better choice over Poisson distribution. Non-randomized PIT was proposed by Czado et al. (2009) to check the statistical consistency between the predictive distribution and the distribution of observations for count data, and it was applied by Christou and Fokianos (2014) to show that their negative binomial autoregressive model is better than the Poisson autoregressive model by Fokianos et al. (2009) when fitted to the data of transactions on the stock market.

Firstly we fit the NBTNGARCH model to the data of car accidents in New York City with $K = 30$, and obtain the MLE $\hat{\theta}$ and $\hat{r}$. With observed data, then we can generate the estimated means $\left\{\hat{\mu}_{it}(\hat{\theta}, \hat{r}) : i = 1, 2, ..., N; t = 1, 2, ..., T\right\}$ according to the process (6.5.2). The PIT is based on following conditional cumulative distribution function:

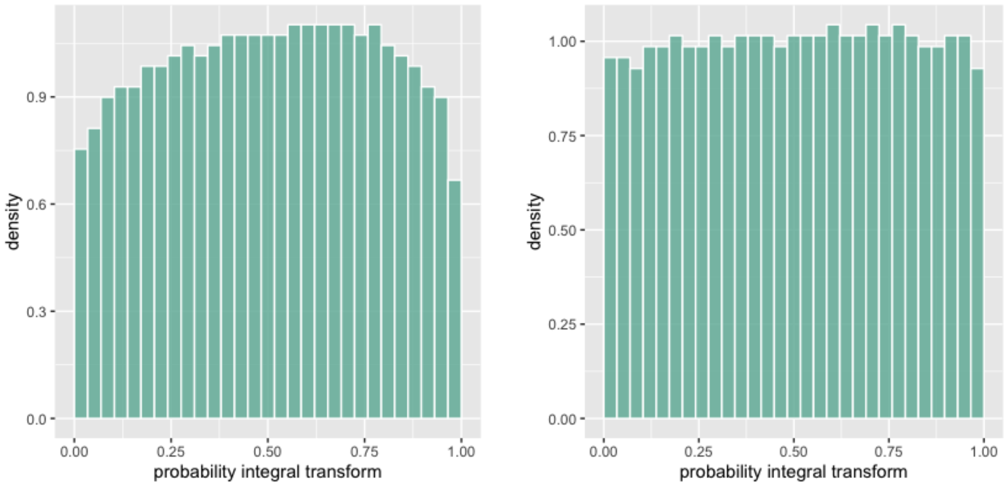
$$F(u|y_{it} = y) = \begin{cases} 0 & u \leq P(y-1), \\ \frac{u-P(y-1)}{P(y)-P(y-1)} & P(y-1) < u \leq P(y), \\ 1 & u > P(y). \end{cases} \tag{6.5.6}$$

P is the conditional cumulative distribution function of non-negative binomial distribution evaluated at $\hat{\mu}_{it}(\hat{\theta}, \hat{r})$, and the PIT (6.5.6) should be a cumulative distribution function of standard uniform distribution if $y \sim P$. Similar to Christou and Fokianos (2014), we obtain the mean PIT by

$$\bar{F}(u) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} F_{it}(u|y_{it}), \quad 0 \leq u \leq 1, \tag{6.5.7}$$

where F_{it} is based on predictive distribution P_{it} evaluated at $\hat{\mu}_{it}(\hat{\theta}, \hat{r})$ and the data y_{it}. The mean PIT for PTNGARCH can be obtained similarly. We generated 1000 evenly space values of $u \in [0, 1]$, and obtain 1000 samples of mean PIT according to (6.5.6) and (6.5.7) for PTNGARCH and NBTNGARCH respectively. These two groups of PIT samples are plotted as histograms in Figure 6.5, which suggests that the PIT samples of NBTNGARCH are more likely to follow a standard uniform distribution. Therefore we choose non-negative binomial threshold network GARCH over the Poisson threshold network GARCH when analysing the data of car accidents in New York City.

The estimation results of NBTNGARCH are reported in Table 6.2. Firstly, it is



(a). PTNGARCH (b). NBTNGARCH($K = 30$)

Figure 6.5: PIT histograms for PTNGARCH and NBTNGARCH.

worthy of note that $\alpha^{(2)} < 0$, which means that the conditional mean of the number of car accidents in these neighbourhoods are less affected by the number on the previous day if it is above the threshold $r = 8$. Secondly, the conditional mean of the number of car accidents in one area is also affected by its geometrically neighboured areas. Besides, the estimated value of β is significantly larger than other coefficients, indicating a strong persistence in conditional mean. These characteristics align with what we have under PTNGARCH in Section 5.4.2.

	ω	$\alpha^{(1)}$	$\alpha^{(2)}$	λ	β	r
Estimation	0.0195	0.1375	-0.0148	0.0028	0.8596	8
SE	0.0038	0.0032	0.0051	0.0011	0.0032	\

Table 6.2: Estimation results based on daily number of car accidents in 41 neighbourhoods of NYC.

Chapter 7

Further Work

7.1 On the network specification

In this research, we use a pre-specified network (denoted by row-normalized adjacency matrix W) to describe the relations between N individuals. For example, in model 4.1.1, the network effect on each individual $i = 1, 2, ..., N$ is denoted by

$$\sum_{j=1}^{N} w_{ij} y_{j,t-1}^2 \tag{7.1.1}$$

in the sense of Cliff and Ord (1972). However, this setting can be improved considering two cases in practice:

- There exist multiple networks that can potentially describe the relationships between these individuals;
- The formation of the network is dynamic over time rather than being static.

To deal with multiple networks, it is worth considering the weight matrix fusion technique of Lu et al. (2024). The authors suggested a weighted average of multiple row-normalized adjacency matrices, with the weights being unknown parameters. If we incorporate the fusion network effect of Lu et al. (2024) instead of (7.1.1) in the network GARCH models, we may be able to analyze the effects of different networks by estimating the weight parameters.

The study of network formation dates back to Erdős and Rényi (1960), and there has been a vast literature on random network models, including dynamic models that describe how networks change over time (see Caldarelli and Vespignani (2007) and Kim et al. (2018) for literature reviews). However, little has been done to incorporate dynamic networks in high-dimensional time series models. Therefore, it would be very meaningful to consider a dynamic network in a network GARCH model.

7.2 Heterogeneous parameters

In this research, we explore time-varying parameters via a self-excited threshold effect. For instance, in model (4.1.1), the coefficient of $y_{i,t-1}^2$ switches between $\alpha^{(1)}$ and $\alpha^{(2)}$ based on the value of $y_{i,t-1}$. Moreover, it is beneficial to examine cases where parameters switch according to an exogenous random process. Cai (1994) and Hamilton and Susmel (1994) introduced ARCH models with Markov-switching parameters to depict sudden shifts in the conditional variance. In these models, parameters change according to a multi-state Markov process. Incorporating Markov-switching parameters in a network GARCH model is both meaningful and challenging. Even in the univariate case, estimating Markov-switching GARCH models poses challenges due to the path dependence problem, which stems from the latent Markov process. Despite Bayesian methods proposed to address this issue (Bauwens et al., 2010; Augustyniak et al., 2018), it remains uncertain whether these methods are applicable for estimating a Markov-switching network GARCH model, particularly when N is large.

Apart from considering time-varying parameters, another extension to current network GARCH models is to accommodate nodal variant parameters. In model 4.1.1, the parameters are the same for each $i = 1, 2, ..., N$. By allowing the parameters to change across each node i in the network, the model will be able to capture nodal variations, providing a more accurate and flexible representation of the underlying processes. In fact, spatially variant parameters have been successfully incorporated by several spatio-temporal autoregressive models (see Rao (2008), Al-Sulami et al. (2017) among others), inspiring us to extend network GARCH models in this direction in the future.

Appendix A

Proofs of Theoretical Results

A.1 Proofs of results in Chapter 3

We use C and ρ uniformly over different contexts to represent constants where $0 < C < \infty$ and $0 < \rho < 1$. In some cases we use C_i and ρ_i with subscript i to distinguish between different constants.

Lemma A.1.1. *Let $\{X_{i,n} : i \in D_n, n \geq 1\}$ be a $\mathbb{R}$-valued η-weakly dependent random field. If $\sup_n \sup_{i \in D_n} \|X_{i,n}\|_p < \infty$ for some $p > 2$, then for any $i, j \in D_n$:*

$$|\operatorname{Cov}(X_{i,n}, X_{j,n})| \leq C\|X\|_p^{\frac{p}{p-1}} [\eta_{n,1,1}(\rho(i,j))]^{\frac{p-2}{p-1}}, \tag{A.1.1}$$

where $\|X\|_p := \sup_n \sup_{i \in D_n} \|X_{i,n}\|_p$. The same result holds for θ-dependence.

Proof. Let $X_{i,n}(k) = -k \vee X_{i,n} \wedge k$ be a truncation of $X_{i,n}$ at level $k > 0$, where $\vee$ and $\wedge$ mean *maximum* and *minimum* respectively. Then for any $i \in D_n$ and $a \in (0, p)$:

$$\begin{aligned}
\mathbb{E}|X_{i,n} - X_{i,n}(k)|^a \leq & \mathbb{E}\left[|X_{i,n}|^a 1(|X_{i,n}| \geq k)\right] \\
\leq & (\mathbb{E}|X_{i,n}|^p)^{a/p} \left[\mathbb{P}(|X_{i,n}| \geq k)\right]^{1-(a/p)} \\
\leq & \|X\|_p^a \left[\frac{\|X\|_p^p}{k^p}\right]^{1-(a/p)} = \|X\|_p^p k^{a-p},
\end{aligned}$$

where the second line and the third line come from Hölder's inequality and Markov inequality respectively. Hence $\sup_n \sup_{i\in D_n} \|X_{i,n} - X_{i,n}(k)\|_a \le \|X\|_p^{p/a} k^{1-(p/a)}$.

For any $i, j \in D_n$ we have

$$\begin{aligned}|\operatorname{Cov}(X_{i,n}, X_{j,n})| \le& |\operatorname{Cov}(X_{i,n}(k), X_{j,n}(k))| \\ &+ |\operatorname{Cov}(X_{i,n} - X_{i,n}(k), X_{j,n}(k))| \\ &+ |\operatorname{Cov}(X_{i,n}, X_{j,n} - X_{j,n}(k))|.\end{aligned}$$

For the last term on the right-hand-side (RHS), we could find $a \in (1, p)$ such that $1/a + 1/p = 1$ and therefore

$$\begin{aligned}&|\operatorname{Cov}(X_{i,n}, X_{j,n} - X_{j,n}(k))| \\ \le& |\mathbb{E}[X_{i,n}(X_{j,n} - X_{j,n}(k))]| + |\mathbb{E}(X_{i,n})||\mathbb{E}(X_{j,n} - X_{j,n}(k))| \\ \le& \|X_{i,n}\|_p \|X_{j,n} - X_{j,n}(k)\|_a + \|X_{i,n}\|_1 \|X_{j,n} - X_{j,n}(k)\|_1 \\ \le& 2\|X_{i,n}\|_p \|X_{j,n} - X_{j,n}(k)\|_a \\ \le& 2\|X\|_p^p k^{2-p}.\end{aligned}$$

Same bound could also be derived for the second term on the RHS. As for the first term, note that $X(k)$ is a function of X with bound k and Lipschitz constant 1. Then by (3.2.4) we have $|\operatorname{Cov}(X_{i,n}(k), X_{j,n}(k))| \le 2k\eta_{n,1,1}(\rho(i,j))$, then

$$|\operatorname{Cov}(X_{i,n}, X_{j,n})| \le 6\|X\|_p^{\frac{p}{p-1}} [\eta_{n,1,1}(\rho(i,j))]^{\frac{p-2}{p-1}}$$

by choosing $k = \left[\frac{\|X\|_p^p}{\eta_{n,1,1}(\rho(i,j))}\right]^{\frac{1}{p-1}}$. The proof under θ-dependence follows similar arguments.

□

Lemma A.1.2. *(Proposition 6.3.9 in Brockwell and Davis (2009)) Let $(Z_n)_{n\ge1}$ and $(V_{n,k})_{n,k\in\mathbb{N}_+}$ be sequences of random vectors. $Z_n \xrightarrow{d} V$ if the following statements are true:*

1. *For each $k \in \mathbb{N}_+$, there exists V_k such that $V_{n,k} \xrightarrow{d} V_k$ as $n \to \infty$;*

2. $V_k \xrightarrow{d} V$ *as* $k \to \infty$;

3. $\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(|Z_n - V_{n,k}| > \delta) = 0$ *for any* $\delta > 0$.

Lemma A.1.3. *(Lemma 2 in Bolthausen (1982)) Let* $(\nu_n)_{n\in\mathbb{N}_+}$ *be a sequence of probability measures over* $\mathbb{R}$ *with*

1. $\sup_n \int x^2 \nu_n(dx) < \infty$,

2. $\lim_{n\to\infty} \int (\mathbf{i}\lambda - x)e^{\mathbf{i}\lambda x}\nu_n(dx) = 0$ *for all* $\lambda \in \mathbb{R}$.

Then $\nu_n \xrightarrow{d} N(0,1)$ *as* $n \to \infty$.

Lemma A.1.4. *(Lemma A.1 in Jenish and Prucha (2009)) For any* $i \in D \subset \mathbb{R}^d$ *and* $h \geq 1$, *let*

$$N_i(h) := |\{j \in D : h \leq |j - i| < h + 1\}|_c$$

be the number of all elements of D *located at any distance in* $[h, h+1)$ *from* i. *Then* $\sup_i N_i(h) \leq Ch^{d-1}$.

Lemma A.1.5. *(Lemma A.4 in Xu et al. (2024)) For any* $\alpha > 0$ *and* $s \geq 2$,

$$\sum_{h=[s]}^{\infty} h^{-\alpha-1} < \frac{2^{\alpha+1}}{\alpha} s^{-\alpha},$$

where $[s]$ *denotes the largest integer less than or equal to* s.

A.1.1 Proof of results in Section 3.2

Proof of Proposition 3.2

Firstly we prove that $X_{i,n} = H_{i,n}((\varepsilon_j)_{j\in D})$ is well-defined in $\mathbb{L}^1$. For any $s \in \mathbb{N}$, let $X_{i,n}^{(s)} = H_{i,n}((\varepsilon_j 1_{\{\rho(i,j)\leq s\}})_{j\in D})$. Then by (3.2.6) we have

$$\begin{aligned}&\left|X_{i,n}^{(s+m)} - X_{i,n}^{(s)}\right| \\ \leq &\sum_{k=1}^{m} \left|X_{i,n}^{(s+k)} - X_{i,n}^{(s+k-1)}\right|\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{m} \left| H_{i,n}((\varepsilon_j 1_{\{\rho(i,j)\leq s+k\}})_{j\in D}) - H_{i,n}((\varepsilon_j 1_{\{\rho(i,j)\leq s+k-1\}})_{j\in D}) \right| \\
&\leq \sum_{k=1}^{m} B_{i,n}(s+k)(\max_{\rho(i,j)\leq s+k-1} |\varepsilon_j|^l \vee 1) \sum_{\rho(i,j)=s+k} |\varepsilon_j|.
\end{aligned}
$$

Since $\sup_{i\in D} \|\varepsilon_i\|_p < \infty$ with $p > l+1$, by Hölder's inequality and Lemma A.1.4 we obtain that

$$
\left\| X_{i,n}^{(s+m)} - X_{i,n}^{(s)} \right\|_1 \leq C \sum_{k=1}^{m} (s+k)^{d-1} B_{i,n}(s+k). \tag{A.1.2}
$$

Notice that $(s+k)^{d-1}B_{i,n}(s+k) \to 0$ as $s \to \infty$, according to (3.2.7). Then if m is fixed, $\left\| X_{i,n}^{(s+m)} - X_{i,n}^{(s)} \right\|_1 \to 0$ as $s \to \infty$. Therefore $\{X_{i,n}^{(s)} : s \geq 0\}$ is a Cauchy sequence in $\mathbb{L}^1$, and $X_{i,n} = \lim_{s\to\infty} X_{i,n}^{(s)}$ is well-defined.

Let $U_n, V_n \subset D_n$ be two arbitrary sub-lattices of D_n with $|U_n|_c = u$, $|V_n|_c = v$ and $\rho(U_n, V_n) \geq r$. $f \in \mathcal{F}_u$ and $g \in \mathcal{G}_v$ are two arbitrary Lipschitz functions with $\|f\|_\infty = \|g\|_\infty = 1$. For an arbitrary threshold value $T > 0$, define $\varepsilon_i(T) := -T \vee \varepsilon_i \wedge T$, and $X_{i,n}^{(s)}(T) := H_{i,n}((\varepsilon_j(T) 1_{\{\rho(i,j)\leq s\}})_{j\in D})$. Notice that

$$
\begin{aligned}
&|\mathrm{Cov}[f((X_{i,n})_{i\in U_n}), g((X_{i,n})_{i\in V_n})]| \\
\leq &\left| \mathrm{Cov}[f((X_{i,n})_{i\in U_n}) - f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n})_{i\in V_n})] \right| \\
&+ \left| \mathrm{Cov}[f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n})_{i\in V_n}) - g((X_{i,n}^{(s)}(T))_{i\in V_n})] \right| \\
&+ \left| \mathrm{Cov}[f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n}^{(s)}(T))_{i\in V_n})] \right|.
\end{aligned} \tag{A.1.3}
$$

We start with the first term in the right-hand-side (RHS) of (A.1.3):

$$
\begin{aligned}
&\left| \mathrm{Cov}[f((X_{i,n})_{i\in U_n}) - f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n})_{i\in V_n})] \right| \\
\leq &2 \,\mathrm{Lip}(f) \sum_{i\in U_n} \mathbb{E}|X_{i,n} - X_{i,n}^{(s)}(T)| \\
\leq &2u \,\mathrm{Lip}(f) \left[\sup_{i\in U_n} \mathbb{E}|X_{i,n} - X_{i,n}^{(s)}| + \sup_{i\in U_n} \mathbb{E}|X_{i,n}^{(s)} - X_{i,n}^{(s)}(T)| \right].
\end{aligned}
$$

Same as (A.1.2), we can prove that

$$\mathbb{E}|X_{i,n} - X_{i,n}^{(s)}| \leq C \sum_{h \geq s} h^{d-1} B_{i,n}(h).$$

Notice that $\sum_{\rho(i,j)\leq s} = \mathcal{O}(s^d)$ according to Lemma A.1(ii) in Jenish and Prucha (2009), then by using (3.2.6) repeatedly we can also prove that (ignoring a constant factor):

$$\begin{aligned}
&\mathbb{E}|X_{i,n}^{(s)} - X_{i,n}^{(s)}(T)| \\
=&\mathbb{E}\left|H_{i,n}((\varepsilon_j 1_{\{\rho(i,j)\leq s\}})_{j\in D}) - H_{i,n}((\varepsilon_j(T) 1_{\{\rho(i,j)\leq s\}})_{j\in D})\right| \\
\leq&\left(\sum_{h=0}^{\infty} B_{i,n}(h)h^{d-1}\right)\mathbb{E}\left|(\max_{\rho(i,j)\leq s}|\varepsilon_j|^l)\sum_{\rho(i,j)\leq s}|\varepsilon_j|1_{\{|\varepsilon_j|\geq T\}}\right| \\
\leq&\left(\sum_{h=0}^{\infty} B_{i,n}(h)h^{d-1}\right)s^d\mathbb{E}\left|\max_{\rho(i,j)\leq s}|\varepsilon_j|^{l+1}1_{\{|\varepsilon_j|\geq T\}}\right|.
\end{aligned}$$

Since $\mathbb{E}|\varepsilon_i|^p < \infty$ and $p > l+1$, by Hölder's inequality we have

$$\mathbb{E}|X_{i,n}^{(s)} - X_{i,n}^{(s)}(T)| \leq \left(\sum_{h=0}^{\infty} B_{i,n}(h)h^{d-1}\right) C s^d T^{l+1-p}.$$

Then we obtain the bound

$$\begin{aligned}
&\left|\mathrm{Cov}[f((X_{i,n})_{i\in U_n}) - f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n})_{i\in V_n})]\right| \\
\leq&2u\,\mathrm{Lip}(f)\left[C_1C(s) + C_2C_B s^d T^{l+1-p}\right].
\end{aligned} \tag{A.1.4}$$

The bound of the second term on RHS of (A.1.3) follows analogously:

$$\begin{aligned}
&\left|\mathrm{Cov}[f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n})_{i\in V_n}) - g((X_{i,n}^{(s)}(T))_{i\in V_n})]\right| \\
\leq&2v\,\mathrm{Lip}(g)\left[C_1C(s) + C_2C_B s^d T^{l+1-p}\right].
\end{aligned} \tag{A.1.5}$$

To obtain the bound for the last term on RHS of (A.1.3), define following functions $F_T : \mathbb{R}^{s^d u} \mapsto \mathbb{R}$ and $G_T : \mathbb{R}^{s^d v} \mapsto \mathbb{R}$ as follows:

$$F_T(((\varepsilon_j)_{\rho(i,j)\leq s})_{i\in U_n}) := f((H_{i,n}((\varepsilon_j(T)1_{\{\rho(i,j)\leq s\}})_{j\in D}))_{i\in U_n}) = f((X_{i,n}^{(s)}(T))_{i\in U_n});$$

$$G_T(((\varepsilon_j)_{\rho(i,j)\le s})_{i\in V_n}) := g((H_{i,n}((\varepsilon_j(T)1_{\{\rho(i,j)\le s\}})_{j\in D}))_{i\in V_n}) = g((X_{i,n}^{(s)}(T))_{i\in V_n}).$$

By the η-weak dependence of $\{\varepsilon_i : i \in D\}$, if $r \ge 2s$ we have

$$\begin{aligned}&\left|\operatorname{Cov}[f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n}^{(s)}(T))_{i\in V_n})]\right|\\ \le&[s^d u \operatorname{Lip}(F_T) + s^d v \operatorname{Lip}(G_T)]\bar{\eta}_\varepsilon(r-2s).\end{aligned}$$

Notice that for any $\mathfrak{X} = (\mathbb{X}_{i,n})_{i\in U_n}, \mathfrak{Y} = (\mathbb{Y}_{i,n})_{i\in U_n} \in \mathbb{R}^{s^d u}$:

$$\begin{aligned}&\frac{|F_T(\mathfrak{X}) - F_T(\mathfrak{Y})|}{\sum_{i\in U_n}|\mathbb{X}_{i,n} - \mathbb{Y}_{i,n}|}\\ \le&\operatorname{Lip}(f)\frac{\sum_{i\in U_n}\left|H_{i,n}((X_{ij,n}(T)1_{\{\rho(i,j)\le s\}})_{j\in D}) - H_{i,n}((Y_{ij,n}(T)1_{\{\rho(i,j)\le s\}})_{j\in D})\right|}{\sum_{i\in U_n}\sum_{\rho(i,j)\le s}|X_{ij,n} - Y_{ij,n}|}\\ \le&\operatorname{Lip}(f)\left(\sum_{h=0}^{\infty}B_{i,n}(h)h^{d-1}\right)T^l\frac{\sum_{i\in U_n}\sum_{\rho(i,j)\le s}|X_{ij,n}(T) - Y_{ij,n}(T)|}{\sum_{i\in U_n}\sum_{\rho(i,j)\le s}|X_{ij,n} - Y_{ij,n}|}\end{aligned}$$

by using (3.2.6) repeatedly. Therefore we can prove that:

$$\begin{aligned}\operatorname{Lip}(F_T) &\le C_B T^l \operatorname{Lip}(f);\\ \operatorname{Lip}(G_T) &\le C_B T^l \operatorname{Lip}(g).\end{aligned}$$

Then we can bound the last term on RHS of (A.1.3) by

$$\begin{aligned}&\left|\operatorname{Cov}[f((X_{i,n}^{(s)}(T))_{i\in U_n}), g((X_{i,n}^{(s)}(T))_{i\in V_n})]\right|\\ \le&[u\operatorname{Lip}(f) + v\operatorname{Lip}(g)]C_B s^d T^l \bar{\eta}_\varepsilon(r-2s).\end{aligned} \tag{A.1.6}$$

for any $r \ge 2s$.

Combining (A.1.4), (A.1.5), (A.1.6) we can prove Proposition 3.2 by setting the threshold value $T = \bar{\eta}_\varepsilon(r-2s)^{-\frac{1}{p-1}}$. The result under θ-coefficients could be verified analogously.

Verification of (3.2.10) and (3.2.11)

Assume that $\bar{\eta}_\varepsilon = \mathcal{O}(r^{-\mu})$ for some $\mu > \frac{p-1}{p-1-l}d$ and $B_{i,n}(h) = \mathcal{O}(h^{-b})$ for some $b \geq \frac{p-1-l}{p-1}\mu$. Notice that:

$$\sum_{h=s}^{\infty} h^{-k} \leq \int_{s-1}^{\infty} \frac{1}{x^k} dx = \frac{1}{k-1}(s-1)^{1-k}$$

if $k > 1$. Then we have $C(s) = \sum_{h=s}^{\infty} B_{i,n}(h)h^{d-1} \leq Cs^{-b+d}$. (3.2.10) follows by letting $s = [r/3] < [r/2]$ in (3.2.9).

Assume that $d = 2$, $\bar{\eta}_\varepsilon = \mathcal{O}(r^{-\mu})$ for some $\mu > 0$ and $B_{i,n}(h) = \mathcal{O}(e^{-bh})$ for some $b \geq \frac{p-1-l}{p-1}\mu$. Notice that:

$$\sum_{h=s}^{\infty} he^{-bh} = \frac{se^{-bs} - (s-1)e^{-b(s+1)}}{(1-e^{-b})^2}.$$

Then $C(s) = \sum_{h=s}^{\infty} B_{i,n}(h)h = \mathcal{O}(se^{-bs})$, and (3.2.11) follows by letting $s = [\log r] < [r/2]$ in (3.2.9).

A.1.2 Proof of Theorem 3.1

Let $Y_{i,n} = \frac{X_{i,n}}{M_n}$ where $M_n = \sup_{i \in D_n} c_{i,n}$. From (A.1.9) and Claim A.1.1, we could verify that $Y_{i,n}$ also satisfies Assumption 3.3.1 and Assumption 3.3.2 if $X_{i,n}$ does.

Again in the proof of LLN, we still use the decomposition $Y_{i,n}(k)$ and $\tilde{Y}_{i,n}(k)$ in (A.1.10), which are continuous transformations of $Y_{i,n}$ with Lipschitz constants 1. From Proposition 3.1 we know that $Y_{i,n}(k)$ and $\tilde{Y}_{i,n}(k)$ also inherit the dependence coefficient from $Y_{i,n}$.

Since

$$\begin{aligned}\mathbb{E}\left|\sum_{i \in D_n}(Y_{i,n} - \mathbb{E}Y_{i,n})\right| \leq& \mathbb{E}\left|\sum_{i \in D_n}(Y_{i,n}(k) - \mathbb{E}Y_{i,n}(k))\right| + \mathbb{E}\left|\sum_{i \in D_n}(\tilde{Y}_{i,n}(k) - \mathbb{E}\tilde{Y}_{i,n}(k))\right| \\ \leq& \mathbb{E}\left|\sum_{i \in D_n}(Y_{i,n}(k) - \mathbb{E}Y_{i,n}(k))\right| + 2\sum_{i \in D_n}\mathbb{E}|\tilde{Y}_{i,n}(k)|,\end{aligned}$$

we have

$$\left\|(|D_n|_c)^{-1}\sum_{i\in D_n}(Y_{i,n}-\mathbb{E}Y_{i,n})\right\|_1 \le \left\|(|D_n|_c)^{-1}\sum_{i\in D_n}(Y_{i,n}(k)-\mathbb{E}Y_{i,n}(k))\right\|_1 + 2\sup_n\sup_{i\in D_n}\mathbb{E}|\tilde{Y}_{i,n}(k)|$$

Note that $\sup_n\sup_{i\in D_n}\mathbb{E}|\tilde{Y}_{i,n}(k)| \le \sup_n\sup_{i\in D_n}\mathbb{E}\left[|Y_{i,n}|\,1\left(|Y_{i,n}|\ge k\right)\right]$ for any $k>0$, then according to (3.3.2), it suffices to show that

$$\lim_{k\to\infty}\lim_{n\to\infty}\left\|(|D_n|_c)^{-1}\sum_{i\in D_n}(Y_{i,n}(k)-\mathbb{E}Y_{i,n}(k))\right\|_1 = 0 \tag{A.1.7}$$

in order to prove that

$$\lim_{n\to\infty}\left\|(|D_n|_c)^{-1}\sum_{i\in D_n}(Y_{i,n}-\mathbb{E}Y_{i,n})\right\|_1 = 0.$$

Let $\sigma_n^2(k) = \mathrm{Var}\left[\sum_{i\in D_n}Y_{i,n}(k)\right]$, then

$$\left\|(|D_n|_c)^{-1}\sum_{i\in D_n}(Y_{i,n}(k)-\mathbb{E}Y_{i,n}(k))\right\|_1 \le (|D_n|_c)^{-1}\sigma_n(k)$$

by Lyapunov's inequality. Since $Y_{i,n}(k)$ is a bounded function of $Y_{i,n}$ with Lipschitz constant 1, then by Lemma A.1.(iii) in Jenish and Prucha (2009) and (3.2.2) we have

$$\sigma_n^2(k) \le C|D_n|_c\sum_{s=0}^{\infty}s^{d-1}\bar{\epsilon}_{1,1}(s).$$

Recall from Assumption 3.3.2 that $\bar{\epsilon}_{1,1}(s) = \mathcal{O}(s^{-\alpha})$ with $\alpha > d$, therefore $\lim_{n\to\infty}(|D_n|_c)^{-1}\sigma_n(k) = 0$ for each $k>0$. This completes the proof.

A.1.3 Proof of Theorem 3.2

Let $Y_{i,n} = \frac{X_{i,n}}{M_n}$ where $M_n = \sup_{i \in D_n} c_{i,n}$, and denote $S_{n,Y} = \sum_{i \in D_n} Y_{i,n}$ and $\sigma_{n,Y}^2 = \mathrm{Var}(S_{n,Y})$. Then it could be easily verified that

$$\sigma_n^{-1} S_n = \sigma_{n,Y}^{-1} S_{n,Y}.$$

Therefore it suffices to prove the CLT for $\{Y_{i,n} : i \in D_n, n \geq 1\}$. In what follows, we would denote for simplicity that $S_n = \sum_{i \in D_n} Y_{i,n}$ and $\sigma_n^2 = \mathrm{Var}(S_n)$.

In this new setting, Assumption 3.3.6 becomes

$$\liminf_{n \to \infty} (|D_n|_c)^{-1} \sigma_n^2 > 0. \tag{A.1.8}$$

Assumption 3.3.3 implies that $Y_{i,n}$ are uniformly L_m-bounded as

$$\sup_n \sup_{i \in D_n} \mathbb{E}|Y_{i,n}|^m \leq \sup_n \sup_{i \in D_n} \mathbb{E} \left| \frac{X_{i,n}}{c_{i,n}} \right|^m \leq \infty \tag{A.1.9}$$

for some $m > 2$. Then we will show that Assumption 3.3.4 and Assumption 3.3.5 about dependence coefficient of $X_{i,n}$ covers the dependence coefficient of $Y_{i,n}$ in the following claim:

Claim A.1.1. *The dependence coefficients $\epsilon^*_{n,u,v}(s)$ of $Y_{i,n}$ and the dependence coefficients $\epsilon_{n,u,v}(s)$ of $X_{i,n}$ satisfy*

$$\epsilon^*_{n,u,v}(s) \leq \frac{1}{M_n} \epsilon_{n,u,v}(s).$$

Proof. Let $f \in \mathcal{F}_u : \mathbb{R}^u \mapsto \mathbb{R}$ and $g \in \mathcal{G}_v : \mathbb{R}^v \mapsto \mathbb{R}$ be two arbitrary Lipschitz bounded functions. Define

$$\mathfrak{X}_{U_n} := \{X_{i,n} : i \in U_n, U_n \subset D_n\}, \quad \mathfrak{X}_{V_n} := \{X_{i,n} : i \in V_n, V_n \subset D_n\},$$

and

$$\mathfrak{Y}_{U_n} := \{Y_{i,n} : i \in U_n, U_n \subset D_n\}, \quad \mathfrak{Y}_{V_n} := \{Y_{i,n} : i \in V_n, V_n \subset D_n\}.$$

Then we could define functions $F : \mathbb{R}^u \mapsto \mathbb{R}$ and $G : \mathbb{R}^v \mapsto \mathbb{R}$:

$$F(\mathfrak{X}_{U_n}) := f((X_{i,n}/M_n)_{i \in U_n}),$$
$$G(\mathfrak{X}_{V_n}) := g((X_{i,n}/M_n)_{i \in V_n}).$$

For the Lipschitz constants of F and G we have

$$\begin{aligned}
\operatorname{Lip}(F) &= \sup_{(x_1,\dots,x_u)\neq(y_1,\dots,y_u)\in\mathbb{R}^u} \frac{|F(x_1,\dots,x_u) - F(y_1,\dots,y_u)|}{|x_1 - y_1| + \dots + |x_u - y_u|} \\
&= \sup_{(x_1,\dots,x_u)\neq(y_1,\dots,y_u)\in\mathbb{R}^u} \frac{\left|f(\frac{x_1}{M_n},\dots,\frac{x_u}{M_n}) - f(\frac{y_1}{M_n},\dots,\frac{y_u}{M_n})\right|}{|x_1 - y_1| + \dots + |x_u - y_u|} \\
&= \frac{1}{M_n} \sup_{(x_1,\dots,x_u)\neq(y_1,\dots,y_u)\in\mathbb{R}^u} \frac{\left|f(\frac{x_1}{M_n},\dots,\frac{x_u}{M_n}) - f(\frac{y_1}{M_n},\dots,\frac{y_u}{M_n})\right|}{|\frac{x_1}{M_n} - \frac{y_1}{M_n}| + \dots + |\frac{x_u}{M_n} - \frac{y_u}{M_n}|} \\
&\leq \frac{1}{M_n} \operatorname{Lip}(f).
\end{aligned}$$

And similarly $\operatorname{Lip}(G) \leq \frac{1}{M_n} \operatorname{Lip}(g)$. Obviously, we also have $\|F\|_\infty \leq \|f\|_\infty$ and $\|G\|_\infty \leq \|g\|_\infty$. Consequently, for case when $X_{i,n}$ are η-dependent, we have

$$\begin{aligned}
|\operatorname{Cov}(f(\mathfrak{Y}_{U_n}), g(\mathfrak{Y}_{V_n}))| &= |\operatorname{Cov}(F(\mathfrak{X}_{U_n}), G(\mathfrak{X}_{V_n}))| \\
&\leq [u\|G\|_\infty \operatorname{Lip}(F) + v\|F\|_\infty \operatorname{Lip}(G)]\eta_{n,u,v}(s) \\
&\leq [u\|g\|_\infty \operatorname{Lip}(f) + v\|f\|_\infty \operatorname{Lip}(g)]\frac{1}{M_n}\eta_{n,u,v}(s).
\end{aligned}$$

Hence $\eta^*_{n,u,v}(s) \leq \frac{1}{M_n}\eta_{n,u,v}(s)$, and same results hold for θ-dependence as well.

□

For $k > 0$, we decompose $Y_{i,n}$ into two parts:

$$\begin{aligned}
Y_{i,n}(k) &= -k \vee Y_{i,n} \wedge k, \\
\tilde{Y}_{i,n}(k) &= Y_{i,n} - Y_{i,n}(k).
\end{aligned} \tag{A.1.10}$$

Their variances are

$$\sigma_n^2(k) = \mathrm{Var}\left[\sum_{i\in D_n} Y_{i,n}(k)\right], \qquad \tilde{\sigma}_n^2(k) = \mathrm{Var}\left[\sum_{i\in D_n} \tilde{Y}_{i,n}(k)\right].$$

Claim A.1.2. $|\sigma_n - \sigma_n(k)| \leq \tilde{\sigma}_n(k)$.

Proof. Let

$$S_n(k) = \sum_{i\in D_n} [Y_{i,n}(k) - \mathbb{E}Y_{i,n}(k)], \qquad \tilde{S}_n(k) = \sum_{i\in D_n} [\tilde{Y}_{i,n}(k) - \mathbb{E}\tilde{Y}_{i,n}(k)].$$

Note that $S_n = S_n(k) + \tilde{S}_n(k)$, $\sigma_n = \|S_n\|_2$, $\sigma_n(k) = \|S_n(k)\|_2$ and $\tilde{\sigma}_n(k) = \|\tilde{S}_n(k)\|_2$, then the inequality could be derived according to Minkowski's inequality.

□

Recalling from (A.1.9) that $\|Y\|_m := \sup_n \sup_{i\in D_n} \|Y_{i,n}\|_m < \infty$ for some $m > 2$, then for each $k > 0$,

$$\|Y(k)\|_m := \sup_n \sup_{i\in D_n} \|Y_{i,n}(k)\|_m \leq \|Y\|_m,$$

and

$$\|\tilde{Y}(k)\|_m := \sup_n \sup_{i\in D_n} \|\tilde{Y}_{i,n}(k)\|_m \leq \|Y\|_m.$$

Claim A.1.3. *There exists constants* $0 < C_* \leq C^* < \infty$ *and* $0 < N < \infty$ *such that*

$$C_*|D_n|_c \leq \sigma_n^2 \leq C^*|D_n|_c,$$

for all $n \geq N$.

Proof. (A.1.8) implies that, there exists $C_* > 0$ and $N > 0$ such that $C_*|D_n|_c \leq \sigma_n^2$ for all $n \geq N$, which proves the lower bound.

For the upper bound, according to the covariance inequalities (A.1.1) derived in

Lemma A.1.1 with $p = m$:

$$\begin{aligned}\sigma_n^2 \leq& \sum_{i\in D_n} \mathbb{E}Y_{i,n}^2 + \sum_{\substack{i,j\in D_n\\ i\neq j}} |\operatorname{Cov}(Y_{i,n}, Y_{j,n})| \\ \leq& \|Y\|_m^2 |D_n|_c + C_1 \|Y\|_m^{\frac{m}{m-1}} \sum_{\substack{i,j\in D_n\\ i\neq j}} [\bar{\epsilon}_{1,1}(\rho(i,j))]^{\frac{m-2}{m-1}} \\ \leq& \|Y\|_m^2 |D_n|_c + C_1 \|Y\|_m^{\frac{m}{m-1}} \sum_{i\in D_n} \sum_{s=1}^{\infty} \sum_{\substack{j\in D_n\\ \rho(i,j)\in[s,s+1)}} [\bar{\epsilon}_{1,1}(\rho(i,j))]^{\frac{p-2}{p-1}}.\end{aligned}$$

Lemma A.1 (iii) in Jenish and Prucha (2009) gives

$$\sup_{i\in D} |\{j \in D : \rho(i,j) \in [s, s+1)\}|_c \leq C_2 s^{d-1}$$

for $s \geq 1$. Therefore, there exists constant $C^* > 0$ such that

$$\begin{aligned}\sigma_n^2 \leq& \left\{ \|Y\|_m^2 + C_1 C_2 \|Y\|_m^{\frac{m}{m-1}} \sum_{s=1}^{\infty} s^{d-1} [\bar{\epsilon}_{1,1}(s)]^{\frac{m-2}{m-1}} \right\} |D_n|_c \\ :=& C^* |D_n|_c,\end{aligned}$$

where the last equality follows from Assumption 3.3.4(a) and Assumption 3.3.5(a).

□

Observe that $\tilde{Y}_{i,n}(k)$ is a continuous function of $Y_{i,n}$ with Lipschitz constant 1, therefore $\tilde{Y}_{i,n}(k)$ inherits the dependence coefficient from $Y_{i,n}$ according to Proposition 3.1. For each $k > 0$,

$$\begin{aligned}\tilde{\sigma}_n^2(k) \leq& \sum_{i,j\in D_n} |\operatorname{Cov}(Y_{i,n} - Y_{i,n}(k), Y_{j,n} - Y_{j,n}(k))| \\ \leq& \sum_{\substack{i,j\in D_n\\ \rho(i,j)\leq r}} [|\operatorname{Cov}(Y_{i,n}, Y_{j,n} - Y_{j,n}(k))| + |\operatorname{Cov}(Y_{i,n}(k), Y_{j,n} - Y_{j,n}(k))|] \\ &+ \sum_{\substack{i,j\in D_n\\ \rho(i,j)> r}} [|\operatorname{Cov}(Y_{i,n}, Y_{j,n} - Y_{j,n}(k))| + |\operatorname{Cov}(Y_{i,n}(k), Y_{j,n} - Y_{j,n}(k))|]\end{aligned}$$

$$\leq C_1 r^d \|Y\|_m^m k^{2-m} |D_n|_c + \left\{ C_2 \|Y\|_m^{\frac{m}{m-1}} \sum_{s=r+1}^{\infty} s^{d-1} [\bar{\epsilon}_{1,1}(s)]^{\frac{m-2}{m-1}} \right\} |D_n|_c.$$

The last inequality follows from similar arguments in the proof of Lemma A.1.1, combining with Lemma A.1 (ii), (iii) in Jenish and Prucha (2009). Let $r = k^\delta$ where $\delta \in (0, \frac{m-2}{d})$, together with the lower bound of σ_n^2 in Claim A.1.3, there exists $N > 0$ such that

$$\lim_{k\to\infty} \sup_{n\geq N} \frac{\tilde{\sigma}_n^2(k)}{\sigma_n^2} = 0. \tag{A.1.11}$$

Combining Claim A.1.2 with (A.1.11) we get

$$\lim_{k\to\infty} \sup_{n\geq N} \left| 1 - \frac{\sigma_n(k)}{\sigma_n} \right| \leq \lim_{k\to\infty} \sup_{n\geq N} \frac{\tilde{\sigma}_n(k)}{\sigma_n} = 0 \tag{A.1.12}$$

for some $N > 0$.

On the other hand, note that $Y_{i,n}(k)$ is a bounded function of $Y_{i,n}$ with Lipschitz constant 1. By (3.2.2) we have

$$\sigma_n^2(k) \leq C_1 C_2 |D_n|_c \sum_{s=0}^{\infty} s^{d-1} \bar{\epsilon}_{1,1}(s) \tag{A.1.13}$$

for each $k > 0$. With the lower bounds for σ_n^2, we have for each $k > 0$, there exists constants $N > 0$ and $C > 0$ such that

$$\frac{\sigma_n(k)}{\sigma_n} \leq C < \infty \tag{A.1.14}$$

for all $n \geq N$. This result, together with (A.1.12) play a key role in the commencing arguments.

For the next step, we will adopt Lemma A.1.2 to reduce the problem of proving CLT for $Y_{i,n}$ to the problem of proving CLT for the bounded random field $Y_{i,n}(k)$.

Claim A.1.4. *We have*

$$\sigma_n^{-1} \sum_{i\in D_n} Y_{i,n} \xrightarrow{d} N(0,1) \tag{A.1.15}$$

if

$$\sigma_n^{-1}(k) \sum_{i \in D_n} [Y_{i,n}(k) - \mathbb{E}Y_{i,n}(k)] \xrightarrow{d} N(0,1) \tag{A.1.16}$$

for each $k \in \mathbb{N}_+$.

Proof. Let $Z_n = \sigma_n^{-1} \sum_{i \in D_n} Y_{i,n}$, and $V_{n,k} = \sigma_n^{-1} \sum_{i \in D_n} [Y_{i,n}(k) - \mathbb{E}Y_{i,n}(k)]$. Let μ_n and ν be the probability measures of Z_n and V respectively. If Z_n does not converge to V in distribution, then the Lévy-Prokhorov metric $d(\mu_n, \nu)$ does not converge to 0 as $n \to \infty$, i.e. for any $\delta > 0$, there always exist sub-indices $(n_m)_{m \in \mathbb{N}_+}$ such that $d(\mu_{n_m}, \nu) > \delta$ for all n_m. Next, we will find a sub-sequence of (Z_{n_m}) such that it converges to V, contradicting with that $d(\mu_{n_m}, \nu) > \delta$ for all n_m. Recalling from (A.1.14), there exists $N > 0$, $C(k) > 0$ such that $\frac{\sigma_n(k)}{\sigma_n} \leq C(k)$ for each $k \in \mathbb{N}_+$ and all $n \geq N$. Assume that $n_m \geq N$, by Bolzano–Weierstrass theorem we have:

- For $k = 1$, there exists sub-sub-indices $(n_{m(l_1)})_{l_1 \in \mathbb{N}_+}$ such that

$$\lim_{l_1 \to \infty} \frac{\sigma_{n_{m(l_1)}}(1)}{\sigma_{n_{m(l_1)}}} = \alpha(1);$$

- For $k = 2$, there exists sub-sub-sub-indices $(n_{m(l_1(l_2))})_{l_2 \in \mathbb{N}_+}$ such that

$$\lim_{l_2 \to \infty} \frac{\sigma_{n_{m(l_1(l_2))}}(2)}{\sigma_{n_{m(l_1(l_2))}}} = \alpha(2);$$

$$\cdots$$

Now we could find a sub-sequence (n_r^*) of (n_m) by letting $n_1^* = n_{m(1)}$, $n_2^* = n_{m(l_1(2))}$, $\cdots$ such that

$$\lim_{r \to \infty} \frac{\sigma_{n_r^*}(k)}{\sigma_{n_r^*}} = \alpha(k)$$

for each $k \in \mathbb{N}_+$.

Observe that

$$V_{n_r^*,k} = \frac{\sigma_{n_r^*}(k)}{\sigma_{n_r^*}} \left\{ \sigma_{n_r^*}^{-1}(k) \sum_{i \in D_{n_r^*}} \left[Y_{i,n_r^*}(k) - \mathbb{E}Y_{i,n_r^*}(k) \right] \right\}.$$

If (A.1.16) holds, then the first condition in Lemma A.1.2 is satisfied since

$$V_{n_r^*,k} \xrightarrow{d} V_k \sim N(0, \alpha^2(k))$$

as $r \to \infty$. Recalling from (A.1.12),

$$\lim_{k\to\infty} |\alpha(k) - 1| \leq \lim_{k\to\infty} \lim_{r\to\infty} \left| \alpha(k) - \frac{\sigma_{n_r^*}(k)}{\sigma_{n_r^*}} \right| + \lim_{k\to\infty} \sup_{n\geq N} \left| \frac{\sigma_n(k)}{\sigma_n} - 1 \right| = 0,$$

hence the second condition in Lemma A.1.2 is also verified.

Using Markov's inequality,

$$\mathbb{P}(|Z_n - V_{n,k}| > \delta) = \mathbb{P}\left(\left| \sigma_n^{-1} \sum_{i\in D_n} (\tilde{Y}_{i,n}(k) - \mathbb{E}\tilde{Y}_{i,n}(k)) \right| > \delta \right) \leq \frac{\tilde{\sigma}_n^2(k)}{\delta^2 \sigma_n^2},$$

for any $\delta > 0$. Hence condition 3 in Lemma A.1.2 holds for Z_n and $V_{n,k}$ because of (A.1.11), obviously it also holds for the sub-sequences $Z_{n_r^*}$ and $V_{n_r^*,k}$.

Applying Lemma A.1.2 on the sub-sequences $Z_{n_r^*}$ and $V_{n_r^*,k}$, we have $Z_{n_r^*} \xrightarrow{d} V$ as $r \to \infty$. Since (n_r^*) is a sub-sequence of (n_m), $Z_{n_r^*} \xrightarrow{d} V$ contradicts with former assumption that Z_n does not converge weakly to V.

□

Now we consider the case when $(Y_{i,n})$ are bounded as $\sup_n \sup_{i\in D_n} |Y_{i,n}| \leq C_Y$. Let $(d_n)_{n\geq 1}$ be a sequence such that $\lim_{n\to\infty} d_n = \infty$, $\lim_{n\to\infty} \frac{d_n^d}{(|D_n|_c)^{1/2}} = 0$, and

1. $\lim_{n\to\infty} \bar{\theta}_{\infty,1}(d_n)(|D_n|_c)^{1/2} = 0$ for θ-coefficients;

2. $\lim_{n\to\infty} \bar{\eta}_{\infty,1}(d_n)(|D_n|_c) = 0$ for η-coefficients.

According to Assumption 3.3.4(b), we could set $d_n = (|D_n|_c)^p$ with $p \in (\frac{1}{2\beta}, \frac{1}{2d})$ since $\beta > d$ for case 1 above. As for case 2, we could set $d_n = (|D_n|_c)^q$ with $q \in (\frac{1}{\beta}, \frac{1}{2d})$ since $\beta > 2d$ in Assumption 3.3.5(b).

Define

$$a_n = \sum_{\substack{i,j\in D_n \\ \rho(i,j)\leq d_n}} \mathrm{Cov}(Y_{i,n}, Y_{j,n}).$$

Recalling from the covariance inequality for bounded random variables (3.2.2), there exists constant $C > 0$ such that

$$\begin{aligned}|\sigma_n^2 - a_n| &= \sum_{\substack{i,j\in D_n \\ \rho(i,j)>d_n}} |\operatorname{Cov}(Y_{i,n}, Y_{j,n})| \\ &\le \sum_{i\in D_n} \sum_{s=d_n}^{\infty} \sum_{\substack{j\in D_n \\ \rho(i,j)\in[s,s+1)}} C s^{d-1} \bar{\epsilon}_{1,1}(s) \\ &\le C C_2 |D_n|_c \sum_{s=d_n}^{\infty} s^{d-1} \bar{\epsilon}_{1,1}(s) \\ &= o(|D_n|_c).\end{aligned} \tag{A.1.17}$$

Then we have

$$0 < \liminf_{n\to\infty} (|D_n|_c)^{-1} \sigma_n^2 \le \liminf_{n\to\infty} (|D_n|_c)^{-1} a_n + \liminf_{n\to\infty} (|D_n|_c)^{-1} o(|D_n|_c).$$

Through similar arguments in the proof of Claim A.1.3, we have $\sup_{n\ge N} a_n = \mathcal{O}(|D_n|_c)$ for some $N > 0$. Consequently, $\sigma_n^2 = a_n + o(|D_n|_c) = a_n[1 + o(1)]$ for sufficiently large n. Define

$$\bar{S}_n = a_n^{-1/2} \sum_{i\in D_n} Y_{i,n} = \frac{\sigma_n}{a_n^{1/2}} \sigma_n^{-1} \sum_{i\in D_n} Y_{i,n},$$

then it remains for us to show following convergence, which could be verified using Lemma A.1.3.

Claim A.1.5. *$\bar{S}_n \xrightarrow{d} N(0,1)$ as $n \to \infty$.*

Proof. The first condition in Lemma A.1.3 is satisfied since $a_n = \mathcal{O}(|D_n|_c)$ and $\sigma_n^2 = \mathcal{O}(|D_n|_c)$ for sufficiently large n. Then it suffices to verify the second condition, i.e.

$$\lim_{n\to\infty} \mathbb{E}\left[(\mathbf{i}\lambda - \bar{S}_n) e^{\mathbf{i}\lambda \bar{S}_n}\right] = 0 \tag{A.1.18}$$

for all $\lambda \in \mathbb{R}$.

Let

$$S_{i,n} = \sum_{\substack{j\in D_n \\ \rho(i,j)\le d_n}} Y_{j,n}, \qquad \bar{S}_{i,n} = a_n^{-1/2} S_{i,n}.$$

Then we can make decomposition as follows

$$(\mathbf{i}\lambda - \bar{S}_n)e^{\mathbf{i}\lambda\bar{S}_n} = T_{1,n} + T_{2,n} + T_{3,n},$$

where

$$\begin{aligned}
T_{1,n} &= \mathbf{i}\lambda e^{\mathbf{i}\lambda\bar{S}_n}\left(1 - a_n^{-1}\sum_{i\in D_n} Y_{i,n}S_{i,n}\right),\\
T_{2,n} &= a_n^{-1/2}e^{\mathbf{i}\lambda\bar{S}_n}\sum_{i\in D_n} Y_{i,n}\left(e^{-\mathbf{i}\lambda\bar{S}_{i,n}} + \mathbf{i}\lambda\bar{S}_{i,n} - 1\right),\\
T_{3,n} &= -a_n^{-1/2}\sum_{i\in D_n} Y_{i,n}e^{\mathbf{i}\lambda(\bar{S}_n - \bar{S}_{i,n})}.
\end{aligned}$$

For the next step, we will prove that $\lim_{n\to\infty}\mathbb{E}|T_{k,n}| = 0$ for each $k = 1, 2, 3$.

We firstly consider the term $T_{1,n}$. Note that $\sum_{i\in D_n}\mathbb{E}(Y_{i,n}S_{i,n}) = a_n$, then for sufficiently large n we have

$$\begin{aligned}
\mathbb{E}|T_{1,n}|^2 =& \lambda^2\left[1 - 2a_n^{-1}\sum_{i\in D_n}\mathbb{E}(Y_{i,n}S_{i,n}) + a_n^{-2}\mathbb{E}\left(\sum_{i\in D_n} Y_{i,n}S_{i,n}\right)^2\right]\\
=& \lambda^2\left[1 - 2a_n^{-1}a_n + a_n^{-2}\operatorname{Var}\left(\sum_{i\in D_n} Y_{i,n}S_{i,n}\right) + a_n^{-2}a_n^2\right]\\
=& \lambda^2 a_n^{-2}\operatorname{Var}\left(\sum_{\substack{i,j\in D_n\\ \rho(i,j)\le d_n}} Y_{i,n}Y_{j,n}\right)\\
=& \lambda^2 a_n^{-2}\sum_{\substack{i,j,k,l\in D_n\\ \rho(i,j)\le d_n\\ \rho(k,l)\le d_n}}\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})\\
\le& C_\lambda|D_n|_c^{-2}\sum_{\substack{i,j,k,l\in D_n\\ \rho(i,j)\le d_n\\ \rho(k,l)\le d_n\\ \rho(i,k)>3d_n}}|\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})|\\
&+ C_\lambda|D_n|_c^{-2}\sum_{\substack{i,j,k,l\in D_n\\ \rho(i,j)\le d_n\\ \rho(k,l)\le d_n\\ \rho(i,k)\le 3d_n}}|\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})|,
\end{aligned}$$

for some $0 < C_\lambda < \infty$.

Define function $f_u : \mathbb{R}^u \mapsto \mathbb{R}$ as

$$f(x_1, \cdots, x_u) = -C_Y^u \vee x_1 \cdots x_u \wedge C_Y^u, \tag{A.1.19}$$

then f_u is a bounded Lipschitz function. Recalling from (3.2.2) we have

$$|\operatorname{Cov}(Y_{i,n}, Y_{j,n})| = |\operatorname{Cov}(f_1(Y_{i,n}), f_1(Y_{j,n})| \le C_1 \bar{\epsilon}_{1,1}(\rho(i,j)),$$
$$|\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})| = |\operatorname{Cov}(f_2(Y_{i,n}, Y_{j,n}), f_2(Y_{k,n}, Y_{l,n}))| \le C_2 \bar{\epsilon}_{2,2}(\rho(\{i,j\}, \{k,l\})),$$
$$|\operatorname{Cov}(Y_{i,n}, Y_{j,n}Y_{k,n}Y_{l,n})| = |\operatorname{Cov}(f_1(Y_{i,n}), f_3(Y_{j,n}, Y_{k,n}, Y_{l,n}))| \le C_3 \bar{\epsilon}_{1,3}(\rho(i, \{j,k,l\})),$$

for some positive constants C_1, C_2 and C_3.

When $\rho(i,k) > 3d_n$, we have $\rho(\{i,j\}, \{k,l\}) > \rho(i,k) - 2d_n$. Let

$$N_i(r) = |\{(j,k,l) : \rho(i,j) \le d_n, \rho(k,l) \le d_n, 3d_n < r \le \rho(i,k) < r+1\}|_c .$$

Then according to Lemma A.1(ii), (iv) in Jenish and Prucha (2009) we have $\sup_{i \in \mathbb{R}^d} N_i(r) \le C_4 d_n^{2d} r^{d-1}$ for some constant $C_4 > 0$. Then we have for each $i \in D_n$,

$$\begin{aligned}
&\sum_{\substack{j,k,l \in D_n \\ \rho(i,j) \le d_n \\ \rho(k,l) \le d_n \\ \rho(i,k) > 3d_n}} |\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})| \\
\le & C_2 \sum_{r=3d_n}^{\infty} \sup_{i \in \mathbb{R}^d} N_i(r) \bar{\epsilon}_{2,2}(r - 2d_n) \\
\le & \left[C_2 C_4 \sum_{r=3d_n}^{\infty} r^{d-1} \bar{\epsilon}_{2,2}(r - 2d_n) \right] d_n^{2d} \\
\le & \left[C_2 C_4 3^{d-1} \sum_{r=d_n}^{\infty} r^{d-1} \bar{\epsilon}_{2,2}(r) \right] d_n^{2d}.
\end{aligned}$$

Therefore, there exists constant $C_5 > 0$ such that

$$\sum_{\substack{i,j,k,l \in D_n \\ \rho(i,j) \le d_n \\ \rho(k,l) \le d_n \\ \rho(i,k) > 3d_n}} |\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})| \le C_5 |D_n|_c d_n^{2d}. \tag{A.1.20}$$

When $\rho(i,k) \le 3d_n$, let $V_i(r)$ be a ball centered at i with radius of r, then $V_i(4d_n)$ includes all (j,k,l) such that $\rho(i,j) \le d_n$, $\rho(k,l) \le d_n$ and $\rho(i,k) \le 3d_n$. Let

$$M_i(r) = |\{(j,k,l) : j,k,l \in V_i(4d_n), r \le \rho(i,\{j,k,l\}) < r+1\}|_c .$$

Then by Lemma A.1(ii), (v) in Jenish and Prucha (2009) we have $\sup_{i \in \mathbb{R}^d} M_i(r) \le C_6 d_n^{2d} r^{d-1}$ for some constant $C_6 > 0$. Then for each $i \in D_n$,

$$\begin{aligned}
&\sum_{\substack{j,k,l \in D_n \\ \rho(i,j) \le d_n \\ \rho(k,l) \le d_n \\ \rho(i,k) \le 3d_n}} |\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})| \\
\le & \sum_{j,k,l \in V_i(4d_n)} |\operatorname{Cov}(Y_{i,n}Y_{j,n}, Y_{k,n}Y_{l,n})| \\
\le & \sum_{j,k,l \in V_i(4d_n)} [|\mathbb{E}(Y_{i,n}Y_{j,n}Y_{k,n}Y_{l,n})| + |\mathbb{E}(Y_{i,n}Y_{j,n})||\mathbb{E}(Y_{k,n}Y_{l,n})|] \\
\le & \sum_{j,k,l \in V_i(4d_n)} \left[C_3 \bar{\epsilon}_{1,3}(\rho(i,\{j,k,l\})) + C_1^2 \bar{\epsilon}_{1,1}(\rho(i,\{j,k,l\}))\bar{\epsilon}_{1,1}(\rho(d_n))\right] \\
\le & (C_3 + C_1^2) \sum_{j,k,l \in V_i(4d_n)} \bar{\epsilon}_{1,3}(\rho(i,\{j,k,l\})) \\
\le & (C_3 + C_1^2) \sum_{r=1}^{4d_n} M_i(r) \bar{\epsilon}_{1,3}(r) \\
\le & (C_3 + C_1^2) C_6 d_n^{2d} \sum_{r=1}^{4d_n} r^{d-1} \bar{\epsilon}_{1,3}(r).
\end{aligned}$$

By Assumption 3.3.4(a) and Assumption 3.3.5(a) we have

$$\sum_{\substack{i,j,k,l\in D_n\\ \rho(i,j)\le d_n\\ \rho(k,l)\le d_n\\ \rho(i,k)\le 3d_n}} |\operatorname{Cov}(Y_{i,n}Y_{j,n},Y_{k,n}Y_{l,n})| \le C_7|D_n|_c d_n^{2d} \tag{A.1.21}$$

for some $C_7 > 0$.

Note that $\lim_{n\to\infty}\frac{d_n^{2d}}{|D_n|_c} = 0$, then (A.1.20) and (A.1.21) imply that

$$\mathbb{E}|T_{1,n}|^2 \le C_\lambda(C_5+C_7)\frac{d_n^{2d}}{|D_n|_c} \to 0$$

as $n\to\infty$.

Now we consider the second term, for sufficiently large n we have

$$\begin{aligned}|T_{2,n}| =&|a_n^{-1/2}||\sum_{i\in D_n} Y_{i,n}\left(e^{-\mathbf{i}\lambda\bar{S}_{i,n}}+\mathbf{i}\lambda\bar{S}_{i,n}-1\right)|\\ \le& C_8(|D_n|_c)^{-1/2}C_Y\sum_{i\in D_n}|e^{-\mathbf{i}\lambda\bar{S}_{i,n}}+\mathbf{i}\lambda\bar{S}_{i,n}-1|\end{aligned}$$

for some constant $C_8 > 0$. Note that

$$\begin{aligned}|\bar{S}_{i,n}| \le& a_n^{-1/2}\sum_{\substack{j\in D_n\\ \rho(i,j)\le d_n}}|Y_{j,n}|\\ \le& C_9 C_Y a_n^{-1/2}d_n^d\\ =&\mathcal{O}((|D_n|_c)^{-1/2})d_n^d.\end{aligned}$$

for some $C_9 > 0$. The second inequality adopts Lemma A.1(ii) in Jenish and Prucha (2009). Then $\lim_{n\to\infty}|\bar{S}_{i,n}| = 0$, hence $|\mathbf{i}\lambda\bar{S}_{i,n}| < 1/2$ for sufficiently large n. Since $|e^{-z}+z-1| \le |z|^2$ for complex number $|z| < 1/2$, $|e^{-\mathbf{i}\lambda\bar{S}_{i,n}}+\mathbf{i}\lambda\bar{S}_{i,n}-1| \le \lambda^2|\bar{S}_{i,n}|^2$ a.s. for sufficiently large n.

Now we have

$$
\begin{aligned}
\mathbb{E}|T_{2,n}| \leq & C_8(|D_n|_c)^{-1/2} C_Y \sum_{i \in D_n} \lambda^2 \mathbb{E}|\bar{S}_{i,n}|^2 \\
\leq & C_8(|D_n|_c)^{1/2} C_Y \lambda^2 \sup_{i \in D_n} \mathbb{E}|\bar{S}_{i,n}|^2 \\
\leq & C_8(|D_n|_c)^{1/2} C_Y \lambda^2 a_n^{-1} \sup_{i \in D_n} \sum_{\substack{j,k \in D_n \\ \rho(i,j) \leq d_n \\ \rho(i,k) \leq d_n}} \mathbb{E}|Y_{j,n} Y_{k,n}| \\
\leq & C_{10}((|D_n|_c)^{-1/2}) \sup_{i \in D_n} \sum_{\substack{j,k \in D_n \\ \rho(i,j) \leq d_n \\ \rho(i,k) \leq d_n}} \bar{\epsilon}_{1,1}(\rho(j,k)) \\
\leq & C_{10}((|D_n|_c)^{-1/2}) \sup_{i \in D_n} \sum_{\substack{j \in D_n \\ \rho(i,j) \leq d_n}} \sum_{r=1}^{2d_n} N_j(r) \bar{\epsilon}_{1,1}(r) \\
\leq & C_{11}((|D_n|_c)^{-1/2}) d_n^d \sum_{r=1}^{2d_n} r^{d-1} \bar{\epsilon}_{1,1}(r) \\
\leq & C_{12}(|D_n|_c)^{-1/2} d_n^d,
\end{aligned}
$$

where $N_j(r) = |\{i : r \leq \rho(i,j) < r+1\}|_c$, the last two inequalities come from Lemma A.1 (iii) in Jenish and Prucha (2009), Assumption 3.3.4(a) and Assumption 3.3.5(a). Then $\lim_{n\to\infty} \mathbb{E}|T_{2,n}| = 0$.

As for the third term, we want to prove that $\lim_{n\to\infty} |\mathbb{E}T_{3,n}| = 0$. Firstly note that

$$
\begin{aligned}
|\mathbb{E}T_{3,n}| = & \left| \mathbb{E}\left[a_n^{-1/2} \sum_{i \in D_n} Y_{i,n} e^{\mathbf{i}\lambda(\bar{S}_n - \bar{S}_{i,n})} \right] \right| \\
\leq & C(|D_n|_c)^{-1/2} \sum_{i \in D_n} \left| \mathbb{E} Y_{i,n} e^{\mathbf{i}\lambda(\bar{S}_n - \bar{S}_{i,n})} \right| \\
\leq & C(|D_n|_c)^{-1/2} \sum_{i \in D_n} \left(\left|\mathbb{E}Y_{i,n} \cos\lambda(\bar{S}_n - \bar{S}_{i,n})\right| + \left|\mathbb{E}Y_{i,n} \sin\lambda(\bar{S}_n - \bar{S}_{i,n})\right| \right).
\end{aligned}
$$

Let $f^*(\mathfrak{Y}_i) = \cos\lambda(\bar{S}_n - \bar{S}_{i,n})$ where $\mathfrak{Y}_i = (Y_{j,n})_{j \in D_n, \rho(i,j) > d_n}$. f^* is bounded with Lipschitz constant $\mathrm{Lip}(f^*) = |\lambda| {a_n}^{-1/2}$, with domain $\mathbb{R}^u$ for some $u \leq |D_n|_c$. Another bounded Lipschitz function f_1 is defined in (A.1.19). By (3.2.3) we have:

$$\begin{aligned}|\mathbb{E}[Y_{i,n}\cos\lambda(\bar{S}_n-\bar{S}_{i,n})]| =& |\operatorname{Cov}[f^*(\mathfrak{Y}_i), f_1(Y_{i,n})]| \\ \leq& \bar{\theta}_{\infty,1}(d_n).\end{aligned}$$

Same holds if $f^*(\mathfrak{Y}_i)=\sin\lambda(\bar{S}_n-\bar{S}_{i,n})$. Since $\lim_{n\to\infty}\bar{\theta}_{\infty,1}(d_n)(|D_n|_c)^{1/2}=0$, we have

$$|\mathbb{E}T_{3,n}| \leq C(|D_n|_c)^{-1/2}\sum_{i\in D_n}\bar{\theta}_{\infty,1}(d_n) \leq C(|D_n|_c)^{1/2}\bar{\theta}_{\infty,1}(d_n) \to 0$$

as $n\to\infty$. Similarly by (3.2.4) we have:

$$\begin{aligned}|\mathbb{E}[Y_{i,n}\cos\lambda(\bar{S}_n-\bar{S}_{i,n})]| =& |\operatorname{Cov}[f^*(\mathfrak{Y}_i), f_1(Y_{i,n})]| \\ \leq& (C_Y|D_n|_c|\lambda|a_n^{-1/2}+1)\bar{\eta}_{\infty,1}(d_n).\end{aligned}$$

Same holds if $f^*(\mathfrak{Y}_i)=\sin\lambda(\bar{S}_n-\bar{S}_{i,n})$. Since $\lim_{n\to\infty}\bar{\eta}_{\infty,1}(d_n)|D_n|_c=0$, we have

$$|\mathbb{E}T_{3,n}| \leq C(|D_n|_c)^{-1/2}\sum_{i\in D_n}(|D_n|_c)^{1/2}\bar{\eta}_{\infty,1}(d_n) \leq C|D_n|_c\bar{\eta}_{\infty,1}(d_n) \to 0$$

as $n\to\infty$.

$\square$

The proof of Theorem 3.2 is completed as Claim A.1.5 is verified.

A.1.4 Proof of results in Section 3.4

Proof of Proposition 3.3

For any $r>0$ and $U\subset D_{NT}$, let $\mathfrak{X}_U=(X_{it})_{(i,t)\in U}$ and $\mathfrak{X}_U^{(r)}=(X_{it}^{(r)})_{(i,t)\in U}$. $f\in\mathcal{F}_u, g\in\mathcal{G}_v$ are two arbitrary bounded Lipschitz functions, then for any $V\subset D_{NT}$ such that $\rho(U,V)>2r$, $f(\mathfrak{X}_U^{(r)})$ is independent from $g(\mathfrak{X}_V^{(r)})$. By Assumption 3.4.2 we have

$$\begin{aligned}&|\operatorname{Cov}[f(\mathfrak{X}_U), g(\mathfrak{X}_V)]| \\ \leq& \left|\operatorname{Cov}\left[f(\mathfrak{X}_U)-f(\mathfrak{X}_U^{(r)}), g(\mathfrak{X}_V)\right]\right| + \left|\operatorname{Cov}\left[f(\mathfrak{X}_U^{(r)}), g(\mathfrak{X}_V)-g(\mathfrak{X}_V^{(r)})\right]\right|\end{aligned}$$

$$\le 2\left\|g\right\|_\infty \mathbb{E}\left|f(\mathfrak{X}_U)-f(\mathfrak{X}_U^{(r)})\right|+2\left\|f\right\|_\infty \mathbb{E}\left|g(\mathfrak{X}_V)-g(\mathfrak{X}_V^{(r)})\right|$$
$$\le 2\left\|g\right\|_\infty \mathrm{Lip}(f)\sum_{(i,t)\in U}\mathbb{E}\left|X_{it}-X_{it}^{(r)}\right|+2\left\|f\right\|_\infty \mathrm{Lip}(g)\sum_{(i,t)\in V}\mathbb{E}\left|X_{it}-X_{it}^{(r)}\right|$$
$$\le 2C_1\left[u\left\|g\right\|_\infty \mathrm{Lip}(f)+v\left\|f\right\|_\infty \mathrm{Lip}(g)\right]\delta(r),$$

for some constant $C_1>0$. Therefore, $\bar{\eta}(s)\le C_2\delta(s/2)$ for some constant $C_2>0$ by letting $s=2r+1$.

Proof of Proposition 3.4

Assumption 3.4.5 allows us to adopt Theorem 3.1 on functions $l_{it}(\theta)$. i.e. for any $\theta\in\Theta$,

$$\frac{1}{NT}\sum_{(it)\in D_{NT}}\left[l_{it}(\theta)-\mathbb{E}l_{it}(\theta)\right]\xrightarrow{p}0 \tag{A.1.22}$$

as $NT\to\infty$. Together with Assumption 3.4.3(c) we have

$$\begin{aligned} & \lim_{NT\to\infty}[L_{NT}(\theta)-L_{NT}(\theta_0)] \\ = & \lim_{NT\to\infty}\{\mathbb{E}[L_{NT}(\theta)]-\mathbb{E}[L_{NT}(\theta_0)]\} \\ \le & 0, \end{aligned} \tag{A.1.23}$$

and the equality holds only if $\theta=\theta_0$, which means θ_0 is uniquely identifiable.

Note that Assumption 3.4.4 implies that

$$\lim_{NT\to\infty}\mathbb{P}\left[|L_{NT}(\hat{\theta}_{NT})-\tilde{L}_{NT}(\hat{\theta}_{NT})|<\frac{\delta}{3}\right]=1$$

for any $\delta>0$, hence

$$\lim_{NT\to\infty}\mathbb{P}\left[L_{NT}(\hat{\theta}_{NT})>\tilde{L}_{NT}(\hat{\theta}_{NT})-\frac{\delta}{3}\right]=1.$$

Since $\hat{\theta}_{NT}$ maximizes $\tilde{L}_{NT}(\theta)$, we have

$$\lim_{NT\to\infty}\mathbb{P}\left[\tilde{L}_{NT}(\hat{\theta}_{NT})>\tilde{L}_{NT}(\theta_0)-\frac{\delta}{3}\right]=1.$$

So

$$\lim_{NT\to\infty} \mathbb{P}\left[L_{NT}(\hat{\theta}_{NT}) > \tilde{L}_{NT}(\theta_0) - \frac{2\delta}{3}\right] = 1.$$

Furthermore, from Assumption 3.4.4,

$$\lim_{NT\to\infty} \mathbb{P}\left[\tilde{L}_{NT}(\theta_0) > L_{NT}(\theta_0) - \frac{\delta}{3}\right] = 1.$$

Therefore we have

$$\lim_{NT\to\infty} \mathbb{P}\left[0 \leq L_{NT}(\theta_0) - L_{NT}(\hat{\theta}_{NT}) < \delta\right] = 1. \tag{A.1.24}$$

Let $V_k(\theta)$ be an open sphere with centre θ and radius $1/k$. Note that $L_{NT}(\theta)$ is continuous in θ and $\Theta \setminus V_k(\theta_0)$ is a closed set according to Assumption 3.4.3. By (A.1.23), we could find

$$\delta = \inf_{\theta\in\Theta\setminus V_k(\theta_0)} \left[L_{NT}(\theta_0) - L_{NT}(\theta)\right] > 0.$$

Then by (A.1.24),

$$\lim_{NT\to\infty} \mathbb{P}\left\{0 \leq L_{NT}(\theta_0) - L_{NT}(\hat{\theta}_{NT}) < \inf_{\theta\in\Theta\setminus V_k(\theta_0)} \left[L_{NT}(\theta_0) - L_{NT}(\theta)\right]\right\} = 1.$$

This implies that

$$\lim_{NT\to\infty} \mathbb{P}\left[\hat{\theta}_{NT} \in V_k(\theta_0)\right] = 1$$

for any given $k > 0$, which means $\hat{\theta}_{NT} \xrightarrow{p} \theta_0$ as $NT \to \infty$.

Proof of Proposition 3.5

Based on Assumptions 3.4.7(a) and 3.4.7(b), Theorem 3.1 facilitates the convergence

$$\frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left\{\frac{\partial^2}{\partial\theta\partial\theta'} l_{it}(\theta_0) - \mathbb{E}\left[\frac{\partial^2}{\partial\theta\partial\theta'} l_{it}(\theta_0)\right]\right\} \xrightarrow{p} 0$$

as $NT \to \infty$, hence

$$\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'} + A_{NT} \xrightarrow{p} 0. \tag{A.1.25}$$

By Assumption 3.4.7(e) we have

$$-B_{NT}^{-1/2}\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}A_{NT}^{-1}B_{NT}^{1/2} = I_k + o_p(1). \tag{A.1.26}$$

On the other hand, with Assumptions 3.4.7(c), 3.4.7(d) and 3.4.7(e), we can prove that

$$\sqrt{NT}B_{NT}^{-1/2}\frac{\partial L_{NT}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, I_k). \tag{A.1.27}$$

By the Taylor expansion, for some θ^* between $\hat{\theta}_{NT}$ and θ_0 we have

$$\frac{\partial \tilde{L}_{NT}(\hat{\theta}_{NT})}{\partial\theta} = \frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial\theta} + \frac{\partial^2 \tilde{L}_{NT}(\theta^*)}{\partial\theta\partial\theta'}(\hat{\theta}_{NT} - \theta_0).$$

Since $\frac{\partial \tilde{L}_{NT}(\hat{\theta}_{NT})}{\partial\theta} = 0$, we have

$$\begin{aligned}
&\sqrt{NT}(B_{NT}^{-1/2}A_{NT})(\hat{\theta}_{NT} - \theta_0)\\
=& -(B_{NT}^{-1/2}A_{NT})\left(\frac{\partial^2 \tilde{L}_{NT}(\theta^*)}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{NT}\frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial\theta}\\
=& -(B_{NT}^{-1/2}A_{NT})\left(\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}B_{NT}^{1/2}\sqrt{NT}B_{NT}^{-1/2}\frac{\partial L_{NT}(\theta_0)}{\partial\theta} + o_p(1),
\end{aligned}$$

according to Assumption 3.4.6 and the fact that $\hat{\theta}_{NT} \xrightarrow{p} \theta_0$. Therefore, combining (A.1.26) with (A.1.27), we can prove the asymptotic distribution of $\hat{\theta}_{NT}$ as follows:

$$\sqrt{NT}(B_{NT}^{-1/2}A_{NT})(\hat{\theta}_{NT} - \theta_0) \xrightarrow{d} N(0, I_k).$$

Proof of Proposition 3.6

By (3.4.8) we have

$$\sqrt{NT}(\hat{\theta}_{NT} - \theta_0) = -\left(\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta},$$

with

$$\frac{\partial L_{NT}(\theta_0)}{\partial \theta} = \frac{2}{NT} \sum_{(i,t)\in D_{NT}} \varepsilon_{it}\mathbf{x}_{i,t-1},$$
$$\frac{\partial^2 L_{NT}(\theta_0)}{\partial \theta \partial \theta'} = -\frac{2}{NT} \sum_{(i,t)\in D_{NT}} \mathbf{x}_{i,t-1}\mathbf{x}'_{i,t-1}.$$

To prove Proposition 3.6, it suffices to verify following statements:

(i). $\frac{1}{\sqrt{NT}}\Sigma_{NT}^{-1/2}\sum_{(i,t)\in D_{NT}} \varepsilon_{it}\mathbf{x}_{i,t-1} \xrightarrow{d} N(0,\sigma^2)$;

(ii). $\frac{1}{NT}\sum_{(i,t)\in D_{NT}} \left[\mathbf{x}_{i,t-1}\mathbf{x}'_{i,t-1} - \mathbb{E}\left(\mathbf{x}_{i,t-1}\mathbf{x}'_{i,t-1}\right)\right] \xrightarrow{p} 0.$

To prove (i), we will prove in Claim A.1.7 that $\{\varepsilon_{it}\mathbf{x}_{i,t-1} : (i,t) \in D_{NT}, NT \geq 1\}$ satisfies the conditions of Corollary 3.2.1. Particularly in proving weak dependence, we will make use of Proposition 3.3, hence we need to prove Claim A.1.6 at first. Notice that the weak dependence and asymptotic properties are derived conditioning on $\mathcal{Z}$ in this proof.

Claim A.1.6. *For any $s \geq 0$, let $\mathcal{F}_{it}(s) = \sigma\{\varepsilon_{j\tau} : |i-j| \leq s, |\tau - t| \leq s\}$. Under Assumptions 3.4.8, 3.4.9 and 3.4.10 we have*

$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}} \|y_{it} - \mathbb{E}\left[y_{it}|\mathcal{F}_{it}(s)\right]\|_2 \leq C\delta(s) \tag{A.1.28}$$

and

$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}} \left\|\sum_{j=1}^{N} w_{ij}y_{jt} - \mathbb{E}\left[\sum_{j=1}^{N} w_{ij}y_{jt}|\mathcal{F}_{it}(s)\right]\right\|_2 \leq C\delta(s) \tag{A.1.29}$$

with $\delta(s) = \mathcal{O}(s^{-\alpha})$ for some $\alpha > 4 \vee \frac{2p-2}{p-2}$.

Proof. Let $\mathcal{E}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})$ and $Z = (Z_1, Z_2, ..., Z_N)'$. By (2.3) in Zhu et al. (2017), under Assumption 3.4.8(b) we can rewrite (3.4.4) as

$$y_{it} = \mathbf{e}_i'\left[(I_N - G)^{-1}\mathcal{B}_0 + \sum_{k=0}^{\infty} G^k \mathcal{E}_{t-k}\right],$$

where $\mathcal{B}_0 = \beta_0\mathbf{1}_N + Z\gamma$, $G = \beta_1 W + \beta_2 I_N$, I_N is an $N \times N$ identity matrix, $\mathbf{1}_N$ is an N-dimensional vector with all elements being 1, and $\mathbf{e}_i$ is an N-dimensional vector with the i-th element being 1 and others being zero. Then we have

$$\begin{aligned}
&\|y_{it} - \mathbb{E}\left[y_{it}|\mathcal{F}_{it}(s)\right]\|_2 \\
\leq &\left\| \mathbf{e}_i' \left\{ \sum_{k=0}^{\infty} G^k \mathcal{E}_{t-k} - \sum_{k=0}^{\infty} \mathbb{E}\left[G^k \mathcal{E}_{t-k} | \mathcal{F}_{it}(s) \right] \right\} \right\|_2 \\
\leq &\left\| \mathbf{e}_i' \left\{ \sum_{k=0}^{s} \left[G^k \mathcal{E}_{t-k} - \mathbb{E}(G^k \mathcal{E}_{t-k} | \mathcal{F}_{it}(s)) \right] \right\} \right\|_2 \\
&+ \left\| \mathbf{e}_i' \left\{ \sum_{k=s+1}^{\infty} \left[G^k \mathcal{E}_{t-k} - \mathbb{E}(G^k \mathcal{E}_{t-k} | \mathcal{F}_{it}(s)) \right] \right\} \right\|_2 \\
= &T_1 + T_2.
\end{aligned}$$

Note that $\mathcal{E}_{t-k}$ is independent from $\mathcal{F}_{it}(s)$ when $k > s$. Then by Assumption 3.4.9(b) we have

$$\begin{aligned}
T_2^2 =&\mathbb{E} \left| \mathbf{e}_i' \sum_{k=s+1}^{\infty} G^k \mathcal{E}_{t-k} \right|^2 \\
=&\mathbb{E} \left[\left(\sum_{k=s+1}^{\infty} \mathbf{e}_i' G^k \mathcal{E}_{t-k} \right) \left(\sum_{k=s+1}^{\infty} \mathcal{E}_{t-k}' (G')^k \mathbf{e}_i \right) \right] \\
=&\sigma^2 \sum_{k=s+1}^{\infty} \mathbf{e}_i' (GG')^k \mathbf{e}_i \\
\leq &C \sum_{k=s+1}^{\infty} \rho^k,
\end{aligned}$$

which converges to zero exponentially as $s \to \infty$ since $0 < \rho < 1$.

Moreover, by Assumption 3.4.9(a),

$$\begin{aligned}
T_1 =& \left\| \sum_{k=0}^{s} \sum_{j=1}^{N} \mathbf{e}_i' G^k \mathbf{e}_j \left[\varepsilon_{j,t-k} - \mathbb{E}(\varepsilon_{j,t-k} | \mathcal{F}_{it}(s)) \right] \right\|_2 \\
\leq & \sum_{k=0}^{s} \sum_{j=1}^{N} \left\| \mathbf{e}_i' G^k \mathbf{e}_j \left[\varepsilon_{j,t-k} - \mathbb{E}(\varepsilon_{j,t-k} | \mathcal{F}_{it}(s)) \right] \right\|_2 \\
=& \sum_{k=0}^{s} \sum_{|j-i|>s} \left\| \mathbf{e}_i' G^k \mathbf{e}_j \varepsilon_{j,t-k} \right\|_2
\end{aligned}$$

$$\leq C \sum_{k=0}^{s} \rho^k \sum_{|j-i|>s} |j-i|^{-\alpha-2}.$$

According to Lemma A.1.4 and Lemma A.1.5, we have

$$\begin{aligned}\sum_{|j-i|>s} |j-i|^{-\alpha-2} &= \sum_{h=[s]}^{\infty} \sum_{h\leq|j-i|<h+1} |j-i|^{-\alpha-2} \\ &\leq \sum_{h=[s]}^{\infty} Ch^{-\alpha-1} \\ &< C\frac{2^{\alpha+1}}{\alpha} s^{-\alpha}.\end{aligned}$$

Therefore we complete the proof of (A.1.28).

Now we prove (A.1.29). According to Assumption 3.4.9(a), we can verify that

$$\max_{i\neq j} w_{ij} < C|j-i|^{-\alpha-2}.$$

Based on (A.1.28), Lemma A.1.4 and Lemma A.1.5, we have

$$\begin{aligned}&\left\| \sum_{j=1}^{N} w_{ij}y_{jt} - \mathbb{E}\left[\sum_{j=1}^{N} w_{ij}y_{jt}|\mathcal{F}_{it}(s)\right]\right\|_2 \\ \leq& \sum_{|j-i|\leq s/2} w_{ij}\left\|y_{jt} - \mathbb{E}\left[y_{jt}|\mathcal{F}_{it}(s)\right]\right\|_2 + \sum_{|j-i|>s/2} w_{ij}\left\|y_{jt} - \mathbb{E}\left[y_{jt}|\mathcal{F}_{it}(s)\right]\right\|_2 \\ \leq& \sum_{|j-i|\leq s/2} w_{ij}\left\|y_{jt} - \mathbb{E}\left[y_{jt}|\mathcal{F}_{jt}(s/2)\right]\right\|_2 + \sum_{h=[s/2]}^{\infty} \sum_{h\leq|j-i|<h+1} w_{ij}\left\|y_{jt}\right\|_2 \\ \leq& C_1(s/2)^{-\alpha} + C_2(s/2)^{-\alpha}.\end{aligned}$$

□

Claim A.1.7. *Under Assumptions 3.4.8, 3.4.9 and 3.4.10,*

(a). $\sup_{NT\geq1}\sup_{(i,t)\in D_{NT}} \mathbb{E}\left\|\varepsilon_{it}\mathbf{x}_{i,t-1}\right\|^p < \infty$ *for some* $p > 2$*;*

(b). $\{\varepsilon_{it}\mathbf{x}_{i,t-1} : (i,t) \in D_{NT}, NT \geq 1\}$ *are* η*-weakly dependent with* $\bar{\eta}(s) = \mathcal{O}(s^{-\alpha})$ *for some* $\alpha > 4 \vee \frac{2p-2}{p-2}$.

Proof. Claim A.1.7(a) can be easily derived from Assumption 3.4.8(a) and Assumption 3.4.10. As for Claim A.1.7(b), notice that

$$\varepsilon_{it}\mathbf{x}_{i,t-1} = \begin{pmatrix} \varepsilon_{it} \\ \varepsilon_{it}\sum_{j=1}^N w_{ij}y_{j,t-1} \\ \varepsilon_{it}y_{i,t-1} \\ \varepsilon_{it}Z_i \end{pmatrix}.$$

Since

$$\begin{aligned} &\mathbb{E}\left\|\varepsilon_{it}\mathbf{x}_{i,t-1} - \mathbb{E}(\varepsilon_{it}\mathbf{x}_{i,t-1}|\mathcal{F}_{it}(s))\right\|^2 \\ \leq & 2\mathbb{E}\left|\varepsilon_{it}\left[\sum_{j=1}^N w_{ij}y_{j,t-1} - \mathbb{E}\left(\sum_{j=1}^N w_{ij}y_{j,t-1}|\mathcal{F}_{it}(s)\right)\right]\right|^2 \\ &+ 2\mathbb{E}\left|\varepsilon_{it}\left[y_{i,t-1} - \mathbb{E}\left(y_{i,t-1}|\mathcal{F}_{it}(s)\right)\right]\right|^2. \end{aligned}$$

Then by Claim A.1.6 and Proposition 3.3 we complete the proof. □

Notice that $\mathbf{x}_{i,t-1}\mathbf{x}_{i,t-1}'$ is a $(m+3)\times(m+3)$ matrix as follows:

$$\begin{pmatrix} 1 & \sum_{j=1}^N w_{ij}y_{j,t-1} & y_{i,t-1} & Z_i' \\ \sum_{j=1}^N w_{ij}y_{j,t-1} & \left(\sum_{j=1}^N w_{ij}y_{j,t-1}\right)^2 & \left(\sum_{j=1}^N w_{ij}y_{j,t-1}\right)y_{i,t-1} & \left(\sum_{j=1}^N w_{ij}y_{j,t-1}\right)Z_i' \\ y_{i,t-1} & \left(\sum_{j=1}^N w_{ij}y_{j,t-1}\right)y_{i,t-1} & y_{i,t-1}^2 & y_{i,t-1}Z_i' \\ Z_i & \left(\sum_{j=1}^N w_{ij}y_{j,t-1}\right)Z_i & y_{i,t-1}Z_i & Z_iZ_i' \end{pmatrix}.$$

To prove statement (ii), we need to verify that each element of $\mathbf{x}_{i,t-1}\mathbf{x}_{i,t-1}$ satisfies the conditions of Theorem 3.1. By Assumption 3.4.8(a), Assumption 3.4.10 and Claim A.1.6, LLN already holds for elements $\sum_{j=1}^N w_{ij}y_{j,t-1}$, $y_{i,t-1}$, $\left(\sum_{j=1}^N w_{ij}y_{j,t-1}\right)Z_i$ and $y_{i,t-1}Z_i$. The LLN of the rest of the elements in $\mathbf{x}_{i,t-1}\mathbf{x}_{i,t-1}$ will be proved with the support of Claim A.1.8 below.

Claim A.1.8. *Under Assumptions 3.4.8, 3.4.9 and 3.4.10, following arrays of random fields*

$$\left\{y_{it}^2 : (i,t) \in D_{NT}, NT \geq 1\right\},$$

$$\left\{\left(\sum_{j=1}^{N} w_{ij}y_{jt}\right)^2 : (i,t) \in D_{NT}, NT \geq 1\right\},$$
$$\left\{\left(\sum_{j=1}^{N} w_{ij}y_{jt}\right) y_{it} : (i,t) \in D_{NT}, NT \geq 1\right\}$$

are η-dependent with $\bar{\eta}(s) = \mathcal{O}(s^{-\mu})$ for some $\mu > 2$.

Proof. By triangle inequality and Cauchy-Schwartz inequality we have

$$\begin{aligned}
&\left\|\left(\sum_{j=1}^{N} w_{ij}y_{jt}\right) y_{it} - \mathbb{E}\left[\left(\sum_{j=1}^{N} w_{ij}y_{jt}\right) y_{it}|\mathcal{F}_{it}(s)\right]\right\|_1 \\
\leq &\left\|\left(\sum_{j=1}^{N} w_{ij}y_{jt}\right) y_{it} - \left(\sum_{j=1}^{N} w_{ij}y_{jt}\right)\mathbb{E}\left[y_{it}|\mathcal{F}_{it}(s)\right]\right\|_1 \\
&+ \left\|\left(\sum_{j=1}^{N} w_{ij}y_{jt}\right)\mathbb{E}\left[y_{it}|\mathcal{F}_{it}(s)\right] - \mathbb{E}\left[\sum_{j=1}^{N} w_{ij}y_{jt}|\mathcal{F}_{it}(s)\right]\mathbb{E}\left[y_{it}|\mathcal{F}_{it}(s)\right]\right\|_1 \\
&+ \left\|\mathbb{E}\left\{\left[y_{it} - \mathbb{E}(y_{it}|\mathcal{F}_{it}(s))\right]\left[\sum_{j=1}^{N} w_{ij}y_{jt} - \mathbb{E}\left[\sum_{j=1}^{N} w_{ij}y_{jt}|\mathcal{F}_{it}(s)\right]\right]|\mathcal{F}_{it}(s)\right\}\right\|_1 \\
\leq &\left\|\sum_{j=1}^{N} w_{ij}y_{jt}\right\|_2 \left\|y_{it} - \mathbb{E}(y_{it}|\mathcal{F}_{it}(s))\right\|_2 \\
&+ \left\|\mathbb{E}(y_{it}|\mathcal{F}_{it}(s))\right\|_2 \left\|\sum_{j=1}^{N} w_{ij}y_{jt} - \mathbb{E}\left[\sum_{j=1}^{N} w_{ij}y_{jt}|\mathcal{F}_{it}(s)\right]\right\|_2 \\
&+ \left\|y_{it} - \mathbb{E}(y_{it}|\mathcal{F}_{it}(s))\right\|_2 \left\|\sum_{j=1}^{N} w_{ij}y_{jt} - \mathbb{E}\left[\sum_{j=1}^{N} w_{ij}y_{jt}|\mathcal{F}_{it}(s)\right]\right\|_2 .
\end{aligned}$$

Then by (A.1.28), (A.1.29) and Proposition 3.3, the array of random fields

$$\left\{\left(\sum_{j=1}^{N} w_{ij}y_{jt}\right) y_{it} : (i,t) \in D_{NT}, NT \geq 1\right\}$$

is η-dependent with $\bar{\eta}(s) = \mathcal{O}(s^{-\alpha})$ for $\alpha > 4 \vee \frac{2p-2}{p-2} > 2$. Using similar arguments we can also verify the η-dependence of the other two in Claim A.1.8. $\square$

With statements (i) and (ii) we complete the proof of Proposition 3.6.

A.2 Proofs of results in Chapter 4

Lemma A.1.4 and Lemma A.1.5 are needed in the proof of Lemma 4.3.1. In the proofs of asymptotic properties of MLE, we rely on Proposition 3.3 to prove η-weak dependence by verifying near-epoch-dependence instead (see the definition of near-epoch-dependence in Section 2.4.2). A useful property of near-epoch dependence is that it is preserved under summation, multiplication and finite shift, as what will be shown respectively in Lemma A.2.1 to Lemma A.2.3 below. For the proof of these lemmas we refer to Davidson (1994). Comparing to the AR-type models, the likelihood functions in GARCH-type cases are evaluated in a iterative way. Therefore we also need Lemma A.2.4 at last, followed by its proof.

Lemma A.2.1. *If $\{x_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ and $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ are $\mathbb{L}^p$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-μ_x and size-μ_y respectively. Then $\{x_{it}+y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is $\mathbb{L}^p$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-$\min\{\mu_x, \mu_y\}$.*

Lemma A.2.2. *If $\{x_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ and $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ are $\mathbb{L}^2$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-μ.*

(a). $\{x_{it}y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is $\mathbb{L}^1$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-μ;

(b). $\{x_{it}y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is $\mathbb{L}^2$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-$\frac{r-2}{2r-2}\mu$, if $\sup_{i,t} \|x_{it}\|_{2r} < \infty$ and $\sup_{i,t} \|y_{it}\|_{2r} < \infty$ for some $r > 2$;

(c). $\{x_{it}y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is $\mathbb{L}^2$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-μ, if $\sup_{i,t} \|x_{it}\| < \infty$ and $\sup_{i,t} \|y_{it}\| < \infty$ almost surely.

Lemma A.2.3. *If $\{x_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is $\mathbb{L}^p$-NED ($p \geq 1$) on $\{\varepsilon_{it} : (i,t) \in D\}$, so is $\{x_{j\tau} : (j,\tau) \in D_{NT}, NT \geq 1\}$ with $\rho((i,t),(j,\tau)) < \infty$.*

Lemma A.2.4. *$\{x_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is an array of random fields being uniformly $\mathbb{L}^p$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-μ, and uniformly $\mathbb{L}^q$-bounded ($q \geq p \geq 1$) in the sense that $\sup_{N,T} \sup_{(i,t)\in D_{NT}} \|x_{it}\|_q < \infty$. Let $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ be another uniformly $\mathbb{L}^q$-bounded array of random fields where $y_{it} = x_{it} + \phi y_{i,t-1}$ with $|\phi| < 1$, then $\{y_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ is also uniformly $\mathbb{L}^p$-NED on $\{\varepsilon_{it} : (i,t) \in D\}$ of size-μ.*

Proof. With $|\phi| < 1$, the solution of $y_{it} = x_{it} + \phi y_{i,t-1}$ converges in $\mathbb{L}^q$ to $y_{it} = \sum_{k=0}^{\infty} \phi^k x_{i,t-k}$. By triangle inequality we have

$$\|y_{it} - \mathbb{E}(y_{it}|\mathcal{F}_{it}(s))\|_p \leq \sum_{k=0}^{\infty} |\phi|^k \|x_{i,t-k} - \mathbb{E}(x_{i,t-k}|\mathcal{F}_{it}(s))\|_p . \tag{A.2.1}$$

When $k \leq s$, according to Theorem 10.28 in Davidson (1994) we have

$$\begin{aligned} \|x_{i,t-k} - \mathbb{E}(x_{i,t-k}|\mathcal{F}_{it}(s))\|_p \leq & 2 \|x_{i,t-k} - \mathbb{E}(x_{i,t-k}|\mathcal{F}_{i,t-k}(s-k))\|_p \\ \leq & C_1 \psi_x(s-k) \end{aligned}$$

where C_1 is some positive constant and ψ_x are the NED coefficients of x_{it}. When $k > s$,

$$\|x_{i,t-k} - \mathbb{E}(x_{i,t-k}|\mathcal{F}_{it}(s))\|_p = \|x_{i,t-k} - \mathbb{E}(x_{i,t-k})\|_p \leq C_2$$

where C_2 is some finite constant as $\|x_{it}\|_p < \infty$. Therefore (A.2.1) becomes

$$\|y_{it} - \mathbb{E}(y_{it}|\mathcal{F}_{it}(s))\|_p \leq C_1 \sum_{k=0}^{s} |\phi|^k \psi_x(s-k) + C_2 \sum_{k=s+1}^{\infty} |\phi|^k. \tag{A.2.2}$$

The second term on the right-hand-side of (A.2.2) decays exponentially as $s \to \infty$, and is therefore neglectable compared to the polynomial term of $\psi_x(s) = \mathcal{O}(s^{-\mu})$. As for the first term, since

$$\begin{aligned} \sum_{k=0}^{s} |\phi|^k \frac{\psi_x(s-k)}{s^{-\mu}} = & \sum_{k=0}^{s} |\phi|^k \frac{\mathcal{O}((s-k)^{-\mu})}{s^{-\mu}} \\ \leq & C_3 \sum_{k=0}^{s} |\phi|^k, \end{aligned}$$

we have $\sum_{k=0}^{s} |\phi|^k \psi_x(s-k) = \mathcal{O}(s^{-\mu})$ and complete the proof. □

Lemma A.2.5. *If* $0 < \beta < 1$, $\sup_{NT \geq 1} \sup_{(i,t) \in D_{NT}} \mathbb{E} \left| \sigma_{it}^2(\theta) \right| < \infty$, *then*

$$\sigma_{it}^2(\theta) = \sum_{k=1}^{\infty} \beta^{k-1} c_{i,t-k}(\theta) \tag{A.2.3}$$

almost surely, where

$$c_{i,t-k}(\theta) = \omega + \alpha^{(1)} y_{i,t-k}^2 1_{\{y_{i,t-k} \geq 0\}} + \alpha^{(2)} y_{i,t-k}^2 1_{\{y_{i,t-k} < 0\}} + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-k}^2.$$

Proof. Since $y_{it} = \varepsilon_{it}\sigma_{it}(\theta_0)$ and ε_{it} is independent from $\sigma_{it}(\theta_0)$, we also have

$$\sup_{NT \geq 1} \sup_{(i,t) \in D_{NT}} \mathbb{E} |y_{it}|^2 < \infty.$$

Let $\log^+(x) = \log(x)$ if $x > 1$ and 0 otherwise, by Jensen's inequality we have

$$\begin{aligned} &\mathbb{E} \log^+ |c_{i,t-k}(\theta)| \\ \leq &\log^+ \mathbb{E} \left| \omega + \alpha^{(1)} y_{i,t-k}^2 1_{\{y_{i,t-k} \geq 0\}} + \alpha^{(2)} y_{i,t-k}^2 1_{\{y_{i,t-k} < 0\}} + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-k}^2 \right| \\ < &\infty. \end{aligned}$$

By Lemma 2.2 in Berkes et al. (2003) we have $\sum_{k=1}^{\infty} \mathbb{P}\left[|c_{i,t-k}(\theta)| > \zeta^k\right] < \infty$ for any $\zeta > 1$. Therefore $|c_{i,t-k}(\theta)| \leq \zeta^k$ almost surely by the Borel-Cantelli lemma. Letting $1 < \zeta < \frac{1}{|\beta|}$, we can prove that the right-hand-side of (A.2.3) converges almost surely.

It remains for us to show that

$$\sigma_{it}^2(\theta) = \sum_{k=1}^{\infty} \beta^{k-1} c_{i,t-k}(\theta).$$

Notice that

$$\sigma_{it}^2(\theta) - \beta^k \sigma_{i,t-k}^2(\theta) = c_{i,t-1}(\theta) + \beta c_{i,t-2}(\theta) + ... + \beta^{k-1} c_{i,t-k}(\theta).$$

Using Markov's inequality we obtain that $\sum_{k=1}^{\infty} \mathbb{P}\left\{|\beta^k \sigma_{i,t-k}^2(\theta)| > \delta\right\} < \infty$ for any $\delta > 0$, then by Borel-Cantelli lemma $|\beta^k \sigma_{i,t-k}^2(\theta)| \xrightarrow{a.s.} 0$ as $k \to \infty$. Letting $k \to \infty$ on both sides of above equation we complete the proof. □

A.2.1 Proof of Theorem 4.1

Recall that (4.2.1) is a generalized autoregressive equation since the random matrices $\{B_t\}$ are i.i.d.. According to Theorem 3.2 in Bougerol and Picard (1992), there exists a unique strictly stationary solution of model (4.2.1) if and only if the Lyapunov exponent $\gamma < 0$.

By the sub-additive ergodic theorem (see Kingman (1973)),

$$\gamma = \lim_{t\to\infty} \frac{1}{t+1} \log \|B_t B_{t-1} ... B_0\|$$

almost surely, according to (1.4) in Kesten and Spitzer (1984), we know that the Lyapunov exponent associated with i.i.d. random matrices $\{B_t\}$ satisfies

$$\gamma \le \log \rho[\mathbb{E}(B_t)],$$

where $\rho(\cdot)$ denotes the spectral radius of a matrix. Hence the condition of $\gamma < 0$ is implied by a stronger condition that $\rho[\mathbb{E}(B_t)] < 1$. Denoting $\alpha^* := \max\{\alpha^{(1)}, \alpha^{(2)}\}$ and $d_i = \sum_{k=1}^{N} a_{ik}$, we have

$$\begin{aligned}
\mathbb{E}(B_t) &= \mathbb{E}\left\{\alpha^{(1)} R_t E_t + \alpha^{(2)}(I_N - R_t)E_t + \lambda W E_t + \beta I_N\right\} \\
&\le \mathbb{E}\left\{\alpha^* R_t E_t + \alpha^*(I_N - R_t)E_t + \lambda W E_t + \beta I_N\right\} \\
&= \alpha^* I_N + \lambda W + \beta I_N \\
&= \begin{pmatrix} \alpha^* + \beta + \lambda\frac{a_{11}}{d_1} & \lambda\frac{a_{12}}{d_1} & \dots & \lambda\frac{a_{1N}}{d_1} \\ \lambda\frac{a_{21}}{d_2} & \alpha^* + \beta + \lambda\frac{a_{22}}{d_2} & \dots & \lambda\frac{a_{2N}}{d_2} \\ \lambda\frac{a_{N1}}{d_N} & \lambda\frac{a_{N2}}{d_N} & \dots & \alpha^* + \beta + \lambda\frac{a_{NN}}{d_N} \end{pmatrix},
\end{aligned}$$

and then $\rho[\mathbb{E}(B_t)] \le \alpha^* + \beta + \lambda$ according to the Gershgorin circle theorem (see Horn and Johnson (2012)). Consequently, it suffices to verify that $\alpha^* + \beta + \lambda < 1$ to ensure the strict stationarity of model (4.2.1).

A.2.2 Proof of Lemma 4.3.1

For simplicity we use the notation $X^{(i,t,s)} := \mathbb{E}(X|\mathcal{F}_{it}(s))$ for an arbitrary random variable X, $B := \mathbb{E}(B_t)$, $B^{(2)} := \mathbb{E}(B_t \otimes B_t)$, and $|\cdot|_{\max}(|\cdot|_{\min})$ denotes the maximum (minimum) element of vectors and matrices.

Claim A.2.1. *$\bar{c}_1 := |B\mathbf{1}_N|_{\max} < 1$ and $\bar{c}_2 := |B^{(2)}(\mathbf{1}_N \otimes \mathbf{1}_N)|_{\max} < 1$.*

Proof. Since $\mathbb{E}B_t \preceq \alpha^* I_N + \lambda W + \beta I_N$ and $W\mathbf{1}_N = \mathbf{1}_N$, one can easily verify that

$$\bar{c}_1 \leq \alpha^* + \beta + \lambda < 1.$$

For arbitrary i_1, i_2, j_1, j_2,

$$\begin{aligned}&\mathbb{E}\left[\left(\alpha^*\varepsilon_{i_1t}^2 + \lambda\sum_{j_1} w_{i_1j_1}\varepsilon_{j_1t}^2 + \beta\right) \otimes \left(\alpha^*\varepsilon_{i_2t}^2 + \lambda\sum_{j_2} w_{i_2j_2}\varepsilon_{j_2t}^2 + \beta\right)\right]\\ \leq&\kappa_4(\alpha^* + \beta + \lambda) < 1\end{aligned}$$

by Assumption 4.3.1. Then

$$\begin{aligned}&|B^{(2)}(\mathbf{1}_N \otimes \mathbf{1}_N)|_{\max}\\ =&|\mathbb{E}\left[(B_t\mathbf{1}_N) \otimes (B_t\mathbf{1}_N)\right]|_{\max}\\ \preceq&|\mathbb{E}\left[(\alpha^* E_t\mathbf{1}_N + \lambda W E_t\mathbf{1}_N + \beta\mathbf{1}_N) \otimes (\alpha^* E_t\mathbf{1}_N + \lambda W E_t\mathbf{1}_N + \beta\mathbf{1}_N)\right]|_{\max}\\ \leq&\kappa_4(\alpha^* + \beta + \lambda) < 1.\end{aligned}$$

□

Denote the element (i,j) of B_t as

$$b_{ij}^{(t)} = \begin{cases} \alpha^{(1)}1_{\{\varepsilon_{it}\geq 0\}}\varepsilon_{it}^2 + \alpha^{(2)}1_{\{\varepsilon_{it}<0\}}\varepsilon_{it}^2 + \beta & i = j, \\ \lambda w_{ij}\varepsilon_{jt}^2 & \textit{otherwise}. \end{cases}$$

Notice that the stochastic part in B_t consists of $\{\varepsilon_{it} : i = 1, 2, ..., N\}$. Therefore for any $k > s$, B_{t-k} is independent from $\mathcal{F}_{it}(s)$ and $B_{t-k}^{(i,t,s)} = B$. When $k \leq s$, only some of the elements of B_{t-k} are independent from $\mathcal{F}_{it}(s)$, and we will handle this case carefully in Claim A.2.2 below.

Claim A.2.2. *When $k \leq s$ we have*

$$\mathbb{E}\left[B_{t-k} \otimes B_{t-k}^{(i,t,s)}\right] = \mathbb{E}\left[B_{t-k}^{(i,t,s)} \otimes B_{t-k}\right] \preceq B^{(2)},$$
$$\mathbb{E}\left[B_{t-k}^{(i,t,s)} \otimes B_{t-k}^{(i,t,s)}\right] \preceq B^{(2)}.$$

Proof. For arbitrary i_1, i_2, j_1, j_2,

$$\mathbb{E}\left[b_{i_1 j_1}^{(t-k)} \mathbb{E}(b_{i_2 j_2}^{(t-k)} | \mathcal{F}_{it}(s))\right] = \begin{cases} \mathbb{E}\left[b_{i_1 j_1}^{(t-k)}\right] \mathbb{E}\left[b_{i_2 j_2}^{(t-k)}\right] & j_1 = j_2 \ \& \ |j_2 - i| > s, \\ \mathbb{E}\left[b_{i_1 j_1}^{(t-k)} b_{i_2 j_2}^{(t-k)}\right] & otherwise. \end{cases} \tag{A.2.4}$$

And

$$\mathbb{E}\left[\mathbb{E}(b_{i_1 j_1}^{(t-k)} | \mathcal{F}_{it}(s)) \mathbb{E}(b_{i_2 j_2}^{(t-k)} | \mathcal{F}_{it}(s))\right] = \begin{cases} \mathbb{E}\left[b_{i_1 j_1}^{(t-k)}\right] \mathbb{E}\left[b_{i_2 j_2}^{(t-k)}\right] & j_1 = j_2 \ \& \ |j_2 - i| > s, \\ \mathbb{E}\left[b_{i_1 j_1}^{(t-k)} b_{i_2 j_2}^{(t-k)}\right] & otherwise. \end{cases} \tag{A.2.5}$$

Since $\mathbb{E}\left[b_{i_1 j_1}^{(t-k)}\right] \mathbb{E}\left[b_{i_2 j_2}^{(t-k)}\right] \leq \mathbb{E}\left[b_{i_1 j_1}^{(t-k)} b_{i_2 j_2}^{(t-k)}\right]$ when $j_1 = j_2$, we complete the proof. □

Let $\mathbf{e}_i$ be an N-dimensional vector with the i-th component being 1 and others being 0. By (4.2.3) we have

$$h_{it} = \mathbf{e}_i' \mathbf{h}_t = \omega \sum_{k=0}^{\infty} \mathbf{e}_i' \Pi_{t-1,k} \mathbf{1}_N,$$

where $\Pi_{t-1,k} = B_{t-1} ... B_{t-k}$ for $k \geq 1$ and $\Pi_{t-1,0} = I_N$. Now we will show that

$$\sup_{NT \geq 1} \sup_{(i,t) \in D_{NT}} \|h_{it} - \mathbb{E}(h_{it} | \mathcal{F}_{it}(s)\|_2 = \mathcal{O}(s^{-\mu}) \tag{A.2.6}$$

with Assumption 4.3.3(a); Or

$$\sup_{NT \geq 1} \sup_{(i,t) \in D_{NT}} \|h_{it} - \mathbb{E}(h_{it} | \mathcal{F}_{it}(s)\|_2 = \mathcal{O}(\rho^{s}) \tag{A.2.7}$$

with Assumption 4.3.3(b).

By (4.2.3) we have

$$\begin{aligned}
&\|h_{it} - \mathbb{E}(h_{it}|\mathcal{F}_{it}(s)\|_2 \\
= &\left\|\mathbf{e}_i' \textstyle\sum_{k\geq 0} \Pi_{t-1,k}(\omega \mathbf{1}_N) - \mathbb{E}\left[\mathbf{e}_i' \textstyle\sum_{k\geq 0} \Pi_{t-1,k}(\omega \mathbf{1}_N)|\mathcal{F}_{it}(s)\right]\right\|_2 \\
\leq &\left\|\mathbf{e}_i' \textstyle\sum_{k\leq s} \left[\Pi_{t-1,k} - \Pi_{t-1,k}^{(i,t,s)}\right](\omega \mathbf{1}_N)\right\|_2 \\
&+ \left\|\mathbf{e}_i' \textstyle\sum_{k> s} \left[\Pi_{t-1,k} - \Pi_{t-1,s}^{(i,t,s)} B^{k-s}\right](\omega \mathbf{1}_N)\right\|_2 .
\end{aligned} \tag{A.2.8}$$

Then we handle the two terms separately.

When $k > s$,

$$\begin{aligned}
&\mathbb{E}\left|\mathbf{e}_i' \sum_{k>s}\left[\Pi_{t-1,k} - \Pi_{t-1,s}^{(i,t,s)} B^{k-s}\right](\omega \mathbf{1}_N)\right|^2 \\
\leq &\sum_{k>s} \omega^2 (\mathbf{e}' \otimes \mathbf{e}')\mathbb{E}\left\{\left[\Pi_{t-1,k} - \Pi_{t-1,s}^{(i,t,s)} B^{k-s}\right] \otimes \left[\Pi_{t-1,k} - \Pi_{t-1,s}^{(i,t,s)} B^{k-s}\right]\right\}(\mathbf{1}_N \otimes \mathbf{1}_N) \\
\leq &\sum_{k>s} \omega^2 (\mathbf{e}' \otimes \mathbf{e}')\mathbb{E}\left[\Pi_{t-1,k} \otimes \Pi_{t-1,k}\right](\mathbf{1}_N \otimes \mathbf{1}_N) \\
&+ \sum_{k>s} \omega^2 (\mathbf{e}' \otimes \mathbf{e}')\mathbb{E}\left[\Pi_{t-1,k} \otimes (\Pi_{t-1,s}^{(i,t,s)} B^{k-s})\right](\mathbf{1}_N \otimes \mathbf{1}_N) \\
&+ \sum_{k>s} \omega^2 (\mathbf{e}' \otimes \mathbf{e}')\mathbb{E}\left[(\Pi_{t-1,s}^{(i,t,s)} B^{k-s}) \otimes \Pi_{t-1,k}\right](\mathbf{1}_N \otimes \mathbf{1}_N) \\
&+ \sum_{k>s} \omega^2 (\mathbf{e}' \otimes \mathbf{e}')\mathbb{E}\left[(\Pi_{t-1,s}^{(i,t,s)} B^{k-s}) \otimes (\Pi_{t-1,s}^{(i,t,s)} B^{k-s})\right](\mathbf{1}_N \otimes \mathbf{1}_N).
\end{aligned}$$

For the first term, we use Claim A.2.1:

$$\begin{aligned}
&(\mathbf{e}' \otimes \mathbf{e}')\mathbb{E}\left[\Pi_{t-1,k} \otimes \Pi_{t-1,k}\right](\mathbf{1}_N \otimes \mathbf{1}_N) \\
=&(\mathbf{e}' \otimes \mathbf{e}')\mathbb{E}\left[(B_{t-1} \otimes B_{t-1})(B_{t-2} \otimes B_{t-2})...(B_{t-k} \otimes B_{t-k})\right](\mathbf{1}_N \otimes \mathbf{1}_N) \\
=&(\mathbf{e}' \otimes \mathbf{e}')\left[B^{(2)}\right]^k(\mathbf{1}_N \otimes \mathbf{1}_N) \\
\leq&\bar{c}_2^k.
\end{aligned}$$

For the second term, notice that

$$
\begin{aligned}
&\mathbb{E}\left[\Pi_{t-1,k}\otimes(\Pi_{t-1,s}^{(i,t,s)}B^{k-s})\right]\\
=&\mathbb{E}\left[(B_{t-1}\otimes B_{t-1}^{(i,t,s)})(B_{t-2}\otimes B_{t-2}^{(i,t,s)})...(B_{t-s}\otimes B_{t-s}^{(i,t,s)})(B_{t-s-1}\otimes B)...(B_{t-k}\otimes B)\right]\\
\preceq&\left[B^{(2)}\right]^{s}(B\otimes B)^{k-s}.
\end{aligned}
$$

In the second line above, the first s terms are in forms $B_{t-k}\otimes B_{t-k}^{(i,t,s)}$ with $k\leq s$, and then the third line is obtained by applying Claim A.2.2. Therefore we have

$$
\begin{aligned}
&(\mathbf{e}'\otimes\mathbf{e}')\mathbb{E}\left[\Pi_{t-1,k}\otimes(\Pi_{t-1,s}^{(i,t,s)}B^{k-s})\right](\mathbf{1}_N\otimes\mathbf{1}_N)\\
\leq&(\mathbf{e}'\otimes\mathbf{e}')\left[B^{(2)}\right]^{s}(B\otimes B)^{k-s}(\mathbf{1}_N\otimes\mathbf{1}_N)\\
\leq&\bar{c}_2^{s}[\bar{c}_1^{2}]^{k-s}.
\end{aligned}
$$

Similarly, we could also verify that

$$
(\mathbf{e}'\otimes\mathbf{e}')\mathbb{E}\left[(\Pi_{t-1,s}^{(i,t,s)}B^{k-s})\otimes\Pi_{t-1,k}\right](\mathbf{1}_N\otimes\mathbf{1}_N)\leq\bar{c}_2^{k-s}[\bar{c}_1^{2}]^{s}
$$

and

$$
(\mathbf{e}'\otimes\mathbf{e}')\mathbb{E}\left[(\Pi_{t-1,s}^{(i,t,s)}B^{k-s})\otimes(\Pi_{t-1,s}^{(i,t,s)}B^{k-s})\right](\mathbf{1}_N\otimes\mathbf{1}_N)\leq[\bar{c}_1^{2}]^{k}.
$$

Therefore,

$$
\mathbb{E}\left|\mathbf{e}_i'\sum_{k>s}\left[\Pi_{t-1,k}-\Pi_{t-1,s}^{(i,t,s)}B^{k-s}\right](\omega\mathbf{1}_N)\right|^2\leq 4\sum_{k>s}\omega^2\rho^k=\mathcal{O}(\rho^s),\tag{A.2.9}
$$

where $\rho=\max\{\bar{c}_1^2,\bar{c}_2\}<1$ by Claim A.2.1. Hence the second term on the right-hand-side of (A.2.8) decays exponentially as $s\to\infty$. It remains for us to deal with the first term

$$
\left\|\mathbf{e}_i'\sum_{k\leq s}\left[\Pi_{t-1,k}-\Pi_{t-1,k}^{(i,t,s)}\right](\omega\mathbf{1}_N)\right\|_2.
$$

When $k \leq s$,

$$
\begin{aligned}
&\mathbb{E}\left|\mathbf{e}_i' \sum_{k\leq s}\left[\Pi_{t-1,k} - \Pi_{t-1,k}^{(i,t,s)}\right](\omega \mathbf{1}_N)\right|^2 \\
\leq &\sum_{k\leq s}\omega^2(\mathbf{e}'\otimes\mathbf{e}')\mathbb{E}\left\{\left[\Pi_{t-1,k} - \Pi_{t-1,k}^{(i,t,s)}\right]\otimes\left[\Pi_{t-1,k} - \Pi_{t-1,k}^{(i,t,s)}\right]\right\}(\mathbf{1}_N\otimes\mathbf{1}_N) \\
= &\sum_{k\leq s}\omega^2(\mathbf{e}'\otimes\mathbf{e}')\mathbb{E}\left[\Pi_{t-1,k}\otimes\Pi_{t-1,k} - \Pi_{t-1,k}\otimes\Pi_{t-1,k}^{(i,t,s)}\right](\mathbf{1}_N\otimes\mathbf{1}_N) \\
&+\sum_{k\leq s}\omega^2(\mathbf{e}'\otimes\mathbf{e}')\mathbb{E}\left[\Pi_{t-1,k}^{(i,t,s)}\otimes\Pi_{t-1,k}^{(i,t,s)} - \Pi_{t-1,k}^{(i,t,s)}\otimes\Pi_{t-1,k}\right](\mathbf{1}_N\otimes\mathbf{1}_N) \\
:= &T_1 + T_2.
\end{aligned}
$$

By (A.2.4), $\mathbb{E}\left[b_{ij_1}^{(t-k)}b_{ij_2}^{(t-k)}\right] \neq \mathbb{E}\left[b_{ij_1}^{(t-k)}\mathbb{E}(b_{ij_2}^{(t-k)}|\mathcal{F}_{it}(s))\right]$ only if $j_1 = j_2$ and $|j_1 - i| > s$, then

$$
\begin{aligned}
&(\mathbf{e}_i'\otimes\mathbf{e}_i')\mathbb{E}\left[B_{t-k}\otimes B_{t-k} - B_{t-k}\otimes B_{t-k}^{(i,t,s)}\right](\mathbf{1}_N\otimes\mathbf{1}_N) \\
= &\sum_{j_1}\sum_{j_2}\left\{\mathbb{E}\left[b_{ij_1}^{(t-k)}b_{ij_2}^{(t-k)}\right] - \mathbb{E}\left[b_{ij_1}^{(t-k)}\mathbb{E}(b_{ij_2}^{(t-k)}|\mathcal{F}_{it}(s))\right]\right\} \\
= &\sum_{|j-i|>s}\left\{\mathbb{E}\left[b_{ij}^{(t-k)}\right]^2 - \mathbb{E}b_{ij}^{(t-k)}\mathbb{E}b_{ij}^{(t-k)}\right\}.
\end{aligned}
$$

Since

$$
\mathbb{E}b_{ij}^{(t-k)} = \begin{cases} \alpha^{(1)}\kappa_2^+ + \alpha^{(2)}\kappa_2^- + \beta & i = j, \\ \lambda w_{ij} & otherwise, \end{cases}
$$

and

$$
\mathbb{E}\left[b_{ij}^{(t-k)}\right]^2 = \begin{cases} \left[\alpha^{(1)}\right]^2\kappa_4^+ + \left[\alpha^{(2)}\right]^2\kappa_4^- + 2\alpha^{(1)}\beta\kappa_2^+ + 2\alpha^{(2)}\beta\kappa_2^- + \beta^2 & i = j, \\ \lambda^2 w_{ij}^2\kappa_4 & otherwise, \end{cases}
$$

where $\kappa_2^+ := \mathbb{E}\left[\varepsilon_{it}^2 1_{\{\varepsilon_{it}\geq 0\}}\right]$, $\kappa_2^- := \mathbb{E}\left[\varepsilon_{it}^2 1_{\{\varepsilon_{it}<0\}}\right]$, $\kappa_4^+ := \mathbb{E}\left[\varepsilon_{it}^4 1_{\{\varepsilon_{it}\geq 0\}}\right]$ and $\kappa_4^- := \mathbb{E}\left[\varepsilon_{it}^4 1_{\{\varepsilon_{it}<0\}}\right]$. Then we obtain that

$$
\mathbb{E}\left[b_{ij}^{(t-k)}\right]^2 - \mathbb{E}b_{ij}^{(t-k)}\mathbb{E}b_{ij}^{(t-k)} = (\kappa_4 - 1)\lambda^2 w_{ij}^2 \quad if \quad i \neq j. \tag{A.2.10}
$$

In light of Claim A.2.2 we also have

$$\begin{aligned}\mathbb{E}\left(\Pi_{t-1,k-1} \otimes \Pi^{(i,t,s)}_{t-1,k-1}\right) &\preceq \left[B^{(2)}\right]^{k-1}, \\ \mathbb{E}\left(\Pi^{(i,t,s)}_{t-1,k-1} \otimes \Pi^{(i,t,s)}_{t-1,k-1}\right) &\preceq \left[B^{(2)}\right]^{k-1}.\end{aligned} \tag{A.2.11}$$

Note that (A.2.10), (A.2.11) and Claim A.2.1 allow us to derive that

$$\begin{aligned}T_1 \leq &\sum_{k\leq s}(\mathbf{e}_i' \otimes \mathbf{e}_i')\left[\mathbb{E}\left(\Pi_{t-1,k-1} \otimes \Pi_{t-1,k-1}\right) + \mathbb{E}\left(\Pi_{t-1,k-1} \otimes \Pi^{(i,t,s)}_{t-1,k-1}\right)\right] \\ &\times \mathbb{E}\left[B_{t-k} \otimes B_{t-k} - B_{t-k} \otimes B^{(i,t,s)}_{t-k}\right](\mathbf{1}_N \otimes \mathbf{1}_N) \\ \leq &2\sum_{k\leq s}[\bar{c}_2]^{k-1}\sum_{|j-i|>s}\left\{\mathbb{E}\left[b^{(t-k)}_{ij} b^{(t-k)}_{ij}\right] - \mathbb{E}b^{(t-k)}_{ij}\mathbb{E}b^{(t-k)}_{ij}\right\} \\ \leq &C\sum_{|j-i|>s} w_{ij}^2\end{aligned} \tag{A.2.12}$$

for some constant $C > 0$. Hence, with Assumption 4.3.3(a) we obtain $T_1 \leq C\sum_{|j-i|>s}|j-i|^{-\mu-2}$. This could also be verified for T_2 by following similar arguments. Then according to Lemma A.1.4 and Lemma A.1.5,

$$\begin{aligned}\sum_{|j-i|>s}|j-i|^{-\mu-2} &= \sum_{h=[s]}^{\infty}\sum_{h\leq|j-i|<h+1}|j-i|^{-\mu-2} \\ &\leq \sum_{h=[s]}^{\infty} Ch^{-\mu-1} \\ &< C\frac{2^{\mu+1}}{\mu}s^{-\mu}\end{aligned}$$

for some constant $C > 0$. Together with (A.2.8) and (A.2.9) we prove (A.2.6).

As for the proof of (A.2.7), using Assumption 4.3.3(b) and letting $s \geq K$ in (A.2.12) we can verify that the first term of (A.2.8)

$$\left\|\mathbf{e}_i'\sum_{k\leq s}\left[\Pi_{t-1,k} - \Pi^{(i,t,s)}_{t-1,k}\right](\omega\mathbf{1}_N)\right\|_2 = 0,$$

while the second term decays exponentially according to (A.2.9). Now we complete the proof of Lemma 4.3.1.

A.2.3 Proof of Theorem 4.2

By Lemma A.2.5 we have

$$\sigma_{it}^2(\theta) = \sum_{k=1}^{\infty} \beta^{k-1} c_{i,t-k}(\theta),$$
$$\tilde{\sigma}_{it}^2(\theta) = \sum_{k=1}^{t} \beta^{k-1} c_{i,t-k}(\theta),$$

almost surely, where

$$c_{i,t-k}(\theta) = \omega + \alpha^{(1)} y_{i,t-k}^2 1_{\{y_{i,t-k} \geq 0\}} + \alpha^{(2)} y_{i,t-k}^2 1_{\{y_{i,t-k} < 0\}} + \lambda \sum_{j=1}^{N} w_{ij} y_{j,t-k}^2.$$

We have

$$\sigma_{it}^2(\theta) - \tilde{\sigma}_{it}^2(\theta) = \beta^t \sigma_{i0}^2(\theta). \tag{A.2.13}$$

The partial derivatives of $\sigma_{it}^2(\theta)$ are

$$\begin{aligned}
\frac{\partial \sigma_{it}^2(\theta)}{\partial \omega} &= \sum_{k=1}^{\infty} \beta^{k-1}, \\
\frac{\partial \sigma_{it}^2(\theta)}{\partial \alpha^{(1)}} &= \sum_{k=1}^{\infty} \beta^{k-1} y_{i,t-k}^2 1_{\{y_{i,t-k} \geq 0\}}, \\
\frac{\partial \sigma_{it}^2(\theta)}{\partial \alpha^{(2)}} &= \sum_{k=1}^{\infty} \beta^{k-1} y_{i,t-k}^2 1_{\{y_{i,t-k} < 0\}}, \\
\frac{\partial \sigma_{it}^2(\theta)}{\partial \lambda} &= \sum_{k=1}^{\infty} \beta^{k-1} \left(\sum_{j=1}^{N} w_{ij} y_{j,t-k}^2 \right), \\
\frac{\partial \sigma_{it}^2(\theta)}{\partial \beta} &= \sum_{k=2}^{\infty} (k-1) \beta^{k-2} c_{i,t-k}(\theta).
\end{aligned} \tag{A.2.14}$$

By similar arguments to the proof of Lemma A.2.5, we can show that the right-hand-sides of (A.2.14) converge almost surely as $0 < \beta < 1$, that is

$$\left\| \frac{\partial \sigma_{it}^2(\theta)}{\partial \theta} \right\| < \infty \qquad a.s. \tag{A.2.15}$$

We also have

$$\frac{\partial \sigma_{it}^2(\theta)}{\partial \theta} - \frac{\partial \tilde{\sigma}_{it}^2(\theta)}{\partial \theta} = t\beta^{t-1}\sigma_{i0}^2(\theta)\mathbf{e}_5 + \beta^t \frac{\partial \sigma_{i0}^2(\theta)}{\partial \theta}, \tag{A.2.16}$$

where $\mathbf{e}_5 = (0, 0, 0, 0, 1)'$.

Now we consider the second order derivatives. For any $\theta_m, \theta_n \in \{\omega, \alpha^{(1)}, \alpha^{(2)}, \lambda\}$,

$$\frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \theta_m \partial \theta_n} = 0.$$

And

$$\begin{aligned}
\frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \omega \partial \beta} &= \sum_{k=2}^{\infty} (k-1)\beta^{k-2}, \\
\frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \alpha^{(1)} \partial \beta} &= \sum_{k=2}^{\infty} (k-1)\beta^{k-2} y_{i,t-k}^2 1_{\{y_{i,t-k} \geq 0\}}, \\
\frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \alpha^{(2)} \partial \beta} &= \sum_{k=2}^{\infty} (k-1)\beta^{k-2} y_{i,t-k}^2 1_{\{y_{i,t-k} < 0\}}, \\
\frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \lambda \partial \beta} &= \sum_{k=2}^{\infty} (k-1)\beta^{k-2} \left(\sum_{j=1}^{N} w_{ij} y_{j,t-k}^2 \right), \\
\frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \beta^2} &= \sum_{k=3}^{\infty} (k-1)(k-2)\beta^{k-3} c_{i,t-k}(\theta).
\end{aligned} \tag{A.2.17}$$

The right-hand-sides of (A.2.17) also converge almost surely as $0 < \beta < 1$, that is

$$\left\| \frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \theta \partial \theta'} \right\| < \infty \qquad a.s. \tag{A.2.18}$$

We also have:

$$\frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\sigma}_{it}^2(\theta)}{\partial \theta \partial \theta'} = t(t-1)\beta^{t-2}\sigma_{i0}^2(\theta)\mathbf{e}_5\mathbf{e}_5' + 2t\beta^{t-1}\frac{\partial \sigma_{i0}^2(\theta)}{\partial \theta}\mathbf{e}_5' + \beta^t \frac{\partial^2 \sigma_{i0}^2(\theta)}{\partial \theta \partial \theta'}, \tag{A.2.19}$$

where $\mathbf{e}_5 = (0, 0, 0, 0, 1)'$.

The third order derivative of $\sigma_{it}^2(\theta)$ is also bounded almost surely, as

$$\begin{aligned}
\frac{\partial^3 \sigma_{it}^2(\theta)}{\partial\omega\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3}, \\
\frac{\partial^3 \sigma_{it}^2(\theta)}{\partial\alpha^{(1)}\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3}y_{i,t-k}^2 1_{\{y_{i,t-k}\geq 0\}}, \\
\frac{\partial^3 \sigma_{it}^2(\theta)}{\partial\alpha^{(2)}\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3}y_{i,t-k}^2 1_{\{y_{i,t-k}< 0\}}, \\
\frac{\partial^3 \sigma_{it}^2(\theta)}{\partial\lambda\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3}\left(\sum_{j=1}^{N} w_{ij}y_{j,t-k}^2\right), \\
\frac{\partial^3 \sigma_{it}^2(\theta)}{\partial\beta^3} &= \sum_{k=4}^{\infty}(k-1)(k-2)(k-3)\beta^{k-4}c_{i,t-k}(\theta).
\end{aligned} \tag{A.2.20}$$

almost surely.

Proof of Consistency

Claim A.2.3 below validates that $\tilde{L}_{NT}(\theta)$ serves as an appropriate approximation of the likelihood function $L_{NT}(\theta)$. Claim A.2.4 and Claim A.2.5 facilitate the adoption of Theorem 3.1 on $l_{it}(\theta)$'s, and Claim A.2.6 assists in showing the unique identifiability of the true parameters θ_0.

Claim A.2.3. *For any $\theta \in \Theta$, $|L_{NT}(\theta) - \tilde{L}_{NT}(\theta)| \xrightarrow{p} 0$ as $T \to \infty$ and $N \to \infty$.*

Proof. By (A.2.13) we have

$$\begin{aligned}
&\mathbb{E}|L_{NT}(\theta) - \tilde{L}_{NT}(\theta)| \\
\leq& \frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\left|\frac{\sigma_{it}^2(\theta) - \tilde{\sigma}_{it}^2(\theta)}{\sigma_{it}^2(\theta)\tilde{\sigma}_{it}^2(\theta)}\right| y_{it}^2 + \left|\log\left(1 + \frac{\tilde{\sigma}_{it}^2(\theta) - \sigma_{it}^2(\theta)}{\sigma_{it}^2(\theta)}\right)\right|\right] \\
\leq& \frac{C_1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\left|\sigma_{it}^2(\theta) - \tilde{\sigma}_{it}^2(\theta)\right| y_{it}^2\right] + \frac{C_2}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left|\sigma_{it}^2(\theta) - \tilde{\sigma}_{it}^2(\theta)\right| \\
=& \frac{C_3}{NT}\sum_{(i,t)\in D_{NT}} \beta^t \mathbb{E}[y_{it}^2] + \frac{C_4}{NT}\sum_{(i,t)\in D_{NT}} \beta^t
\end{aligned}$$

for any $\theta \in \Theta$. Since $\mathbb{E}[y_{it}^2] < \infty$, we have

$$\mathbb{E}|L_{NT}(\theta) - \tilde{L}_{NT}(\theta)| \to 0$$

as $T \to \infty$ and $N \to \infty$. □

Claim A.2.4. $\sup_{(i,t)\in D} \sup_{\theta\in\Theta} \|l_{it}(\theta)\|_p < \infty$ *for some* $p > 1$.

Proof. For any $\theta \in \Theta$ we have

$$\begin{aligned} \|l_{it}(\theta)\|_p &= \left\| \log \sigma_{it}^2(\theta) + \frac{y_{it}^2}{\sigma_{it}^2(\theta)} \right\|_p \\ &\leq \left\| \log \sigma_{it}^2(\theta) \right\|_p + \frac{1}{\omega} \left\| \sigma_{it}^2(\theta_0)\varepsilon_{it}^2 \right\|_p , \end{aligned}$$

where

$$\begin{aligned} \left\| \log \sigma_{it}^2(\theta) \right\|_p &\leq \left\| \log^+ \sigma_{it}^2(\theta) \right\|_p + \left\| \log^- \sigma_{it}^2(\theta) \right\|_p \\ &\leq \left\| \sigma_{it}^2(\theta) \right\|_p + 1 + \max\{0, -\log(\omega)\}. \end{aligned}$$

By Assumption 4.3.1 and Lemma A.2.5 we complete the proof. □

Claim A.2.5. *For any* $\theta \in \Theta$, $\{l_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ *is* η*-weakly dependent with* $\bar{\eta}(s) \leq Cs^{-\mu}$ *for some constants* $C > 0$ *and* $\mu > 2$.

Proof. Note that $\sigma_{it}^2(\theta) \geq \omega$ for any $(i,t) \in D$ and $\theta \in \Theta$, we could easily verify that functions $f(x) = log(x)$ and $g(x) = \frac{1}{x}$ are both Lipschitz on the interval $(\omega, \infty]$. Then by Lemma 4.3.1, Lemma A.2.1, Lemma A.2.2(a) and Proposition 2 in Jenish and Prucha (2012) we can prove that

$$\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \|l_{it}(\theta) - \mathbb{E}(l_{it}(\theta)|\mathcal{F}_{it}(s))\|_1 \leq Cs^{-\mu}. \tag{A.2.21}$$

Then by Proposition 3.3 we complete the proof. □

Claim A.2.6. *If* $\sigma_{it}^2(\theta) = \sigma_{it}^2(\theta_0)$ *for each* $i = 1, 2, ..., N$ *and some* $t \in \mathbb{Z}$, *then* $\theta = \theta_0$.

Proof. For any $i = 1, 2, ..., N$, we have

$$(1-\beta B)\sigma_{it}^2(\theta) = \omega + \alpha B y_{it}^2 + \lambda \sum_{j=1}^{N} w_{ij} B y_{j,t}^2,$$

where B stands for the back-shift operator in the sense that $By_{it}^2 = y_{i,t-1}^2$, and α represents $\alpha^{(1)}$ or $\alpha^{(2)}$ for simplicity. With $\mathbf{h}_t(\theta) := \left(\sigma_{1,t}^2(\theta), ..., \sigma_{N,t}^2(\theta)\right)'$ and $\mathbf{v}_t := \left(y_{1,t}^2, ..., y_{N,t}^2\right)'$, this equation could be vectorized as

$$(1-\beta B)\mathbf{h}_t(\theta) = \omega \mathbf{1}_N + (BR_\theta + \lambda BW)\mathbf{v}_t,$$

where R_θ is a constant matrix (given $\mathbf{y}_{t-1}$) with diagonal elements of either $\alpha^{(1)}$ or $\alpha^{(2)}$, and other entries being zero.

The polynomial $1-\beta x$ has a root $x = 1/\beta$, which lies outside the unit circle since $0 < \beta < 1$. Therefore the inverse $\frac{1}{1-\beta x}$ is well-defined for any $|x| \leq 1$, and we have

$$\mathbf{h}_t(\theta) = \frac{\omega}{1-\beta}\mathbf{1}_N + \mathcal{M}_\theta(B)\mathbf{v}_t$$

with $\mathcal{M}_\theta(B) := \frac{B}{1-\beta B}R_\theta + \frac{\lambda B}{1-\beta B}W$.

If $\sigma_{it}^2(\theta) = \sigma_{it}^2(\theta_0)$ for each $i = 1, 2, ..., N$ and some $t \in \mathbb{Z}$, thus $\mathbf{h}_t(\theta) = \mathbf{h}_t(\theta_0)$, consequently,

$$\{\mathcal{M}_\theta(B) - \mathcal{M}_{\theta_0}(B)\}\mathbf{v}_t = \left(\frac{\omega_0}{1-\beta_0} - \frac{\omega}{1-\beta}\right)\mathbf{1}_N.$$

If $\mathcal{M}_\theta(B) \neq \mathcal{M}_{\theta_0}(B)$, $\mathcal{M}_\theta(B) - \mathcal{M}_{\theta_0}(B)$ could be write in polynomial form $\sum_{k=0}^{\infty} C_k B^k$ with constant matrix coefficients C_k. Therefore,

$$C_0\mathbf{v}_t = \left(\frac{\omega_0}{1-\beta_0} - \frac{\omega}{1-\beta}\right)\mathbf{1}_N - \sum_{k=1}^{\infty} C_k \mathbf{v}_{t-k}.$$

This means that $\mathbf{v}_t$ is measurable w.r.t. $\mathcal{H}_{t-1} := \sigma\{\mathbf{v}_{t-1}, \mathbf{v}_{t-2}, ...\}$, i.e. $\mathbf{v}_t = \mathbb{E}(\mathbf{v}_t|\mathcal{H}_{t-1})$.

However, since $\{\varepsilon_{it}\}$ is a non-degenerate random process according to Assumption 4.1.1, $\mathbf{v}_t - \mathbb{E}_{\theta_0}(\mathbf{v}_t|\mathcal{H}_{t-1}) = \mathbf{v}_t(E_t - I_N)\mathbf{h}_t(\theta_0) \neq 0$, which contradicts with $\mathcal{M}_\theta(B) \neq$

$\mathcal{M}_{\theta_0}(B)$. Hence $\mathcal{M}_\theta(x) = \mathcal{M}_{\theta_0}(x)$ holds for any $|x| \le 1$, i.e.

$$\frac{x}{1-\beta x} R_\theta - \frac{x}{1-\beta_0 x} R_{\theta_0} = \left(\frac{\lambda_0 x}{1-\beta_0 x} - \frac{\lambda x}{1-\beta x} \right) W.$$

Note that the diagonal elements of W are all zeros while the matrix on the left side of the equation has non-zero diagonal elements, so we have

$$\frac{\alpha x}{1-\beta x} = \frac{\alpha_0 x}{1-\beta_0 x},$$
$$\frac{\lambda x}{1-\beta x} = \frac{\lambda_0 x}{1-\beta_0 x},$$

which imply $\alpha^{(1)} = \alpha_0^{(1)}$, $\alpha^{(2)} = \alpha_0^{(2)}$, $\beta = \beta_0$ and $\lambda = \lambda_0$. Besides, $\omega = \omega_0$ could be easily derived from $\frac{\omega}{1-\beta} = \frac{\omega_0}{1-\beta_0}$. □

Claim A.2.4, Claim A.2.5 and Proposition 3.3 allow us to adopt Theorem 3.1 on functions $\{l_{it}(\theta) : (i,t) \in D_{NT}, NT \ge 1\}$. i.e. for any $\theta \in \Theta$,

$$\frac{1}{NT} \sum_{(i,t)\in D_{NT}} [l_{it}(\theta) - \mathbb{E} l_{it}(\theta)] \xrightarrow{p} 0 \tag{A.2.22}$$

as $T \to \infty$ and $N \to \infty$. Therefore we have

$$\begin{aligned}
&\lim_{T,N\to\infty} [L_{NT}(\theta) - L_{NT}(\theta_0)] \\
=&\lim_{T,N\to\infty} \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \mathbb{E} \left\{ \log \frac{\sigma_{it}^2(\theta)}{\sigma_{it}^2(\theta_0)} + \varepsilon_{it}^2 \left[\frac{\sigma_{it}^2(\theta_0)}{\sigma_{it}^2(\theta)} - 1 \right] \right\} \\
\ge&\lim_{T,N\to\infty} \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \mathbb{E} \left[(1-\varepsilon_{it}^2) \log \frac{\sigma_{it}^2(\theta)}{\sigma_{it}^2(\theta_0)} \right] \\
=&0.
\end{aligned}$$

The equality in $\ge$ above holds only if $\theta = \theta_0$ by Claim A.2.6. Following similar arguments in the proof of Proposition 3.4 we can prove the consistency of $\hat{\theta}_{NT}$.

Proof of Asymptotic Normality

Notice that $y_{it} = \varepsilon_{it}\sigma_{it}(\theta_0)$, then we have

$$\begin{aligned}\frac{\partial L_{NT}(\theta_0)}{\partial\theta} =& \frac{1}{NT}\sum_{(i,t)\in D_{NT}}\left[\frac{1}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0) - \frac{y_{it}^2}{\sigma_{it}^4(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)\right] \\ =& \frac{1}{NT}\sum_{(i,t)\in D_{NT}}\left[\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)\right]\end{aligned} \tag{A.2.23}$$

and

$$\Sigma_{NT} = \frac{\kappa_4 - 1}{NT}\sum_{(i,t)\in D_{NT}}\mathbb{E}\left[\frac{1}{\sigma_{it}^4(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)\frac{\partial}{\partial\theta'}\sigma_{it}^2(\theta_0)\right] = (NT)var\left[\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\right].$$

Firstly we need to prove a CLT:

$$(\Sigma_{NT})^{-1/2}\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, I_5). \tag{A.2.24}$$

Since $\sigma_{it}^2(\theta_0) \geq \omega_0$ for any $(i,t) \in D_{NT}$, by Assumption 4.3.4 and (A.2.15) we can prove that

$$\left\|\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)\right\|_r < \infty \tag{A.2.25}$$

for some $r > 2$. With Assumption 4.3.6 and Claim A.2.7 below, we can use Corollary 3.2.1 in Chapter 3 to prove (A.2.24).

Claim A.2.7. *$\left\{\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0) : (i,t) \in D_{NT}, NT \geq 1\right\}$ is η-weakly dependent with $\bar{\eta}(s) \leq Cs^{-\mu}$ for some constants $C > 0$ and $\mu > 4 \vee \frac{2(r-1)}{r-2}$.*

Proof. Note that function $g(x) = \frac{1}{x}$ is Lipschitz continuous on the interval $(\omega_0, \infty]$. Then by Lemma 4.3.1 and Proposition 2 in Jenish and Prucha (2012) we can prove that

$$\left\|\frac{1}{\sigma_{it}^2(\theta_0)} - \mathbb{E}\left(\frac{1}{\sigma_{it}^2(\theta_0)}|\mathcal{F}_{it}(s)\right)\right\|_2 \leq Cs^{-\mu} \tag{A.2.26}$$

As for the term $\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)$, note that $\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0) = \mathbf{u}_{i,t-1} + \beta\frac{\partial}{\partial\theta}\sigma_{i,t-1}^2(\theta_0)$ where

$$\mathbf{u}_{i,t-1} = \left(1, y_{i,t-1}^2 1_{\{\varepsilon_{i,t-1}\geq 0\}}, y_{i,t-1}^2 1_{\{\varepsilon_{i,t-1}<0\}}, \sum_{j=1}^{N} w_{i,j}y_{j,t-1}^2, \sigma_{i,t-1}^2(\theta_0)\right)'.$$

$\left\{y_{it}^2 1_{\{\varepsilon_{it}\geq 0\}} : (i,t) \in D_{NT}, NT \geq 1\right\}$ is uniformly $\mathbb{L}^2$-NED of size-μ on $\mathcal{E}$, since

$$\begin{aligned}
&\left\|y_{it}^2 1_{\{\varepsilon_{it}\geq 0\}} - \mathbb{E}\left(y_{it}^2 1_{\{\varepsilon_{it}\geq 0\}} \mid \mathcal{F}_{it}(s)\right)\right\|_2\\
=&\left\|\varepsilon_{it}^2 1_{\{\varepsilon_{it}\geq 0\}}\sigma_{it}^2(\theta_0) - \varepsilon_{it}^2 1_{\{\varepsilon_{it}\geq 0\}}\mathbb{E}\left[\sigma_{it}^2(\theta_0) \mid \mathcal{F}_{it}(s)\right]\right\|_2\\
\leq& C\left\|\sigma_{it}^2(\theta_0) - \mathbb{E}\left[\sigma_{it}^2(\theta_0) \mid \mathcal{F}_{it}(s)\right]\right\|_2
\end{aligned}$$

for some constant $C > 0$. The inequality holds since $\mathbb{E}\left(\varepsilon_{it}^4 1_{\{\varepsilon_{it}\geq 0\}}\right) \leq \mathbb{E}\varepsilon_{it}^4 < \infty$. By Assumption 4.3.4, we could verify that $\left\{y_{it}^2 1_{\{\varepsilon_{it}\geq 0\}} : (i,t) \in D_{NT}, NT \geq 1\right\}$ is uniformly $\mathbb{L}^r$-bounded ($r > 2$), since $\mathbb{E}\left(y_{it}^{2r} 1_{\{\varepsilon_{it}\geq 0\}}\right) \leq \mathbb{E}\left[\varepsilon_{it}^{2r}\sigma_{it}^{2r}(\theta_0)\right] < \infty$. Moreover, recall that $\sum_{j=1}^N w_{ij} = 1$, then $\left\{\sum_{j=1}^N w_{ij}y_{jt}^2 : (i,t) \in D_{NT}, NT \geq 1\right\}$ is also uniformly $\mathbb{L}^2$-NED of size-μ on $\mathcal{E}$ since

$$\left\|\sum_{j=1}^N w_{ij}y_{jt}^2 - E\left[\sum_{j=1}^N w_{ij}y_{jt}^2 \middle| \mathcal{F}_{it}(s)\right]\right\|_2 \leq \sum_{j=1}^N w_{ij}\left\|y_{jt}^2 - E\left[y_{jt}^2\middle| \mathcal{F}_{it}(s)\right]\right\|_2.$$

By Assumption 4.3.4, we could also verify that $\left\{\sum_{j=1}^N w_{ij}y_{jt}^2\right\}$ is uniformly $\mathbb{L}^r$-bounded. Therefore by Lemma A.2.4 we obtain that:

$$\left\|\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0) - \mathbb{E}\left(\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)|\mathcal{F}_{it}(s)\right)\right\|_2 \leq Cs^{-\mu}. \tag{A.2.27}$$

By Lemma A.2.2(c), (A.2.26) and (A.2.27) lead to

$$\begin{aligned}
&\left\|\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0) - \mathbb{E}\left(\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)|\mathcal{F}_{it}(s)\right)\right\|_2\\
&\leq \left\|1-\varepsilon_{it}^2\right\|_2\left\|\frac{1}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0) - \mathbb{E}\left(\frac{1}{\sigma_{it}^2(\theta_0)}\frac{\partial}{\partial\theta}\sigma_{it}^2(\theta_0)|\mathcal{F}_{it}(s)\right)\right\|_2\\
&\leq Cs^{-\mu}.
\end{aligned} \tag{A.2.28}$$

According to Proposition 3.3 we complete the proof.

□

Notice that $y_{it} = \varepsilon_{it}\sigma_{it}(\theta_0)$, by (A.2.23) we have

$$\begin{aligned}\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'} =& \frac{1}{NT}\sum_{(i,t)\in D_{NT}}\left[\left(\frac{2y_{it}^2}{\sigma_{it}^6(\theta_0)} - \frac{1}{\sigma_{it}^4(\theta_0)}\right)\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}\right.\\ &\left.+\left(\frac{1}{\sigma_{it}^2(\theta_0)} - \frac{y_{it}^2}{\sigma_{it}^4(\theta_0)}\right)\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'}\right]\\ =&\frac{1}{NT}\sum_{(i,t)\in D_{NT}}\left[\frac{2\varepsilon_{it}^2-1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}\right.\\ &\left.+\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'}\right]\\ :=&\frac{1}{NT}\sum_{(i,t)\in D_{NT}}(\eta_{it}+\xi_{it}).\end{aligned}$$

With Claim A.2.8 and Claim A.2.9 below, by Theorem 3.1 we have

$$\begin{aligned}&\frac{1}{NT}\sum_{(i,t)\in D_{NT}}\eta_{it} - \frac{1}{\kappa_4-1}\Sigma_{NT}\xrightarrow{p}0,\\ &\frac{1}{NT}\sum_{(i,t)\in D_{NT}}\xi_{it}\xrightarrow{p}0.\end{aligned}$$

This leads to

$$\frac{\partial^2}{\partial\theta\partial\theta'}L_{NT}(\theta_0) - \frac{1}{\kappa_4-1}\Sigma_{NT}\xrightarrow{p}0. \tag{A.2.29}$$

Claim A.2.8. $\left\{\frac{2\varepsilon_{it}^2-1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'} : (i,t)\in D_{NT}, NT\geq 1\right\}$ *is uniformly $\mathbb{L}^p$-bounded for some $p>1$, and η-weakly dependent with $\bar{\eta}(s)\leq Cs^{-\mu}$ for some constants $C>0$ and $\mu>2$.*

Proof. By Lemma A.2.2(a) and (A.2.28) we can prove the $\mathbb{L}^1$-NED of size μ, which leads to η-weakly dependence with $\bar{\eta}(s)\leq Cs^{-\mu}$ according to Proposition 3.3. And the uniform $\mathbb{L}^p$-boundedness is directly obtained from Assumption 4.3.1 and (A.2.25). □

Claim A.2.9. $\left\{\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'} : (i,t)\in D_{NT}, NT\geq 1\right\}$ *is uniformly $\mathbb{L}^p$-bounded for some $p>1$, and η-weakly dependent with $\bar{\eta}(s)\leq Cs^{-\mu}$ for some constants $C>0$*

and $\mu > 2$.

Proof. Notice that $\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'} = M_{i,t-1} + \beta\frac{\partial^2\sigma_{i,t-1}^2(\theta_0)}{\partial\theta\partial\theta'}$ where

$$M_{i,t-1} = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\omega} \\ 0 & 0 & 0 & 0 & \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\alpha^{(1)}} \\ 0 & 0 & 0 & 0 & \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\alpha^{(2)}} \\ 0 & 0 & 0 & 0 & \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\lambda} \\ \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\omega} & \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\alpha^{(1)}} & \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\alpha^{(2)}} & \frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\lambda} & 2\frac{\partial\sigma_{i,t-1}^2(\theta_0)}{\partial\beta} \end{pmatrix}.$$

Since all entries of M_{it} are components of $\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}$, by (A.2.27), Lemma A.2.4 and Assumption 4.3.4, we have

$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'} - \mathbb{E}\left(\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'}|\mathcal{F}_{it}(s)\right)\right\|_2 \leq Cs^{-\mu}. \tag{A.2.30}$$

Lemma A.2.2(a), (A.2.26) and (A.2.30) imply that

$$\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'} - \mathbb{E}\left(\frac{1-\varepsilon_{it}^2}{\sigma_{it}^2(\theta_0)}\frac{\partial^2\sigma_{it}^2(\theta_0)}{\partial\theta\partial\theta'}|\mathcal{F}_{it}(s)\right)\right\|_1 \leq Cs^{-\mu}. \tag{A.2.31}$$

Then by Proposition 3.3 we complete the proof.

□

Since the target function $\tilde{L}_{NT}(\theta)$ is an approximation of the exact likelihood function $L_{NT}(\theta)$, we need following claim:

Claim A.2.10. *As $T\to\infty$, $N\to\infty$ and $N = o(T)$,*

(a). $\sqrt{NT}\left\|\frac{\partial\tilde{L}_{NT}(\theta_0)}{\partial\theta} - \frac{\partial L_{NT}(\theta_0)}{\partial\theta}\right\| \xrightarrow{p} 0.$

(b). $\sup_{\|\theta-\theta_0\|<\xi}\left\|\frac{\partial^2\tilde{L}_{NT}(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\theta'}\right\| = O_p(\xi).$

Proof. We start with the proof of (a). Notice that

$$\begin{aligned}&\sqrt{NT}\left|\frac{\partial\tilde{L}_{NT}(\theta_0)}{\partial\beta} - \frac{\partial L_{NT}(\theta_0)}{\partial\beta}\right| \\ \leq&\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\left|\frac{1}{\tilde{\sigma}_{it}^2(\theta_0)}\frac{\partial\tilde{\sigma}_{it}^2(\theta_0)}{\partial\beta} - \frac{1}{\sigma_{it}^2(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right|\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\left|\frac{y_{it}^2}{\tilde{\sigma}_{it}^4(\theta_0)}\frac{\partial\tilde{\sigma}_{it}^2(\theta_0)}{\partial\beta}-\frac{y_{it}^2}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right| \\
\leq&\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\left|\frac{1}{\tilde{\sigma}_{it}^2(\theta_0)}\right|\left|\frac{\partial\tilde{\sigma}_{it}^2(\theta_0)}{\partial\beta}-\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right| \\
&+\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\left|\frac{1}{\sigma_{it}^2(\theta_0)\tilde{\sigma}_{it}^2(\theta_0)}\right|\left|\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right|\left|\sigma_{it}^2(\theta_0)-\tilde{\sigma}_{it}^2(\theta_0)\right| \\
&+\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}|y_{it}^2|\left|\frac{1}{\tilde{\sigma}_{it}^4(\theta_0)}\right|\left|\frac{\partial\tilde{\sigma}_{it}^2(\theta_0)}{\partial\beta}-\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right| \\
&+\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}|y_{it}^2|\left|\frac{1}{\sigma_{it}^2(\theta_0)\tilde{\sigma}_{it}^4(\theta_0)}+\frac{1}{\sigma_{it}^4(\theta_0)\tilde{\sigma}_{it}^2(\theta_0)}\right|\left|\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right|\left|\sigma_{it}^2(\theta_0)-\tilde{\sigma}_{it}^2(\theta_0)\right| \\
\leq&\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\frac{1}{\omega_0}\left|\frac{\partial\tilde{\sigma}_{it}^2(\theta_0)}{\partial\beta}-\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right| \\
&+\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\frac{1}{\omega_0^2}\left|\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right|\left|\sigma_{it}^2(\theta_0)-\tilde{\sigma}_{it}^2(\theta_0)\right| \\
&+\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\frac{1}{\omega_0^2}|y_{it}^2|\left|\frac{\partial\tilde{\sigma}_{it}^2(\theta_0)}{\partial\beta}-\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right| \\
&+\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\frac{2}{\omega_0^3}|y_{it}^2|\left|\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}\right|\left|\sigma_{it}^2(\theta_0)-\tilde{\sigma}_{it}^2(\theta_0)\right|
\end{aligned}
$$

Firstly, in view of (A.2.16) we have:

$$
\begin{aligned}
&\left\|\frac{1}{\sqrt{NT}}\sum_{(i,t)\in D_{NT}}\left(\frac{1}{\omega_0}+\frac{y_{it}^2}{\omega_0^2}\right)\left|\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\beta}-\frac{\partial\tilde{\sigma}_{it}^2(\theta_0)}{\partial\beta}\right|\right\|_1 \\
\leq&\frac{C_1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T t\beta_0^{t-1}\left\|\frac{1}{\omega_0}+\frac{y_{it}^2}{\omega_0^2}\right\|_1+\frac{C_2}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\beta_0^t\left\|\frac{1}{\omega_0}+\frac{y_{it}^2}{\omega_0^2}\right\|_1 \\
\leq&\frac{C_3}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T t\beta_0^{t-1}+\frac{C_4}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\beta_0^t \\
\leq&\frac{C_3}{\sqrt{NT}}\sum_{i=1}^N\frac{1}{(1-\beta_0)^2}+\frac{C_4}{\sqrt{NT}}\sum_{i=1}^N\frac{\beta_0}{1-\beta_0}\to 0
\end{aligned}
\tag{A.2.32}
$$

when $T\to\infty$, $N\to\infty$ and $N=o(T)$.

Then by (A.2.13) and (A.2.15) we have

$$\begin{aligned} &\left\| \frac{1}{\sqrt{NT}} \sum_{(i,t)\in D_{NT}} \left(\frac{1}{\omega_0^2} + \frac{2y_{it}^2}{\omega_0^3} \right) \left| \sigma_{it}^2(\theta_0) - \tilde{\sigma}_{it}^2(\theta_0) \right| \left| \frac{\partial \tilde{\sigma}_{it}^2(\theta_0)}{\partial \beta} \right| \right\|_1 \\ \leq & \frac{C_1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \beta_0^t + \frac{C_2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \beta_0^t \left\| y_{it} \right\|_1 \\ \leq & \frac{C_3}{\sqrt{NT}} \sum_{i=1}^{N} \frac{\beta_0}{1-\beta_0} \to 0 \end{aligned} \tag{A.2.33}$$

when $T \to \infty$, $N \to \infty$ and $N = o(T)$. In light of (A.2.32) and (A.2.33) we can prove that

$$\sqrt{NT} \left| \frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial \beta} - \frac{\partial L_{NT}(\theta_0)}{\partial \beta} \right| \xrightarrow{p} 0.$$

The proof regarding partial derivatives w.r.t. ω, $\alpha^{(1)}$, $\alpha^{(2)}$ and λ follows similar arguments and is therefore omitted.

Now we turn to the proof of (b). For any $\theta_m, \theta_n \in \{\omega, \alpha^{(1)}, \alpha^{(2)}, \lambda, \beta\}$,

$$\begin{aligned} \frac{\partial^2 L_{NT}(\theta)}{\partial \theta_m \partial \theta_n} =& \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(\frac{2y_{it}^2}{\sigma_{it}^6(\theta)} - \frac{1}{\sigma_{it}^4(\theta)} \right) \frac{\partial \sigma_{it}^2(\theta)}{\partial \theta_m} \frac{\partial \sigma_{it}^2(\theta)}{\partial \theta_n} \right] \\ &+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(\frac{1}{\sigma_{it}^2(\theta)} - \frac{y_{it}^2}{\sigma_{it}^4(\theta)} \right) \frac{\partial^2 \sigma_{it}^2(\theta)}{\partial \theta_m \partial \theta_n} \right] \end{aligned} \tag{A.2.34}$$

Since

$$\begin{aligned} &\sup_{\|\theta-\theta_0\|<\xi} \left| \frac{\partial^2 \tilde{L}_{NT}(\theta)}{\partial \theta_m \partial \theta_n} - \frac{\partial^2 L_{NT}(\theta_0)}{\partial \theta_m \partial \theta_n} \right| \\ \leq & \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sup_{\theta\in\Theta} \left| \frac{\partial^2 \tilde{l}_{it}(\theta)}{\partial \theta_m \partial \theta_n} - \frac{\partial^2 l_{it}(\theta)}{\partial \theta_m \partial \theta_n} \right| \\ &+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sup_{\|\theta-\theta_0\|<\xi} \left| \frac{\partial^2 l_{it}(\theta)}{\partial \theta_m \partial \theta_n} - \frac{\partial^2 l_{it}(\theta_0)}{\partial \theta_m \partial \theta_n} \right|, \end{aligned} \tag{A.2.35}$$

we will handle above two terms separately.

For the first term on the right-hand-side of (A.2.35), we have

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sup_{\theta\in\Theta} \left| \frac{\partial^2 \tilde{l}_{it}(\theta)}{\partial \theta_m \partial \theta_n} - \frac{\partial^2 l_{it}(\theta)}{\partial \theta_m \partial \theta_n} \right|$$

$$
\begin{aligned}
&\le \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_{it}^2(\theta)}\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}-\frac{1}{\tilde{\sigma}_{it}^2(\theta)}\frac{\partial^2\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}\right| \\
&\quad+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_{it}^4(\theta)}\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_m}\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_n}-\frac{1}{\tilde{\sigma}_{it}^4(\theta)}\frac{\partial\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_m}\frac{\partial\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_n}\right| \\
&\quad+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\left|\frac{y_{it}^2}{\sigma_{it}^4(\theta)}\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}-\frac{y_{it}^2}{\tilde{\sigma}_{it}^4(\theta)}\frac{\partial^2\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}\right| \\
&\quad+\frac{2}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\left|\frac{2y_{it}^2}{\sigma_{it}^6(\theta)}\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_m}\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_n}-\frac{2y_{it}^2}{\tilde{\sigma}_{it}^6(\theta)}\frac{\partial\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_m}\frac{\partial\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_n}\right| \\
&:= T_1+T_2+T_3+T_4.
\end{aligned}
\tag{A.2.36}
$$

According to (A.2.13) and (A.2.19) we have:

$$
\begin{aligned}
T_3 \le&\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta} y_{it}^2\left|\frac{1}{\sigma_{it}^2(\theta)}+\frac{1}{\tilde{\sigma}_{it}^2(\theta)}\right|\left|\frac{\sigma_{it}^2(\theta)-\tilde{\sigma}_{it}^2(\theta)}{\tilde{\sigma}_{it}^2(\theta)\sigma_{it}^2(\theta)}\right|\left|\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}\right| \\
&+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\left|\frac{y_{it}^2}{\tilde{\sigma}_{it}^4(\theta)}\right|\left|\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}-\frac{\partial^2\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}\right| \\
\le&\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\frac{2y_{it}^2}{\omega^3}\left|\sigma_{it}^2(\theta)-\tilde{\sigma}_{it}^2(\theta)\right|\left|\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}\right| \\
&+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\frac{y_{it}^2}{\omega^2}\left|\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}-\frac{\partial^2\tilde{\sigma}_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}\right| \\
\le&\frac{C_1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}y_{it}^2\beta^t+\frac{C_2}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}y_{it}^2 t\beta^{t-1}+\frac{C_3}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}y_{it}^2 t(t-1)\beta^{t-2}.
\end{aligned}
$$

Therefore we have $T_3 \xrightarrow{p} 0$ as $T\to\infty$, $N\to\infty$ and $N=o(T)$. And this convergence could be derived similarly for T_1, T_2, T_4. Hence

$$
\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sup_{\theta\in\Theta}\left|\frac{\partial^2\tilde{l}_{it}(\theta)}{\partial\theta_m\partial\theta_n}-\frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n}\right| \xrightarrow{p} 0.
$$

For the second term on the right-hand-side of (A.2.35), a Taylor expansion of $\frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n}$ at θ_0 yields that

$$
\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n}-\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n}\right| \le \xi \sup_{\|\theta-\theta_0\|<\xi}\left|\frac{\partial^3 l_{it}(\theta)}{\partial\theta_m\partial\theta_n\partial\theta_l}\right|.
$$

Therefore we have

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n}-\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n}\right| \\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{\partial^3 l_{it}(\theta)}{\partial\theta_m\partial\theta_n\partial\theta_l}\right| \\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|-\frac{6y_{it}^2}{\sigma_{it}^8(\theta)}+\frac{2}{\sigma_{it}^6(\theta)}\right|\left|\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_m}\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_n}\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_l}\right| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{2y_{it}^2}{\sigma_{it}^6(\theta)}-\frac{1}{\sigma_{it}^4(\theta)}\right|\left|\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_l}\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_n}\right| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{2y_{it}^2}{\sigma_{it}^6(\theta)}-\frac{1}{\sigma_{it}^4(\theta)}\right|\left|\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_m}\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_n\partial\theta_l}\right| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{2y_{it}^2}{\sigma_{it}^6(\theta)}-\frac{1}{\sigma_{it}^4(\theta)}\right|\left|\frac{\partial\sigma_{it}^2(\theta)}{\partial\theta_l}\frac{\partial^2\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n}\right| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|-\frac{y_{it}^2}{\sigma_{it}^4(\theta)}+\frac{1}{\sigma_{it}^2(\theta)}\right|\left|\frac{\partial^3\sigma_{it}^2(\theta)}{\partial\theta_m\partial\theta_n\partial\theta_l}\right| \\
\leq&\frac{C_1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{6y_{it}^2}{\omega^8}+\frac{2}{\omega^6}\right|+\frac{C_2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{2y_{it}^2}{\omega^6}+\frac{1}{\omega^4}\right| \\
&+\frac{C_3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{2y_{it}^2}{\omega^6}+\frac{1}{\omega^4}\right|+\frac{C_4}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|\frac{2y_{it}^2}{\omega^6}+\frac{1}{\omega^4}\right| \\
&+\frac{C_5}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{\|\theta-\theta_0\|<\xi}\left|-\frac{y_{it}^2}{\omega^4}+\frac{1}{\omega^2}\right|
\end{aligned}
\tag{A.2.37}
$$

a.s. for any $\theta_l, \theta_m, \theta_n \in \{\omega, \alpha^{(1)}, \alpha^{(2)}, \xi, \beta\}$. Therefore the second term on the right-hand-side of (A.2.35) is $\mathcal{O}(\xi)$ in probability. With (A.2.36) converging to 0 in probability, we prove Claim A.2.10(b).

□

By the Taylor expansion, for some θ^* between $\hat{\theta}_{NT}$ and θ_0 we have

$$\frac{\partial\tilde{L}_{NT}(\hat{\theta}_{NT})}{\partial\theta}=\frac{\partial\tilde{L}_{NT}(\theta_0)}{\partial\theta}+\frac{\partial^2\tilde{L}_{NT}(\theta^*)}{\partial\theta\partial\theta'}(\hat{\theta}_{NT}-\theta_0).$$

Since $\frac{\partial \tilde{L}_{NT}(\hat{\theta}_{NT})}{\partial \theta} = 0$, we have

$$\sqrt{NT}\Sigma_{NT}^{1/2}(\hat{\theta}_{NT} - \theta_0) = -\Sigma_{NT}^{1/2}\left(\frac{\partial^2 \tilde{L}_{NT}(\theta^*)}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{NT}\frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial\theta}$$
$$= -\Sigma_{NT}^{1/2}\left(\Sigma_{NT}^{-1/2}\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}\Sigma_{NT}^{-1/2}\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta} + o_p(1) \tag{A.2.38}$$

according to Claim A.2.10. By Assumption 4.3.6 and (A.2.29) we have

$$\Sigma_{NT}^{-1/2}\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'} = \frac{1}{\kappa_4 - 1}\Sigma_{NT}^{1/2} + o_p(1).$$

Therefore

$$(\kappa_4-1)\left(\Sigma_{NT}^{-1/2}\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right)\Sigma_{NT}^{-1/2} = (\kappa_4-1)\left(\frac{1}{\kappa_4 - 1}\Sigma_{NT}^{1/2} + o_p(1)\right)\Sigma_{NT}^{-1/2} = I_5 + o_p(1). \tag{A.2.39}$$

Combining (A.2.24), (A.2.38) and (A.2.39) we complete the proof of Theorem 4.2.

A.2.4 Proof of Proposition 4.1

Note that

$$\hat{\kappa}_4 := \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{y_{it}^4}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})}$$

and

$$\widehat{\Sigma}_{NT} := \frac{\hat{\kappa}_4 - 1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{1}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta'}\right].$$

By the asymptotic normality of $\hat{\theta}_{NT}$ in Theorem 4.2, it suffices to prove that

$$\hat{\kappa}_4 := \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{y_{it}^4}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})} \xrightarrow{p} \kappa_4 \tag{A.2.40}$$

and

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{1}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta'} - \mathbb{E}\left[\frac{1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}\right]\right\} \xrightarrow{p} 0. \tag{A.2.41}$$

$$\hat{\kappa}_4 = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\frac{y_{it}^4}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})} - \frac{y_{it}^4}{\sigma_{it}^4(\theta_0)}\right\} + \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{y_{it}^4}{\sigma_{it}^4(\theta_0)}$$
$$:= S_1 + S_2$$

By the law of large numbers and the fact that ε_{it} is i.i.d. across i and t, we have:

$$S_2 = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}^4 \xrightarrow{p} \kappa_4.$$

Then it remains to show that $S_1 \xrightarrow{p} 0$ to prove (A.2.40).

$$S_1 = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\frac{y_{it}^4}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})} - \frac{y_{it}^4}{\sigma_{it}^4(\hat{\theta}_{NT})}\right\} + \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\frac{y_{it}^4}{\sigma_{it}^4(\hat{\theta}_{NT})} - \frac{y_{it}^4}{\sigma_{it}^4(\theta_0)}\right\}$$
$$:= S_{11} + S_{12}$$

where $S_{12} \xrightarrow{p} 0$ since $\hat{\theta}_{NT} \xrightarrow{p} \theta_0$. Meanwhile

$$\begin{aligned}
|S_{11}| \leq & \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{y_{it}^4}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})\sigma_{it}^4(\hat{\theta}_{NT})}\left|\sigma_{it}^2(\hat{\theta}_{NT}) - \tilde{\sigma}_{it}^2(\hat{\theta}_{NT})\right|\left|\sigma_{it}^2(\hat{\theta}_{NT}) + \tilde{\sigma}_{it}^2(\hat{\theta}_{NT})\right| \\
= & \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\sigma_{i0}^2(\hat{\theta}_{NT})}{\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})\sigma_{it}^2(\hat{\theta}_{NT})}\left|\frac{1}{\sigma_{it}^2(\hat{\theta}_{NT})} + \frac{1}{\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}\right|\hat{\beta}^t y_{it}^4 \\
\leq & \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{2\sigma_{i0}^2(\hat{\theta}_{NT})}{\hat{\omega}^3}\hat{\beta}^t y_{it}^4 \\
\leq & \frac{C}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho^t y_{it}^4.
\end{aligned}$$

Then $S_{11} \xrightarrow{p} 0$ according to the remark following Assumption 4.3.4. Thus (A.2.40) is proved.

As for (A.2.41), note that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{1}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta'}-\mathbb{E}\left[\frac{1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}\right]\right\}\\
=&\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\frac{1}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta'}-\frac{1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}\right\}\\
&+\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\frac{1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}-\mathbb{E}\left[\frac{1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}\right]\right\}\\
:=&T_1+T_2
\end{aligned}$$

where $T_2 \xrightarrow{p} 0$ by Claim A.2.8, and

$$\begin{aligned}
T_1=&\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\frac{1}{\tilde{\sigma}_{it}^4(\hat{\theta}_{NT})}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\tilde{\sigma}_{it}^2(\hat{\theta}_{NT})}{\partial\theta'}-\frac{1}{\sigma_{it}^4(\hat{\theta}_{NT})}\frac{\partial\sigma_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\sigma_{it}^2(\hat{\theta}_{NT})}{\partial\theta'}\right\}\\
&+\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\frac{1}{\sigma_{it}^4(\hat{\theta}_{NT})}\frac{\partial\sigma_{it}^2(\hat{\theta}_{NT})}{\partial\theta}\frac{\partial\sigma_{it}^2(\hat{\theta}_{NT})}{\partial\theta'}-\frac{1}{\sigma_{it}^4(\theta_0)}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_{it}^2(\theta_0)}{\partial\theta'}\right\}\\
:=&T_{11}+T_{12}.
\end{aligned}$$

$T_{11} \xrightarrow{p} 0$ has been proved in (A.2.36), and $T_{12} \xrightarrow{p} 0$ as $\hat{\theta}_{NT} \xrightarrow{p} \theta_0$. With (A.2.40) and (A.2.41) we complete the proof.

A.3 Proofs of results in Chapter 5

Lemma A.3.1. *If* $0 < \beta < 1$, $\mathbb{E}|y_{it}| < \infty$ *and* $\mathbb{E}|\lambda_{it}(\nu)| < \infty$ *uniformly for all* $(i,t) \in D_{NT}, NT \geq 1$, *then*

$$\lambda_{it}(\nu) = \sum_{k=1}^{\infty} \beta^{k-1} \left[\omega + \alpha_{i,t-k} y_{i,t-k} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-k} \right] \tag{A.3.1}$$

with probability one.

Proof. Let $\log^+(x) = \log(x)$ if $x > 1$ and 0 otherwise, $u_{i,t-k}(\nu) := \omega + \alpha_{i,t-k} y_{i,t-k} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-k}$. By Jensen's inequality we have

$$\begin{aligned} &\mathbb{E} \log^+ |u_{i,t-k}(\nu)| \\ \leq &\log^+ \mathbb{E} \left| \omega + \alpha_{i,t-k} y_{i,t-k} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-k} \right| \\ < &\infty. \end{aligned}$$

By Lemma 2.2 in Berkes et al. (2003) we have $\sum_{k=1}^{\infty} \mathbb{P}\left[|u_{i,t-k}(\nu)| > \zeta^k\right] < \infty$ for any $\zeta > 1$. Therefore $|u_{i,t-k}(\nu)| \leq \zeta^k$ almost surely by Borel-Cantelli lemma. Letting $1 < \zeta < \frac{1}{|\beta|}$, we can prove that the right-hand-side of (A.3.1) converges almost surely.

It remains for us to show that

$$\lambda_{it}(\nu) = \sum_{k=1}^{\infty} \beta^{k-1} u_{i,t-k}(\nu).$$

From (5.3.2) we have

$$\lambda_{it}(\nu) - \beta^k \lambda_{i,t-k-1}(\nu) = u_{i,t-1}(\nu) + \beta u_{i,t-2}(\nu) + ... + \beta^{k-1} u_{i,t-k}(\nu).$$

Using Markov's inequality we obtain that $\sum_{k=1}^{\infty} \mathbb{P}\left\{|\beta^k \lambda_{i,t-k-1}(\nu)| > \delta\right\} < \infty$ for any $\delta > 0$, then by the Borel-Cantelli lemma $|\beta^k \lambda_{i,t-k-1}(\nu)| \xrightarrow{a.s.} 0$ as $k \to \infty$. Letting $k \to \infty$ on both sides of above equation we complete the proof. $\square$

A.3.1 Proof of Theorem 5.1

Our proof of Theorem 5.1 relies on the arguments given by Doukhan et al. (2006) in their proof of Theorem 2.1. Let

$$\Lambda_t^{(0)} := \left(\lambda_{1t}^{(0)}, \lambda_{2t}^{(0)}, ..., \lambda_{Nt}^{(0)}\right)';$$
$$\mathbb{Y}_t^{(0)} := \left(M_{1t}(\lambda_{1t}^{(0)}), M_{2t}(\lambda_{2t}^{(0)}), ..., M_{Nt}(\lambda_{Nt}^{(0)})\right)',$$

where $\{\lambda_{it}^{(0)} : i = 1, 2, ..., N, t \in \mathbb{Z}\}$ are IID positive random variables with mean 1. For each $n \geq 1$, we define $\{\mathbb{Y}_t^{(n)} : t \in \mathbb{Z}\}$ and $\{\Lambda_t^{(n)} : t \in \mathbb{Z}\}$ through following recursion:

$$\begin{aligned} \mathbb{Y}_t^{(n)} &= \mathbb{M}_t(\Lambda_t^{(n)}); \\ \Lambda_t^{(n)} &= \omega \mathbf{1}_N + A(\mathbb{Y}_{t-1}^{(n-1)})\mathbb{Y}_{t-1}^{(n-1)} + \beta\Lambda_{t-1}^{(n-1)}. \end{aligned} \tag{A.3.2}$$

Claim A.3.1. *$\{\mathbb{Y}_t^{(n)} : t \in \mathbb{Z}\}$ is strictly stationary for each $n \geq 0$.*

Proof. Since $\{M_{it}(\cdot) : i = 1, 2, ..., N, t \in \mathbb{Z}\}$ are independent Poisson processes with unit intensity, then for any t and h we have

$$\begin{aligned} &\mathbb{P}\left\{\mathbb{Y}_{1+h}^{(n)} = \mathbf{y}_1, ..., \mathbb{Y}_{t+h}^{(n)} = \mathbf{y}_t\right\} \\ =&\mathbb{E}\left(\mathbb{P}\left\{\mathbb{Y}_{1+h}^{(n)} = \mathbf{y}_1, ..., \mathbb{Y}_{t+h}^{(n)} = \mathbf{y}_t \,\middle|\, \Lambda_{1+h}^{(n)}, ..., \Lambda_{t+h}^{(n)}\right\}\right) \\ =&\mathbb{E}\left(\mathbb{P}\left\{\mathbb{M}_{1+h}(\Lambda_{1+h}^{(n)}) = \mathbf{y}_1, ..., \mathbb{M}_{t+h}(\Lambda_{t+h}^{(n)}) = \mathbf{y}_t \,\middle|\, \Lambda_{1+h}^{(n)}, ..., \Lambda_{t+h}^{(n)}\right\}\right) \\ =&\mathbb{E}\left(\prod_{k=1}^{t}\prod_{i=1}^{N} \frac{\left(\lambda_{i,k+h}^{(n)}\right)^{y_{ik}}}{y_{ik}!} e^{-\lambda_{i,k+h}^{(n)}}\right). \end{aligned} \tag{A.3.3}$$

When $n = 0$, $\mathbb{P}\left\{\mathbb{Y}_{1+h}^{(0)} = \mathbf{y}_1, ..., \mathbb{Y}_{t+h}^{(0)} = \mathbf{y}_t\right\}$ is h-invariant for any t and h, by (A.3.3) and the IID of $\{\lambda_{it}^{(0)} : i = 1, 2, ..., N, t \in \mathbb{Z}\}$. Therefore $\{\mathbb{Y}_t^{(0)} : t \in \mathbb{Z}\}$ is strictly stationary. Assume that $\{\mathbb{Y}_t^{(n-1)} : t \in \mathbb{Z}\}$ and $\{\Lambda_t^{(n-1)} : t \in \mathbb{Z}\}$ are strictly stationary, then $\{\Lambda_t^{(n)} : t \in \mathbb{Z}\}$ is also strictly stationary since $\Lambda_t^{(n)} = \omega\mathbf{1}_N + A(\mathbb{Y}_{t-1}^{(n-1)})\mathbb{Y}_{t-1}^{(n-1)} + \beta\Lambda_{t-1}^{(n-1)}$. According to (A.3.3) and the strict stationarity of $\{\Lambda_t^{(n)} : t \in \mathbb{Z}\}$, we have $\{\mathbb{Y}_t^{(n)} : t \in \mathbb{Z}\}$ being strictly stationary too. Claim A.3.1 can be proved by induction.

□

Claim A.3.2. *$\mathbb{E}\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| \leq C\rho^n$ for some constants $C > 0$ and $0 < \rho < 1$.*

Proof. It is easy to verify that

$$\mathbb{E}\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| = \mathbb{E}\left[\mathbb{E}\left(\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| \middle| \Lambda_t^{(n+1)}, \Lambda_t^{(n)}\right)\right] = \mathbb{E}\left\|\Lambda_t^{(n+1)} - \Lambda_t^{(n)}\right\|.$$

Recall from (A.3.2) that

$$\Lambda_t^{(n)} = \omega \mathbf{1}_N + A(\mathbb{Y}_{t-1}^{(n-1)})\mathbb{Y}_{t-1}^{(n-1)} + \beta\Lambda_{t-1}^{(n-1)},$$

then

$$\begin{aligned}&\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| \\ \leq &\left\|A(\mathbb{Y}_{t-1}^{(n)})\mathbb{Y}_{t-1}^{(n)} - A(\mathbb{Y}_{t-1}^{(n-1)})\mathbb{Y}_{t-1}^{(n-1)}\right\| + \beta\left\|\Lambda_{t-1}^{(n)} - \Lambda_{t-1}^{(n-1)}\right\|.\end{aligned} \tag{A.3.4}$$

Define a function $\psi(y) = \alpha^{(1)} 1_{\{y \geq r\}} y + \alpha^{(2)} 1_{\{y<r\}} y$ for $y \in \mathbb{N}$, then Assumption 5.2.1(b) assures that $\psi(y)$ is non-decreasing on $\mathbb{N}$. Let $y', y \in \mathbb{N}$ such that $y' > y$:

- If $y' > y \geq r$, we have $0 < \psi(y') - \psi(y) = \alpha^{(1)}(y' - y) \leq \alpha^*(y' - y)$ where $\alpha^* = \max\{\alpha^{(1)}, \alpha^{(2)}\}$;
- If $r > y' > y$, we have $0 < \psi(y') - \psi(y) = \alpha^{(2)}(y' - y) \leq \alpha^*(y' - y)$;
- If $y' \geq r > y$, we have $0 < \psi(y') - \psi(y) = \alpha^{(1)}y' - \alpha^{(2)}y \leq \alpha^{(1)}(y' - y) \leq \alpha^*(y' - y)$ by Assumption 5.2.1(a).

Similarly when $y' \leq y$, we have $0 \geq \psi(y') - \psi(y) \geq \alpha^*(y' - y)$. Therefore we obtain that:

$$|\psi(y') - \psi(y)| \leq \alpha^*|y' - y| \tag{A.3.5}$$

for any $y', y \in \mathbb{N}$. Then we have:

$$\begin{aligned}&\left|\left(A(\mathbb{Y}_{t-1}^{(n)})\mathbb{Y}_{t-1}^{(n)} - A(\mathbb{Y}_{t-1}^{(n-1)})\mathbb{Y}_{t-1}^{(n-1)}\right)_i\right| \\ =&\left|\psi(y_{i,t-1}^{(n)}) - \psi(y_{i,t-1}^{(n-1)}) + \xi\sum_{j=1}^N w_{ij}(y_{j,t-1}^{(n)} - y_{j,t-1}^{(n-1)})\right| \\ \leq&\alpha^*\left|y_{i,t-1}^{(n)} - y_{i,t-1}^{(n-1)}\right| + \xi\sum_{j=1}^N w_{ij}\left|y_{j,t-1}^{(n)} - y_{j,t-1}^{(n-1)}\right|\end{aligned} \tag{A.3.6}$$

for $i = 1, 2, ..., N$, where $(\mathbb{Y})_i$ is the i-th element of $\mathbb{Y}$.

Combining (A.3.4) and (A.3.6) we have

$$\begin{aligned}
&\mathbb{E}\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| \\
\leq &\mathbb{E}\left\|(\alpha^* I_N + \xi W + \beta I_N)(\mathbb{Y}_{t-1}^{(n)} - \mathbb{Y}_{t-1}^{(n-1)})\right\| \\
\leq &\rho(\alpha^* I_N + \xi W + \beta I_N)\mathbb{E}\left\|\mathbb{Y}_{t-1}^{(n)} - \mathbb{Y}_{t-1}^{(n-1)}\right\| \\
\leq &|\alpha^* + \xi + \beta|\mathbb{E}\left\|\mathbb{Y}_{t-1}^{(n)} - \mathbb{Y}_{t-1}^{(n-1)}\right\|
\end{aligned}$$

where $\rho(\cdot)$ denotes the spectral radius, and the last inequality is due to the Gershgorin circle theorem. Let $\rho := |\alpha^* + \xi + \beta|$, we have:

$$\begin{aligned}
&\mathbb{E}\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| \\
\leq &\rho\mathbb{E}\left\|\mathbb{Y}_{t-1}^{(n)} - \mathbb{Y}_{t-1}^{(n-1)}\right\| \\
\leq &\rho^n\mathbb{E}\left\|\mathbb{Y}_{t-n}^{(1)} - \mathbb{Y}_{t-n}^{(0)}\right\| \\
= &\rho^n\mathbb{E}\left\|\Lambda_{t-n}^{(1)} - \Lambda_{t-n}^{(0)}\right\| \\
\leq &C\rho^n
\end{aligned}$$

for some $0 < \rho < 1$ and $C = \mathbb{E}\left\|\Lambda_{t-n}^{(1)} - \Lambda_{t-n}^{(0)}\right\| < \infty$.

□

By Claim A.3.2,

$$\begin{aligned}
\mathbb{P}\left\{\mathbb{Y}_t^{(n+1)} \neq \mathbb{Y}_t^{(n)}\right\} = &\sum_{h=1}^{\infty}\mathbb{P}\left\{\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| = h\right\} \\
\leq &\mathbb{E}\left\|\mathbb{Y}_t^{(n+1)} - \mathbb{Y}_t^{(n)}\right\| \\
\leq &C\rho^n.
\end{aligned}$$

Therefore $\sum_{n=1}^{\infty}\mathbb{P}\left\{\mathbb{Y}_t^{(n+1)} \neq \mathbb{Y}_t^{(n)}\right\} < \infty$, and

$$\mathbb{P}\left\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\left[\mathbb{Y}_t^{(k+1)} \neq \mathbb{Y}_t^{(k)}\right]\right\} = 0$$

according to Borel-Cantelli lemma. This indicates that, there exists M such that for all $n > M$, $\mathbb{Y}_t^{(n)}$ equals (almost surely) to some $\mathbb{Y}_t$ with integer components. i.e. $\mathbb{Y}_t = \lim_{n\to\infty} \mathbb{Y}_t^{(n)}$ exists almost surely. Apparently, $\{\mathbb{Y}_t : t \in \mathbb{Z}\}$ is strictly stationary since $\{\mathbb{Y}_t^{(n)} : t \in \mathbb{Z}\}$ is strictly stationary for each $n \geq 0$, according to Claim A.3.1.

At last, by Claim A.3.2 we also have:

$$\mathbb{E}\left\|\mathbb{Y}_t^{(n+m)} - \mathbb{Y}_t^{(n)}\right\| \leq \sum_{k=0}^{m-1} \mathbb{E}\left\|\mathbb{Y}_t^{(n+k+1)} - \mathbb{Y}_t^{(n+k)}\right\| \leq C\rho^n \sum_{k=0}^{m-1} \rho^k,$$

for any $n, m \in \mathbb{N}$. Therefore $\{\mathbb{Y}_t^{(n)} : n \geq 0\}$ is a Cauchy sequence in $\mathbb{L}^1$, hence $\mathbb{E}\|\mathbb{Y}_t\| < \infty$.

A.3.2 Proof of Theorem 5.2

By Lemma A.3.1 we have

$$\lambda_{it}(\nu) = \sum_{k=1}^{\infty} \beta^{k-1} \left[\omega + \alpha_{i,t-k} y_{i,t-k} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-k}\right]$$

and

$$\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \sup_{\nu\in\Theta\times\mathbb{Z}_+} |\lambda_{it}(\nu)| < \infty \tag{A.3.7}$$

with probability one, where $\alpha_{i,t-k} = \alpha^{(1)} 1_{\{y_{i,t-k}\geq r\}} + \alpha^{(2)} 1_{\{y_{i,t-k}<r\}}$. Given initial values $\tilde{\lambda}_{i0} = 0$ for $i = 1, 2, ..., N$, we could replace $\lambda_{it}(\nu)$ with $\tilde{\lambda}_{it}(\nu)$ and get

$$\tilde{\lambda}_{it}(\nu) = \sum_{k=1}^{t} \beta^{k-1} \left[\omega + \alpha_{i,t-k} y_{i,t-k} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-k}\right]$$

for $i = 1, 2, ..., N, t \geq 1$. Therefore we have

$$\lambda_{it}(\nu) - \tilde{\lambda}_{it}(\nu) = \beta^t \lambda_{i0}(\nu). \tag{A.3.8}$$

Now we are ready to prove the consistency of $\hat{\nu}_{NT}$ when $T \to \infty$ and $N \to \infty$. The proof is broken up into Claim A.3.3 to Claim A.3.6 below: Claim A.3.3 shows that the choice of initial values is asymptotically negligible; Claims A.3.4 and A.3.5 verify the weak dependence of $\{l_{it}(\nu) : (i,t) \in D_{NT}, NT \geq 1\}$, and facilitate the adoption of LLN; Claim A.3.6 is concerned with the identifiability of the true parameters ν_0.

Claim A.3.3. *For any $\nu \in \Theta \times \mathbb{Z}_+$, $|L_{NT}(\nu) - \tilde{L}_{NT}(\nu)| \xrightarrow{p} 0$ as $T \to \infty$ and $N \to \infty$.*

Proof. The proof is similar to that of Claim A.2.3 and is omitted here. □

Claim A.3.4. *The functions $l_{it}(\nu)$ are uniformly $\mathbb{L}^p$-bounded for some $p > 1$, i.e.*

$$\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \sup_{\nu\in\Theta\times\mathbb{Z}_+} \|l_{it}(\nu)\|_p < \infty.$$

Proof. According to Hölder's inequality, we have

$$\begin{aligned}\|l_{it}(\nu)\|_p &= \|y_{it}\log\lambda_{it}(\nu) - \lambda_{it}(\nu)\|_p \\ &\leq \|y_{it}\log\lambda_{it}(\nu)\|_p + \|\lambda_{it}(\nu)\|_p \\ &\leq \|y_{it}\|_{2p}\|\log\lambda_{it}(\nu)\|_{2p} + \|\lambda_{it}(\nu)\|_p .\end{aligned}$$

Notice that

$$\begin{aligned}&\sup_{\nu\in\Theta\times\mathbb{Z}_+} \|\log\lambda_{it}(\nu)\|_{2p} \\ \leq& \sup_{\nu\in\Theta\times\mathbb{Z}_+} \left\|\log^+\lambda_{it}(\nu)\right\|_{2p} + \sup_{\nu\in\Theta\times\mathbb{Z}_+} \left\|\log^-\lambda_{it}(\nu)\right\|_{2p} \\ \leq& \sup_{\nu\in\Theta\times\mathbb{Z}_+} \|\lambda_{it}(\nu)+1\|_{2p} + \sup_{\nu\in\Theta\times\mathbb{Z}_+} \max\{-\log(\omega), 0\},\end{aligned}$$

where $\log^-(x) = 0$ if $x \geq 1$ and $\log^-(x) = -\log(x)$ if $0 < x < 1$. Then by Assumption 5.3.2(a) and (A.3.7) we complete the proof. □

Claim A.3.5. *For any $\nu \in \Theta \times \mathbb{Z}_+$, the array of random fields $\{l_{it}(\nu) : (i,t) \in D_{NT}, NT \geq 1\}$ is η-weakly dependent with coefficients $\bar{\eta}_0(r) \leq Cr^{-\mu_0}$ where $\mu_0 > 2$.*

Proof. For each $(i,t) \in D_{NT}$ and $h = 1, 2, ...$, define $\{y^{(h)}_{j\tau} : (j,\tau) \in D_{NT}, NT \geq 1\}$ such that $y^{(h)}_{j\tau} \neq y_{j\tau}$ if and only if $\rho((i,t),(j,\tau)) = h$.

$$\lambda^{(h)}_{it}(\nu) = \sum_{k=1}^{\infty} \beta^{k-1} \left[\omega + \alpha^{(h)}_{i,t-k} y^{(h)}_{i,t-k} + \xi \sum_{j=1}^{N} w_{ij} y^{(h)}_{j,t-k} \right],$$

where

$$\alpha^{(h)}_{i,t-k} = \alpha^{(1)} 1_{\{y^{(h)}_{i,t-k} \geq r\}} + \alpha^{(2)} 1_{\{y^{(h)}_{i,t-k} < r\}}.$$

Then by (A.3.5) and Assumption 5.3.3 we have

$$\begin{aligned}
&|\lambda_{it}(\nu) - \lambda^{(h)}_{it}(\nu)| \\
\leq & \sum_{k=1}^{\infty} \beta^{k-1} |\alpha_{i,t-k} y_{i,t-k} - \alpha^{(h)}_{i,t-k} y^{(h)}_{i,t-k}| + \sum_{k=1}^{\infty} \sum_{j=1}^{N} \beta^{k-1} \xi w_{ij} |y_{j,t-k} - y^{(h)}_{j,t-k}| \\
= & \beta^{h-1} |\alpha_{i,t-h} y_{i,t-h} - \alpha^{(h)}_{i,t-h} y^{(h)}_{i,t-h}| + \xi \beta^{h-1} \sum_{1 \leq |j-i| \leq h} w_{ij} |y_{j,t-h} - y^{(h)}_{j,t-h}| \\
& + \xi w_{i,i\pm h} \sum_{k=1}^{h} \beta^{k-1} |y_{i\pm h,t-k} - y^{(h)}_{i\pm h,t-k}| \\
\leq & \alpha^{*} \beta^{h-1} |y_{i,t-h} - y^{(h)}_{i,t-h}| + \xi \beta^{h-1} \sum_{1 \leq |j-i| \leq h} |y_{j,t-h} - y^{(h)}_{j,t-h}| \\
& + C \xi h^{-b} \sum_{k=1}^{h} |y_{i\pm h,t-k} - y^{(h)}_{i\pm h,t-k}|.
\end{aligned} \tag{A.3.9}$$

Therefore $\lambda_{it}(\nu)$ satisfies condition (3.2.6) with $B_{(i,t),NT}(h) \leq Ch^{-b}$ and $l = 0$. By Proposition 3.2 and (3.2.10), the array of random fields $\{\lambda_{it}(\nu) : (i,t) \in D_{NT}, NT \geq 1\}$ is η-weakly dependent with coefficients $\bar{\eta}_{\lambda}(r) \leq Cr^{-\mu_y + 2}$.

Similarly we can define

$$l^{(h)}_{it}(\nu) = y^{(h)}_{it} \log \lambda^{(h)}_{it}(\nu) - \lambda^{(h)}_{it}(\nu).$$

Since

$$|l_{it}(\nu) - l^{(h)}_{it}(\nu)| \leq y_{it} \left| \log \frac{\lambda_{it}(\nu)}{\lambda^{(h)}_{it}(\nu)} \right| + |\lambda_{it}(\nu) - \lambda^{(h)}_{it}(\nu)|$$

$$
\begin{aligned}
&\le y_{it}\left|\frac{\lambda_{it}(\nu)}{\lambda_{it}^{(h)}(\nu)}-1\right|+|\lambda_{it}(\nu)-\lambda_{it}^{(h)}(\nu)| \\
&\le \frac{y_{it}}{\omega}|\lambda_{it}(\nu)-\lambda_{it}^{(h)}(\nu)|+|\lambda_{it}(\nu)-\lambda_{it}^{(h)}(\nu)|,
\end{aligned}
$$

$l_{it}(\nu)$ also satisfies condition 3.2.6 with $B_{(i,t),NT}(h) \le Ch^{-b}$ and $l=1$ by (A.3.9), the array of random fields $\{l_{it}(\nu) : (i,t) \in D_{NT}, NT \ge 1\}$ is η-weakly dependent with coefficients $\bar{\eta}_0(r) \le Cr^{-\frac{2p-2}{2p-1}\mu_y+2}$. Notice that $\frac{2p-2}{2p-1}\mu_y - 2 > 2$ since $\mu_y > \frac{4p-2}{p-1}$. □

Claim A.3.6. *$\lambda_{it}(\nu) = \lambda_{it}(\nu_0)$ for all $(i,t) \in D_{NT}$ if and only if $\nu = \nu_0$.*

Proof. The *if* part is obvious, it remains for us to prove the *only if* part. Observe that

$$
(1-\beta B)\lambda_{it}(\nu) = \omega + \alpha B y_{it} + \xi \sum_{j=1}^{N} w_{ij} B y_{jt},
$$

where B stands for the back-shift operator in the sense that $By_{it}^2 = y_{i,t-1}^2$, and α represents either $\alpha^{(1)}$ or $\alpha^{(2)}$ according to the value of α_{it} at time t. Therefore we have

$$
(1-\beta B)\Lambda_t(\nu) = \omega \mathbf{1}_N + (\alpha B I_N + \xi B W)\mathbb{Y}_t.
$$

The polynomial $1-\beta x$ has a root $x = 1/\beta$, which lies outside the unit circle since $0 < \beta < 1$. Therefore the inverse $\frac{1}{1-\beta x}$ is well-defined for any $|x| \le 1$, and we have

$$
\Lambda_t(\nu) = \frac{\omega}{1-\beta}\mathbf{1}_N + \mathcal{P}_\nu(B)\mathbb{Y}_t
$$

with $\mathcal{P}_\nu(B) := \frac{\alpha B}{1-\beta B} I_N + \frac{\xi B}{1-\beta B} W$. As $\lambda_{it}(\nu) = \lambda_{it}(\nu_0)$ for each $i = 1, 2, ..., N$,

$$
[\mathcal{P}_\nu(B) - \mathcal{P}_{\nu_0}(B)]\,\mathbb{Y}_t = \left(\frac{\omega_0}{1-\beta_0} - \frac{\omega}{1-\beta}\right)\mathbf{1}_N.
$$

We can deduce from above equation that $\mathcal{P}_\nu(x) = \mathcal{P}_{\nu_0}(x)$ for any $|x| \le 1$, otherwise $\mathbb{Y}_t$ will be degenerated to a deterministic vector given $\mathcal{H}_{t-1}$. $\mathcal{P}_\nu(x) = \mathcal{P}_{\nu_0}(x)$ implies that

$$
\frac{\alpha x}{1-\beta x} I_N - \frac{\alpha_0 x}{1-\beta_0 x} I_N = \left(\frac{\xi_0 x}{1-\beta_0 x} - \frac{\xi x}{1-\beta x}\right) W.
$$

The diagonal elements of W are all zeros while the matrix on the left side of above equation has non-zero diagonal elements, so we have

$$\frac{\alpha x}{1-\beta x} = \frac{\alpha_0 x}{1-\beta_0 x},$$
$$\frac{\xi x}{1-\beta x} = \frac{\xi_0 x}{1-\beta_0 x},$$

which imply $\alpha = \alpha_0$, $\beta = \beta_0$ and $\xi = \xi_0$. Besides, $\omega = \omega_0$ could be easily derived from $\frac{\omega}{1-\beta} = \frac{\omega_0}{1-\beta_0}$.

□

With Claim A.3.4 and Claim A.3.5, we can apply Theorem 3.1 and prove the consistency of $\hat{\nu}_{NT}$ following similar arguments in the proof of Theorem 4.2.

A.3.3 Proof of Theorem 5.3

With a fixed threshold parameter $r = r_0$, we will rewrite $\hat{\theta}_{NT} := \hat{\theta}_{NT}^{(r_0)}$, $\lambda_{it}(\theta) := \lambda_{it}(\theta, r_0)$ and $l_{it}(\theta) := l_{it}(\theta, r_0)$ etc., in succeeding proofs for notation simplicity. Before we prove the asymptotic normality, we derive some intermediate results regarding the first, second and third order derivatives of $\lambda_{it}(\theta)$. These results are repeatedly used in later proofs.

Since

$$\lambda_{it}(\theta) = \sum_{k=1}^{\infty} \beta^{k-1} \left[\omega + \left(\alpha^{(1)} 1_{\{y_{i,t-k} \geq r\}} + \alpha^{(2)} 1_{\{y_{i,t-k} < r\}} \right) y_{i,t-k} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-k} \right]$$

almost surely, the partial derivative of $\lambda_{it}(\theta)$ are

$$\frac{\partial \lambda_{it}(\theta)}{\partial \omega} = \sum_{k=1}^{\infty} \beta^{k-1},$$
$$\frac{\partial \lambda_{it}(\theta)}{\partial \alpha^{(1)}} = \sum_{k=1}^{\infty} \beta^{k-1} y_{i,t-k} 1_{\{y_{i,t-k} \geq r\}},$$
$$\frac{\partial \lambda_{it}(\theta)}{\partial \alpha^{(2)}} = \sum_{k=1}^{\infty} \beta^{k-1} y_{i,t-k} 1_{\{y_{i,t-k} < r\}}, \tag{A.3.10}$$

$$\begin{aligned}
\frac{\partial \lambda_{it}(\theta)}{\partial \xi} &= \sum_{k=1}^{\infty} \beta^{k-1} \left(\sum_{j=1}^{N} w_{ij} y_{j,t-k} \right), \\
\frac{\partial \lambda_{it}(\theta)}{\partial \beta} &= \sum_{k=2}^{\infty} (k-1) \beta^{k-2} u_{i,t-k}(\theta),
\end{aligned}$$

where

$$u_{i,t-k}(\theta) = \omega + \alpha^{(1)} y_{i,t-k} 1_{\{y_{i,t-k} \geq r\}} + \alpha^{(2)} y_{i,t-k} 1_{\{y_{i,t-k} < r\}} + \xi \sum_{j=1}^{N} w_{ij} y_{j,t-k}.$$

We also notice that

$$\frac{\partial \lambda_{it}(\theta)}{\partial \theta} - \frac{\partial \tilde{\lambda}_{it}(\theta)}{\partial \theta} = t\beta^{t-1} \lambda_{i0}(\nu) \mathbf{e}_5 + \beta^t \frac{\partial \lambda_{i0}(\theta)}{\partial \theta}, \tag{A.3.11}$$

where $\mathbf{e}_5 = (0, 0, 0, 0, 1)'$.

Now we consider the second order derivatives. For any $\theta_m, \theta_n \in \{\omega, \alpha^{(1)}, \alpha^{(2)}, \xi\}$,

$$\frac{\partial^2 \lambda_{it}(\theta)}{\partial \theta_m \partial \theta_n} = 0.$$

Also

$$\begin{aligned}
\frac{\partial^2 \lambda_{it}(\theta)}{\partial \omega \partial \beta} &= \sum_{k=2}^{\infty} (k-1) \beta^{k-2}, \\
\frac{\partial^2 \lambda_{it}(\theta)}{\partial \alpha^{(1)} \partial \beta} &= \sum_{k=2}^{\infty} (k-1) \beta^{k-2} y_{i,t-k} 1_{\{y_{i,t-k} \geq r\}}, \\
\frac{\partial^2 \lambda_{it}(\theta)}{\partial \alpha^{(2)} \partial \beta} &= \sum_{k=2}^{\infty} (k-1) \beta^{k-2} y_{i,t-k} 1_{\{y_{i,t-k} < r\}}, \qquad \text{(A.3.12)} \\
\frac{\partial^2 \lambda_{it}(\theta)}{\partial \xi \partial \beta} &= \sum_{k=2}^{\infty} (k-1) \beta^{k-2} \left(\sum_{j=1}^{N} w_{ij} y_{j,t-k} \right), \\
\frac{\partial^2 \lambda_{it}(\theta)}{\partial \beta^2} &= \sum_{k=3}^{\infty} (k-1)(k-2) \beta^{k-3} u_{i,t-k}(\theta).
\end{aligned}$$

We also have:

$$\frac{\partial^2\lambda_{it}(\nu)}{\partial\theta\partial\theta'}-\frac{\partial^2\tilde{\lambda}_{it}(\nu)}{\partial\theta\partial\theta'}=t(t-1)\beta^{t-2}\lambda_{i0}(\nu)\mathbf{e}_5\mathbf{e}_5'+2t\beta^{t-1}\frac{\partial\lambda_{i0}(\nu)}{\partial\theta}\mathbf{e}_5'+\beta^t\frac{\partial^2\lambda_{i0}(\nu)}{\partial\theta\partial\theta'}, \quad \text{(A.3.13)}$$

where $\mathbf{e}_5=(0,0,0,0,1)'$.

As for the third order derivatives of $\lambda_{it}(\theta)$,

$$\begin{aligned}
\frac{\partial^3\lambda_{it}(\theta)}{\partial\omega\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3},\\
\frac{\partial^3\lambda_{it}(\theta)}{\partial\alpha^{(1)}\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3}y_{i,t-k}1_{\{y_{i,t-k}\geq r\}},\\
\frac{\partial^3\lambda_{it}(\theta)}{\partial\alpha^{(2)}\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3}y_{i,t-k}1_{\{y_{i,t-k}< r\}},\\
\frac{\partial^3\lambda_{it}(\theta)}{\partial\xi\partial\beta^2} &= \sum_{k=3}^{\infty}(k-1)(k-2)\beta^{k-3}\left(\sum_{j=1}^{N}w_{ij}y_{j,t-k}\right),\\
\frac{\partial^3\lambda_{it}(\theta)}{\partial\beta^3} &= \sum_{k=4}^{\infty}(k-1)(k-2)(k-3)\beta^{k-4}u_{i,t-k}(\theta).
\end{aligned} \quad \text{(A.3.14)}$$

Based on the consistency of $\hat{\theta}_{NT}$, we are now ready to prove asymptotic normality. We split the proof into Claim A.3.7 to Claim A.3.10 below.

Claim A.3.7. $\sqrt{NT}\left|\frac{\partial\tilde{L}_{NT}(\theta_0)}{\partial\theta}-\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\right|\xrightarrow{p}0$ *as* $\min\{N,T\}\to\infty$ *and* $T/N\to\infty$.

Proof.

$$\begin{cases}\dfrac{\partial L_{NT}(\theta)}{\partial\theta}=\dfrac{1}{NT}\displaystyle\sum_{(i,t)\in D_{NT}}\left(\frac{y_{it}}{\lambda_{it}(\theta)}-1\right)\frac{\partial\lambda_{it}(\theta)}{\partial\theta},\\[2ex] \dfrac{\partial\lambda_{it}(\theta)}{\partial\theta}=\mathbf{h}_{i,t-1}+\beta\dfrac{\partial\lambda_{i,t-1}(\theta)}{\partial\theta},\end{cases} \quad \text{(A.3.15)}$$

where

$$\mathbf{h}_{i,t-1}:=\left(1,y_{i,t-1}1_{\{y_{i,t-1}\geq r\}},y_{i,t-1}1_{\{y_{i,t-1}<r\}},\sum_{j=1}^{N}w_{ij}y_{j,t-1},\lambda_{i,t-1}\right)'.$$

Similarly we have

$$\begin{cases} \dfrac{\partial \tilde{L}_{NT}(\theta)}{\partial \theta} = \dfrac{1}{NT} \displaystyle\sum_{(i,t)\in D_{NT}} \left(\dfrac{y_{it}}{\tilde{\lambda}_{it}(\theta)} - 1 \right) \dfrac{\partial \tilde{\lambda}_{it}(\theta)}{\partial \theta}, \\ \dfrac{\partial \tilde{\lambda}_{it}(\theta)}{\partial \theta} = \tilde{\mathbf{h}}_{i,t-1} + \beta \dfrac{\partial \tilde{\lambda}_{i,t-1}(\theta)}{\partial \theta}. \end{cases} \tag{A.3.16}$$

Therefore we have

$$\begin{aligned} &\sqrt{NT} \left| \frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial \beta} - \frac{\partial L_{NT}(\theta_0)}{\partial \beta} \right| \\ \leq & \frac{1}{\sqrt{NT}} \sum_{(i,t)\in D_{NT}} \left| y_{it} \left[\frac{\lambda_{it}(\theta_0) - \tilde{\lambda}_{it}(\theta_0)}{\tilde{\lambda}_{it}(\theta_0)\lambda_{it}(\theta_0)} \frac{\partial \tilde{\lambda}_{it}(\theta_0)}{\partial \beta} \right. \right. \\ & \left. \left. + \frac{1}{\lambda_{it}(\theta_0)} \left(\frac{\partial \tilde{\lambda}_{it}(\theta_0)}{\partial \beta} - \frac{\partial \lambda_{it}(\theta_0)}{\partial \beta} \right) \right] - \left(\frac{\partial \tilde{\lambda}_{it}(\theta_0)}{\partial \beta} - \frac{\partial \lambda_{it}(\theta_0)}{\partial \beta} \right) \right| \\ \leq & \frac{1}{\sqrt{NT}} \sum_{(i,t)\in D_{NT}} \frac{y_{it}}{\omega_0^2} \left| \lambda_{it}(\theta_0) - \tilde{\lambda}_{it}(\theta_0) \right| \left| \frac{\partial \tilde{\lambda}_{it}(\theta_0)}{\partial \beta} \right| \\ & + \frac{1}{\sqrt{NT}} \sum_{(i,t)\in D_{NT}} \left(\frac{y_{it}}{\omega_0} + 1 \right) \left| \frac{\partial \lambda_{it}(\theta_0)}{\partial \beta} - \frac{\partial \tilde{\lambda}_{it}(\theta_0)}{\partial \beta} \right|. \end{aligned}$$

Firstly, by Assumption 5.3.2(a) and (A.3.8) we have

$$\begin{aligned} & \left\| \frac{1}{\sqrt{NT}} \sum_{(i,t)\in D_{NT}} \frac{y_{it}}{\omega_0^2} \left| \lambda_{it}(\theta_0) - \tilde{\lambda}_{it}(\theta_0) \right| \left| \frac{\partial \tilde{\lambda}_{it}(\theta_0)}{\partial \beta} \right| \right\|_1 \\ \leq & \frac{C_1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \beta_0^t \left\| y_{it} \right\|_1 \\ \leq & \frac{C_2}{\sqrt{NT}} \sum_{i=1}^{N} \frac{\beta_0}{1-\beta_0} \to 0 \end{aligned} \tag{A.3.17}$$

when $\min\{N, T\} \to \infty$ and $T/N \to \infty$. Then in view of (A.3.11):

$$\begin{aligned} & \left\| \frac{1}{\sqrt{NT}} \sum_{(i,t)\in D_{NT}} \left(\frac{y_{it}}{\omega_0} + 1 \right) \left| \frac{\partial \lambda_{it}(\theta_0)}{\partial \beta} - \frac{\partial \tilde{\lambda}_{it}(\theta_0)}{\partial \beta} \right| \right\|_1 \\ \leq & \frac{C_1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} t\beta_0^{t-1} \left\| \frac{y_{it}}{\omega_0} + 1 \right\|_1 + \frac{C_2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \beta_0^t \left\| \frac{y_{it}}{\omega_0} + 1 \right\|_1 \end{aligned} \tag{A.3.18}$$

$$\leq \frac{C_3}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} t\beta_0^{t-1} + \frac{C_4}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \beta_0^t$$
$$\leq \frac{C_3}{\sqrt{NT}} \sum_{i=1}^{N} \frac{1}{(1-\beta_0)^2} + \frac{C_4}{\sqrt{NT}} \sum_{i=1}^{N} \frac{\beta_0}{1-\beta_0} \to 0$$

when $\min\{N,T\} \to \infty$ and $T/N \to \infty$. In light of (A.3.17) and (A.3.18) we can prove that

$$\sqrt{NT} \left| \frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial \beta} - \frac{\partial L_{NT}(\theta_0)}{\partial \beta} \right| \xrightarrow{p} 0.$$

The proofs regarding partial derivatives w.r.t. ω, $\alpha^{(1)}$, $\alpha^{(2)}$ and ξ follow similar arguments and are therefore omitted.

□

Claim A.3.8. $\sup_{|\theta-\theta_0|<\xi} \left| \frac{\partial^2 \tilde{L}_{NT}(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\theta'} \right| = \mathcal{O}_p(\xi)$ *as* $\min\{N,T\} \to \infty$ *and* $T/N \to \infty$*.*

Proof. For any $\theta_m, \theta_n \in \{\omega, \alpha^{(1)}, \alpha^{(2)}, \xi, \beta\}$,

$$\begin{aligned} &\frac{\partial^2 L_{NT}(\theta)}{\partial\theta_m\partial\theta_n} \\ =&\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(\frac{y_{it}}{\lambda_{it}(\theta)} - 1 \right) \frac{\partial^2 \lambda_{it}(\theta)}{\partial\theta_m\partial\theta_n} - \frac{y_{it}}{\lambda_{it}^2(\theta)} \frac{\partial \lambda_{it}(\theta)}{\partial\theta_m} \frac{\partial \lambda_{it}(\theta)}{\partial\theta_n} \right], \end{aligned} \quad \text{(A.3.19)}$$

and

$$\begin{aligned} &\frac{\partial^2 \tilde{L}_{NT}(\theta)}{\partial\theta_m\partial\theta_n} \\ =&\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left(\frac{y_{it}}{\tilde{\lambda}_{it}(\theta)} - 1 \right) \frac{\partial^2 \tilde{\lambda}_{it}(\theta)}{\partial\theta_m\partial\theta_n} - \frac{y_{it}}{\tilde{\lambda}_{it}^2(\theta)} \frac{\partial \tilde{\lambda}_{it}(\theta)}{\partial\theta_m} \frac{\partial \tilde{\lambda}_{it}(\theta)}{\partial\theta_n} \right]. \end{aligned} \quad \text{(A.3.20)}$$

Since

$$\begin{aligned} &\sup_{|\theta-\theta_0|<\xi} \left| \frac{\partial^2 \tilde{L}_{NT}(\theta)}{\partial\theta_m\partial\theta_n} - \frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta_m\partial\theta_n} \right| \\ \leq&\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sup_{\theta\in\Theta} \left| \frac{\partial^2 \tilde{l}_{it}(\theta)}{\partial\theta_m\partial\theta_n} - \frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n} \right| \\ &+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sup_{|\theta-\theta_0|<\xi} \left| \frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n} - \frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} \right|, \end{aligned} \quad \text{(A.3.21)}$$

we will handle above two terms separately.

For the first term on the right-hand-side of (A.3.21), we have

$$\begin{aligned}
&\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left|\frac{\partial^2\tilde{l}_{it}(\theta)}{\partial\theta_m\partial\theta_n}-\frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n}\right|\right\|_1\\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|y_{it}\sup_{\theta\in\Theta}\left(\frac{1}{\lambda_{it}}-\frac{1}{\tilde{\lambda}_{it}}\right)\sup_{\theta\in\Theta}\frac{\partial^2\lambda_{it}}{\partial\theta_m\partial\theta_n}\right\|_1\\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\sup_{\theta\in\Theta}\left(\frac{y_{it}}{\tilde{\lambda}_{it}}-1\right)\sup_{\theta\in\Theta}\left(\frac{\partial^2\lambda_{it}}{\partial\theta_m\partial\theta_n}-\frac{\partial^2\tilde{\lambda}_{it}}{\partial\theta_m\partial\theta_n}\right)\right\|_1\\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|y_{it}\sup_{\theta\in\Theta}\left(\frac{\lambda_{it}^2}{\tilde{\lambda}_{it}^2}-1\right)\sup_{\theta\in\Theta}\frac{1}{\lambda_{it}^2}\frac{\partial\lambda_{it}}{\partial\theta_m}\frac{\partial\lambda_{it}}{\partial\theta_n}\right\|_1 \qquad \text{(A.3.22)}\\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\sup_{\theta\in\Theta}\frac{y_{it}}{\tilde{\lambda}_{it}^2}\left[\frac{\partial\tilde{\lambda}_{it}}{\partial\theta_m}\left(\frac{\partial\tilde{\lambda}_{it}}{\partial\theta_n}-\frac{\partial\lambda_{it}}{\partial\theta_n}\right)\right]\right\|_1\\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\sup_{\theta\in\Theta}\frac{y_{it}}{\tilde{\lambda}_{it}^2}\left[\frac{\partial\lambda_{it}}{\partial\theta_n}\left(\frac{\partial\tilde{\lambda}_{it}}{\partial\theta_m}-\frac{\partial\lambda_{it}}{\partial\theta_m}\right)\right]\right\|_1\\
:=&T_1+T_2+T_3+T_4+T_5
\end{aligned}$$

Analogous to the proof of (A.3.17) we can show that $T_1\to 0$ and $T_3\to 0$ as $\min\{N,T\}\to\infty$ and $T/N\to\infty$. In light of (A.3.13), we can also verify that

$$T_2\leq\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[C_1t(t-1)\rho^{t-2}+C_2t\rho^{t-1}+C_3\rho^t]\left\|\sup_{\theta\in\Theta}\left(\frac{y_{it}}{\tilde{\lambda}_{it}}-1\right)\right\|_1.$$

Then $T_2\to 0$ as well. Similarly, using (A.3.11) we obtain that $T_4\to 0$ and $T_5\to 0$. Then it remains to investigate the second term in the right-hand-side of (A.3.21).

A Taylor expansion of $\frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n}$ at θ_0 yields that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sup_{|\theta-\theta_0|<\xi}\left|\frac{\partial^2 l_{it}(\theta)}{\partial\theta_m\partial\theta_n}-\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n}\right|\\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{|\theta-\theta_0|<\xi}\left|\frac{\partial^3 l_{it}(\theta)}{\partial\theta_m\partial\theta_n\partial\theta_l}\right|\\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{|\theta-\theta_0|<\xi}\left|\frac{y_{it}}{\lambda_{it}}-1\right|\left|\frac{\partial^3\lambda_{it}}{\partial\theta_m\partial\theta_n\partial\theta_l}\right|
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{|\theta-\theta_0|<\xi}\left|\frac{2y_{it}}{\lambda_{it}^3}\right|\left|\frac{\partial\lambda_{it}}{\partial\theta_l}\frac{\partial\lambda_{it}}{\partial\theta_m}\frac{\partial\lambda_{it}}{\partial\theta_n}\right| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{|\theta-\theta_0|<\xi}\left|\frac{y_{it}}{\lambda_{it}^2}\right|\left|\frac{\partial\lambda_{it}}{\partial\theta_l}\frac{\partial^2\lambda_{it}}{\partial\theta_m\partial\theta_n}\right| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{|\theta-\theta_0|<\xi}\left|\frac{y_{it}}{\lambda_{it}^2}\right|\left|\frac{\partial\lambda_{it}}{\partial\theta_n}\frac{\partial^2\lambda_{it}}{\partial\theta_l\partial\theta_m}\right| \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi\sup_{|\theta-\theta_0|<\xi}\left|\frac{y_{it}}{\lambda_{it}^2}\right|\left|\frac{\partial\lambda_{it}}{\partial\theta_m}\frac{\partial^2\lambda_{it}}{\partial\theta_n\partial\theta_l}\right| \\
&:= B_1+B_2+B_3+B_4+B_5
\end{aligned}
\tag{A.3.23}
$$

for any $\theta_l,\theta_m,\theta_n\in\{\omega,\alpha^{(1)},\alpha^{(2)},\xi,\beta\}$. According to Assumption 5.3.2(a), (A.3.14) we can verify that

$$
\mathbb{E}\left|\frac{y_{it}}{\lambda_{it}}-1\right|\left|\frac{\partial^3\lambda_{it}}{\partial\theta_m\partial\theta_n\partial\theta_l}\right|<\infty,
$$

hence $B_1=\mathcal{O}(\xi)$ in probability. The other terms could be verified following similar lines, in light of (A.3.10) and (A.3.12).

Taking (A.3.22) and (A.3.23) back to (A.3.21), we complete the proof.

□

Claim A.3.9. *(a).* $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{\partial l_{it}(\theta_0)}{\partial\theta}\right\|_{2p}<\infty$ *for some* $p>1$*;*

(b). For each $\mathbf{v}\in\mathbb{R}^5$ *such that* $|\mathbf{v}|=1$, $\left\{\mathbf{v}'\frac{\partial l_{it}(\theta_0)}{\partial\theta}:(i,t)\in D_{NT},NT\geq 1\right\}$ *are* η*-weakly dependent, with dependence coefficients* $\bar{\eta}_1(r)\leq Cr^{-\mu_1}$ *where* $\mu_1>4\vee\frac{2p-1}{p-1}$.

Proof. Recall from (A.3.15) that

$$
\frac{\partial l_{it}(\theta_0)}{\partial\theta}=\frac{y_{it}}{\lambda_{it}(\theta_0)}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta}-\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta}.
$$

By Assumption 5.3.2 we could prove (a).

Now we verify (b). In the proof of Claim A.3.5, for each $(i,t)\in D_{NT}$ and $h=1,2,...$, we defined $\{y_{j\tau}^{(h)}:(j,\tau)\in D_{NT},NT\geq 1\}$ such that $y_{j\tau}^{(h)}\neq y_{j\tau}$ if and only if $\rho((i,t),(j,\tau))=h$. At first, we verify that $\frac{\partial l_{it}(\theta_0)}{\partial\beta}$ satisfies condition (3.2.6). Notice that

$$
\begin{aligned}
&\left|\frac{\partial l_{it}(\theta_0)}{\partial\beta}-\frac{\partial l_{it}^{(h)}(\theta_0)}{\partial\beta}\right| \\
\leq & y_{it}\left|\frac{1}{\lambda_{it}(\theta_0)}\frac{\partial\lambda_{it}(\theta_0)}{\partial\beta}-\frac{1}{\lambda_{it}^{(h)}(\theta_0)}\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\beta}\right|+\left|\frac{\partial\lambda_{it}(\theta_0)}{\partial\beta}-\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\beta}\right| \\
\leq & \left|\frac{y_{it}}{\omega_0}+1\right|\left|\frac{\partial\lambda_{it}(\theta_0)}{\partial\beta}-\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\beta}\right|+\frac{y_{it}}{\omega_0^2}\left|\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\beta}\right|\left|\lambda_{it}(\theta_0)-\lambda_{it}^{(h)}(\theta_0)\right|.
\end{aligned}
\tag{A.3.24}
$$

Since

$$
\frac{\partial\lambda_{it}(\theta_0)}{\partial\beta}=\sum_{k=2}^{\infty}(k-1)\beta_0^{k-2}u_{i,t-k}(\theta_0),
$$

where

$$
u_{i,t-k}(\theta_0)=\omega_0+\alpha_0^{(1)}y_{i,t-k}1_{\{y_{i,t-k}\geq r_0\}}+\alpha_0^{(2)}y_{i,t-k}1_{\{y_{i,t-k}<r_0\}}+\xi_0\sum_{j=1}^{N}w_{ij}y_{j,t-k}.
$$

Following analogous arguments in (A.3.9), we obtain that

$$
\begin{aligned}
\left|\frac{\partial\lambda_{it}(\theta_0)}{\partial\beta}-\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\beta}\right| \leq & \alpha_0^*(h-1)\beta_0^{h-2}|y_{i,t-h}-y_{i,t-h}^{(h)}| \\
&+\xi_0(h-1)\beta_0^{h-2}\sum_{1\leq|i-j|\leq h}|y_{j,t-h}-y_{j,t-h}^{(h)}| \\
&+Ch^{-b}\sum_{k=2}^{h}|y_{i\pm h,t-k}-y_{i\pm h,t-k}^{(h)}|.
\end{aligned}
\tag{A.3.25}
$$

Combining (A.3.9), (A.3.24) and (A.3.25) we can verify that $\frac{\partial l_{it}(\theta_0)}{\partial\beta}$ satisfies condition (3.2.6) with $B_{(i,t),NT}(h)\leq Ch^{-b}$ and $l=1$. Partial derivatives of $l_{it}(\theta_0)$ with respect to other parameters in θ_0 follows similarly. Therefore $\mathbf{v}'\frac{\partial l_{it}(\theta_0)}{\partial\theta}$ satisfies condition (3.2.6) with $B_{(i,t),NT}(h)\leq Ch^{-b}$ and $l=1$ for each $\mathbf{v}\in\mathbb{R}^5$.

According to Proposition 3.2 and (3.2.10), the array of random fields $\{\mathbf{v}'\frac{\partial l_{it}(\theta_0)}{\partial\theta}:(i,t)\in D_{NT},NT\geq 1\}$ is η-weakly dependent with coefficients $\bar{\eta}_1(r)\leq Cr^{-\frac{2p-2}{2p-1}\mu_y+2}$. Notice that $\frac{2p-2}{2p-1}\mu_y-2>4\vee\frac{2p-1}{p-1}$ since $\mu_y>\frac{6p-3}{p-1}\vee\frac{(4p-3)(2p-1)}{2(p-1)^2}$.

□

Claim A.3.10. *(a).* $\sup_{NT\geq 1}\sup_{(i,t)\in D_{NT}}\left\|\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta\partial\theta'}\right\|_p < \infty$ *for some* $p > 1$*;*

(b). With respect to all $\theta_m, \theta_n \in \{\omega, \alpha^{(1)}, \alpha^{(2)}, \xi, \beta\}$, $\left\{\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} : (i,t)\in D_{NT}, NT\geq 1\right\}$ *are* η*-weakly dependent, with dependence coefficients* $\bar{\eta}_2(r) \leq Cr^{-\mu_2}$ *where* $\mu_2 > 2$*.*

Proof. Recall from (A.3.19) that

$$\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} = \left(\frac{y_{it}}{\lambda_{it}(\theta_0)} - 1\right)\frac{\partial^2\lambda_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} - \frac{y_{it}}{\lambda_{it}^2(\theta_0)}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta_m}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta_n}.$$

Then Claim A.3.10(a) could be directly obtained by Assumption 5.3.2(a).

Same as previous proofs, for each $(i,t)\in D_{NT}$ and $h = 1, 2, ...$, we defined $\{y_{j\tau}^{(h)} : (j,\tau)\in D_{NT}, NT\geq 1\}$ such that $y_{j\tau}^{(h)} \neq y_{j\tau}$ if and only if $\rho((i,t),(j,\tau)) = h$. To prove (b), we verify that $\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n}$ satisfies condition(3.2.6). Firstly we have:

$$\begin{aligned}
&\left|\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} - \frac{\partial^2 l_{it}^{(h)}(\theta_0)}{\partial\theta_m\partial\theta_n}\right| \\
\leq&\left|\frac{y_{it}}{\lambda_{it}(\theta_0)} + 1\right|\left|\frac{\partial^2\lambda_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} - \frac{\partial^2\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m\partial\theta_n}\right| + y_{it}\left|\frac{\partial^2\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m\partial\theta_n}\right|\left|\frac{1}{\lambda_{it}(\theta_0)} - \frac{1}{\lambda_{it}^{(h)}(\theta_0)}\right| \\
&+ \frac{y_{it}}{\lambda_{it}^2(\theta_0)}\left|\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta_m}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta_n} - \frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m}\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_n}\right| \\
&+ \left|\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m}\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_n}\right|\left|\frac{y_{it}}{\lambda_{it}^2(\theta_0)} - \frac{y_{it}}{(\lambda_{it}^{(h)}(\theta_0))^2}\right| \qquad \text{(A.3.26)} \\
\leq&\left|\frac{y_{it}}{\lambda_{it}(\theta_0)} + 1\right|\left|\frac{\partial^2\lambda_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} - \frac{\partial^2\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m\partial\theta_n}\right| \\
&+ \frac{y_{it}}{\lambda_{it}(\theta_0)\lambda_{it}^{(h)}(\theta_0)}\left|\frac{\partial^2\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m\partial\theta_n}\right|\left|\lambda_{it}(\theta_0) - \lambda_{it}^{(h)}(\theta_0)\right| \\
&+ \frac{y_{it}}{\lambda_{it}^2(\theta_0)}\left|\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta_m}\right|\left|\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta_n} - \frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_n}\right| \\
&+ \frac{y_{it}}{\lambda_{it}^2(\theta_0)}\left|\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_n}\right|\left|\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta_m} - \frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m}\right| \\
&+ \frac{y_{it}}{\lambda_{it}(\theta_0)\lambda_{it}^{(h)}(\theta_0)}\left|\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m}\frac{\partial\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_n}\right|\left|\frac{1}{\lambda_{it}(\theta_0)} + \frac{1}{\lambda_{it}^{(h)}(\theta_0)}\right|\left|\lambda_{it}(\theta_0) - \lambda_{it}^{(h)}(\theta_0)\right| \\
\leq&\left(\frac{y_{it}}{\omega_0} + 1\right)\left|\frac{\partial^2\lambda_{it}(\theta_0)}{\partial\theta_m\partial\theta_n} - \frac{\partial^2\lambda_{it}^{(h)}(\theta_0)}{\partial\theta_m\partial\theta_n}\right| + C_1\frac{y_{it}}{\omega_0^2}\left|\lambda_{it}(\theta_0) - \lambda_{it}^{(h)}(\theta_0)\right|
\end{aligned}$$

$$+ C_2 \frac{y_{it}}{\omega_0^2} \left| \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta_n} - \frac{\partial \lambda_{it}^{(h)}(\theta_0)}{\partial \theta_n} \right| + C_3 \frac{y_{it}}{\omega_0^2} \left| \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta_m} - \frac{\partial \lambda_{it}^{(h)}(\theta_0)}{\partial \theta_m} \right|$$
$$+ C_4 \frac{y_{it}}{\omega_0^3} \left| \lambda_{it}(\theta_0) - \lambda_{it}^{(h)}(\theta_0) \right|.$$

Taking the second order derivative with respect to ξ and β as an example, analogous to (A.3.9) and (A.3.25) we have:

$$\begin{aligned} &\left| \frac{\partial^2 \lambda_{it}(\theta_0)}{\partial \xi \partial \beta} - \frac{\partial^2 \lambda_{it}^{(h)}(\theta_0)}{\partial \xi \partial \beta} \right| \\ \leq &\sum_{k=2}^{\infty} (k-1)\beta^{k-2} \left| \sum_{j=1}^{N} w_{ij} y_{j,t-k} - \sum_{j=1}^{N} w_{ij} y_{j,t-k}^{(h)} \right| \\ \leq &(h-1)\beta_0^{h-2} \sum_{|i-j|\leq h} |y_{j,t-h} - y_{j,t-h}^{(h)}| \\ &+ Ch^{-b} \sum_{k=2}^{h} |y_{i\pm h,t-k} - y_{i\pm h,t-k}^{(h)}|. \end{aligned} \tag{A.3.27}$$

Proofs regarding second order derivatives with respect to other parameters follow similar arguments and are omitted. Substituting (A.3.9), (A.3.25) and (A.3.27) back to (A.3.26), we have that $\frac{\partial^2 l_{it}(\theta_0)}{\partial \theta_m \partial \theta_n}$ satisfies condition (3.2.6) with $B_{(i,t),NT}(h) \leq Ch^{-b}$ and $l = 1$.

According to Proposition 3.2 and (3.2.10), the array of random fields $\{\frac{\partial^2 l_{it}(\theta_0)}{\partial \theta_m \partial \theta_n} : (i,t) \in D_{NT}, NT \geq 1\}$ is η-weakly dependent with coefficients $\bar{\eta}_1(r) \leq Cr^{-\frac{2p-2}{2p-1}\mu_y + 2}$, and $\frac{2p-2}{2p-1}\mu_y - 2 > 2$.

□

By the Taylor expansion, for some θ^* between $\hat{\theta}_{NT}$ and θ_0 we have

$$\frac{\partial \tilde{L}_{NT}(\hat{\theta}_{NT})}{\partial \theta} = \frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial \theta} + \frac{\partial^2 \tilde{L}_{NT}(\theta^*)}{\partial \theta \partial \theta'} (\hat{\theta}_{NT} - \theta_0).$$

Since $\frac{\partial \tilde{L}_{NT}(\hat{\theta}_{NT})}{\partial \theta} = 0$, we have

$$\sqrt{NT}\Sigma_{NT}^{1/2}(\hat{\theta}_{NT} - \theta_0)$$

$$= -\Sigma_{NT}^{1/2}\left(\frac{\partial^2 \tilde{L}_{NT}(\theta^*)}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{NT}\frac{\partial \tilde{L}_{NT}(\theta_0)}{\partial\theta} \tag{A.3.28}$$
$$= -\Sigma_{NT}^{1/2}\left(\Sigma_{NT}^{-1/2}\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}\Sigma_{NT}^{-1/2}\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta} + o_p(1)$$

according to Claims A.3.7 and A.3.8.

Notice that $y_{it} = M_{it}(\lambda_{it}(\theta_0))$ is Poisson distributed with mean $\lambda_{it}(\theta_0)$ conditioning on historical information $\mathcal{H}_{t-1}$, with $\{M_{it} : (i,t) \in D_{NT}, NT \geq 1\}$ being IID Poisson point processes with intensity 1. Therefore we have

$$\begin{aligned}
&\mathbb{E}\left(\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left\{\mathbb{E}\left[\left(\frac{M_{it}(\lambda_{it}(\theta_0))}{\lambda_{it}(\theta_0)} - 1\right)\frac{\partial^2\lambda_{it}(\theta_0)}{\partial\theta\partial\theta'}|\mathcal{H}_{t-1}\right]\right\}\\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left\{\mathbb{E}\left[\frac{M_{it}(\lambda_{it}(\theta_0))}{\lambda_{it}^2(\theta_0)}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta'}|\mathcal{H}_{t-1}\right]\right\}\\
=&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{\lambda_{it}(\theta_0)}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta'}\right]\\
=&-\Sigma_{NT}.
\end{aligned}$$

By Claim A.3.10 and Theorem 3.1 we have

$$\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'} + \Sigma_{NT} \xrightarrow{p} 0. \tag{A.3.29}$$

According to condition (5.3.5) we can further prove that

$$-\left(\Sigma_{NT}^{-1/2}\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right)\Sigma_{NT}^{-1/2} = \left(\Sigma_{NT}^{1/2} + o_p(1)\right)\Sigma_{NT}^{-1/2} = I_5 + o_p(1). \tag{A.3.30}$$

When $\tau \neq t$ or $j \neq i$ we have

$$\mathbb{E}\left[\left(\frac{M_{it}(\lambda_{it}(\theta_0))}{\lambda_{it}(\theta_0)} - 1\right)\left(\frac{M_{j\tau}(\lambda_{j\tau}(\theta_0))}{\lambda_{j\tau}(\theta_0)} - 1\right)\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta}\frac{\partial\lambda_{j\tau}(\theta_0)}{\partial\theta'}|\mathcal{H}_{t-1}\right] = 0$$

assuming $\tau < t$. Then we can verify that

$$\begin{aligned}
&\text{Var}\left(\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\right)\\
=&\frac{1}{NT}\mathbb{E}\left\{\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{M_{it}(\lambda_{it}(\theta_0))}{\lambda_{it}(\theta_0)}-1\right)\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta}\right]\right.\\
&\left.\times\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{M_{it}(\lambda_{it}(\theta_0))}{\lambda_{it}(\theta_0)}-1\right)\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta'}\right]\right\}\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[\left(\frac{M_{it}(\lambda_{it}(\theta_0))}{\lambda_{it}(\theta_0)}-1\right)^2\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta}\frac{\partial\lambda_{it}(\theta_0)}{\partial\theta'}\right]\\
=&\Sigma_{NT}.
\end{aligned}$$

For each $\mathbf{v}\in\mathbb{R}^5$, $\text{Var}\left(\sum_{(i,t)\in D_{NT}}\mathbf{v}'\frac{\partial l_{it}(\theta_0)}{\partial\theta}\right)=(NT)\mathbf{v}'\Sigma_{NT}\mathbf{v}$. By (5.3.5) and the symmetry of Σ_{NT},

$$\inf_{NT\geq 1}\mathbf{v}'\Sigma_{NT}\mathbf{v}>0.$$

Then by Claim A.3.9 and Theorem 3.2 we can prove that

$$[(NT)\mathbf{v}'\Sigma_{NT}\mathbf{v}]^{-1/2}\mathbf{v}'(NT)\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\xrightarrow{d}N(0,1).$$

According to the Cramér-Wold theorem, we have:

$$(\Sigma_{NT})^{-1/2}\sqrt{NT}\frac{\partial L_{NT}(\theta_0)}{\partial\theta}\xrightarrow{d}N(0,I_5). \tag{A.3.31}$$

Combining (A.3.28), (A.3.30) and (A.3.31) we complete the proof of Theorem 5.3.

A.3.4 Proof of Proposition 5.1

Recalling from (5.3.7), the Wald statistic is

$$W_{NT}:=(\Gamma\hat{\theta}_{NT}-\eta)'\left\{\frac{\Gamma}{NT}\widehat{\Sigma}_{NT}^{-1}\Gamma'\right\}^{-1}(\Gamma\hat{\theta}_{NT}-\eta),$$

where

$$\widehat{\Sigma}_{NT} := \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\theta}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta'} \right].$$

It suffices to show that

$$\frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\theta}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta'} \right] \xrightarrow{p} \Sigma_{NT}. \tag{A.3.32}$$

Firstly,

$$\begin{aligned}
&\frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\theta}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta'} \right] - \Sigma_{NT} \\
=&\frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\theta}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta'} - \mathbb{E}\left(\frac{1}{\lambda_{it}(\theta_0)} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta'} \right) \right] \\
=&\frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\theta}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta'} - \frac{1}{\lambda_{it}(\theta_0)} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta'} \right] \\
&+ \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\lambda_{it}(\theta_0)} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta'} - \mathbb{E}\left(\frac{1}{\lambda_{it}(\theta_0)} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta'} \right) \right] \\
:=&T_1 + T_2.
\end{aligned}$$

Similar to the proof of Claim A.3.10, we can verify that the LLN Theorem 3.1 applies to $\left\{ \frac{1}{\lambda_{it}(\theta_0)} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta'} : (i,t) \in D_{NT}, NT \geq 1 \right\}$ and therefore $T_2 \xrightarrow{p} 0$.

T_1 can be further decomposed as follows:

$$\begin{aligned}
&\frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\theta}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta'} - \frac{1}{\lambda_{it}(\theta_0)} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta'} \right] \\
=&\frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\tilde{\lambda}_{it}(\hat{\theta}_{NT})} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \tilde{\lambda}_{it}(\hat{\theta}_{NT})}{\partial \theta'} - \frac{1}{\lambda_{it}(\hat{\theta}_{NT})} \frac{\partial \lambda_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \lambda_{it}(\hat{\theta}_{NT})}{\partial \theta'} \right] \\
&+ \frac{1}{NT} \sum_{(i,t)\in D_{NT}} \left[\frac{1}{\lambda_{it}(\hat{\theta}_{NT})} \frac{\partial \lambda_{it}(\hat{\theta}_{NT})}{\partial \theta} \frac{\partial \lambda_{it}(\hat{\theta}_{NT})}{\partial \theta'} - \frac{1}{\lambda_{it}(\theta_0)} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{it}(\theta_0)}{\partial \theta'} \right] \\
:=&S_1 + S_2.
\end{aligned}$$

$S_2 \xrightarrow{p} 0$ since $\hat{\theta}_{NT} \xrightarrow{p} \theta_0$. And the proof of $S_1 \xrightarrow{p} 0$ is similar to the proof of (A.3.22), therefore omitted.

A.4 Proofs of results in Chapter 6

A.4.1 Proof of Theorem 6.1

Firstly, in Claims A.4.1 and A.4.2 below we will verify Conditions 1 and 2 in Wu and Shao (2004) respectively.

Claim A.4.1. *There exists an* $\mathbf{x} \in \mathbb{R}_+^N$ *such that* $\mathbb{E}\left|\mathbf{x} - \mathbb{G}(\mathbf{x}, \mathbf{u})\right|_\infty < \infty$.

Proof. Assume that $|\mathbf{x} - \mathbb{G}(\mathbf{x}, \mathbf{u})|_\infty = \left|x_i - g_{\theta_0}\left(F_{x_i}^{-1}(u_i), \sum_{j=1}^N w_{ij} F_{x_j}^{-1}(u_j), x_i\right)\right|$ for some $i \in \{1, 2, ..., N\}$ without loss of generality. Then by Assumption 6.3.1 we have

$$
\begin{aligned}
&\mathbb{E}\left|\mathbf{x} - \mathbb{G}(\mathbf{x}, \mathbf{u})\right|_\infty \\
=&\int_0^1 \int_0^1 ... \int_0^1 \left|x_i - g_{\theta_0}\left(F_{x_i}^{-1}(u_i), \sum_{j=1}^N w_{ij} F_{x_j}^{-1}(u_j), x_i\right)\right| du_1 du_2 ... du_N \\
\leq& x_i + g_{\theta_0}(0,0,0) + \rho_1 \int_0^1 F_{x_i}^{-1}(u_i) du_i \\
&+ \rho_2 \int_0^1 \int_0^1 ... \int_0^1 \sum_{j=1}^N w_{ij} F_{x_j}^{-1}(u_j) du_1 du_2 ... du_N + \rho_3 x_i \\
\leq& g_{\theta_0}(0,0,0) + (1 + \rho_1 + \rho_3) x_i + \rho_2 \sum_{j=1}^N w_{ij} x_j,
\end{aligned}
$$

where $\int_0^1 F_x^{-1}(u) du = x$ since u follows a uniform distribution on $(0, 1)$. By choosing $\mathbf{x}$ such that $|\mathbf{x}|_\infty < \infty$ we complete the proof. □

Claim A.4.2. *There exists an* $\mathbf{x}' \in \mathbb{R}_+^N$*, constants* $C > 0$ *and* $\rho \in (0, 1)$ *such that*

$$
\mathbb{E}\left|\mathbb{X}_t(\mathbf{x}) - \mathbb{X}_t(\mathbf{x}')\right|_\infty \leq C\rho^t |\mathbf{x} - \mathbf{x}'|_\infty
$$

for all $\mathbf{x} \in \mathbb{R}_+^N$ *and* $t \in \mathbb{N}$.

Proof. According to (6.3.4), for some $i \in \{1, 2, ..., N\}$:

$$
\begin{aligned}
&\mathbb{E}\left|\mathbb{X}_1(\mathbf{x}) - \mathbb{X}_1(\mathbf{x}')\right|_\infty \\
=&\mathbb{E}\left|\mathbb{G}_{\mathbb{U}_1}(\mathbf{x}) - \mathbb{G}_{\mathbb{U}_1}(\mathbf{x}')\right|_\infty
\end{aligned}
$$

$$
\begin{aligned}
=&\int_0^1\int_0^1 ... \int_0^1 \left| g_{\theta_0}\left(F_{x_i}^{-1}(u_i), \sum_{j=1}^N w_{ij}F_{x_j}^{-1}(u_j), x_i\right)\right.\\
&\left. -g_{\theta_0}\left(F_{x_i'}^{-1}(u_i), \sum_{j=1}^N w_{ij}F_{x_j'}^{-1}(u_j), x_i'\right)\right| du_1du_2...du_N\\
\leq&\rho_1\int_0^1 \left|F_{x_i}^{-1}(u_i) - F_{x_i'}^{-1}(u_i)\right| du_i\\
&+\rho_2\int_0^1\int_0^1 ... \int_0^1 \sum_{j=1}^N w_{ij}\left|F_{x_j}^{-1}(u_j) - F_{x_j'}^{-1}(u_j)\right| du_1du_2...du_N\\
&+\rho_3|x_i - x_i'|\\
\leq&(\rho_1+\rho_3)|x_i - x_i'| + \rho_2\sum_{j=1}^N w_{ij}|x_j - x_j'|\\
\leq&(\rho_1+\rho_2+\rho_3)|\mathbf{x}-\mathbf{x}'|_\infty.
\end{aligned}
$$

Assume that $\mathbb{E}\left|\mathbb{X}_t(\mathbf{x}) - \mathbb{X}_t(\mathbf{x}')\right|_\infty \leq (\rho_1+\rho_2+\rho_3)^t|\mathbf{x}-\mathbf{x}'|_\infty$, then we obtain that

$$
\begin{aligned}
&\mathbb{E}\left|\mathbb{X}_{t+1}(\mathbf{x}) - \mathbb{X}_{t+1}(\mathbf{x}')\right|_\infty\\
=&\mathbb{E}\left|\mathbb{G}_{\mathbb{U}_{t+1}}(\mathbb{X}_t(\mathbf{x})) - \mathbb{G}_{\mathbb{U}_{t+1}}(\mathbb{X}_t(\mathbf{x}'))\right|_\infty\\
=&\mathbb{E}\left[\mathbb{E}\left(\left|\mathbb{G}_{\mathbb{U}_{t+1}}(\mathbb{X}_t(\mathbf{x})) - \mathbb{G}_{\mathbb{U}_{t+1}}(\mathbb{X}_t(\mathbf{x}'))\right|_\infty |\mathbb{U}_t, \mathbb{U}_{t-1}, ..., \mathbb{U}_1\right)\right]\\
\leq&\mathbb{E}\left[(\rho_1+\rho_2+\rho_3)\left|\mathbb{X}_t(\mathbf{x}) - \mathbb{X}_t(\mathbf{x}')\right|_\infty\right]\\
=&(\rho_1+\rho_2+\rho_3)^{t+1}|\mathbf{x}-\mathbf{x}'|_\infty.
\end{aligned}
\tag{A.4.1}
$$

Therefore Claim A.4.2 could be proved by induction. □

Claims A.4.1 and A.4.2 allow us to apply Theorem 2 in Wu and Shao (2004) on the backward iteration process $\mathbb{Z}_t$, hence Theorem 6.1(a) is proved, and $\{\mathbb{X}_t : t \geq 0\}$ is geometric moment contracting with unique stationary distribution π.

To prove (b) we still need to verify that $\mathbb{E}_\pi|\mathbb{X}_t|_\infty < \infty$. By (6.3.2) and (6.3.4), there exists some $i \in \{1, 2, ..., N\}$ such that

$$
\mathbb{E}|\mathbb{X}_1(\mathbf{x})|_\infty = \mathbb{E}(|\mathbb{X}_1|_\infty|\mathbb{X}_0 = \mathbf{x})
$$

$$
\begin{aligned}
&= \mathbb{E}\left[g_{\theta_0}\left(F_{x_i}^{-1}(u_{i1}), \sum_{j=1}^{N} w_{ij} F_{x_j}^{-1}(u_{j1}), x_i \right) \middle| \mathbb{X}_0 = \mathbf{x} \right] \\
&\leq g_{\theta_0}(0,0,0) + (\rho_1 + \rho_2 + \rho_3)|\mathbf{x}|_\infty.
\end{aligned}
$$

Notice that for any $i = 1, 2, ..., N$, $\mathbb{E}\left(F_{x_{i1}}^{-1}(u_{i2})|\mathbb{X}_0 = \mathbf{x}\right) = \mathbb{E}\left(x_{i1}|\mathbb{X}_0 = \mathbf{x}\right)$ since $\mathbb{U}_t$'s belong to another space that is independent from $\mathcal{X}$. Then there exists some $i \in \{1, 2, ..., N\}$ such that

$$
\begin{aligned}
\mathbb{E}|\mathbb{X}_2(\mathbf{x})|_\infty =& \mathbb{E}(|\mathbb{X}_2|_\infty|\mathbb{X}_0 = \mathbf{x}) \\
& \mathbb{E}\left[g_{\theta_0}\left(F_{x_{i1}}^{-1}(u_{i2}), \sum_{j=1}^{N} w_{ij} F_{x_{j1}}^{-1}(u_{j2}), x_{i1} \right) \middle| \mathbb{X}_0 = \mathbf{x} \right] \\
\leq & g_{\theta_0}(0,0,0) + (\rho_1 + \rho_2 + \rho_3)\mathbb{E}|\mathbb{X}_1(\mathbf{x})|_\infty.
\end{aligned}
$$

Iterative calculation leads to

$$
\mathbb{E}|\mathbb{X}_t(\mathbf{x})|_\infty \leq \frac{1 - (\rho_1 + \rho_2 + \rho_3)^t}{1 - (\rho_1 + \rho_2 + \rho_3)} g_{\theta_0}(0,0,0) + (\rho_1 + \rho_2 + \rho_3)^t |\mathbf{x}|_\infty.
$$

By Theorem 6.1(a) we have $\mathbb{X}_t(\mathbf{x}) \xrightarrow{d} \mathbb{Z}_\infty \sim \pi$ for all $\mathbf{x} \in \mathcal{X}$ as $t \to \infty$. Choosing $\mathbf{x}$ such that $|\mathbf{x}|_\infty < \infty$, we have

$$
\mathbb{E}_\pi|\mathbb{X}_t|_\infty \leq \liminf_{t\to\infty} \mathbb{E}|\mathbb{X}_t(\mathbf{x})|_\infty \leq \frac{g_{\theta_0}(0,0,0)}{1 - (\rho_1 + \rho_2 + \rho_3)} < \infty,
$$

according to Theorem 3.4 in Billingsley (1999).

A.4.2 Proof of results in Section 6.4

Proof of Lemma 6.4.1

In this proof we utilize the property that every Cauchy sequence in the Banach space L_{2p}, $p > 1$ converges to a limit within the space. For any $i = 1, 2, ..., N$, $t \in \mathbb{Z}$ and

$s \in \mathbb{N}$, define

$$\tilde{y}_{kl}^{(i,t,s)} = \begin{cases} y_{kl} \quad if \ \max\{|i-k|, |t-l|\} \le s; \\ 0 \quad otherwise. \end{cases} \tag{A.4.2}$$

Correspondingly, let

$$\tilde{\mu}_{kl}^{(i,t,s)}(\theta) = g_\theta\left(\tilde{y}_{k,l-1}^{(i,t,s)}, \sum_{j=1}^N w_{kj}\tilde{y}_{j,l-1}^{(i,t,s)}, \tilde{\mu}_{k,l-1}^{(i,t,s)}(\theta)\right) \tag{A.4.3}$$

For $s > 1$, by iteration there exits a function $g_\theta^{(s)}$ such that

$$\begin{aligned} \tilde{\mu}_{it}^{(i,t,s)}(\theta) &= g_\theta\left(\tilde{y}_{i,t-1}^{(i,t,s)}, \sum_{j=1}^N w_{ij}\tilde{y}_{j,t-1}^{(i,t,s)}, \tilde{\mu}_{i,t-1}^{(i,t,s)}(\theta)\right) \\ &= g_\theta\left(y_{i,t-1}, \sum_{|i-j|\le s} w_{ij}y_{j,t-1}, g_\theta\left(\tilde{y}_{i,t-2}^{(i,t,s)}, \sum_{j=1}^N w_{ij}\tilde{y}_{j,t-2}^{(i,t,s)}, \tilde{\mu}_{i,t-2}^{(i,t,s)}(\theta)\right)\right) \\ &\cdots \\ &= g_\theta^{(s)}\left(\left(y_{i,t-k}, \sum_{|i-j|\le s} w_{ij}y_{j,t-k}\right)_{1\le k\le s}\right). \end{aligned} \tag{A.4.4}$$

Claim A.4.3. *For any $i = 1, 2, ..., N$, $t \in \mathbb{Z}$, $\mu_{it}(\theta) = \lim_{s\to\infty} \tilde{\mu}_{it}^{(i,t,s)}(\theta)$ is well-defined in L_{2p}.*

Proof. Fixing an integer $m \ge 0$, by Assumption 6.4.4 we have

$$\begin{aligned} &\left|\tilde{\mu}_{it}^{(i,t,s+m)}(\theta) - \tilde{\mu}_{it}^{(i,t,s)}(\theta)\right| \\ \le& C_2 \sum_{s<|i-j|\le s+m} w_{ij}|y_{j,t-1}| + \rho\left|\tilde{\mu}_{i,t-1}^{(i,t,s+m)}(\theta) - \tilde{\mu}_{i,t-1}^{(i,t,s)}(\theta)\right| \\ &\cdots \\ \le& \rho^s C_1|y_{i,t-s-1}| + \sum_{k=0}^{s-1}\rho^k C_2 \sum_{s<|i-j|\le s+m} w_{ij}|y_{j,t-k-1}| \\ &+ \rho^s C_2 \sum_{|i-j|\le s+m} w_{ij}|y_{j,t-s-1}| + \rho^{s+1}\left|\tilde{\mu}_{i,t-s-1}^{(i,t,s+m)}(\theta) - \tilde{\mu}_{i,t-s-1}^{(i,t,s)}(\theta)\right| \\ &\cdots \end{aligned}$$

$$\leq \sum_{k=0}^{m-1} \rho^{s+k} C_1 |y_{i,t-s-1-k}| + \sum_{k=0}^{s-1} \rho^k C_2 \sum_{s<|i-j|\leq s+m} w_{ij} |y_{j,t-k-1}|$$
$$+ \sum_{k=0}^{m-1} \rho^{s+k} C_2 \sum_{|i-j|\leq s+m} w_{ij} |y_{j,t-s-1-k}|.$$

By Assumption 6.4.2 and Assumption 6.4.3 we have $C_y := \sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \|y_{it}\|_{2p} < \infty$ for some $p > 1$ and $w_{ij} \leq C|i-j|^{-\alpha}$ for some $\alpha > 2$. By Lemma A.1.(iii) in Jenish and Prucha (2009) we obtain that:

$$\left\|\tilde{\mu}_{it}^{(i,t,s+m)}(\theta) - \tilde{\mu}_{it}^{(i,t,s)}(\theta)\right\|_p$$
$$\leq C_y C_1 \rho^s \left(\sum_{k=0}^{m-1} \rho^k\right) + C_y C_2 \left(\sum_{k=0}^{s-1} \rho^k\right) \left(\sum_{h=s}^{s+m-1} C h^{1-\alpha}\right)$$
$$+ C_y C_2 \rho^s \left(\sum_{k=0}^{m-1} \rho^k\right) \left(\sum_{h=0}^{s+m-1} C h^{1-\alpha}\right),$$

which converges to 0 as $s \to \infty$. Therefore $\left\{\tilde{\mu}_{it}^{(i,t,s)}(\theta) : s \geq 0\right\}$ is a Cauchy sequence in L_{2p}, completing the proof of Claim A.4.3. □

By Claim A.4.3, there exists a function $g_\theta^{(\infty)}$ such that

$$\mu_{it}(\theta) = g_\theta^{(\infty)} \left(\left(y_{i,t-k}, \sum_{j=1}^{N} w_{ij} y_{j,t-k} \right)_{k\geq 1} \right)$$

is well-defined in L_{2p}.

Claim A.4.4. *The η-coefficients of $\{\mu_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ satisfy*

$$\bar{\eta}_\mu^{(0)}(r) \leq C r^{2-\mu}$$

for some constant $C > 0$.

Proof. For any $s \in \mathbb{N}$, we take

$$y_{kl}^{(i,t,s)} = y_{kl} \quad \textit{if and only if} \quad \max\{|i-k|, |t-l|\} \neq s \tag{A.4.5}$$

Correspondingly, let

$$\mu_{it}^{(i,t,s)}(\theta) = g_\theta^{(\infty)}\left(\left(y_{i,t-k}^{(i,t,s)}, \sum_{j=1}^{N} w_{ij} y_{j,t-k}^{(i,t,s)}\right)_{k\geq 1}\right). \tag{A.4.6}$$

Then we have

$$\begin{aligned}
&\left|\mu_{it}(\theta) - \mu_{it}^{(i,t,s)}(\theta)\right| \\
\leq &\rho^{s-1} C_1 \left|y_{i,t-s} - y_{i,t-s}^{(i,t,s)}\right| \\
&+ C s^{-\alpha} \sum_{k=0}^{s-2} \rho^k C_2 \left|y_{i\pm s,t-k-1} - y_{i\pm s,t-k-1}^{(i,t,s)}\right| \\
&+ \rho^{s-1} C_2 \sum_{|i-j|\leq s} w_{ij} \left|y_{j,t-s} - y_{j,t-s}^{(i,t,s)}\right|.
\end{aligned} \tag{A.4.7}$$

By Example 3.2.1 in Chapter 3 we complete the proof. □

Proof of Lemma 6.4.2

For any $i = 1, 2, ..., N$, $t \in \mathbb{Z}$ and $s \in \mathbb{N}$, define

$$l_{it}^{(i,t,s)}(\theta) = \mathcal{B}^{-1}(\mu_{it}^{(i,t,s)}(\theta)) y_{it}^{(i,t,s)} - \mathcal{A} \circ \mathcal{B}^{-1}(\mu_{it}^{(i,t,s)}(\theta)), \tag{A.4.8}$$

where $y_{it}^{(i,t,s)}$ is defined by (A.4.5) and $\mu_{it}^{(i,t,s)}(\theta)$ is defined by (A.4.6). By Assumption 6.4.5 we have

$$\left|l_{it}(\theta) - l_{it}^{(i,t,s)}(\theta)\right| \leq C(|y_{it}| + 1) \left|\mu_{it}(\theta) - \mu_{it}^{(i,t,s)}(\theta)\right|.$$

Then by the result we have in (A.4.7), we can prove that the η-coefficients of $\{l_{it}(\theta) : (i,t) \in D_{NT}, NT \geq 1\}$ satisfy

$$\bar{\eta}_l^{(0)}(r) \leq C r^{2 - \frac{p-4}{p-2}\mu}$$

for some constant $C > 0$.

At last, by Assumption 6.4.2 and the Lipschitz continuity of $\mathcal{B}^{-1}$ and $\mathcal{A} \circ \mathcal{B}^{-1}$ in Assumption 6.4.5, it is easy to verify that $\sup_{NT\geq 1} \sup_{(i,t)\in D_{NT}} \sup_{\theta\in\Theta} \|l_{it}(\theta)\|_{\frac{p}{2}} < \infty$

for some $p > 4$, using Hölder's inequality.

Proof of Lemma 6.4.3

The proof of Lemma 6.4.3 is similar to the proof of Lemma 6.4.1, therefore it is omitted here.

Proof of Lemma 6.4.4

For any $i = 1, 2, ..., N$, $t \in \mathbb{Z}$ and $s \in \mathbb{N}$, by (A.4.8) we have

$$\frac{\partial l_{it}^{(i,t,s)}(\theta_0)}{\partial \theta} = \left[(\mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))y_{it}^{(i,t,s)} - (\mathcal{A} \circ \mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\right] \frac{\partial \mu_{it}^{(i,t,s)}(\theta_0)}{\partial \theta} \tag{A.4.9}$$

where $y_{it}^{(i,t,s)}$ is defined by (A.4.5) and $\mu_{it}^{(i,t,s)}(\theta)$ is defined by (A.4.6). By Assumption 6.4.9 we have

$$\begin{aligned}
&\left|\frac{\partial l_{it}(\theta_0)}{\partial \theta} - \frac{\partial l_{it}^{(i,t,s)}(\theta_0)}{\partial \theta}\right| \\
\leq &\left|(\mathcal{B}^{-1})'(\mu_{it}(\theta_0))\frac{\partial \mu_{it}(\theta_0)}{\partial \theta}y_{it} - (\mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\frac{\partial \mu_{it}(\theta_0)}{\partial \theta}y_{it}\right| \\
&+ \left|(\mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\frac{\partial \mu_{it}(\theta_0)}{\partial \theta}y_{it} - (\mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\frac{\partial \mu_{it}^{(i,t,s)}(\theta_0)}{\partial \theta}y_{it}\right| \\
&+ \left|(\mathcal{A} \circ \mathcal{B}^{-1})'(\mu_{it}(\theta_0))\frac{\partial \mu_{it}(\theta_0)}{\partial \theta} - (\mathcal{A} \circ \mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\frac{\partial \mu_{it}(\theta_0)}{\partial \theta}\right| \\
&+ \left|(\mathcal{A} \circ \mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\frac{\partial \mu_{it}(\theta_0)}{\partial \theta} - (\mathcal{A} \circ \mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\frac{\partial \mu_{it}^{(i,t,s)}(\theta_0)}{\partial \theta}\right| \qquad \text{(A.4.10)} \\
\leq &C\left|\mu_{it}(\theta_0) - \mu_{it}^{(i,t,s)}(\theta_0)\right|\left|\frac{\partial \mu_{it}(\theta_0)}{\partial \theta}\right||y_{it}| \\
&+ \left|\frac{\partial \mu_{it}(\theta_0)}{\partial \theta} - \frac{\partial \mu_{it}^{(i,t,s)}(\theta_0)}{\partial \theta}\right|\left|(\mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\right||y_{it}| \\
&+ C\left|\mu_{it}(\theta_0) - \mu_{it}^{(i,t,s)}(\theta_0)\right|\left|\frac{\partial \mu_{it}(\theta_0)}{\partial \theta}\right| \\
&+ \left|\frac{\partial \mu_{it}(\theta_0)}{\partial \theta} - \frac{\partial \mu_{it}^{(i,t,s)}(\theta_0)}{\partial \theta}\right|\left|(\mathcal{A} \circ \mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\right|.
\end{aligned}$$

The rest of the proof follows similarly as the proof of Lemma 6.4.2.

Proof of Lemma 6.4.5

The proof of Lemma 6.4.5 is similar to the proof of Lemma 6.4.1, therefore it is omitted here.

Proof of Lemma 6.4.6

For any $i = 1, 2, ..., N$, $t \in \mathbb{Z}$ and $s \in \mathbb{N}$, similar to (A.4.9) we have

$$\begin{aligned}
&\frac{\partial^2 l_{it}^{(i,t,s)}(\theta_0)}{\partial\theta\partial\theta'} \\
&= \left[(\mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))y_{it}^{(i,t,s)} - (\mathcal{A}\circ\mathcal{B}^{-1})'(\mu_{it}^{(i,t,s)}(\theta_0))\right]\frac{\partial^2\mu_{it}^{(i,t,s)}(\theta_0)}{\partial\theta\partial\theta'} \\
&\quad + \left[(\mathcal{B}^{-1})''(\mu_{it}^{(i,t,s)}(\theta_0))y_{it}^{(i,t,s)} - (\mathcal{A}\circ\mathcal{B}^{-1})''(\mu_{it}^{(i,t,s)}(\theta_0))\right]\frac{\partial\mu_{it}^{(i,t,s)}(\theta_0)}{\partial\theta}\frac{\partial\mu_{it}^{(i,t,s)}(\theta_0)}{\partial\theta'}
\end{aligned} \tag{A.4.11}$$

where $y_{it}^{(i,t,s)}$ is defined by (A.4.5) and $\mu_{it}^{(i,t,s)}(\theta)$ is defined by (A.4.6). By Assumption 6.4.12 we can decompose $\left|\frac{\partial^2 l_{it}(\theta_0)}{\partial\theta\partial\theta'} - \frac{\partial^2 l_{it}^{(i,t,s)}(\theta_0)}{\partial\theta\partial\theta'}\right|$ similarly as (A.4.10). With Assumption 6.4.13 it is not hard to prove Lemma 6.4.6.

A.4.3 Proof of Proposition 6.1

We need to verify that the function g_θ defined in (6.5.2) satisfies Assumption 6.3.1. That is, the function

$$g_\theta(x, y, z) = \omega + \alpha^{(1)}x + \alpha^{(2)}(x - r)^+ + \lambda y + \beta z$$

satisfies condition (6.3.4). For any $a, a', b, b', c, c' \in \mathbb{R}$, we have

$$\left|g_\theta(a, b, c) - g_\theta(a', b', c')\right| \le \max\{\alpha^{(1)}, \alpha^{(1)} + \alpha^{(2)}\}\left|a - a'\right| + \lambda\left|b - b'\right| + \beta\left|c - c'\right|.$$

Since

$$\max\{\alpha^{(1)}, \alpha^{(1)} + \alpha^{(2)}\} + \lambda + \beta < 1,$$

by Theorem 6.1 we complete the proof.

A.4.4 Proof of Proposition 6.2

According to Theorem 6.2, Assumptions 6.4.1 to 6.4.6 are sufficient to obtain consistency. Verifying Assumption 6.4.1 and Assumption 6.4.6 is similar to the proof of Claims A.2.3 and A.2.6, therefore omitted here. Assumption 6.4.2 and Assumption 6.4.3 follow directly with (NB2), (NB3) and (NB4). Notice that in the case of (6.5.2),

$$g_\theta(a,b,c) = \omega + \alpha^{(1)}a + \alpha^{(2)}(a-r)^+ + \lambda b + \beta c.$$

For any (a,b,c) and (a',b',c') in $\mathbb{S}_0$:

$$\begin{aligned}&|g_\theta(a,b,c) - g_\theta(a',b',c')| \\ \leq &\max\{\alpha^{(1)}, \alpha^{(1)} + \alpha^{(2)}\}|a-a'| + \lambda|b-b'| + \beta|c-c'|.\end{aligned}$$

By (NB1) we can easily verify Assumption 6.4.4. Now it remains to verify Assumption 6.4.5. Firstly, notice that $\mu_{it}(\theta) \geq \omega > 0$. By the compactness of Θ in (NB1), there exists a constant $\omega^* > 0$ such that $\sup_{\theta\in\Theta}\sup_{(i,t)\in D_{NT}}\sup_{NT\geq 1}\mu_{it}(\theta) \geq \omega^*$. That is, for all $x \in \mathbb{S}_\mu$, $x \geq \omega^*$. Therefore functions

$$\begin{aligned}\mathcal{B}^{-1}(x) &= \log(x) - \log(x+K), \\ \mathcal{A}\circ\mathcal{B}^{-1}(x) &= K\log(x+K) - K\log(K)\end{aligned}$$

are Lipschitz continuous on $\mathbb{S}_\mu$. Now we have proved the consistency part of Proposition 6.2 with assumptions (NB1) to (NB4).

According to Theorem 6.3, we still need to verify Assumptions 6.4.7 to 6.4.15, in order to prove the asymptotic normality part. The verification of Assumption 6.4.7 is omitted here since it is similar to the proof of Claim A.3.7 and Claim A.3.8 if $N = o(T)$.

As for Assumption 6.4.8, since

$$\frac{\partial \mu_{it}(\theta, r)}{\partial \theta} = \begin{pmatrix} 1 \\ y_{i,t-1} \\ (y_{i,t-1} - r)^+ \\ \sum_{j=1}^N w_{i,j} y_{j,t-1} \\ \mu_{i,t-1}(\theta, r) \end{pmatrix} + \beta \frac{\partial \mu_{i,t-1}(\theta, r)}{\partial \theta},$$

that is

$$g_\theta^{(1)}(a, b, c, d) = \begin{pmatrix} 1 \\ a \\ (a-r)^+ \\ b \\ c \end{pmatrix} + \beta d,$$

for $(a, b, c, d) \in \mathbb{S}_1$. Then Assumption 6.4.8 is satisfied since $0 < \beta < 1$. We also have that

$$\frac{\partial^2 \mu_{it}(\theta, r)}{\partial \theta \partial \theta'} = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \omega} \\ 0 & 0 & 0 & 0 & \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \alpha^{(1)}} \\ 0 & 0 & 0 & 0 & \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \alpha^{(2)}} \\ 0 & 0 & 0 & 0 & \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \lambda} \\ \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \omega} & \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \alpha^{(1)}} & \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \alpha^{(2)}} & \frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \lambda} & 2\frac{\partial \mu_{i,t-1}(\theta,r)}{\partial \beta} \end{pmatrix} + \beta \frac{\partial^2 \mu_{i,t-1}(\theta, r)}{\partial \theta \partial \theta'},$$

and similarly we can verify Assumption 6.4.11.

As for Assumption 6.4.9, since

$$\frac{d}{dx}\mathcal{B}^{-1}(x) = \frac{1}{x} - \frac{1}{x+K},$$
$$\frac{d}{dx}\mathcal{A} \circ \mathcal{B}^{-1}(x) = \frac{K}{x+K},$$

and $x \geq \omega^*$ for all $x \in \mathbb{S}_\mu$, then $\frac{d}{dx}\mathcal{B}^{-1}(x)$ and $\frac{d}{dx}\mathcal{A} \circ \mathcal{B}^{-1}(x)$ are Lipschitz continuous on $\mathbb{S}_\mu$. The first bound in Assumption 6.4.10 can be proved similarly to (A.2.14) and

(A.3.10). The other two bounds also hold since

$$\left|\frac{d}{dx}\mathcal{B}^{-1}\left(\mu_{it}(\theta_0,r_0)\right)\right| \le \left|\frac{1}{\mu_{it}(\theta_0,r_0)}\right| + \left|\frac{1}{\mu_{it}(\theta_0,r_0)+K}\right| \le \frac{1}{\omega^*} + \frac{1}{\omega^*+K},$$
$$\left|\frac{d}{dx}\mathcal{A}\circ\mathcal{B}^{-1}\left(\mu_{it}(\theta_0,r_0)\right)\right| = \left|\frac{K}{\mu_{it}(\theta_0,r_0)+K}\right| \le \frac{K}{\omega^*+K}.$$

Assumptions 6.4.12 and 6.4.13 can be verified similarly, noticing that

$$\frac{d^2}{dx^2}\mathcal{B}^{-1}(x) = -\frac{1}{x^2} + \frac{1}{(x+K)^2},$$
$$\frac{d^2}{dx^2}\mathcal{A}\circ\mathcal{B}^{-1}(x) = -\frac{K}{(x+K)^2}.$$

Assumption 6.4.14 is supported by (NB1), and it remains to verified Assumption 6.4.15. Firstly notice that

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{y_{it}}{\mu_{it}(\theta_0,r_0)} - \frac{y_{it}+K}{\mu_{it}(\theta_0,r_0)+K}\right)\left(\frac{y_{j\tau}}{\mu_{j\tau}(\theta_0,r_0)} - \frac{y_{j\tau}+K}{\mu_{j\tau}(\theta_0,r_0)+K}\right)\right] \\
=&\mathbb{E}\left\{\mathbb{E}\left[\left(\frac{y_{it}}{\mu_{it}(\theta_0,r_0)} - \frac{y_{it}+K}{\mu_{it}(\theta_0,r_0)+K}\right)\right.\right. \\
&\quad\times\left.\left.\left(\frac{y_{j\tau}}{\mu_{j\tau}(\theta_0,r_0)} - \frac{y_{j\tau}+K}{\mu_{j\tau}(\theta_0,r_0)+K}\right)\right|\mathcal{H}_{t-1}\right]\right\} \\
=&0
\end{aligned}$$

if either $i \neq j$ or $t \neq \tau (t > \tau$ without loss of generality).

Since

$$\mathbb{E}(y_{it}|\mathcal{H}_{t-1}) = \mu_{it}(\theta_0,r_0),$$
$$\mathbb{E}(y_{it}^2|\mathcal{H}_{t-1}) = \frac{\mu_{it}^2(\theta_0,r_0) + K\mu_{it}(\theta_0,r_0)}{K} + \mu_{it}^2(\theta_0,r_0).$$

Then

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{y_{it}}{\mu_{it}(\theta_0,r_0)} - \frac{y_{it}+K}{\mu_{it}(\theta_0,r_0)+K}\right)^2 \frac{\partial\mu_{it}(\theta_0,r_0)}{\partial\theta}\frac{\partial\mu_{it}(\theta_0,r_0)}{\partial\theta'}\right|\mathcal{H}_{t-1}\Bigg] \\
=&\frac{\partial\mu_{it}(\theta_0,r_0)}{\partial\theta}\frac{\partial\mu_{it}(\theta_0,r_0)}{\partial\theta'}\left[\frac{\mathbb{E}(y_{it}^2|\mathcal{H}_{t-1})}{\mu_{it}^2(\theta_0,r_0)} + \frac{\mathbb{E}(y_{it}^2+2Ky_{it}+K^2|\mathcal{H}_{t-1})}{(\mu_{it}(\theta_0,r_0)+K)^2}\right.
\end{aligned}$$

$$
\begin{aligned}
&-\frac{2\mathbb{E}(y_{it}^2 + Ky_{it}|\mathcal{H}_{t-1})}{\mu_{it}(\theta_0, r_0)(\mu_{it}(\theta_0, r_0) + K)}\Bigg] \\
=&\frac{K}{\mu_{it}^2(\theta_0, r_0) + K\mu_{it}(\theta_0, r_0)} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta'}.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right] \\
=&\frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left\{\mathbb{E}\left[\left(\frac{y_{it}}{\mu_{it}(\theta_0, r_0)} - \frac{y_{it}+K}{\mu_{it}(\theta_0, r_0)+K}\right)\frac{\partial^2 \mu_{it}(\theta_0, r_0)}{\partial\theta\partial\theta'}\right.\right. \\
&\left.\left. - \left(\frac{y_{it}}{\mu_{it}^2(\theta_0, r_0)} - \frac{y_{it}+K}{(\mu_{it}(\theta_0, r_0)+K)^2}\right)\frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta'}\right| \mathcal{H}_{t-1}\right]\right\} \\
=&\frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\left(\frac{1}{(\mu_{it}\theta_0, r_0)+K} - \frac{1}{\mu_{it}(\theta_0, r_0)}\right)\frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta'}\right].
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
&\mathrm{Var}\left[\sqrt{NT}\frac{\partial L_{NT}(\theta_0, r_0)}{\partial \theta}\right] \\
=&\frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\left(\frac{y_{it}}{\mu_{it}(\theta_0, r_0)} - \frac{y_{it}+K}{\mu_{it}(\theta_0, r_0)+K}\right)^2\frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta'}\right] \\
=&\frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left\{\mathbb{E}\left[\left(\frac{y_{it}}{\mu_{it}(\theta_0, r_0)} - \frac{y_{it}+K}{\mu_{it}(\theta_0, r_0)+K}\right)^2\frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta'}\right| \mathcal{H}_{t-1}\right]\right\} \\
=&\frac{1}{NT}\sum_{(i,t)\in D_{NT}} \mathbb{E}\left[\frac{K}{\mu_{it}^2(\theta_0, r_0) + K\mu_{it}(\theta_0, r_0)} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta} \frac{\partial \mu_{it}(\theta_0, r_0)}{\partial \theta'}\right] \\
=&-\mathbb{E}\left[\frac{\partial^2 L_{NT}(\theta_0)}{\partial\theta\partial\theta'}\right].
\end{aligned}
$$

By (NB5) we can verify Assumption 6.4.15. Now we complete the proof of Proposition 6.2.

Bibliography

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, 8(3):261–275.

Al-Osh, M. A. and Alzaid, A. A. (1988). Integer-valued moving average (INMA) process. *Statistical Papers*, 29:281–300.

Al-Sulami, D., Jiang, Z., Lu, Z., and Zhu, J. (2017). Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data. *Econometrics and Statistics*, 2:22–35.

Andrews, D. W. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4):930–934.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer Dordrecht, 1st edition.

Armillotta, M. and Fokianos, K. (2024). Count network autoregression. *Journal of Time Series Analysis*, 45(4):584–612.

Augustyniak, M., Boudreault, M., and Morales, M. (2018). Maximum likelihood estimation of the Markov-switching GARCH model based on a general collapsing procedure. *Methodology and Computing in Applied Probability*, 20(1):165–188.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Bauwens, L., Preminger, A., and Rombouts, J. V. (2010). Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13(2):218–244.

Berkes, I., Horváth, L., and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli*, 9(2):201–227.

Billingsley, P. (1999). *Convergence of Probability Measures*. John Wiley & Sons, 2nd edition.

Billingsley, P. (2008). *Probability and Measure*. John Wiley & Sons.

Billio, M., Caporin, M., Frattarolo, L., and Pelizzon, L. (2023). Networks in risk spillovers: A multivariate GARCH perspective. *Econometrics and Statistics*, 28:1–29.

Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business & Economic Statistics Section, 1976*, pages 177–181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The Review of Economics and Statistics*, 72(3):498–505.

Bollerslev, T. and Engle, R. F. (1993). Common persistence in conditional variances. *Econometrica*, 61(1):167–186.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, pages 1047–1050.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52(1):115–127.

Brockwell, P. J. and Davis, R. A. (2009). *Time Series: Theory and Methods*. Springer Science & Business Media.

Cai, J. (1994). A markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, 12(3):309–316.

Caldarelli, G. and Vespignani, A. (2007). *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science*. World Scientific.

Christou, V. and Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35(1):55–78.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

Cliff, A. and Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4(3):267–284.

Cui, Y., Li, Q., and Zhu, F. (2020). Flexible bivariate Poisson integer-valued GARCH model. *Annals of the Institute of Statistical Mathematics*, 72(6):1449–1477.

Cui, Y. and Zhu, F. (2018). A new bivariate integer-valued GARCH model allowing for negative cross-correlation. *Test*, 27:428–452.

Curato, I. V., Stelzer, R., and Ströh, B. (2022). Central limit theorems for stationary random fields under weak dependence with application to ambit and mixed moving average fields. *The Annals of Applied Probability*, 32(3):1814–1861.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. OUP Oxford.

Davis, R. A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, 26:1673–1707.

De Wit, E. R., Englund, P., and Francke, M. K. (2013). Price and transaction volume in the Dutch housing market. *Regional Science and Urban Economics*, 43(2):220–241.

Dedecker, J. (1998). A central limit theorem for stationary random fields. *Probability Theory and Related Fields*, 110:397–426.

Dedecker, J. and Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Processes and their Applications*, 106(1):63–80.

Dedecker, J., Doukhan, P., Lang, G., José Rafael, L. R., Louhichi, S., and Prieur, C. (2007). *Weak Dependence: With Examples and Applications.* Springer New York.

Doukhan, P., Fokianos, K., and Li, X. (2012). On weak dependence conditions: The case of discrete valued processes. *Statistics & Probability Letters*, 82(11):1941–1948.

Doukhan, P., Latour, A., and Oraichi, D. (2006). A simple integer-valued bilinear time series model. *Advances in Applied Probability*, 38(2):559–578.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84(2):313–342.

Doukhan, P. and Truquet, L. (2007). Weakly dependent random fields with infinite interactions. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 3:111–132.

El Machkouri, M., Volný, D., and Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications*, 123(1):1–14.

Engle, R. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.

Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics*, 45(1):213–237.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–60.

Fan, J., Yao, Q., and Cai, Z. (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(1):57–80. Publisher: [Royal Statistical Society, Wiley].

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of time series analysis*, 27(6):923–942.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

Fokianos, K., Støve, B., Tjøstheim, D., and Doukhan, P. (2020). Multivariate count autoregression. *Bernoulli*, 26(1):471 – 499.

Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637.

French, K. R., Schwert, G., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Gorodetskii, V. (1978). On the strong mixing property for linear sequences. *Theory of Probability & Its Applications*, 22(2):411–413.

Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics, and Applications.* Springer Science & Business Media.

Hamilton, J. (2020). *Time Series Analysis.* Princeton University Press.

Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2):307–333.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. *Available at SSRN 1117187.*

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis.* Cambridge University Press, 2nd edition.

Hu, Y.-P. and Tsay, R. S. (2014). Principal volatility component analysis. *Journal of Business & Economic Statistics*, 32(2):153–164.

Hølleland, S. and Karlsen, H. A. (2020). A stationary spatio-temporal GARCH model. *Journal of Time Series Analysis*, 41(2):177–209.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150(1):86–98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, 170(1):178–190.

Jones, C. M., Kaul, G., and Lipson, M. L. (1994). Transactions, volume, and volatility. *The Review of Financial Studies*, 7(4):631–651.

Karolyi, G. A. (1995). A multivariate GARCH model of international transmissions of stock returns and volatility: the case of the United States and Canada. *Journal of Business & Economic Statistics*, 13(1):11–25.

Kesten, H. and Spitzer, F. (1984). Convergence in distribution of products of random matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(4):363–386.

Kim, B., Lee, K. H., Xue, L., and Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, 12:105.

Kingman, J. F. C. (1973). Subadditive ergodic theory. *Annals of Probability*, 1(6):883–899.

Lee, S., Kim, D., and Kim, B. (2023). Modeling and inference for multivariate time series of counts based on the INGARCH scheme. *Computational Statistics & Data Analysis*, 177:107579.

Lee, Y., Lee, S., and Tjøstheim, D. (2018). Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. *Test*, 27:52–69.

Li, W., Gao, J., Li, K., and Yao, Q. (2016). Modeling multivariate volatilities via latent common factors. *Journal of Business & Economic Statistics*, 34(4):564–573.

Li, Z., Lam, C., Yao, J., and Yao, Q. (2019). On testing for high-dimensional white noise. *The Annals of Statistics*, 47(6):3382–3412.

Lu, Z., Ren, X., and Zhang, R. (2024). On semiparametrically dynamic functional-coefficient autoregressive spatio-temporal models with irregular location wide nonstationarity. *Journal of the American Statistical Association*, 119(546):1032–1043.

Lu, Z., Steinskog, D. J., Tjøstheim, D., and Yao, Q. (2009). Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(4):859–880.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, 21(4):645–650.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20(4):822–835.

McNees, S. K. (1979). The forecasting record of the 1970s. *New Enaland Economic Review*, (September/October):33–53.

Merlevède, F., Peligrad, M., and Utev, S. (2019). Functional CLT for martingale-like nonstationary dependent structures. *Bernoulli*, 25(4B):3203 – 3233.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59(2):347–370.

Neumann, M. H. (2013). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, 17:120–134.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic block structures. *Journal of the American Statistical Association*, 96(455):1077–1087.

Pan, J., Polonik, W., and Yao, Q. (2010). Estimating factor models for multivariate volatilities: an innovation expansion method. In *Proceedings of COMPSTAT'2010*, pages 305–314. Physica-Verlag Heidelberg.

Pan, Y. and Pan, J. (2024). Threshold network GARCH model. *Journal of Time Series Analysis*. to appear.

Pham, H. (2020). On estimating the number of deaths related to Covid-19. *Mathematics*, 8(5):655.

Rao, S. S. (2008). Statistical analysis of a spatio-temporal model with location-dependent parameters and a test for spatial stationarity. *Journal of Time Series Analysis*, 29(4):673–694.

Ristić, M. M., Weiß, C. H., and Janjić, A. D. (2016). A binomial integer-valued ARCH model. *The International Journal of Biostatistics*, 12(2):20150051.

Samia, N. I. and Chan, K.-S. (2011). Maximum likelihood estimation of a generalized threshold stochastic regression model. *Biometrika*, 98(2):433–448.

Tao, Y., Li, D., and Niu, X. (2024). Grouped network Poisson autoregressive model. *Statistica Sinica*, 34:1603–1624.

Teräsvirta, T. (2009). An introduction to univariate GARCH models. In *Handbook of Financial Time Series*, pages 17–42. Springer Berlin.

Tse, Y. and Tsui, A. K. (2001). A multivariate GARCH model with time-varying correlations. *Mathematics Preprint Archive*, 2001(6):1184–1217.

Wang, C., Liu, H., Yao, J.-F., Davis, R. A., and Li, W. K. (2014). Self-excited threshold Poisson autoregression. *Journal of the American Statistical Association*, 109(506):777–787.

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436.

Xu, H.-Y., Xie, M., Goh, T. N., and Fu, X. (2012). A model for integer-valued time series with conditional overdispersion. *Computational Statistics & Data Analysis*, 56(12):4229–4242.

Xu, X. and Lee, L. (2015). Maximum likelihood estimation of a spatial autoregressive tobit model. *Journal of Econometrics*, 188(1):264–280.

Xu, X., Wang, W., Shin, Y., and Zheng, C. (2024). Dynamic network quantile regression model. *Journal of Business & Economic Statistics*, 42(2):407–421.

Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.

Zhou, J., Li, D., Pan, R., and Wang, H. (2020). Network GARCH model. *Statistica Sinica*, 30:1–18.

Zhu, F. (2010). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, 32(1):54–67.

Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, 389(1):58–71.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *Annals of Statistics*, 45(3):1096–1123.

Zhu, X., Wang, W., Wang, H., and Härdle, W. K. (2019). Network quantile autoregression. *Journal of Econometrics*, 212(1):345–358.